

UCLA

Working Papers in Phonetics

Title

WPP, No. 52: Temporal Neural Coding of Pitch and Vowel Quality

Permalink

<https://escholarship.org/uc/item/0vp3920c>

Author

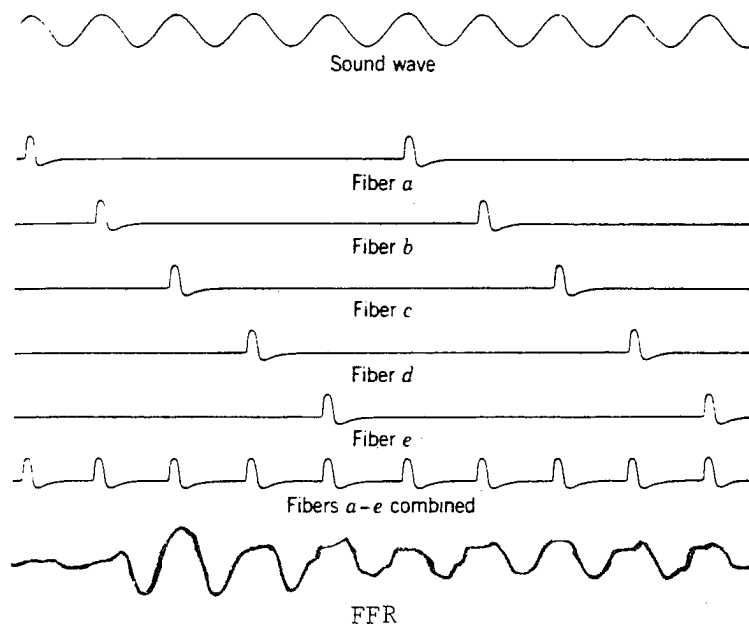
Greenberg, Steven

Publication Date

1980-12-01

TEMPORAL NEURAL CODING of PITCH AND VOWEL QUALITY

Steven Greenberg



UCLA WORKING PAPERS IN PHONETICS

52

DECEMBER, 1980

Ex Libris
Manuel Godínez Jr.

FRONT COVER

Figure illustrates the volley principle of frequency coding and the presumed relationship of the frequency-following response to single unit discharge activity. The figure is from Wever (1949) and is reproduced with permission from Dover Publications, Inc., and E.G. Wever.

Neural Temporal Coding of Pitch and Vowel Quality:
Human Frequency-Following Response Studies
of Complex Signals

Steven Greenberg

UCLA Working Papers in Phonetics -- Number 52

December 1980

© Copyright by
Steven Greenberg
1980

Dedication

To my parents,
for their love and support
during the course of my studies

TABLE OF CONTENTS

I.	PSYCHOPHYSICAL AND PHYSIOLOGICAL BASIS OF PITCH AND VOWEL QUALITY	1
1.1	Introduction	1
1.2	Historical Development of Research on Pitch Perception	3
1.21	The Nineteenth Century Outlook on Pitch: Place versus Periodicity	4
1.22	The Rise of Modern Periodicity Theory: The Volley Principle and the Pitch of the Residue	8
1.23	The Pitch of Inharmonic Signals	9
1.24	Mapping the "Existence Region" of the Residue	13
1.25	Doubts about Periodicity Theory: The Dominance of Low Harmonics	13
1.26	The Rise of Modern Place Theory: The "Internal Spectrum" and the Role of Combination Tones	14
1.27	Phase Effects and the Temporal Fine-Structure of the Waveform	15
1.28	Central Origin of Pitch: Doubts about the Relevance of the Temporal Fine-Structure	17
1.29	Summary: The Major Features of Low Pitch	19
1.3	Pattern Recognition Models of Pitch Perception	21
1.31	Pattern Transformation Model	22
1.32	Optimum Processor Model	24
1.4	Physiological Mechanisms Underlying Frequency Coding	27
1.41	Anatomy of the Auditory Pathway	27
1.42	Physiological Correlates of "Place" Coding	29
1.43	Physiological Correlates of "Periodicity" Coding	33
1.5	Low Pitch Based on Neural Periodicity	39
1.6	Auditory Basis of Timbre	42
1.61	Acoustic Basis of Vowel Quality	43
1.62	Auditory Basis of Vowel Quality	45
1.63	Place Representation of Vowel Spectra	50
1.64	The Role of Neural Periodicity in Vowel Perception	51
1.7	Physiological Basis of the Frequency-Following Response	53
1.71	The Relationship of Depth-Recorded Periodic Brainstem Potentials (Frequency-Following Response) to the Temporal Discharge Patterns of Single Units	53
1.72	The Relationship of Scalp-Recorded Frequency-Following Response to Depth-Recorded Activity	55
1.73	Tonotopic Specificity of Scalp-Recorded Frequency-Following Response	58
1.74	Amplitude Behavior of Frequency-Following Response as a Function of Signal Level and Signal-to-Noise Ratio ...	60
1.75	Summary: Basic Properties of the Frequency-Following Response	62

VII.	FREQUENCY-FOLLOWING RESPONSE TO SYNTHETIC VOWELS	122
7.1	Introduction	122
7.2	Experimental Methods	125
7.3	Experimental Results	126
7.4	Discussion	136
7.5	Frequency-Following Response to Synthetic Vowels as a Function of Stimulus Level	138
7.6	Frequency-Following Response to Synthetic Vowels using the Response-Subtraction Procedure	143
7.61	Experimental Methods	143
7.62	Experimental Results	146
7.63	Discussion	148
VIII.	SUMMARY AND CONCLUSIONS	151
8.1	Introduction	151
8.2	Frequency-Following Response Correlates of Low Pitch	152
8.3	Theoretical Implications of the Present Study for Models of Pitch Perception	157
8.4	Frequency-Following Response Correlates of Vowel Quality ..	159
8.5	Theoretical Implications for Models of Vowel Perception ...	161
NOTES		164
REFERENCES		168

II.	GENERAL EXPERIMENTAL METHODS	64
2.1	Signal Generation	64
2.2	Signal Presentation	64
2.3	Frequency-Following Response Recording System	64
2.4	Spectral Analysis of the Frequency-Following Response	67
2.5	Subjects	67
III.	PRELIMINARY STUDIES OF THE FREQUENCY-FOLLOWING RESPONSE TO THE "MISSING" FUNDAMENTAL	68
3.1	Introduction	68
3.2	Experimental Methods	69
3.3	Experimental Results	69
3.4	Discussion	71
3.5	Frequency-Following Response to the "Missing" Fundamental in the Presence of Low-Frequency Masking Noise	71
IV.	SPECTRAL BASIS OF THE FREQUENCY-FOLLOWING RESPONSE TO THE "MISSING" FUNDAMENTAL	74
4.1	Introduction	74
4.2	Experimental Methods	76
4.3	Experimental Results	77
4.4	Discussion	82
V.	PHASE-SENSITIVITY OF THE FREQUENCY-FOLLOWING RESPONSE TO THE "MISSING" FUNDAMENTAL	89
5.1	Introduction	89
5.2	Comparison of Phase-Sensitivity of the FFR to Tones of Low and High Harmonic Rank	91
5.21	Experimental Methods	91
5.22	Experimental Results	92
5.23	Discussion	98
5.3	Phase-Sensitivity of the FFR as a Function of Intensity	98
5.31	Experimental Methods	99
5.32	Experimental Results	99
5.33	Discussion	103
5.4	Phase-Sensitivity of the FFR as a Function of Relative Spectrum Level of the Stimulus Components	103
5.41	Experimental Methods	104
5.42	Experimental Results	104
5.43	Discussion	104
5.5	Summary	107
VI.	FREQUENCY-FOLLOWING RESPONSE TO INHARMONIC SIGNALS	109
6.1	Introduction	109
6.2	Experimental Methods	111
6.3	Experimental Results	112
6.4	Discussion	118

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1.1 Seebeck's demonstration of the "missing" fundamental	6
1.2 Waveforms and spectra of complex and pure tones	7
1.3 Schouten's model of the residue	10
1.4 Waveforms and spectra of harmonic and inharmonic signals ...	11
1.5 The relationship between phase configuration and stimulus waveform	16
1.6 Goldstein's optimum processor model	26
1.7 Anatomy of the ascending auditory pathway	28
1.8 Frequency threshold curve of a cochlear nerve fiber	30
1.9 Iso-intensity plot of cochlear nerve fiber activity	32
1.10 Wever's volley principle	34
1.11 A period histogram	35
1.12 An interval histogram	36
1.13 Dynamic range of cochlear nerve fibers: Rate versus temporal analyses	38
1.14 Articulatory basis of the vocalic spectrum	44
1.15 Representative spectra of four different vowels	45
1.16 Frequency-following responses recorded from the scalp of human and feline subjects in response to a 500 Hz tone	57
1.17 Growth of FFR amplitude as a function of stimulus intensity	61
2.1 Stimulus presentation and FFR recording system	65
2.2 Analog spectrum analysis system	66
3.1 FFR recorded in response to pure and complex tone stimuli ..	70
3.2 FFR recorded to pure and complex tone stimuli presented in silence and in the presence of masking noise	73
4.1 FFR recorded to two-component stimuli of variable harmonic rank: Fundamental frequency = 244 Hz	78
4.2 FFR amplitude as a function of harmonic rank: Comparison with psychoacoustic data	80
4.3 FFR recorded to two-component stimuli of variable harmonic rank: Fundamental frequency = 366 Hz	81
4.4 FFR recorded to two-component stimuli of variable harmonic rank: Fundamental frequency = 488 Hz	83
4.5 FFR amplitude as a function of harmonic rank for three fundamental frequencies	84
4.6 FFR amplitude as a function of stimulus frequency for three fundamental frequencies	85

5.1	FFR recorded in response to three-component tones of variable phase configuration: Harmonics 3-5	93
5.2	FFR recorded in response to three-component tones of variable phase configuration: Harmonics 7-9	95
5.3	Comparison of FFR recorded in response to phase-coherent and quadrature phase stimuli	96
5.4	Relative spectrum level of FFR recorded in response to phase-coherent and quadrature phase stimuli	97
5.5	FFR recorded to phase-coherent and quadrature phase stimuli as a function of stimulus intensity: Subject JTM ..	100
5.6	Relative spectrum level of FFR recorded to phase-coherent and quadrature phase stimuli	101
5.7	FFR recorded to phase-coherent and quadrature phase stimuli as a function of stimulus intensity: Subject AW ...	102
5.8	FFR recorded to three-component tones of variable phase configuration: Modulation index = 1.00	105
5.9	Comparison of FFR recorded to phase-coherent and quadrature phase stimuli for two different modulation indexes.....	106
6.1	Comparison of FFR recorded to three-component inharmonic and harmonic stimuli	113
6.2	Illustration of response-subtraction procedure	117
6.3	Spectral analysis of subtracted and added FFR recorded to an inharmonic stimulus	119
7.1	FFR recorded in response to synthetic vowels	127
7.2	Spectral analysis of FFR recorded to the vowel [I] for four subjects	128
7.3	Spectral analysis of FFR recorded to the vowel [I]	129
7.4	Spectral analysis of FFR recorded to the vowel [E]	130
7.5	Spectral analysis of FFR recorded to the vowel [ae]	131
7.6	Spectral analysis of FFR recorded to the vowel [a]	132
7.7	A possible FFR correlate of synchrony suppression	135
7.8	FFR recorded to the vowel [a] as a function of stimulus intensity	139
7.9	FFR recorded to a filtered vowel presented in silence and in the presence of high-frequency bandpass masking noise: Stimulus Intensity = 55 dB SPL	140
7.10	FFR recorded to a filtered vowel presented in silence and in the presence of high-frequency bandpass masking noise: Stimulus Intensity = 65 dB SPL	141
7.11	Spectral analysis of FFR and subtracted FFR recorded to the vowel [a]	144
7.12	Spectra of vocalic stimuli used in the response-subtraction experiment	145
7.13	Subtracted FFR recorded in response to four synthetic vowels: Response waveforms and spectral analyses	147
7.14	Spectral analysis of subtracted FFR recorded to four synthetic vowels for three subjects	149

LIST OF TABLES

<u>Table</u>	<u>Page</u>
4.1 Region of dominance for low pitch	74
6.1 Interval analysis of FFR recorded to harmonic and inharmonic stimuli	115

ACKNOWLEDGEMENTS

Many persons contributed their time, energy, and ideas during the course of my dissertation research. To all of them I extend my deepest appreciation.

Among those to whom I am especially grateful are Dr. James Marsh and Professor Peter Ladefoged. Both gave unselfishly of their time and energy, even during the busiest of times, and provided the moral support crucial for the dissertation's successful completion.

Dr. Marsh has guided my research over the past three years, always with keen insight, good humor and abundant patience. I thank him as well for the research and financial support provided over during this period.

Professor Ladefoged provided intellectual and financial support over the course of many years, during which time he has displayed exceptional tolerance of my rather aberrant research interests. I am grateful to him, as well, for serving as my dissertation chairman.

I would also like to extend my deepest appreciation to Dr. Peter Narins for his continual insistence on descriptive clarity and for acting as my tutor in single unit auditory physiology, and to Drs. Susan Curtiss and Victoria Fromkin for serving on my committee and for overlooking the shortcomings of this work as a dissertation in linguistics.

I am most grateful to my laboratory colleagues, Warren Brown and Jim Smith, for providing an intellectually enriching and friendly environment in which to work, and for their technical assistance offered on innumerable occasions. The digitally-based stimulus presentation and recording system was designed by Jim Smith. I also thank Bill Bergerson and Stanley de Wied for providing additional technical assistance.

Mary Smith and Diana van Lancker proofread the final manuscript. I thank them for their diligence and patience.

Thanks are extended to Ren Brown, Warren Brown, Lew D'Lia, Linda Galloway, Don Gilbertson, Manuel Godínez, Helen Haig, Hector Javkin, Penny Jennings, Linnea Lagerquist, Priscilla McCoy, James Marsh, Tracy Marsh, and Andreus Wittenstein for their service as subjects. I am especially grateful to Dr. James Marsh, who served as the "guinea pig" for innumerable pilot experiments.

To Mary, I shall always be grateful for her patience and emotional support during the preparation of the dissertation.

This research was supported, in part, by a grant from NIH awarded to Dr. Peter Ladefoged and by NIH Biomedical Research Support Grant RR 05756.

Permission to reproduce the following figures is gratefully acknowledged:

- Figure 1.1 - Schouten, J. (1940) The residue, a new component in subjective sound analysis. Proc. Kon. Ned. Akad. Wet. 43: 356-365. Figure 5, p. 363. By permission of the Koninlijke Nederlandse Akademie van Wetenschappen.
- Figure 1.3 - Schouten, J. (1940) The residue and the mechanism of hearing. Proc. Kon. Ned. Akad. Wet. 43: 991-999. Figure 1, p. 995. By permission of the Koninlijke Nederlandse Akademie van Wetenschappen.
- Figure 1.4 - Green, D.M. (1976) An Introduction to Hearing. Hillsdale, N.J.: Lawrence Erlbaum Associates. Figure 7.6, p. 189. By permission of Lawrence Erlbaum Associates.
- Figure 1.5 - Green, D.M. (1976) An Introduction to Hearing. Hillsdale, N.J.: Lawrence Erlbaum Associates. Figure 7.8, p. 193. By permission of Lawrence Erlbaum Associates.
- Figure 1.6 - Goldstein, J.L. (1978) Mechanisms of signal analysis and pattern perception in periodicity pitch. Audiology 17: 421-445. Figure 8, p. 435. By permission of the International Society of Audiology and Dr. Julius Goldstein.
- Figure 1.8 - Weiss, T.F., Mulroy, M.J., Turner, R.G., and Pike, C.L. (1976) Tuning of single fibers in the cochlear nerve of the alligator lizard: Relation to receptor morphology. Brain Research 115: 71-90. Figure 7, p. 80. By permission of Elsevier/North-Holland Publishing Co.
- Figure 1.9 - Rose, J.E., Hind, J.E., Anderson, D.J., and Brugge, J. (1971) Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. J. Neurophysiology 34: 685-699. Figure 1, p. 686. By permission of the American Physiological Society and Dr. Jerzy Rose.
- Figure 1.10 - Wever, E.G. (1949) Theory of Hearing. New York: Wiley (Reprinted by Dover Publications, Inc.). Figure 33, p. 167. By permission of Dover Publications, Inc. and Dr. E.G. Wever.
- Figure 1.11(a) - Green, D.M. (1976) An Introduction to Hearing. Hillsdale, N.J.: Lawrence Erlbaum Associates. Figure 5.11, p. 118. By permission of Lawrence Erlbaum Associates, Inc.
- Figure 1.11(b) - Rose, J.E., Brugge, J., Anderson, D.J., and Hind, J.E. (1967) Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol. 30: 769-793. Figure 5, p. 773. By permission of the American Physiological Society and Dr. Jerzy Rose.
- Figure 1.12 - Rose, J.E., Brugge, J., Anderson, D.J., and Hind, J.E. (1967) Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol. 30: 769-793. Figure 3, p. 773. By permission of the American Physiological Society and Dr. Jerzy Rose.

- Figure 1.13(a) - Rose, J.E., Brugge, J., Anderson, D.J., and Hind, J.E. (1967) Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol. 30: 769-793. Figure 3, p. 773. By permission of the American Physiological Society and Dr. Jerzy Rose.
- Figure 1.13(b) - Rose, J.E., Kitzes, L., Gibson, M.M., and Hind, J.E. (1974) Observations on phase-sensitive neurons of anteroventral cochlear nucleus of the cat: Nonlinearity of cochlear output. J. Neurophysiology. 37: 218-253. Figure 6, pp. 231. By permission of the American Physiological Society and Dr. Jerzy Rose.
- Figure 1.14 - Minifie, F.D. (1973) Speech Acoustics. In Normal Aspects of Speech, Hearing, and Language. Englewood Cliffs, N.J.: Prentice-Hall. Figure 7.8, p. 246 (Adapted from G. Fant (1960) Acoustic Theory of Speech Production. The Hague: Mouton. Figure 1.2, p. 19). By permission of Prentice-Hall and Mouton Publishing Co.
- Figure 1.15 - Pickett, J.M. (1980) The Sounds of Speech Communication. Baltimore: University Park Press. Figure 26, p. 66. By permission of University Park Press.
- Figure 2.2 - Smith, J.C., Marsh, J.T., Greenberg, S., and Brown, W.S. (1978) Human auditory frequency-following responses to a missing fundamental. Science 201: 639-641. Figure 1, p. 640. By permission of the American Association for the Advancement of Science and Dr. James C. Smith.

ABSTRACT OF THE DISSERTATION

Neural Temporal Coding of Pitch and Vowel Quality:
Human Frequency-Following Response Studies
of Complex Signals

by

Steven Greenberg

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 1980

Professor Peter Ladefoged, Chairman

The auditory analysis of complex signals such as speech and music is based on neural mechanisms of great precision and complexity. For well over a century auditory research has been concerned with the issue of whether this analysis is carried out mainly in terms of "place" or "periodicity" principles of neural coding.

This issue was addressed in the present study by recording a brainstem potential from the scalp of human subjects, which represents the periodic activity of auditory neurons in the upper brainstem pathway. This potential, the frequency-following response (FFR), was used as an index of neural periodicity in order to determine the extent to which the pitch and timbre of complex signals are encoded in the temporal discharge activity of auditory brainstem neurons.

The first part of the study focussed on the pitch of complex tones. A tone composed of harmonics of a common fundamental may, in the absence of the fundamental component, produce a pitch one octave or more below the frequency of the lowest stimulus component. The pitch perceived is equal to the pitch of the "missing" fundamental. Such stimuli generate frequency-following responses that are spectrally similar to FFR recorded in response to pure tones equal in frequency to

the "missing" fundamental. Both types of responses are periodic and match the perceived pitch of the stimuli.

The relationship between the pitch of complex tones and FFR spectrum was explored in greater detail by recording brainstem responses to signals of variable spectral and phase configuration. The results of these experiments indicate that many perceptual attributes of low pitch are also reflected in the scalp-recorded frequency-following response.

In particular, it was found that the largest FFR is generated by harmonics lying in the spectral region dominant for the low pitch of complex tones (500-1000 Hz). It was also determined that the upper spectral limits of the FFR are similar to those for low pitch.

In a separate series of experiments, it was shown that the amplitude and spectrum of the FFR is generally insensitive to the stimulus phase configuration under conditions similar to those used in analogous psychophysical experiments.

It was also found that the "pitch shift" of inharmonic tones is reflected in the FFR spectrum under certain stimulus conditions and that information related to the perceived pitch is contained in the interpeak intervals of the brainstem response.

These results indicate that the low pitch of complex tones is likely to be based on the temporal discharge activity of auditory neurons.

The neural basis of vowel quality was explored in the second part of the study. Spectral analysis of FFR recorded in response to synthetic vowels indicates that the fundamental frequency and first formant are represented in the brainstem response. Occasionally, the second and third formants are also represented. These results suggest that temporal mechanisms of neural coding play an important role in the perception of vowel quality.

CHAPTER I

PSYCHOPHYSICAL AND PHYSIOLOGICAL BASIS OF PITCH AND VOWEL QUALITY

Wodurch kann uber die Frage was zu einem Tone gehore entschieden werden, als eben durch das Ohr? (How else but through study of the ear can we determine how a tone is perceived?)

August Seebeck (1844)

1.1 INTRODUCTION

Auditory analysis of complex sounds plays a vital role in the communication systems of many species, including our own. The verbal exchange of information, the aesthetic appreciation of music, and many other activities common in our daily lives, depend upon the rapid analysis of the auditory signal. And yet our current understanding of the auditory mechanisms underlying the perception of speech and music is extremely rudimentary. For, very little is known about the way in which complex signals, such as speech, are processed at any level of the auditory pathway.

In the following chapters, some of the basic auditory mechanisms of potential relevance to speech communication and musical aesthetics are explored. Of particular concern will be neural mechanisms involved in perception of low pitch and vowel quality. I will endeavor to show that both processes reflect the synchronization of auditory neurons to certain features of the signal and that the transmission code for pitch and vowel quality is based on the temporal discharge patterns of neural aggregates in the auditory brainstem pathway.

Chapter 1 begins with an historical account of research on pitch perception. The course of pitch theory is traced from the classical "place" model of Helmholtz and "periodicity" model of Wundt and Rutherford to contemporary "pattern recognition" models (Goldstein, 1973; Wightman, 1973a; Terhardt, 1974). In the course of this review, the major features of low pitch are discussed in relation to a model based on the synchronous activity of auditory neurons. It is shown that many of these same features are not easily accommodated within current place models of pitch.

Attention is focussed next on auditory mechanisms underlying the perception of vowel quality. It is suggested that the analysis of low-frequency complex spectra is likely to be based on the phase-locking behavior of peripheral auditory neurons and that the auditory representation of timbre is also likely to be based on the temporal discharge pattern at this level of the auditory pathway.

At present, most of our knowledge concerning the auditory coding of complex sounds is limited to studies of the auditory periphery. At the level of the cochlear nerve (and probably cochlear nucleus), the low-frequency region of the spectrum is almost certainly encoded in terms of the temporal pattern of neural activity. At higher levels of the auditory pathway, the transmission code is much less certain. Whereas in the auditory periphery, neural phase-locking has been observed in response to frequencies up to 5 kHz (Rose et al, 1967), phase-locking in more central brainstem nuclei is apparently restricted to frequencies below 2-3 kHz (Moushegian et al, 1972; Boudreau, 1965; Aitkin et al, 1970; Rose et al, 1966). Consequently, it is not clear whether neural periodicity continues to play an important role at this level of the pathway.

In the present study, this issue is addressed by recording the periodic activity of neural aggregates in the upper auditory brainstem pathway in response to complex signals relevant to the processing of pitch and vowel quality. This neural activity was recorded from the scalp of human subjects, using far-field recording techniques. The response obtained under these conditions is periodic and corresponds to the frequency of stimulation. For this reason, this low-frequency brainstem potential is known as the frequency-following response (FFR). The physiological basis of the FFR is discussed in Chapters 1.71 - 1.75. It is shown that the response reflects the phase-locked activity of low-frequency units (probably) located in the inferior colliculus and that it provides a means to monitor the temporal activity of central auditory neurons located in the upper auditory brainstem pathway of human subjects during presentation of music- and speech-like signals.

Chapter 2 provides the methodological background of the current study, including details of stimulus generation, stimulus presentation and signal recording procedures used in the experiments described in subsequent chapters.

The general strategy of the present study is to record FFR to complex tones which produce pitches of variable salience and magnitude in order to determine whether there is some form of correlation between stimulus percept and some attribute or set of attributes of the brainstem potential.

Chapter 3 demonstrates the frequency-following response to the "missing fundamental." Of particular concern is whether the pitch of the signal is correlated with some property of the FFR. It is observed that the scalp response to a complex tone lacking the fundamental component is a periodic signal whose primary frequency corresponds to the pitch or "missing" fundamental of the stimulus. In Chapter 3, I also show that this response is not generated by a combination tone created through non-linear cochlear interaction.

In Chapter 4, it is shown that the amplitude of the FFR varies as a function of the harmonic rank and frequency of the stimulus components.

The largest responses are generated by frequencies between 500 and 1000 Hz. It is in this same spectral region that the sensation of pitch is most salient.

In Chapter 5 the discussion centers around the relationship between the envelope modulation pattern of the stimulus waveform and the amplitude of the FFR. It is shown that the relationship between the two is dependent on both the sound pressure level and harmonic spacing of the stimulus. In particular it is shown that when the stimulus components are well resolved, the FFR is insensitive to the phase configuration of the signal.

Inharmonic signals have played an important role in the development of contemporary pitch theories. In Chapter 6, I present the results of recording the FFR in response to inharmonic complexes to determine if the pitch shift produced by these signals is reflected in the spectrum of the FFR. It is shown that the pitch shift is reflected in the brainstem potential under special conditions. These conditions suggest that the pitch-relevant information is probably conveyed in the discharge patterns of neurons tuned directly to the stimulus components.

In Chapter 7, attention is focussed on the auditory coding of vowel quality and whether the formant regions of the vocalic spectra are reflected in the spectrum of the brainstem response. The results obtained indicate that the first formant is usually represented in the FFR spectrum and that occasionally the FFR contains spectral information related to even higher formants.

Chapter 8 summarizes the main findings of the dissertation and discusses their possible implications for models of pitch and speech perception.

1.2 HISTORICAL DEVELOPMENT OF RESEARCH ON PITCH PERCEPTION

Our current understanding of pitch perception is based on over a century of research. Early theories of pitch perception focussed on simple attributes of the signal (e.g. the fundamental component or waveform envelope) and assumed a fairly direct relationship between these attributes and the sensation of pitch. This emphasis reflected, in part, a focus on the pitch of single frequency signals. Over the years, with the development of more sophisticated psychophysical and physiological techniques, the emphasis has slowly shifted toward more complex signals. As a consequence, it is now evident that the perception of pitch is a multi-stage process involving a peripheral stage of frequency analysis followed by a central stage of synthesis. Yet, despite the progress of the last hundred years the basic issue still remains: Is pitch based on periodicity or place?

1.21 THE NINETEENTH CENTURY OUTLOOK ON PITCH: PLACE VERSUS PERIODICITY

Over a century ago, Hermann von Helmholtz published his treatise On the Sensations of Tone (1863). In his book Helmholtz attempted to integrate all that was then currently known about auditory physiology and anatomy into a coherent theory of hearing, particularly as would be relevant to the perception and appreciation of music. The book quickly became a classic and dominated much of auditory theory well into this century. So influential was his treatise that many of the questions posed by Helmholtz are still central issues in auditory research.

Two of the most important questions addressed by Helmholtz were:

- (1) How does the ear analyze the spectrum of complex sounds?, and
- (2) What is the relationship between the way the ear analyzes a signal and the pitch produced by that signal?

Helmholtz' answer to question (1) was based on Ohm's (1843) acoustic law, which states that each sound is decomposed by the ear into a set of sinusoidal components of specific frequency and amplitude (Note 1.1). According to Helmholtz, this frequency analysis is performed in the inner ear by means of a large number of resonators (identified with the transverse fibers of the basilar membrane [Note 1.2]) tuned to different frequencies. The tuning of these resonators is arranged so that low frequencies stimulate the apical portion and high frequencies the basal region of the membrane. Nerve fibers connected to the resonators transmit their output to the cortex.

To deal with question (2), Helmholtz extended his frequency-to-place hypothesis in the following way: every frequency gives rise to a separate pitch. Because each frequency stimulates only a single resonator, and consequently, only a small number of nerve fibers, the pitch produced by any single frequency can ultimately be traced to a small population of neurons in the auditory pathway.

Thus, Helmholtz' "place" theory contains two separate hypotheses. The first attributes the ear's frequency analysis to a frequency-to-place transformation in the inner ear. The second links the pitch associated with each frequency to a specific place on the membrane. Although these hypotheses are carefully distinguished by Helmholtz, later statements of place theory often confused the two.

Helmholtz' model predicts that a tone composed of n harmonics will produce n distinct pitches, as long as each harmonic is independently analyzed ("resolved") by the ear. Helmholtz confirmed this by distinguishing, under the most careful listening conditions (Note 1.3), the first sixteen harmonics. However, he was also aware that under normal listening conditions only a single pitch - that of the fundamental - is heard. Presumably, the low pitch of a complex tone is based solely on the fundamental.

If the pitch of a complex tone stems from the fundamental, what happens to the pitch when the fundamental is reduced in amplitude or completely eliminated?

This experiment was performed by Seebeck (1841). He used an acoustic siren to produce a series of complex tones such as those reproduced in Figure 1.1. When the amplitude of the fundamental is negligibly small (signals C and D), does the pitch correspond to an octave above the fundamental (signal B), as Helmholtz would have predicted? The answer is no, for Seebeck determined that the pitch of signal D is the same as A, while signal C (with negligible energy at both the fundamental and the third harmonic) produces two basic pitches - one corresponding to the fundamental, the other to a pitch one octave above the fundamental (Note 1.4).

Seebeck's demonstration raised some serious doubts about the role of the fundamental in the perception of pitch. If the pitch of a complex tone does not stem from the fundamental, from where does it arise? Helmholtz, well aware of Seebeck's experiment, suggested that the "missing" fundamental was re-introduced by middle ear distortion.

Seebeck had another idea. Consider the waveform of a single frequency (Figure 1.2(a)). Its period of repetition is equal to the reciprocal of its frequency: $f_0 = 1/t$. When this same frequency is combined with its harmonics $2f_0$, $3f_0$, ... nf_0 , the sum results in a waveform with a period which is still equal to $1/f_0$ (Figure 1.2(b)). The period of the waveform does not change when the fundamental component is removed (Figure 1.2(c)). Seebeck suggested that it was the waveform period, not the fundamental, which determined a tone's pitch. Seebeck, however, did not say how the ear might translate the waveform periodicity into pitch.

With the development of physiological recording techniques, it became evident that nerve fibers innervating the peripheral musculature were capable of synchronizing to pulses of electrical stimulation (Helmholtz, 1868; Bernstein, 1871). This finding suggested to Wundt (1880) that acoustic stimulation might also induce nerve cells to respond periodically, in synchrony with a tone's basic frequency or period. Thus, the pitch of a complex tone would be based on the frequency or rate of neural discharge and not on the identity of an active set of nerve fibers, as place theory implies.

This "periodicity" theory, however, never achieved the popularity and prestige of Helmholtz' place model. In part, this was due to the fact that the theory required neural firing rates far exceeding the maximum rate possible for any single nerve fiber (Wever, 1949). Moreover, the theory assumed that the acoustic waveform underwent little, if any, modification in the cochlea (e.g. Rutherford's (1886) "telephone" theory). This aspect of the theory directly conflicted with Helmholtz' resonator model and made the general concept of pitch based on neural periodicity appear implausible. The association of place theory with frequency selectivity and the concomitant association of

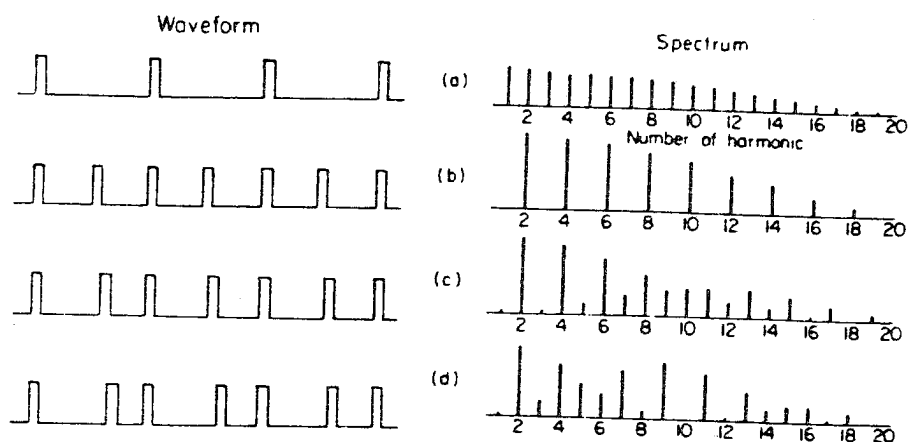


Figure 1.1 Seebeck's demonstration of the "missing" fundamental. The waveforms shown on the left were generated by means of an acoustic siren. Their spectra are shown on the right. Stimulus (b) produces a pitch one octave above that of (a), while stimuli (c) and (d) produce two distinct low pitches, one corresponding to that of signal (a), the other to signal (b). (From Schouten, 1940a. Reproduced by permission of the Koninklijke Nederlandse Akademie van Wetenschappen.)

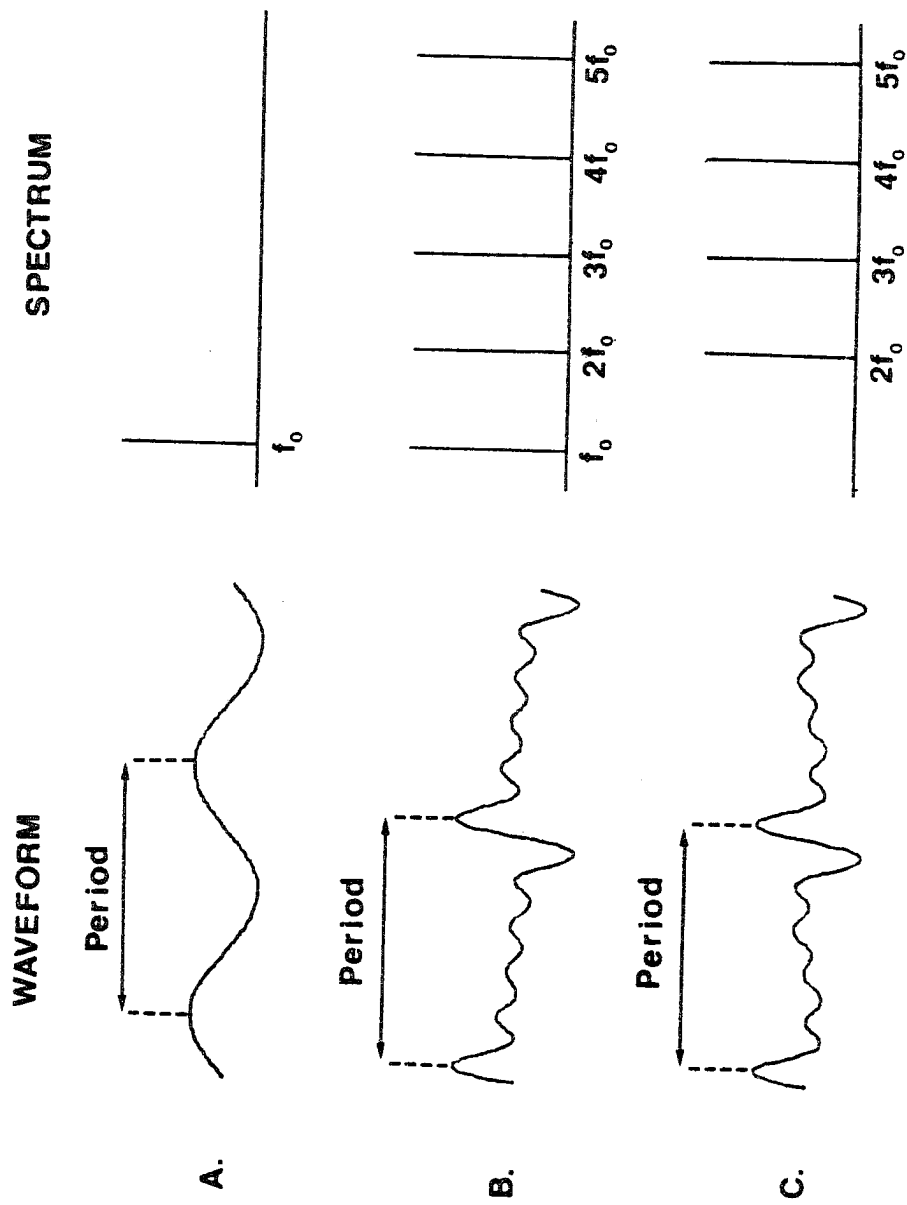


Figure 1.2 The waveform of a pure tone (a) is compared with the waveforms generated by two different complex tones. In (b), the tone is composed of the fundamental and its second through fifth harmonics. Signal (c) is identical to signal (b) except for the absence of the fundamental component. Notice that the waveform periodicity of the two complex tones is similar and that their period is equal to that of the pure tone.

periodicity theory with the absence of frequency analysis, was to profoundly shape the course of future debate on pitch perception. For, any evidence supporting the existence of peripheral frequency selectivity (and by 1930 the evidence was considerable - see von Békésy, 1960) was interpreted as support for Helmholtz' place theory of pitch.

Thus, by the early part of this century, the position of place theory appeared unassailable. Nerve fibers were known not to fire at rates greater than 1000 spikes/sec. Von Békésy (1928) confirmed that the basilar membrane does act as a frequency analyzer, although with much less precision than Helmholtz originally predicted. Psychophysical studies of the ear's ability to perceive more than one frequency at a time (Wegel and Lane, 1924; Fletcher, 1929) suggested that the first ten harmonics of a complex tone are analyzed by the ear as independent components. Consequently, it was assumed that the acoustic waveform would undergo a significant change in the ear due to cochlear frequency analysis, making it virtually impossible for nerve fibers to synchronize to its period.

1.22 THE RISE OF MODERN PERIODICITY THEORY: THE VOLLEY PRINCIPLE AND THE PITCH OF THE RESIDUE

However, within a few years periodicity theory would once more be taken seriously, as the result of two independent developments in auditory research. The first indicated that a group of nerve fibers could synchronize to frequencies much higher than any single fiber. Although the limit of neural periodicity was approximately 500 spikes/sec for single neurons, Wever and Bray (1930a;b) were able to record frequencies as high as 5 kHz with macroelectrodes in the auditory nerve. They concluded that the neural potential recorded represented the aggregate response of many nerve fibers, each locked in phase to the stimulating frequency (see Chapter 1.43). Although Wever and Bray's discovery served to establish the feasibility of neural periodicity as a means of frequency coding, it was still not entirely clear how the pitch of complex tones could be encoded by such means.

A mechanism by which the pitch of complex signals could be encoded in terms of the timing behavior of auditory neurons was proposed by Schouten (1940b). Schouten (1938) first set out to show that the pitch of a complex tone minus the fundamental does not depend on the presence of a distortion component. First, he verified that the pitch of a complex tone does not change upon cancellation of the fundamental. Then, a sinusoid differing slightly from the original fundamental was introduced. If the pitch of the original tone is due to the re-introduction of the fundamental through generation of combination tones produced by cubic ($2f_n - 2f_{n+1}$) or quadratic ($f_{n+1} - f_n$) distortion, a sinusoid 5 Hz higher than the original fundamental should create a sensation of beats at the difference frequency. No such sensation was heard. Schouten concluded that the fundamental was not a necessary cue for pitch perception.

If the pitch of a complex sound does not stem from the fundamental, where does it come from? Schouten's (1940b) proposal attempted to reconcile Ohm's acoustic law with Seebeck's demonstration of the "missing" fundamental. Consonant with place theory and von Békésy's observations, the lower (< 10) harmonics are analyzed as independent frequency components, capable of being perceived under the right conditions as separate tones. In a complex tone containing the fundamental, it is possible to distinguish two sources of low pitch. One, having a deep, hollow timbre, Schouten ascribed to the fundamental, presumably mediated by place mechanisms. The other, with a sharper timbre, was the "residue", derived from the interaction of the high, unresolved harmonics. When the fundamental is absent, the pitch of the complex is based entirely on the residue. The waveform produced by the interaction of the high harmonics has a basic period equal to, in most instances, the period of the fundamental. Schouten suggested that the residue is derived from the output of high frequency neurons synchronized to the amplitude modulation pattern of the cochlear waveform (Figure 1.3):

"For higher harmonics..., which in relative measure come closer to together, the receptors will respond with a complex waveform revealing the periodicity of the total waveform. The overlapping of the excitation curves of various harmonics necessarily leads to the introduction of the fundamental period in the response of the receptors.

"If thus the ear were to perceive pitch by virtue of the periodicity of the excitation of the receptors, the existence of the residue would be explained." Schouten (1940b: 994)

1.23 THE PITCH OF INHARMONIC SIGNALS

One of Schouten's own experiments (1940c) was somewhat at variance with his model. In this experiment, the frequency of each harmonic in a complex was shifted upwards by an additive constant. This upward shift alters the tonality of the signal but leaves the component spacing the same as before. For example, assume the original signal to be a harmonic series composed of integral multiples of 200 Hz. Each harmonic is raised by 50 Hz. The complex is now transformed into an inharmonic signal consisting of frequencies 250 Hz, 450 Hz, 650 Hz, etc. However, the amplitude modulation rate of the waveform is still 200 Hz (Figure 1.4). If the pitch is based on the waveform repetition rate, the pitch of the inharmonic complex will be the same as the original harmonic series. However, Schouten's observations indicated a small, but reliable change in pitch for the inharmonic signal.

In the mid-1950's, de Boer (1956a;b) studied the pitch of inharmonic complex tones in greater detail. Schouten's stimuli had consisted of 20 or more components. De Boer's contained only five or seven. As a consequence, de Boer was able to study the magnitude of the pitch shift in different regions of the spectrum. His findings

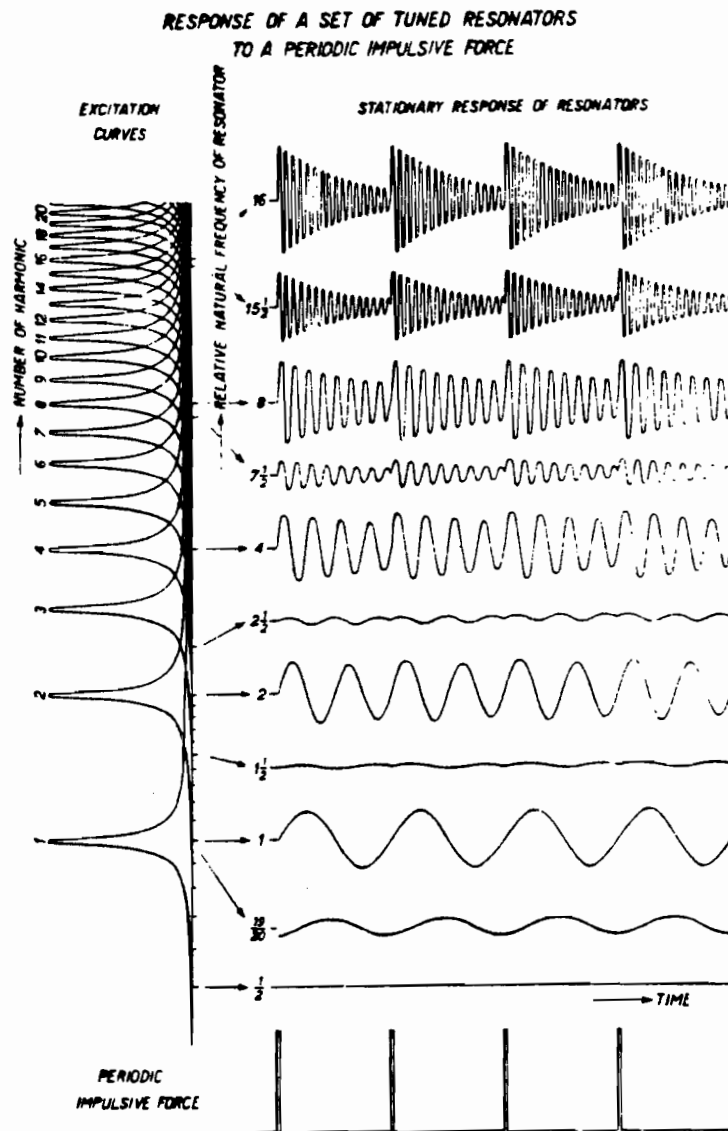


Figure 1.3 Schouten's model of the residue. The cochlea is portrayed as a series of narrowly-tuned bandpass filters whose output, in response to a pulse train, is a series of cochlear waveforms. The cochlear waveforms produced in response to the higher harmonics reproduce the periodicity of the original pulse train due to the interaction of harmonics not resolved by the cochlear filter. (From Schouten, 1940b. Reproduced by permission of of the Koninlijke Nederlandse Akademie van Wetenschappen.)

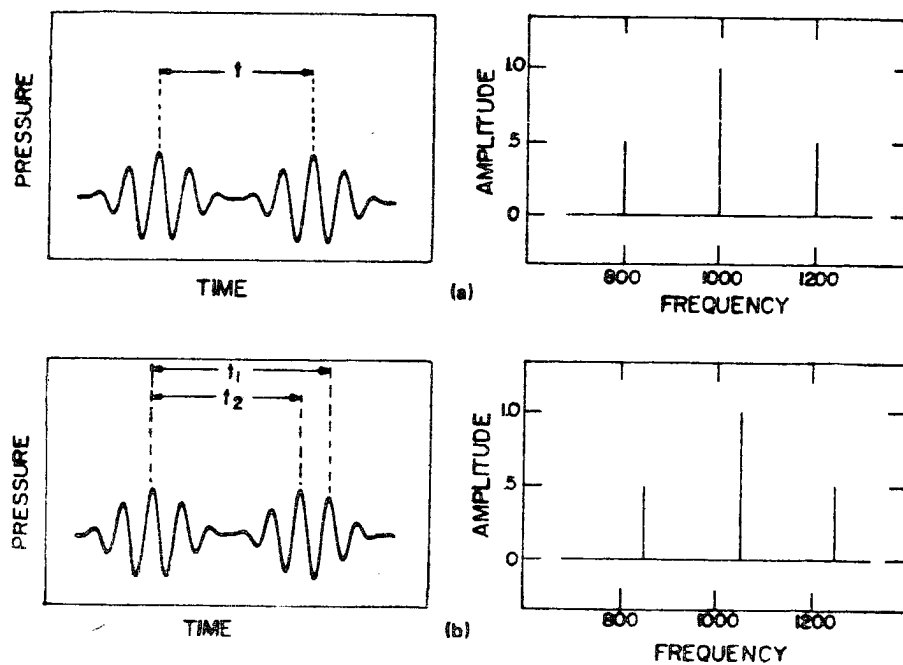


Figure 1.4 The waveform and spectrum of a harmonic signal (top) are compared with those of an inharmonic signal. Although the basic envelope periodicity is the same for the two signals, the fine time-structure (indicated by t), of the signals differ. The actual period of the inharmonic signal is, in this instance, 50 Hz. However, the pitch of the signal is substantially higher than 50 Hz. The pitch corresponds to the reciprocal of one (or both) of the intervals indicated. (From Green, 1976. Reproduced by permission of Lawrence Erlbaum Associates, Inc.)

indicated that the magnitude of the shift is roughly proportional to the harmonic rank of the signal's middle component (the "first effect" of the pitch shift), i.e. the pitch shift is greatest for tones composed of the lower, widely-spaced components. For example, a signal composed of the third through seventh harmonics of a 200 Hz fundamental will have a pitch of 200 Hz. The pitch of the same harmonic series, shifted up by 50 Hz is approximately 210 Hz. This relation can be expressed by the following equation:

$$(1) \quad p = F_c / n$$

where F_c is the frequency of the middle component, n is the component's rank number, and p is the fundamental frequency of the harmonic series to which the inharmonic signal is matched. This formulation implies that each component makes an equal contribution to the computation of the signal's pitch. However, de Boer's data were not entirely consistent with this assumption. The pitches of his inharmonic signals were often higher than predicted by equation (1). De Boer suggested that the computational procedure should be revised to place greater weight on the lower components (see Chapter 1.25).

The magnitude of the pitch shifts observed implied that some attribute of the signal other than the amplitude modulation rate of the waveform must be the basis of the residue. De Boer suggested that the residue generated through interaction of unresolved harmonics is the result of neural discharge activity synchronized to the cochlear waveform's "pseudo-period" - the interval between approximate repetitions of the waveform. The pseudo-period differs from the period of the waveform envelope in the following way: whereas the envelope periodicity is based on the amplitude modulation pattern of the waveform, the pseudo-period is based on the temporal fine-structure of the waveform (Figure 1.4). For a true harmonic series, where each frequency component is an integral multiple of the fundamental, the waveform envelope periodicity and the repetition rate of the fine-structure provide identical estimates of the signal's pitch. However, in an inharmonic series, the two estimates will differ. The envelope periodicity will always be equal to the period of the absolute difference between the component frequencies. The periodicity of the fine time-structure will be ambiguous, in that there are two intervals that describe the pseudo-period equally well. One interval is slightly less than the envelope period, the other slightly longer (Figure 1.4).

The residue generated by lower, minimally-interacting components, was attributed by de Boer to some form of template-matching operation in which the auditory system attempts to find a harmonic series (e.g. using a least-squares error criterion) most closely matching the frequencies of the inharmonic signal (see Chapter 1.32). De Boer showed that mathematically, these two computational procedures are equivalent.

1.24 MAPPING THE "EXISTENCE REGION" OF THE RESIDUE

De Boer's (1956a) study indicated that the residue is not confined to the higher, unresolved harmonics. His results implied the existence of two types of residue - a "classical" residue generated by narrowly-spaced components, and a second generated by the lower, widely-spaced harmonics. Schouten's model suggested that the strongest pitch comes from the narrowly-spaced harmonics. Yet Hoogland (1953) observed that the high harmonics of a 100 Hz fundamental fail to generate a residue when the components are above 3 kHz.

To obtain parametric data on the upper frequency limit of components capable of generating low pitch (the "existence region"), Ritsma (1962) used tones consisting of just three components. His results indicated that under the most favorable circumstances ($f_0 < 250$ Hz), the existence region only encompasses the first 20 to 30 harmonics. Under no circumstance did the existence region exceed 5 kHz.

Ritsma's study was not entirely consistent with Schouten's model. The residue was supposed to be exclusively a product of the high harmonics. However, for tones with high fundamentals (600-800 Hz), the existence region was composed entirely of harmonics which are not supposed to interact in the cochlea. Ritsma attempted to reconcile the disparity by attributing the upper limit to the inability of auditory neurons to synchronize to frequencies above 5 kHz (Wever and Bray, 1930a;b; Galambos and Davis, 1943; Tasaki, 1954) and by suggesting that the residue generated by the lower harmonics (< 6) is less salient than that produced by the less-resolved components. However, this implication was not entirely consistent with de Boer's (1956a) study. Although both the upper and lower components produce a definite sensation of pitch, the quantitative character of the pitch matches clearly implied that the lowest harmonics are more important. However, de Boer's results could be incorporated within Schouten's model given the assumption that different regions of the spectrum could contribute to the generation of the residue. The central question then becomes, which part of the spectrum is most important?

1.25 DOUBTS ABOUT PERIODICITY THEORY: THE DOMINANCE OF LOW HARMONICS

Ritsma (1967) and Plomp (1967) independently performed the crucial experiment. In both studies the pitch of a reference signal (a harmonic complex) was compared to that of a disjoint signal. This signal was composed of two different harmonic series differing in pitch (when presented in isolation) by 6-10%. The compound signal appeared to have only a single pitch. By having subjects decide whether the compound's pitch was higher or lower than the reference signal (whose pitch lay midway between the pitches of the disjoint signal), it was possible to determine which harmonics dominated the pitch of the compound signal. The results of their studies unequivocally indicated that the lower harmonics are more important than the high components. In fact, the pitch of the entire complex appeared to stem from just two components

whose level need be only 10 dB SL to exert their dominance. These components maintained their dominance even when they were as much as 40 dB below the level of the other components (Ritsma, 1967):

The harmonic rank of the "dominant" components tends to lower as the fundamental frequency of a tone increases. For a fundamental of 100 Hz, the dominant region lies between 400 and 500 Hz. For a fundamental of 250 Hz, the dominant harmonics are the third and fourth. And for a fundamental of 700 Hz, the dominant part of the spectrum lies close to 2 kHz (Plomp, 1976: Figure 46).

It is well known that the ear's ability to analyze concurrently-presented frequency components is limited. For a partial to be reliably distinguished from neighboring components, a minimum frequency separation of 15-20% is required (Plomp, 1964). Consequently, only the first five to seven harmonics of a complex are resolved as separate tones. Thus, Plomp's (1967) and Ritsma's (1967) studies implied that aural resolution of stimulus components is a necessary condition for generating a strong, well-defined pitch (Terhardt, 1970). However, if aural resolution were the determining factor, one would expect the fundamental to be dominant (see Chapter 1.31). This is hardly the case. The fundamental is dominant only for fundamentals greater than 1500 Hz (Plomp, 1967).

1.26 THE RISE OF MODERN PLACE THEORY: THE "INTERNAL SPECTRUM" AND THE ROLE OF COMBINATION TONES

The studies of Ritsma (1967), Plomp (1967), and others have shown that the unresolved portion of the spectrum plays a subordinate role in the formation of the residue. Yet, earlier studies (e.g. Schouten, 1940a; de Boer, 1956a; Schouten et al, 1962) had implied that the residue produced by this region of the spectrum was not inconsiderable. The first clue to the resolution of this paradox was uncovered by de Boer (1956a,b), who found that the pitch shift produced by high components was significantly larger than would be predicted on the basis of their harmonic rank. De Boer observed that the deviation was much smaller if one assumed that the pitch stemmed principally from the lowest two of the signal's five (or seven) components (Chapter 1.25). This "second effect" of the pitch shift was confirmed by Schouten et al (1962) using three-component signals. However, the magnitude of the "second effect" was even larger than observed by de Boer. Curiously, the magnitude of the observed pitch shifts could only be related to de Boer's "first effect" if it were assumed that the pitch was derived from a part of the spectrum below the lowest stimulus component.

Goldstein's (1967) study of auditory nonlinearity provided a second clue. For many years, it was assumed that the level of all nonlinear distortion products was significantly below the level of the primaries and that, for low to moderate signal intensities, they could be effectively ignored. Goldstein found, however, that the level of the odd-order combination tones ($f_1 - n(f_2 - f_1)$) could be within 10 to 15 dB of the primaries when $f_2/f_1 = 1.1$ (Note 1.5).

The "second effect" was finally explained by Smoorenburg (1970). Using two-component signals, he demonstrated that the odd-order combination tone series can effectively extend the spectrum of the original signal down several harmonics when $1.07 < F_2/F_1 < 1.10$ and that the magnitude of the "second effect" could be directly related to the presence of this "internal spectrum." Furthermore, he was able to infer, by calculating the effective harmonic rank producing the pitch shift, that the existence region of the residue encompasses only the first eight harmonics. Higher harmonics can produce a residue only by virtue of their combination tone products.

Smoorenburg's (1970) study indicated that the low pitch of complex tones stems entirely from aurally-resolved harmonics. This finding was consistent with place models of pitch, in that only resolved components would be expected to produce distinct peaks in the spatial excitation pattern (see Chapter 1.31). Moreover, Smoorenburg's results appeared to be inconsistent with temporal fine-structure theories, since aurally-resolved components should not interact in the cochlea and would therefore not provide sufficient information about the signal's period to cochlear nerve fibers.

1.27 PHASE EFFECTS AND THE TEMPORAL FINE-STRUCTURE OF THE WAVEFORM

It will be recalled that de Boer (1956a) explained the "first effect" of the pitch shift in terms of two different mechanisms. For high harmonics, the pitch would be based on nerve spikes synchronized to the temporal fine-structure of the waveform. Because widely-spaced harmonics would not be expected to significantly interact in the cochlea, de Boer suggested that their pitch is based on a harmonic template matching operation that minimizes the error between the template and the spectrum of the signal.

Schouten et al (1962) found, using three-component stimuli (harmonic ranks > 5), that each signal was capable of being matched to at least two different pitches and that the values of these pitches correspond to the closest approximations of the fine time-structure to the waveform periodicity.

Now, if the residue is based on the temporal fine-structure, changes in its apparent periodicity should affect both the salience and value of the signal's pitch. The easiest way to induce such changes without affecting the spectrum is through manipulation of the relative phase configuration of the stimulus components. The phase configuration determines the modulation depth of the envelope periodicity. The clearest fine structure periodicity is observed when all components begin in the same phase. A random starting phase configuration results in a much less well-defined periodicity (Figure 1.5). Consequently, the random phase signal should produce a much weaker sensation of pitch.

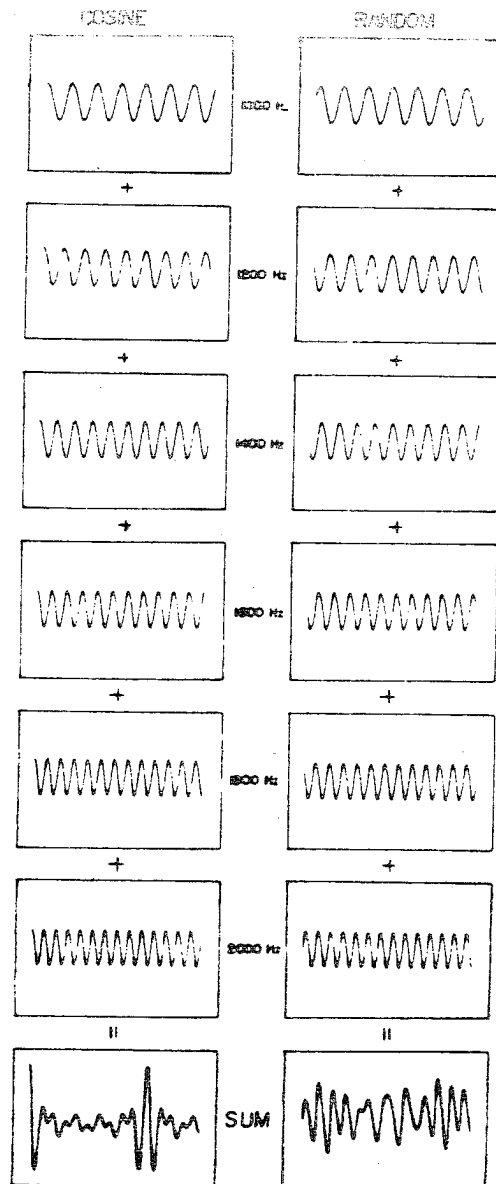


Figure 1.5 The relationship between the phase configuration of the stimulus components and the waveform envelope modulation pattern. When all stimulus components begin in the same phase (phase-coherent), the waveform period is clearly marked. The same stimulus components added in random phase configuration produce a waveform whose periodicity is not as easily discerned. (From Green, 1976. Reproduced by permission of Lawrence Erlbaum Associates, Inc.)

Consistent with this fine-structure hypothesis, a number of investigators have reported that the pitch of a complex signal is sensitive to changes in the phase relation among components (e.g. Mathes and Miller, 1947; Licklider, 1955; Ritsma and Engel, 1964). Licklider, for example, observed that when the phase relationship among 16 components was randomly assigned, the signal's pitch practically disappears. Ritsma and Engel (1964) found that pitch matches to three-component signals (consisting of the ninth, tenth and eleventh harmonics of a common fundamental) with F1 and F3 in cosine phase, and F2 in sine phase were more variable and less reliable than matches to phase-coherent signals.

However, most attempts to replicate these findings have been unsuccessful. Neither Patterson (1973) nor Wightman (1973a) found significant changes of pitch in phase-scrambled signals. De Boer (1956a) observed that, although the pitch of a phase-coherent signal sounds slightly more salient than that derived from phase-scrambled tones, the pitch is unaffected. Consistent with de Boer's observations, Bilsen (1973) reports that the pitch strength (but not its value) of three-component signals is only significantly affected by the phase configuration when all three harmonics are above the eighth. Bilsen suggested that the presence of phase effects for high harmonics is attributable to combination tones. In support of this hypothesis, Buunen *et al* (1974) found a significant correlation between the strength of a signal's pitch and the cancellation level of its combination tones.

These studies, along with Smoorenburg's (1970) account of the "second effect", imply that the pitch of complex tones is never based on the temporal fine-structure. We shall return to this point below, after considering other evidence that appears to support the same conclusion.

1.28 CENTRAL ORIGIN OF PITCH: DOUBTS ABOUT THE RELEVANCE OF THE TEMPORAL FINE-STRUCTURE

Traditionally, place and periodicity theory have focussed on the role of cochlear mechanisms in the formation of pitch. This emphasis on peripheral processes accounts for much of the concern with waveform envelope, temporal fine-structure, and the like. However, a number of studies suggest that under certain circumstances, a sensation of pitch may occur in the absence of any form of peripheral interaction among adjacent components.

Cramer and Huggins (1958) were perhaps the first to systematically explore the relationship of the binaural system to the perception of pitch. They presented a white noise signal to one ear and presented the identical signal to the other ear with a variable delay. This dichotic signal produced a faint pitch equivalent to the reciprocal of the interaural delay.

The relevance of binaural interaction to the perception of the residue is perhaps most clearly delineated in a study by Houtsma and Goldstein (1972). Musically-trained subjects identified musical intervals produced by the residue of two successive pairs of components. The harmonic rank of the components varied from the second to the sixteenth. In one condition, both components were presented to the same ear (monotic presentation). In the other, a single component was played to each ear (dichotic presentation), at a sensation level sufficiently low to preclude bone-conducted peripheral interaction (40-50 dB SPL). Interestingly, performance level was about the same in both conditions (after compensating for the effects of combination tones). In both conditions, musical interval recognition was best for low harmonics, diminishing with increasing harmonic rank (Houtsma and Goldstein, 1972: Figure 1.3). Bilsen (1973), using a masking paradigm, also found the strength of the dichotically-induced residue to be comparable to monotically-generated pitch.

The existence of a dichotically-mediated pitch raises some fundamental questions about the role of peripheral auditory mechanisms in the processing of pitch. For example, if the dominance of low harmonics reflects the limited frequency resolution of the auditory periphery, why are the low harmonics dominant for dichotic pitch as well?

1.29 SUMMARY: THE MAJOR FEATURES OF LOW PITCH

Despite nearly a century and a half of intensive investigation, the physiological and psychological bases of pitch perception remain obscure. Nineteenth century models assumed a rather direct relationship between certain attributes of the signal (such as the fundamental or waveform envelope) and the sensation of pitch. Over the years, it has become evident that the pitch of a complex signal does not stem directly from any single physical attribute of the signal, but rather reflects a multi-stage process, involving both a peripheral stage of frequency analysis and a more central stage of perceptual synthesis. Below, we review some of the features of low pitch that have led to this change in perspective.

(a) The Pitch of Complex Tones does not Depend on the Fundamental Component

For many years, the pitch of complex tones was thought to stem exclusively from the lowest (fundamental) component in the spectrum (Helmholtz, 1863; von Békésy, 1972). This view proved untenable as a consequence of Schouten's (1938) demonstration that a tone's pitch does not change upon cancellation of the fundamental.

(b) The "Missing" Fundamental is not Attributable to Combination Tones Produced by Intracochlear Distortion

The sensation of low pitch in the absence of the fundamental component was attributed by Helmholtz (1863) and Fletcher (1924) to the presence of combination tones produced by middle ear or intracochlear

distortion. This "distortion" hypothesis has proven untenable in light of the following studies. (1) The intensity level of the first-order difference tone ($f_{n+1} - f_n$) and relevant odd-order combination tones $n(f_{n+1} - f_n)$ is at least 30-40 dB below that of the primaries, as measured by either psychophysical (von Békésy, 1934; Plomp, 1965; and Goldstein, 1967; 1970) or physiological (Smoorenburg et al, 1976; Buunen and Rhode, 1978; Goldstein and Kiang, 1968) methods. If the low pitch of a complex tone was based on combination tones generated by intracochlear distortion, one would predict a change in pitch as the signal intensity of the complex is lowered to a level where the combination tones are inaudible. However, this is not the case. A sensation of low pitch is clearly discernable at signal levels at least 20-30 dB below the threshold of the relevant combination tones (Thurlow and Small, 1955; Ritsma, 1967) and does not appreciably change with increasing signal intensity. (2) If the "missing" fundamental were re-introduced through intracochlear distortion, the introduction of a sinusoid differing slightly from the fundamental, should produce a sensation of beats equal to the difference frequency. Schouten (1938), demonstrated that no such sensation occurs. (3) If the low pitch of complex signals were the product of intracochlear distortion, the percept of the "missing" fundamental should be altered in the presence of a masking stimulus whose energy is centered at the fundamental. Licklider (1954) and Thurlow and Small (1955) have shown, however, that the presence of low-frequency masking noise has no appreciable effect on the low pitch of a complex tone.

(c) Spectral Limits of "Low" Pitch

Schouten (1940b) suggested that the pitch of a complex tone is produced by the interaction of the high harmonics. Subsequent research was to show that this region of the spectrum produces a very weak pitch ("rattle" pitch - Rosenberg, 1965; or "roughness" - Terhardt, 1970) that differs both quantitatively and qualitatively from the more salient pitch of low harmonics. Under no circumstance does the "existence" region of low pitch extend above 5 kHz (Ritsma, 1962; 1963). Nor does a low pitch emanate from harmonics above the eighth, except through the contribution of combination tones (Smoorenburg, 1970; Buunen et al, 1974).

(d) The Dominance of Low Harmonics

The pitch of a broadband signal is based on a very narrow region of the spectrum, comprising as few as two components (Ritsma, 1967). This region of "dominance" varies with fundamental frequency as follows (Plomp, 1976): (a) $f_o = 100-200$ Hz: Dominant region = 400-800 Hz: Dominant harmonics = 4-5 (b) $f_o = 200-400$ Hz: Dominant region = 600-1200 Hz: Dominant harmonics = 3-4 (c) $f_o = 400-1000$ Hz: Dominant region = 800-2000 Hz: Dominant harmonics = 2-3. For signals comprising components outside the region of dominance, the pitch is determined by those harmonics nearest the dominance region (Patterson and Wightman, 1976).

(e) Saliency of Low Pitch: Fundamental Frequency

The saliency of low pitch varies as a function of fundamental frequency. The strongest sensation of pitch is produced by a signal with a fundamental of approximately 250 Hz (Plomp, 1975; Bilsen and Ritsma, 1970). Less salient, but still robust pitches are generated by fundamentals spanning an octave above and below this value. Low pitch diminishes considerably in strength for fundamentals less than 100 Hz and greater than 500 Hz. A true low pitch probably does not occur when F_0 is less than 50 Hz or greater than 1000 Hz. For very low fundamentals, the sensation is probably based on amplitude modulation characteristics of the signal (Buunen, 1975). When the fundamental is higher than 1 kHz, the pitch stems from the fundamental component itself (Plomp, 1967).

(f) Saliency of Low Pitch: Number of Components

The saliency of low pitch increases with the number of frequency components. Broadband complexes with many harmonics (e.g. vowels, musical tones) produce a very definite sense of pitch, whereas two-component signals produce a much more ambiguous percept (Plomp, 1976; Evans, 1978). However, the strength of the percept (measured in terms of musical interval recognition (Houtsma and Goldstein, 1972) or masking level (Bilsen, 1973) appears to be determined by the dominance principle and is independent of the number of components (Bilsen, 1973).

(g) Pitch Saliency as a Function of Modulation Depth

Ritsma (1962) found that the upper limit of the "existence" region is dependent on stimulus modulation depth. A modulation depth of 20% (carrier frequency $[F_2]$ is 20 dB SPL above the level of the sidebands $[L_1$ and $L_3]$) effectively halves the existence region. Interestingly, the most salient pitch sensation is produced for a modulation depth of 115% (F_2 5 dB above F_1 and F_3) (Ritsma, 1962).

(h) Phase-Insensitivity of Low Pitch

The pitch of low harmonics (< harmonic 9) is impervious to alterations of the signal's phase configuration (Patterson, 1973; Bilsen, 1973; Buunen *et al.*, 1974). Buunen *et al.* (1974) have shown that previous reports of phase effects are probably due to the presence of combination tones (however, see Moore, 1977).

(i) The Pitch of a Complex is not Based on the Waveform Modulation Frequency

Schouten (1940c) showed that a tone's pitch is not based on the amplitude modulation rate of the stimulus waveform. If it were, raising the frequency of all components by the same additive constant should not alter the signal's pitch (since the difference frequency, and hence the waveform envelope modulation frequency, remains unchanged). Schouten found (subsequently verified by de Boer, 1956a,b; Schouten *et al.*, 1962; Smoorenburg, 1970 and others) that the resultant pitch is slightly lower than or slightly higher than the envelope repetition frequency.

(j) The Low Pitch of Complex Tones is Ambiguous

Schouten et al (1962) have shown that a complex tone is capable of generating more than a single low pitch. This pitch ambiguity is most apparent for inharmonic signals, but is also evident for tones composed solely of high harmonics.

(k) Intensity-Independence of the Dominance Principle

Dominant components determine the pitch of a complex signal even when they are as much as 40 dB below the level of other components. They need only be 10 dB above threshold to exert their dominance (Ritsma, 1967).

(l) Low Pitch in the Absence of Peripheral Interaction

A sensation akin to low pitch may arise when only a single frequency component is presented to each ear. Musically-experienced subjects are capable of recognizing musical intervals based on this "central" pitch. However, these studies should be evaluated with considerable caution. Caution is warranted because of the uncertainty that the sort of information required to successfully identify musical intervals (or a 6% difference in pitch [Bilsen, 1973]) derived from two components signals is comparable to other attributes of low pitch. For example, neither dichotic paradigm required a pitch discrimination finer than 6%. This criterion is comparable to the difference limen for signals producing a very faint pitch, such as modulated noise, and is over an order of magnitude greater than the DL for complex tones (0.1-0.2% - Ritsma and Hoekstra, 1974). Burns and Viemeister (1976) have shown that modulated noise can be used successfully in a musical interval recognition paradigm when subjects have the appropriate musical background and psychophysical training.

1.3 PATTERN RECOGNITION MODELS OF PITCH PERCEPTION

In light of modern psychophysical and physiological research, neither place nor periodicity theory, in their classical form, provides an adequate account of the major features of low pitch. In their place stand models which view pitch perception less deterministically, and more as a probabilistic process akin to visual pattern recognition. In contrast to earlier models, pitch perception is viewed as a multi-stage process, involving both a peripheral stage of frequency analysis and a more central stage of perceptual synthesis.

The present discussion focusses on two contemporary models of pitch - the "pattern transformation" model of Wightman (1973b) and the "optimum processor" model of Goldstein (1973; 1978). These models share certain assumptions. Both assume that the central pitch processor operates exclusively on the spectrum of the signal and disregards whatever component interaction might be reflected in the temporal discharge pattern. This means that the central processor is effectively phase-insensitive. Moreover, the central processor favors well-resolved components. Due to limitations of peripheral auditory resolution,

components separated by less than 10% do not produce distinct peaks in the excitation pattern and are consequently ignored by the central processor.

Each model also assumes a certain degree of indeterminacy in the analysis of the signal's spectrum. This indeterminacy in aural frequency measurement is important, for it ultimately determines the number and strength of the pitches extracted from the peripheral auditory representation.

The models differ in a number of important ways, both with respect to the form of the peripheral auditory representation and the nature of the operations performed by the central pitch processor. Wightman's model is essentially a "place" model of pitch. The peripheral representation of the spectrum is assumed to be tonotopically organized, with component frequencies marked by a local maximum in the neural discharge rate. Goldstein's model, on the other hand, assumes that the signal spectrum is peripherally encoded in terms of the temporal discharge pattern of auditory neurons phase-locked to the component frequencies (Goldstein, 1978). The two models also offer different computational means of extracting the pitch from the peripheral representation. In Wightman's model, the process is basically passive, based on the Fourier transform of the peripheral excitation pattern. The central processor of Goldstein's model operates in a more active manner, seeking to supply a harmonic series that provides the closest fit to a stochastic representation of the signal spectrum.

1.31 PATTERN TRANSFORMATION MODEL

The conceptual basis of Wightman's model is borrowed from a classic problem in visual pattern recognition. The letter "a" is recognized as the same letter as "A", despite the difference in physical form. The problem is similar for models of pitch perception. Two tones may evoke the same pitch despite differences in their spectra. Wightman's model attempts to account for the perceptual invariance by specifying a series of transformations through which dissimilar spectra ultimately generate similar pitches.

The first stage of the model incorporates a frequency analysis of limited precision. This spectrum analysis is represented in terms of the spatial excitation pattern of peripheral neural activity ("peripheral activity pattern"). The frequency of each component is marked by a "peak" in the discharge pattern of the tonotopically-organized neural array. Due to limited cochlear resolution, components separated by less than 10% do not generate separate peaks in the peripheral activity pattern. Consequently, if the original signal consists of the first twelve harmonics of a common fundamental, only the lowest ten harmonics will evoke peaks in the peripheral activity pattern.

The peaks of the peripheral activity pattern serve as the input to a central pitch processor. The central processor performs an operation analogous to a Fourier transform on the peripheral activity pattern. Instead of the conventional Fourier transform performed in the time domain, the central processor performs the analysis in the frequency domain (which is mathematically equivalent to the autocorrelation function of the original signal). The result of this operation is another transform whose major peak is assumed to be the primary pitch of the signal.

The pattern transformation model incorporates several interesting features. For example, the "clarity" of pitch is represented in terms of the height of the peak of the final Fourier transform. This parameter is a function of two basic variables; (a) the number of frequency components and (b) the resolution of the peaks in the peripheral activity pattern. The model consequently makes three rather straight-forward predictions; (1) the pitch strength of a signal is directly correlated with the number of its constituent components (separated by more than 10%) (Wightman, 1973b: Figure 6), (2) the strongest pitch will be produced by the lowest (most widely-spaced) harmonics (ibid: Figure 5), and (3) pitch strength is proportional to fundamental frequency (ibid: Figure 7).

Unfortunately, none of these predictions is empirically supportable. Each prediction is considered in turn:

(1) Pitch strength is proportional to the number of (resolved) components. Although the subjective evaluation of pitch prominence ("pitch salience") increases proportional to the number of components (Fastl and Stoll, 1979), the actual strength of the pitch (as measured in terms of musical interval recognition or accuracy of pitch discrimination) is determined by the harmonic rank of the tonal components. Thus, a two-component tone, whose harmonics lie within the dominance region, generates a pitch equal in strength to a broadband signal composed of twenty or more harmonics (Bilsen, 1973). Moreover, the pitch generated by a two-component signal may, under certain circumstances, be even stronger than that produced by a broadband harmonic complex (see Chapters 1.25 and 1.29 (d); Ritsma, 1967).

(2) The strongest pitch is generated by the most widely-spaced components. This prediction is an inevitable consequence of a place-based model of frequency analysis. However, Ritsma (1967), Plomp (1967), Bilsen (1973), and Houtsma and Goldstein (1972) have all shown that the lowest two harmonics are rarely dominant for pitch, as would be predicted by Wightman's model (Chapter 1.29 (d)).

In a more recent article, Patterson and Wightman (1976) attempt to reconcile this disparity in the following way. The lowest harmonics usually occur in a region of reduced sensitivity (up to 15-20 dB < 1 kHz). Consequently, they are not weighted as heavily by the central pitch processor. This proposed explanation, however, is inconsistent with Ritsma's (1967) demonstration that components in the dominant

region continue to determine the pitch of the complex when they are as much as 40 dB below the level of the other components (Chapter 1.29 (e)).

(3) The strength of pitch increases proportional to the tone's fundamental frequency. This is true only for fundamentals between 50 and 250 Hz. Above 250 Hz, the strength of the pitch actually declines with increasing fundamental frequency, becoming negligible for fundamentals above 1 kHz (Chapter 1.29 (f); Plomp, 1976).

However, other features of Wightman's model are more consistent with psychophysical data. For example, his model predicts rather well the pitch evoked by six- and twelve-component inharmonic tones (the discrepancies between model and data for high harmonics is easily corrected for when combination tones are taken into account). In sum, Wightman's model performs well in terms of predicting the basic pitch value of a stimulus, but fails to accurately predict the relationship between harmonic rank, fundamental frequency, and pitch strength.

1.32 OPTIMUM PROCESSOR MODEL

In Goldstein's model, the auditory system is conceived of as a statistical communication receiver specialized for measuring temporal periods of periodic sounds such as speech and music. The system is designed to function as an "optimal processor" making the best match possible between an internally-stored template comprising a perfect harmonic series and the "noisy" frequency measurements of the signal spectrum.

The initial stage of processing involves a frequency analysis of limited precision. However, in contrast to Wightman's model, the limited precision of aural frequency analysis is not based on the system's peripheral resolution, but rather involves basic constraints on the precision of neural synchronization. Consequently, there is a certain degree of indeterminacy in the frequency representation of each component. This stochastic quality of the peripheral spectrum analysis requires that each component be described in terms of its mean and variance. As in Wightman's model, the peripheral spectrum analysis disregards phase and amplitude information. Only information concerning the component frequencies is conveyed to the central processor.

The central processor "reads" the spectrum of the signal, scanning across the tonotopic array. It estimates the frequency of each component and uses this information to find a harmonic series that optimally fits the estimated spectrum (using a minimum-square error criterion). The signal's pitch is equivalent to the template's lowest component. In determining the harmonic series that best matches the input spectrum, the central processor assigns a harmonic rank to each component. Theoretically, the harmonic rank of only one component need be determined as the processor assumes that the input spectrum is comprised of successive harmonics (Note 1.6).

If the variance of the spectral information is small, the probability of correctly estimating the harmonic rank of the signal components is high. Under this circumstance the central processor will select a harmonic series very close to the original signal. However, as the variance increases, so does the probability of misaligning the harmonic template. When this occurs, the central processor may come up with a very different estimate of the signal's pitch (Figure 1.6).

Goldstein's model implies that ambiguity is inherent in the process of pitch extraction. This ambiguity is not particularly apparent for components of low harmonic rank. However, the probability of misaligning the central template increases significantly for components above the fourth or fifth harmonic. This is true even when the measurement variance of the components is small. This ambiguity is evident in an experiment of Schouten et al (1962). Observers listening to a tone composed of the ninth, tenth, and eleventh harmonics of a 199 Hz fundamental were able to pick out "secondary" pitches in addition to the true fundamental. The value of these secondary pitches deviated from the true fundamental by an amount that is consistent with Goldstein's model (Note 1.7). According to Goldstein's model, the optimum processor treats all periodic signals, including inharmonic tones, as if they were perfect harmonic series. An inharmonic signal is merely assumed to be a harmonic series with a large degree of variance in aural frequency measurement. This means that the pitch of an inharmonic signal will be significantly more ambiguous than the pitch of a harmonic series.

Goldstein's model also predicts that the strength of a signal's pitch is inversely proportional to the amount of ambiguity concerning its pitch. This relationship between pitch strength and pitch ambiguity has a very interesting implication. Because the amount of ambiguity is determined by the variance of the peripheral frequency analysis, and because the strength of the pitch is dependent on the degree of ambiguity, the strength of a tone's low pitch will be directly related to the precision with which the signal spectrum is analyzed by the ear.

Both the existence region and the region of dominance can be explained in such terms. The upper limit of the existence region is the natural consequence of (a) deterioration of the precision of aural frequency analysis (and neural synchronization) above 3 kHz, (b) the increasing probability of template matching errors for components of intermediate and high harmonic rank, and (c) the restriction of peripheral frequency analysis to components separated by more than 10%. The region of dominance reflects (a) the inherently greater precision of peripheral frequency analysis (and neural phase-locking) for frequencies between 800-2000 Hz, and (b) the relatively small probability of template misalignment for low harmonics.

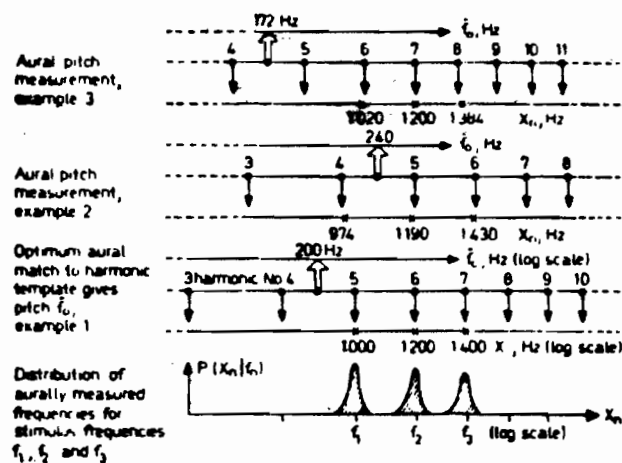


Figure 1.6 Goldstein's optimum processor model. The bottom trace represents the distribution of aural frequency measurement performed in the auditory periphery. Three examples indicate how the optimum processor responds to variability in the aural frequency analysis of the signal spectrum. Misalignment of the harmonic template (examples 2 and 3) results in large errors in estimation of the signal periodicity. (From Goldstein, 1978. Reproduced by permission of the International Society of Audiology and Dr. Julius Goldstein)

1.4 PHYSIOLOGICAL MECHANISMS UNDERLYING FREQUENCY CODING

Most of our knowledge concerning pitch perception is based on psychophysical evidence, as reviewed in Chapters 1.21-1.29. However, it is obvious that the perception of pitch is based on physiological mechanisms of great power and complexity. Unfortunately, very little is known about the physiology of pitch perception. In Chapters 1.41-1.43, some basic auditory anatomy and physiology relevant to the auditory processing of pitch is reviewed. The discussion focusses on two alternative mechanisms by which pitch might be encoded in the auditory system, one based on the spatial pattern of neural excitation, the other based on the periodic activity of auditory neurons.

1.41 ANATOMY OF THE AUDITORY PATHWAY

Figure 1.7 illustrates the primary anatomical pathways of the auditory system. This pathway is divided into four primary zones. The first, encompassing the cochlea and cochlear nerve is commonly referred to as the auditory periphery. The brainstem pathway, composed of the cochlear nucleus, trapezoid body, superior olivary complex, lateral lemniscus, and inferior colliculus, constitutes the second. Directly above the brainstem is the medial geniculate body (auditory nucleus of the thalamus). The pathway proper terminates in the fourth zone, the auditory cortex. Although this study focusses on auditory activity at the level of the brainstem, it should always be kept in mind that the perception of complex sounds involves processing at all levels of the auditory system.

(a) Auditory Periphery

The displacement of the basilar membrane is reflected in the excitation pattern of the cochlear nerve (also known as the auditory nerve, acoustic nerve or VIIIth nerve). In Homo sapiens, the nerve consists of approximately 30,000 fibers. Fibers innervating the basal-most turn of the cochlea are most sensitive ("tuned") to frequencies at the upper end of the audible spectrum (10-20 kHz). Fibers innervating progressively more apical regions of the cochlea are tuned to progressively lower frequencies.

(b) Auditory Brainstem

The afferent fibers of the cochlear nerve project to the ipsilateral cochlear nuclei in two separate branches. The anterior branch terminates in the anteroventral cochlear nucleus (AVCN) and the posterior branch terminates in the posteroventral (PVCN) and dorsal (DCN) nuclei. The tonotopic organization of the cochlear nerve is preserved in its projection to the cochlear nuclei and is, in fact, maintained throughout the ascending pathway (Morest, 1975). Cochlear nerve fibers innervating basal cochlea, and representing high frequencies (>5 kHz), project dorsomedially, while more apical, low-frequency fibers project ventrolaterally.

The primary afferent path leads from the ventral cochlear nuclei via the intermediate and ventral (trapezoid body) acoustic striae, to

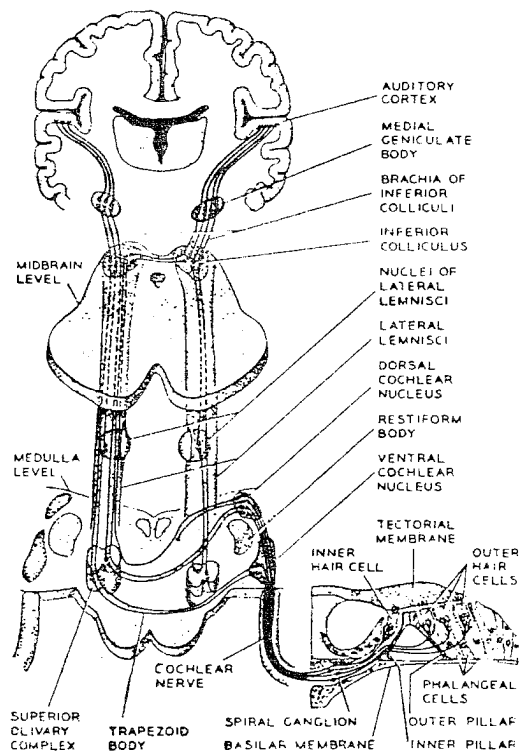


Figure 1.7 The primary anatomical structures and connections of the ascending auditory pathway in the human. Diagram begins at the cochlear level and terminates in the auditory cortex. (Adapted from Flanagan, 1972).

the ipsilateral and contralateral superior olivary nuclei. These nuclei represent the first loci of binaural interaction in the auditory pathway. Fibers originating from the ipsilateral ventral cochlear nucleus (VCN) synapse with lateral dendrites, while axons projecting from the contralateral VCN innervate medial dendrites of the same cell (Stotler, 1953). The superior olivary complex contains two major nuclei. The lateral superior olive (LSO), receiving projections from high and intermediate frequency fibers of the VCN, is thought to be specialized for localization of sound based on interaural intensity cues. The medial superior olive (MSO), innervated primarily by low-frequency fibers, specializes in processing of interaural time differences (Erulkar, 1972).

From MSO, the pathway courses bilaterally, via the lateral lemniscus, to terminate in either the dorsal nucleus of the lateral lemniscus (DNLL) or one of the nuclei of the inferior colliculus (IC). Fibers from the LSO travel via the lateral division of the lateral lemniscus to either ipsilateral DNLL or IC. The IC is composed of several nuclei, the largest being the central nucleus (CNIC).

There is another pathway, which connects the DCN with contralateral DNLL and IC, via the medial division of the lateral lemniscus. A small number of fibers innervating VCN cross the midline and project to contralateral VNLL, DNLL, and IC.

(c) Thalamus and Auditory Cortex

Ascending neurons project from the CNIC, via the brachium of the IC, to the medial geniculate body (MGB) in the thalamus. From the MGB the pathway terminates in auditory cortex (principally the primary cortical area, A1).

1.42 PHYSIOLOGICAL CORRELATES OF "PLACE" CODING

Auditory neurons transmit spectral information by means of their discharge activity. Although the analysis of sound is based upon the activity of many thousands of cells, we shall first consider the response patterns of individual neural units. The auditory system represents frequency in two distinct ways.

As a result of its longitudinal stiffness gradient, the basilar membrane responds differentially along its length, to stimuli of different frequencies. The membrane displacement is greatest at its base for high frequencies and at the apex for low frequencies. Consequently, cochlear nerve fibers innervating hair cells in the basal cochlea will be most sensitive to high frequencies, and fibers innervating apical cochlea will be most responsive to low frequencies. Each cell has a limited range of frequencies to which it will respond at any given intensity. At very low intensities, near threshold, only its characteristic frequency (CF) will excite it. As the signal level is raised, the cell will respond to an increasing range of frequencies. The response area of a unit is traditionally described in terms of its frequency threshold curve (FTC) (Figure 1.8). The FTC indicates the

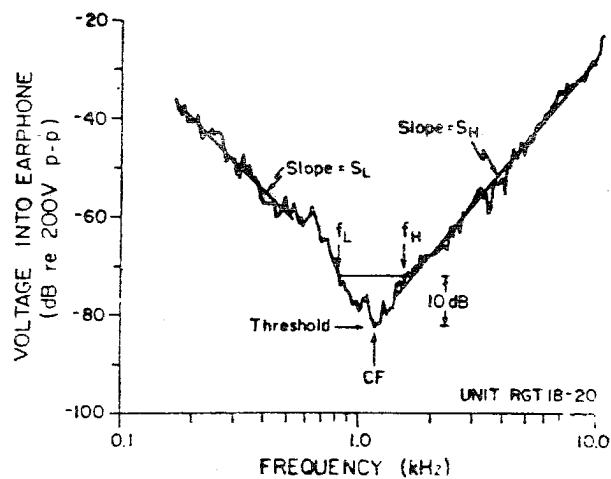


Figure 1.8 A frequency threshold curve of a cochlear nerve fiber. Illustrated are several parameters of the FTC, including its tip (threshold) at the characteristic frequency, the low- and high-frequency slopes, and the bandwidth of the tuning curve. (From Weiss et al., 1976. Reproduced by permission of Elsevier/North Holland Publishing Co.)

signal level required to excite the unit (threshold = 20% above spontaneous rate) as a function of frequency. Although the response area may span several octaves at high signal levels, each cell is still most responsive to a considerably narrower range of frequencies. This is because the unit discharge rate will always be greatest for frequencies lying in the center of its response area. However, for intensities significantly above saturation level, the unit will respond about equally well to a broad range of frequencies.

This aspect of a unit's frequency selectivity is perhaps more clearly illustrated in terms of an iso-intensity curve (Figure 9). At low to moderate intensities, there will always be a single frequency that excites the cell at a higher rate than other frequencies. At higher signal levels (usually 20-40 dB above rate threshold for cochlear nerve fibers - Evans, 1975) the unit will reach its saturation rate, first for the CF and then for surrounding frequencies. Consequently, the frequency selectivity of single units (as measured in terms of discharge rate) begins to deteriorate once it has reached saturation.

The limited dynamic range of peripheral auditory units poses a considerable challenge to place models of pitch perception. For at moderate to high intensities (>60 dB), the ability of the peripheral auditory system to distinguish adjacent components on the basis of discharge rate is seriously compromised (Evans, 1977; Narins and Evans, 1980). Although some degree of frequency selectivity may be recovered in the dorsal cochlear nucleus (Evans and Palmer, 1975), it is not clear whether it is sufficient to account for psychophysical measures of selectivity.

The perceptual attributes of pitch and timbre remain relatively stable over a dynamic range of 100 dB SPL. This stands in stark contrast to the spatial excitation pattern of the cochlear nerve. At high signal levels a significant proportion of cochlear nerve fibers are firing at saturation rate in response to pure tones (Pfeiffer and Kim, 1975; Kim and Molnar, 1979) and broadband signals, such as synthesized vowels (Sachs and Young, 1979).

Place coding of frequency analysis is particularly difficult to reconcile with psychophysical measures of frequency discrimination. The difference limen for single frequencies below 4 kHz is approximately 0.1-0.5%, regardless of intensity (for levels > 10 dB SL). Interestingly, the DL is considerably larger for higher frequencies (~1-2%) and increases with level (Dye and Hafter, 1980). It has been argued that what is important in frequency discrimination is not the bandwidth of the spatial excitation pattern, but the sharpness of the dividing line between active and non-active units (Allanson and Whitfield, 1956; Zwicker, 1975). This suggestion is plausible for high-frequency fibers, where the high frequency slope of the FTC approaches 1000 dB/octave. However, this model would predict much larger DLs for low-frequency signals where the high-frequency slope of the FTC is considerably less steep, when in fact just the opposite is the case.

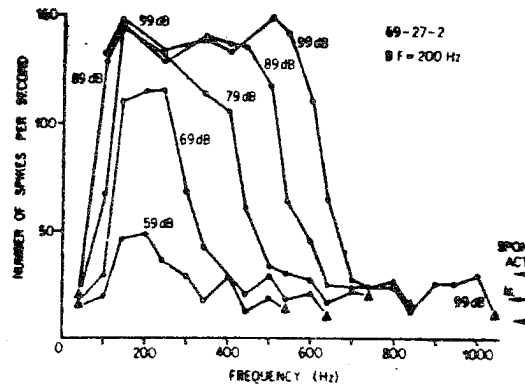


Figure 1.9 An iso-intensity curve for a single cochlear nerve fiber. The discharge rate is plotted as a function of stimulus frequency for several different sound pressure levels. As stimulus intensity increases, the response area of the neuron broadens. (From Rose et al, 1971. Reproduced by permission of the American Physiological Society and Dr. Jerzy Rose)

1.43 PHYSIOLOGICAL CORRELATES OF "PERIODICITY" CODING

Fifty years ago, Wever and Bray (1930a; b) observed that the auditory nerve could synchronize to frequencies up to 5 kHz. The quality of the nerve's spectral synchronizaton was precise enough to pass intelligible speech. Since no single neuron fires at a rate greater than 400 spikes/sec, Wever and Bray (1930b) suggested that the potential recorded represented the summation of many active fibers, each synchronized to the dominant frequency within its response area. For low frequencies (<400 Hz), there would be a one-to-one correspondence between stimulus cycle and neural discharge. Every cycle would elicit a single spike from each nerve. For higher frequencies (up to 5 kHz), each cycle would serve as an effective stimulus for a certain proportion of units. The following cycle would excite another group of cells, the third cycle another group, and so on, until the first group was ready to fire again (Figure 1.10). The upper limit of this "volley" principle would be determined by limits on the ability of a nerve to synchronize to very small time intervals. Above this limit (5 kHz), frequency coding would be based exclusively on place mechanisms.

Evaluation of Wever's theory had to await the development of single unit recording techniques. The earliest microelectrode studies were qualitatively consistent with volley theory (Galambos and Davis, 1943; Tasaki, 1954). They indicated that peripheral auditory neurons preferentially fired during a restricted phase of the stimulus cycle for frequencies up to 5 kHz. Subsequent studies provided considerably more detail about the quantitative aspects of neural periodicity. For example, Rose *et al* (1967) showed that if the time of each spike were plotted relative to the stimulus period, the spikes tended to cluster in the same portion of the period (Figure 1.11), as if the discharge of auditory nerve fibers were "locked" to a restricted phase of the stimulus cycle. This "phase-locked" behavior of auditory neurons is also evident in histograms made of the time intervals between successive discharges (Figure 1.12). This type of plot indicates that the discharges of individual units tend to space themselves out at integral multiples of the stimulus period. The interval histogram clearly shows that not every cycle is an effective stimulus. This "cycle skipping" is not however the result of the fiber's refractory properties. The interval between adjacent spikes is usually considerably longer than the refractory period (approximately 0.5 msec). It can be shown that the conditional firing probability after any number of ineffective cycles is practically the same (Rose, 1970: Table I).

Wever (1949) suggested that the auditory analysis of low frequencies was performed largely in terms of a temporal "code" (Chapter 1.22). His proposal was met with a considerable degree of skepticism, in part because of the pervasive tonotopic organization of the auditory pathway. Although the existence of neural synchronization was beyond dispute, there was (and still is) substantial disagreement concerning its contribution to frequency analysis. However, this

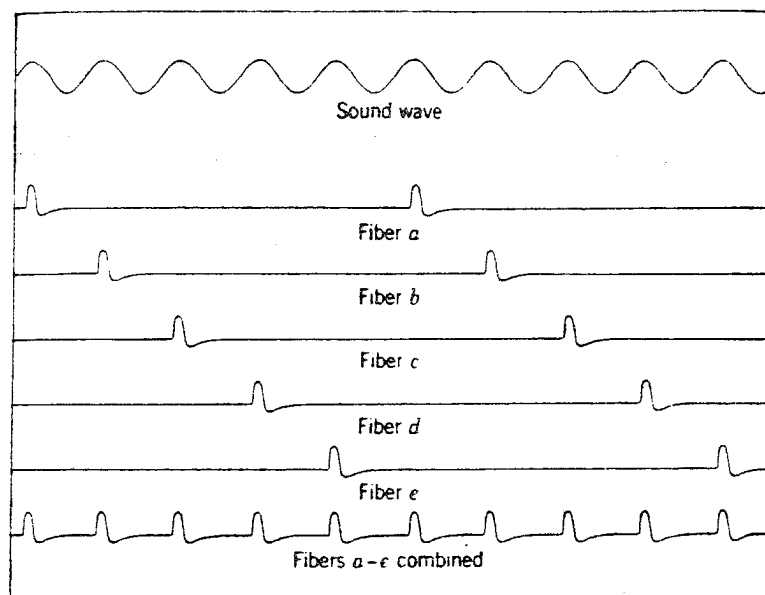


Figure 1.10 The volley principle. A pure tone stimulus (top trace) entrains the activity of a number of nerve fibers which synchronize (phase-lock) to a restricted portion of the stimulus cycle. Although no single fiber fires on every cycle, the frequency of the original stimulus can be reconstructed from the activity of a population of fibers. (From Wever, 1949. Reproduced by permission of Dover Publications, Inc. and Dr. E.G. Wever.)

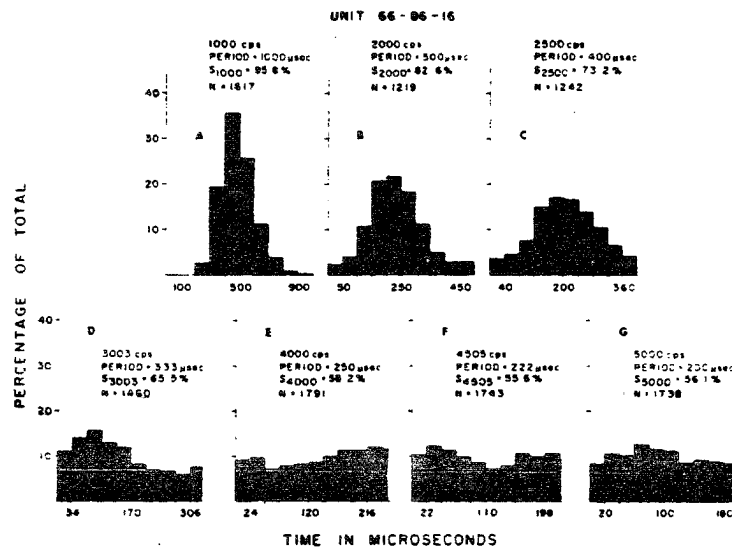
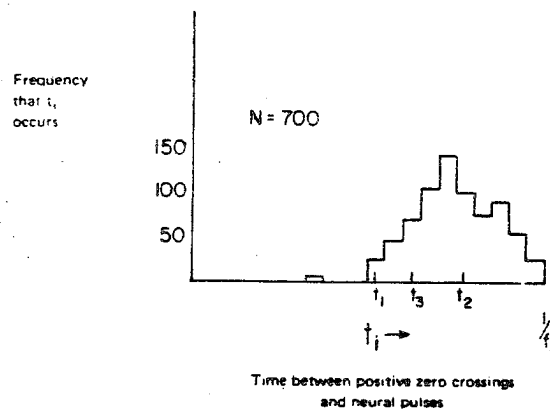
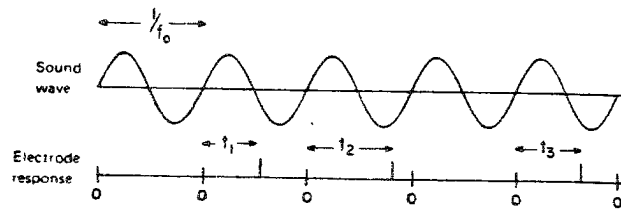


Figure 1.11 The period histogram of a single cochlear nerve fiber is illustrated for a range of stimulus frequencies (bottom figure). Frequencies between 1000 and 2500 Hz are most effective in eliciting phase-locked activity for this fiber. At top is an illustration of the process by which the period histogram is constructed. The running record of spike activity is divided into a series of intervals equal to the period of the stimulus tone. The temporal position of each spike, relative to the stimulus period, is then plotted. (Top figure is from Green, 1976 and is reproduced by permission of Lawrence Erlbaum Associates, Inc. Bottom figure is from Rose *et al.*, 1967 and is reproduced by permission of the American Physiological Society and Dr. Jerzy Rose)

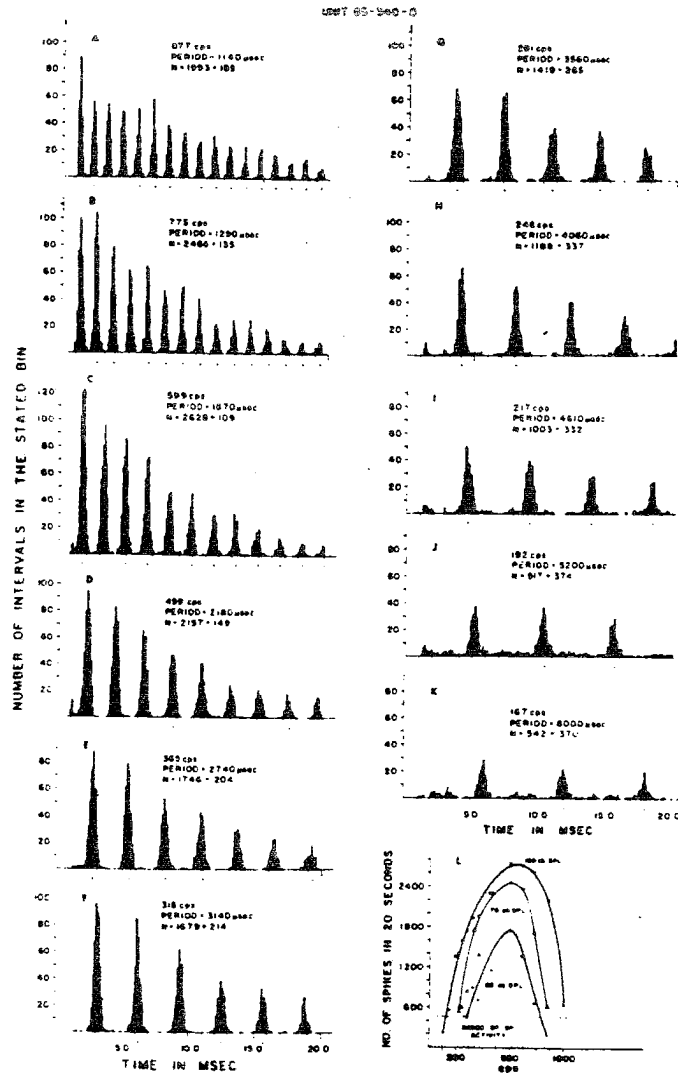


Figure 1.12 An interval histogram of a single cochlear nerve fiber for a series of stimulus frequencies. The histogram records the number of intervals obtained between successive discharges of the fiber. The first peak in each histogram corresponds to the period of the stimulus frequency. Succeeding peaks are integral multiples of the stimulus period. An iso-intensity curve for this same unit is shown in the lower right-hand corner. Notice that phase-locking is most precise for frequencies near the center of the unit's response area. (From Rose *et al.*, 1967. Reproduced by permission of the American Physiological Society and Dr. Jerzy Rose)

skepticism has begun to fade, in recent years, as an increasing amount of physiological and psychophysical research lends plausibility to Wever's hypothesis.

These studies suggest that for signals below 5 kHz, the primary means of frequency analysis is in terms of the temporal discharge pattern of peripheral auditory neurons. Some of the more substantive evidence in support of a temporal code is as follows:

(a) The Precision of Frequency Discrimination

One would expect, on the basis of place coding of frequency, that the DL for frequency discrimination would be proportionally smaller for high frequencies, reflecting the steepened high-frequency slopes of the FTC (Chapter 1.42). However, just the opposite pattern is observed (Moore, 1973; Wier *et al.*, 1977; Dye and Hafter, 1980). Frequency discrimination is most precise for frequencies lying in the optimum region of neural phase-locking (500-2000 Hz) and deteriorates sharply for frequencies above the effective upper limit of synchronization (5 kHz).

(b) The Dynamic Range of Auditory Neurons

Speech and other complex sounds maintain their perceptual identity over a range of 100 dB SPL. Yet the dynamic range of most peripheral auditory neurons (as measured in terms of discharge rate) is restricted to 20-40 dB SPL (Evans, 1975). This discrepancy suggests that frequency information must be conveyed in terms of some other parameter of auditory discharge. One such candidate parameter is the frequency of neural synchronization. The threshold of a fiber, as measured in terms of its phase-locking behavior, is often 20 dB SPL below the rate threshold. Because a fiber maintains its periodic discharge behavior well above rate saturation, the dynamic range of a unit's phase-locked response to a stimulating tone may be as much as 100 dB SPL (Figure 1.13).

(c) Representation of Complex Signals

Because of the limited dynamic range of peripheral auditory fibers, signals presented at moderate-to-high intensities will activate a substantial proportion of these fibers, most of which will be discharging at saturation level (Kim and Molnar, 1979). Consequently, it is difficult to envision a means by which the spectrum of complex signals could be encoded in terms of place principles. Sachs and Young (1979) have shown that information concerning neither spectral components nor spectral peaks is preserved, at high signal levels, in the spatial excitation pattern of the cochlear nerve (Chapter 1.63).

However, these same fibers signal a considerable amount of spectral information in terms of their temporal discharge pattern. This is because the phase-locking behavior of a fiber is usually dominated by that frequency component closest to the center of its response area (Evans, 1978). Consequently, even at high signal levels, there will always be a certain proportion of units whose phase-locking behavior is completely dominated by a single frequency component (Note 1.8).

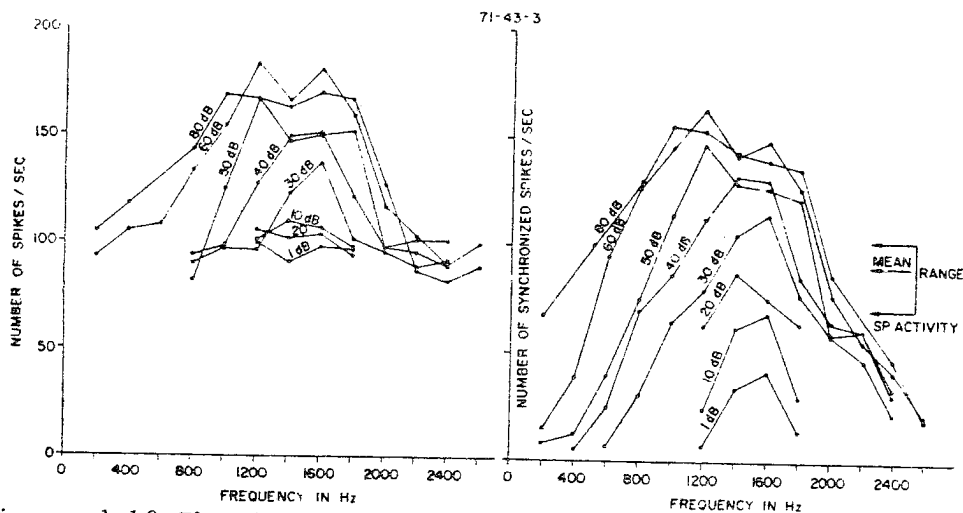
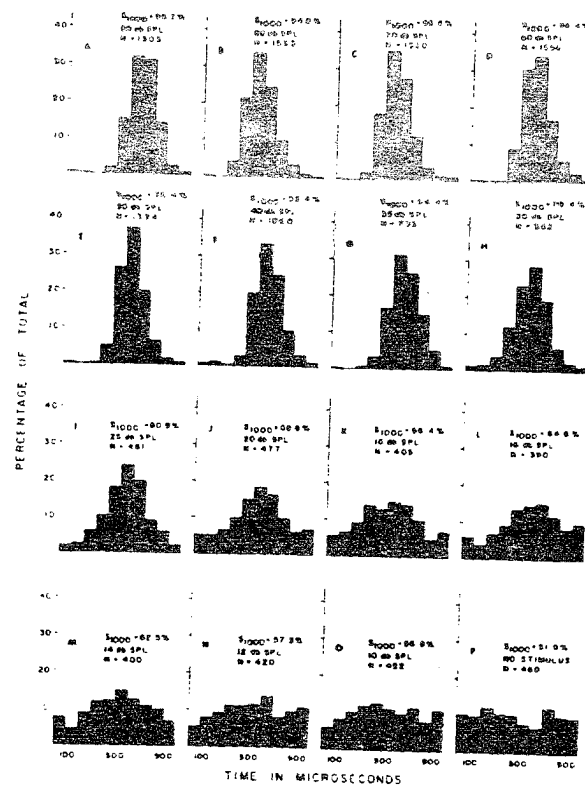


Figure 1.13 The dynamic range of peripheral auditory neurons as illustrated by two different methods. At top is a series of period histograms derived from cochlear nerve activity over a range of 80 dB. Notice that the unit displays phase-locked activity over a range of 80 dB. The bottom trace contains iso-intensity contours for a unit in the AVCN. The curves on the left illustrate the dynamic range in terms of discharge rate per se, while the curves on the right shows the same activity plotted in terms of the synchronized rate. The dynamic range of the fiber is considerably expanded when plotted in terms of the synchronized rate. (The top figure is from Rose et al, 1967 and the lower figure is from Rose et al, 1974. Both figures are reproduced by permission of the American Physiological Society and Dr. Jerzy Rose.)

(d) Distribution of Phase-Sensitive Neurons in the Auditory Pathway

Neural periodicity has been most intensively studied in the cochlear nerve. However, phase-sensitive neurons are present (in varying proportions) at every level of the auditory pathway. All cochlear nerve fibers responsive to frequencies below 5 kHz are capable of phase-locking (Evans, 1975). The phase-locking behavior of AVCN units is very similar to cochlear nerve fibers (Rose et al, 1974; Moushegian and Rupert, 1970). At higher levels of the auditory pathway there is a progressive diminution of the effective upper frequency limit of phase-locking. At the level of the MSO, the upper limit is around 3-4 kHz (Boudreau, 1965; Moushegian et al, 1972), and is approximately 2 kHz in the IC (Rose et al, 1966). Above the level of the brainstem, there are relatively few phase-sensitive units. For example, Rouiller et al (1979) found that only about 1-2% of units in the MGB are capable of phase-locking, and these only to frequencies up to 1 kHz. It is unclear whether "true" phase-locking neurons exist at the level of the auditory cortex. Ribaupierre et al (1972) have recorded from units which are capable of synchronizing to a train of clicks up to 200 clicks/sec. However, the response of these units may represent an "on" response, rather than true phase-locking. The paucity of phase-sensitive neurons in MGB and auditory cortex suggest that timing information is probably converted into some other form at the upper-most levels of the auditory pathway.

1.5 LOW PITCH BASED ON NEURAL PERIODICITY

Certain features of low pitch, difficult to explain in terms of place mechanisms, are consistent with a pitch extraction mechanism based on neural periodicity. The primary motivation for place models of pitch perception - the primacy of aurally resolved components - is not incompatible with a temporal mechanism of frequency coding. One need only assume that the spectrum of a complex sound is measured in terms of neural periodicity rather than the spatial excitation pattern. Evidence supporting this view is presented in Chapters 1.43(a) - 1.43(c) and 1.64.

(a) Spectral Limits of Low Pitch

The existence region of low pitch encompasses the first eight harmonics of a complex tone for fundamentals up to 500-600 Hz. For higher fundamentals the existence region includes only harmonics below 5 kHz. This upper limit of frequency components giving rise to a sensation of low pitch coincides with the upper limit of neural synchronization in the cochlear nerve and AVCN.

(b) The Region of Dominance

The dominance of low harmonics suggests that the pitch extraction mechanism operates directly on the spectrum of a complex signal. It is often assumed that this spectral analysis is based on place mechanisms (Terhardt, 1970; 1974) in which the central pitch processor derives a periodic function from peaks in the spatial excitation pattern

(Wightman, 1973b). However, this conception does not explain why the region of the spectrum producing the most clearly defined peaks in terms of the spatial excitation pattern - the fundamental and second harmonic - rarely determines the low pitch of a complex tone.

An alternative explanation of the dominance region - based on the temporal properties of auditory coding - is incorporated in Goldstein's (1973) "optimum processor" model of pitch perception (Chapter 1.32). In Goldstein's model, the strength of a tone's pitch is a function of its ambiguity, which is, in turn, dependent on the frequency analysis of the signal's spectral components. The sharper the frequency analysis, the smaller the pitch ambiguity and the stronger the pitch. Now, the discrimination of single frequencies is most acute between 600 and 2000 Hz - precisely the region of the spectrum dominant for low pitch. It is in this same region that the precision of neural synchronization (as measured in terms of the dispersion of modes in the interval histogram) is greatest for single cochlear nerve fibers (Rose et al, 1967).

(c) Salience of Low Pitch as a Function of Fundamental Frequency

Place models of pitch perception predict that widely-spaced harmonics will give rise to a strong low pitch. One would therefore expect tones with high fundamentals to produce a more salient low pitch than low fundamentals, when in fact, the salience of low pitch decreases rather than increases for fundamentals above 250 - 500 Hz (Bilsen and Ritsma, 1970; Yost and Hill, 1978; 1979). Moreover, the sensation of low pitch disappears altogether for fundamentals higher than 1 kHz (Ritsma, 1962). Above this limit, the pitch probably stems directly from the fundamental (Plomp, 1967).

Interestingly, relatively few upper auditory brainstem units (DNLL and IC) and no thalamic (MGB) or cortical units (AI) display phase-locked behavior to frequencies above 1 kHz. The upper limit of low pitch may therefore reflect constraints on the phase-locking capabilities of units in the higher levels of the auditory pathway.

(d) Intensity-Independence of the Dominance Principle

Units tuned to the component frequencies phase-lock at intensities that may be as much as 20-30 dB below rate threshold. As a consequence, frequency information can be encoded in terms of neural periodicity at much lower intensities than would be possible in terms of discharge rate. This is significant because the rate threshold of single cochlear fibers is often 10-20 dB above behavioral threshold (of the cat or guinea pig). Low harmonics exert their dominance at levels as low as 10 dB SL. At this level, units tuned to the dominant components have just reached or are still below rate threshold. However, at this same level, these units are firmly synchronized to the component frequencies and maintain this synchrony over a dynamic range of 90-100 dB.

(e) Phase-Insensitivity of Low Pitch

The low pitch of a complex tone is relatively impervious to changes in the signal's phase configuration. This fact is frequently cited against temporal theories of pitch perception (Wightman, 1973a), for changes in the phase configuration also affect the fine temporal structure of the waveform. This argument is invalid because it ignores the operation of cochlear filtering. In order for phase changes to affect the fine time-structure, three or more components must interact. However, within the existence region, cochlear filtering will prevent more than two contiguous components from interacting.

The phase-insensitivity of low pitch is, consequently, far from being incompatible with neural periodicity. In fact, there are two distinct types of information that could be conveyed to a central pitch processor, both of which would be phase-insensitive.

The first is based on units phase-locked exclusively to the signal's frequency components. This population would be able to transmit detailed information concerning the signal's spectrum.

A second source of information could be conveyed by units tuned to frequencies midway between two contiguous signal components. The temporal discharge activity of these units is governed by both components concurrently. The resulting discharge pattern will exhibit peaks at integral multiples of both frequencies (or their mean) and at the period of the compound waveform (i.e. the fundamental) (Rose et al, 1969; 1974; Javel, 1972; 1980).

(f) The Pitch of Inharmonic Signals

The pitch of inharmonic signals is frequently ambiguous. Observers often perceive two distinct pitches, one slightly higher than and one slightly lower than the waveform envelope repetition frequency (Note 1.9). Clearly then, the pitch of such signals is not based on the activity of neurons synchronized to the envelope of the cochlear waveform.

De Boer (1956a) suggested that the pitch of inharmonic signals could be based on either the temporal fine structure of the cochlear waveform or on a harmonic template matched to the signal's spectrum. For low, aurally resolved harmonics, de Boer suggested the latter mechanism would be more likely. However the slopes of the cochlear filter appear to be considerably shallower than would be predicted on the basis of psychophysical measures of frequency selectivity. As a consequence, there is considerable interaction between contiguous components in the temporal discharge patterns of cochlear fibers tuned midway between widely-spaced components. Evans (1978) and Javel (1972; 1980) have shown that the phase-locking behavior of these fibers may, in fact, reflect the pitch ambiguity of inharmonic tones.

(g) Low Pitch in the Absence of Peripheral Interaction

The mere presence of neural activity synchronized to the fine-structure of the cochlear waveform does not, of course, imply that this information is utilized by the auditory system. This caution is necessary in light of studies showing that a sensation of low pitch can be generated through dichotic presentation of two-component signals (Houtsma and Goldstein, 1971; 1972; Bilsen, 1973). In the absence of peripheral interaction among components, the resulting pitch must be derived from units synchronized exclusively to the frequency components. A central pitch extractor could then derive the fundamental through some form of spectral cross-correlation (Licklider, 1959; 1962; Bilsen, 1977).

However, the existence of a dichotically-mediated pitch does not necessarily imply that the process of pitch extraction is exclusively a central phenomenon. For one thing, the upper limit of this type of pitch is considerably lower than monaural pitch (330 Hz vs 1000 Hz - Bilsen and Goldstein, 1974). Second, dichotic pitch is considerably less salient than monaural pitch. Many observers have difficulty perceiving any sensation of pitch whatever (Evans, 1978).

Therefore it seems entirely plausible that two separate pitch extraction mechanisms operate in parallel - one, based on the temporal discharge pattern of peripheral fibers sensitive to the interaction of contiguous components, the other based on the cross-correlation product of a central pitch extractor operating on the discharge activity of neurons synchronized to the component frequencies.

1.6 AUDITORY BASIS OF TIMBRE

The discussion, so far, has focused on the low pitch of complex tones. We have observed that this pitch is present even in the absence of the fundamental and that its presence is, in fact, superfluous for its generation. Yet the fundamental, presented as a separate tone, also produces a distinct pitch. When presented concurrently with its harmonics, its presence can be detected not so much in terms of the tone's low pitch, but rather as a difference in the overall "quality" or "timbre" of the signal. Although a single frequency component produces a salient pitch of its own, this pitch may contribute far more to the signal's timbre than to its low pitch. Thus, any single (resolved) component, other than the fundamental, participates in two different processes concurrently. Its frequency information, in concert with that of other components, is used to infer the periodicity of the signal, while its pitch contributes to the signal's overall quality.

Timbre is a much more elusive quality than pitch. Because of its multidimensional character timbre is much easier to define in terms of what it is not, than in terms of what it is. As defined by the American National Standards Institute (ANSI, S3.20-1973: 56),

"Timbre is that attribute of auditory sensation in terms of which a subject can judge that two sounds similarly presented and having the same loudness and [low] pitch are dissimilar."

Because of this uncertainty about its properties, timbre has received far less attention than pitch. It is perhaps not surprising that timbre has been most intensively studied with respect to vowel perception, where its significance is most apparent.

1.61 ACOUSTIC BASIS OF VOWEL QUALITY

Vowel sounds are produced by the coordinated activity of many different vocal structures. The production of a vowel begins with the repetitive opening and closing of the vocal folds in response to the upward flow of air initiated by contraction of the lungs. This glottal vibration produces a series of pulses with a spectrum slope of approximately -12 dB/octave (Figure 1.14). The fundamental frequency of the glottal spectrum is determined by the rate of vocal fold vibration. The typical rate for an adult male voice ranges between 80 and 150 pulses/sec, while the range for a female voice is approximately an octave higher.

The spectrum of the glottal waveform is modified as the sound propagates through the vocal tract. Certain regions of the spectrum are selectively amplified in accordance with the resonant frequencies (formants) of the vocal tract during production of the vowel (Figure 1.14). These resonances are determined by the size and shape of the pharyngeal and oral cavities. For example, the vowel [i] (as in heed) is produced with the body of the tongue arched forward and the blade almost touching the upper palate. With the tongue in this position, the anterior portion of the oral cavity is relatively small, while the pharyngeal cavity and the posterior area of the oral cavity is relatively large. This particular configuration produces a vowel whose spectrum contains four or five energy maxima (peaks) in the region below 4 kHz. Although, these spectral peaks are, strictly speaking, merely manifestations of the formants of the vocal tract, they are commonly referred to as formants. The first (lowest) formant (F1) of the vowel [i] is very low, about 280 Hz. The second formant (F2), on the other hand, is very high, around 2250 Hz, for a typical male speaker. The higher formants range between 3 and 4.5 kHz (Figure 1.14). When the vocal tract changes its configuration, as in the production of [a] (as in hot), the resonance pattern also changes, and this shift is reflected in the formant pattern of the vocalic spectrum. F1 for [a] is nearly an octave and a half higher than F1 in [i]. F2, on the other hand, is about 1 octave lower than the F2 in [i]. This shift in the vowel's F-pattern reflects a change in the position of the tongue from one that is relatively high and forward to one in which the tongue body retracts and lowers.

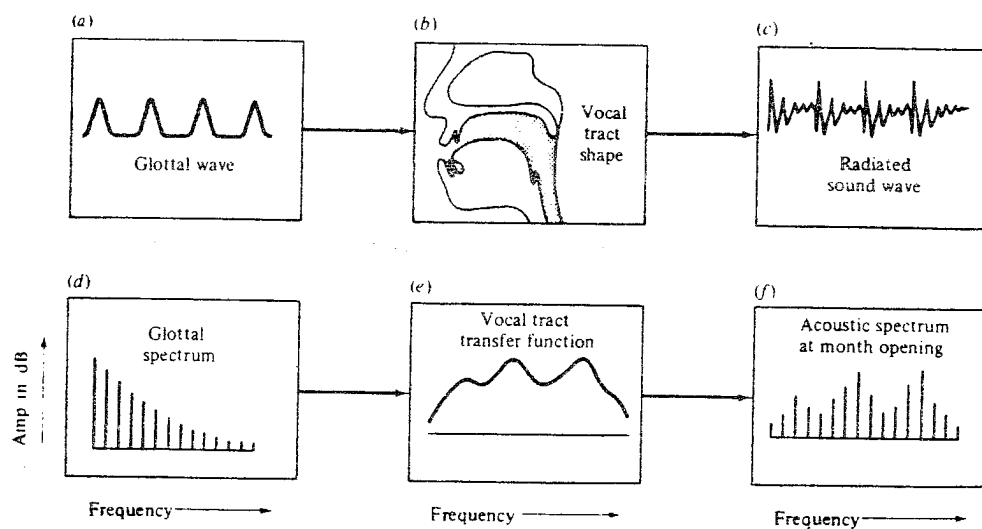


Figure 1.14 The spectrum and waveform of a speech sound are schematically drawn to illustrate how the laryngeal source spectrum is filtered as it is propagated through the vocal tract. (From Minifie, 1973. Adapted from Fant, 1960. Reproduced by permission of Mouton Publishing Co. and Prentice Hall, Inc.)

1.62 AUDITORY BASIS OF VOWEL QUALITY

Because each vowel has a distinctive F-pattern (Figure 1.15), it is reasonable to assume that the acoustic basis of vowel quality is related to the distribution of energy across the spectrum. On this assumption, it is possible to successfully program a computer to discriminate among different vowels spoken by a range of different speakers (Plomp et al, 1967; Klein et al, 1970; Pols, 1970; Pols et al, 1971; Papcun, 1980). However, this does not necessarily imply that the auditory system uses the same "algorithm" to process vowel quality. In particular, we may ask whether the auditory system looks at the entire spectrum, or focusses on only a few relevant parameters?

This question has been addressed by Plomp and his colleagues (Plomp, 1970; Plomp et al, 1967; ; Pols, 1970; Pols et al, 1969; 1973; Klein et al, 1970; Nierop et al, 1973). Their basic technique was as follows. Spectra of Dutch vowels, produced by a range of speakers, were analyzed using a series of 18 contiguous, non-overlapping 1/3 octave band filters. The spectrum of each vowel was thereby reduced to 18 numbers, representing the amount of energy passed through each filter. This information served as input into a principal components analysis (PCA). PCA determined which factors (based, in this instance, on the energy distribution across the spectrum) are most important for the differentiation of the Dutch vowels. Interestingly, the vowels could be distinguished largely on the basis of just three factors. These factors are closely associated with the first three energy maxima of the reduced spectra. These results confirm that vocalic spectra are acoustically distinguishable primarily in terms of their formant locations.

Pols et al (1969) sought to determine whether the same three factors could also account for perceptual judgements of vowel similarity. Using the method of triadic comparisons, the perceptual distance between any two vowels was determined. Pols found that a three-dimensional space was sufficient to describe the perceptual distance configuration and that these dimensions correlated highly with the three main factors determined by the acoustic analysis of the same vowels. Pols' study suggests that the perceptual judgement of vowel quality is based primarily on the lowest three energy maxima in the spectrum.

However, it is not clear from these results whether differences in vowel quality are based on both the amplitude and frequency of the lower formants (i.e. the spectrum "envelope") or on formant frequencies alone. This issue has been addressed by Lindquist and Pauli (1968), Mushnikov and Chistovich (1969), Ainsworth and Millar (1972), Chistovich and Lublinskaja (1979), and Chistovich et al (1979), all of whom have found that vowel identity is remarkably insensitive to formant amplitude over a large intensity range (25-45 dB). For example, Mushnikov and Chistovich (1969) (cited in Chistovich, 1971) found that F2 for the vowel [e] could be as much as 46 dB below the level of F1 (the typical F1-F2 level disparity for this vowel is approximately 15

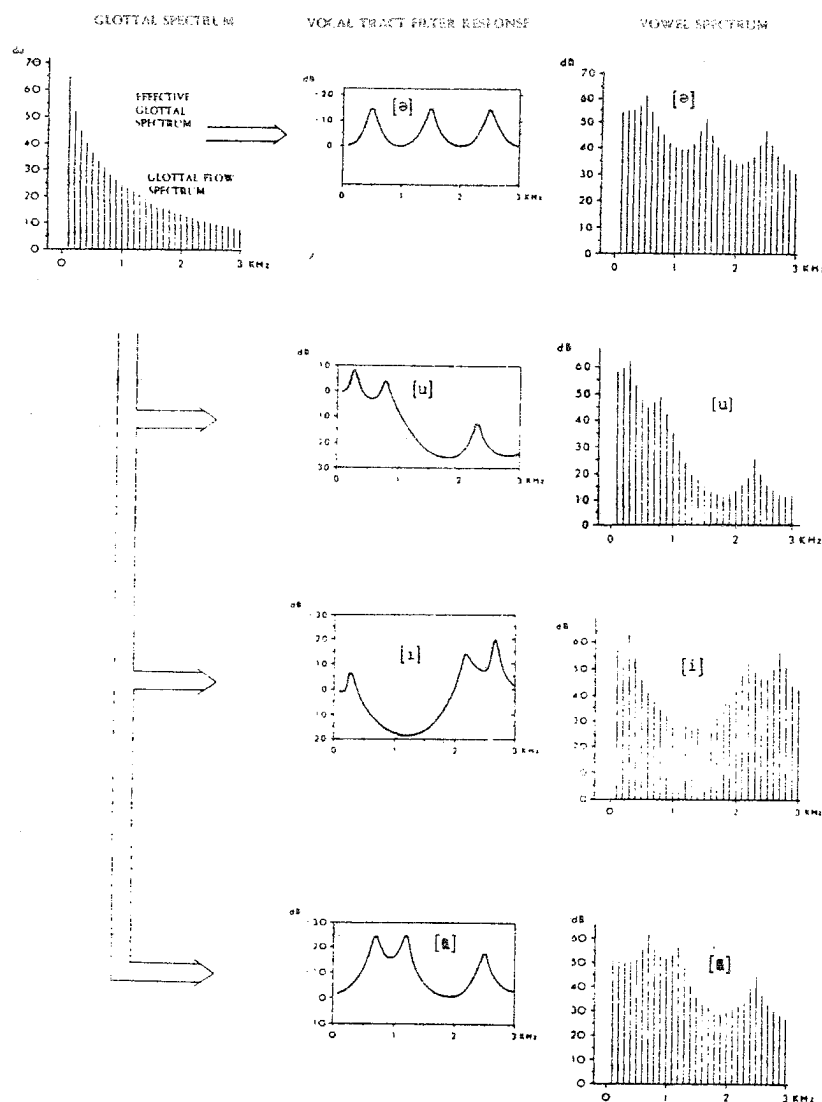


Figure 1.15 The spectra of four different vowels and their associated transfer functions. (From Pickett, 1980. Reproduced by permission of University Park Press.)

dB) before its quality appreciably changed. This insensitivity to relative formant level is particularly pronounced for high, front vowels where the distance between F1 and F2 is particularly large (Ainsworth and Millar, 1972). Chistovich and Lublinskaja (1979) have shown that this insensitivity to formant level becomes manifest when the formant peaks are separated by more than 3-3.5 critical band units ($2/3$ - $3/4$ octave). For smaller formant separations, the relative formant level can have a significant effect on perceived vowel quality. Interestingly, when F2-F1 is small, a reasonable approximation to the vowel can be obtained with a single formant stimulus. The precise location of this single formant depends on the relative F1/F2 level (Chistovich and Lublinskaja, 1979a). Carlson *et al* (1970; 1975) have obtained similar single formant approximations for both F2-F1 and F3-F2. These results suggest that the perception of vowel quality is based largely on the detection of spectral maxima independent of their relative level, as long as there is sufficient separation between adjacent peaks. When the peak-to-peak distance is smaller than this threshold, a level-dependent interaction between formant peaks occurs. Chistovich and Lublinskaja (1979) suggest that this interaction is the product of some form of spatial integration operating over a restricted tonotopic domain.

The studies reviewed up to this point clearly indicate that vowel quality is based on the location of the lowest two or three energy maxima in the spectrum. But what constitutes an energy maximum? Fant (1960) and Carlson *et al* (1975) suggest that the auditory system is able to infer the precise location of the VOCAL TRACT resonance from the amplitudes of individual harmonics in the formant region. In their view, what is being perceived is a region of spectral prominence, in which individual harmonics contribute to the overall quality perceived in proportion to their relative amplitude. This hypothesis implies a certain degree of interaction between vowel quality and fundamental frequency. In particular, it predicts that the lower the fundamental frequency, the more precise the interpolation of spectral maxima should be. There should consequently be a significant difference in the processing of vowel quality based on male and female voices (unless the formant bandwidths of a female voice are significantly broader than a male's).

An alternative hypothesis is offered by Karnickaya *et al* (1975). They suggest that the spectral peak is based on the most prominent harmonic in the formant region (particularly F1). Their conclusion is based on the following evidence. Listeners were asked to determine the phonemic identity of a series of synthetic stimuli, whose spectra contained energy only at the two lowest spectrum peaks. All other harmonics were removed. F2 was represented by several harmonics (centered around 2 kHz), whose amplitudes were comparable to those found in the vowels [i] and [e]. A single frequency component, of variable frequency, was substituted for F1. When the single component was below 400 Hz, the signal was usually identified as [i]. When the sinusoid was greater than 400 Hz, listeners tended to identify the stimulus as [e]. Next, two frequency components (one lower than 400 Hz,

the other higher) were presented with the original F2. Subjects could switch the perceived vocalic identity from [i] to [e] merely by adjusting the relative amplitude of the two frequency components. In a separate study, listeners were asked to make a similar perceptual judgement using slightly different stimuli. F2 was the same as in the first experiment. But the lower region of the spectrum consisted of four harmonics instead of two. Three of the components were close together. The fourth was either 200 Hz higher or lower than the other harmonics. If the perceptual formant is based on the integration of energy across several components, as Carlson *et al* (1975) suggest, then the identity of the vowel will be determined by the three-component complex when its total energy exceeds that of the single component. However, Karnickaya *et al* (1975) found that the phonemic identity of the stimulus was always determined by the spectrum level of the individual components, rather than by their summed energy.

The studies reviewed above do not provide any clear-cut answer concerning the auditory definition of a spectral maximum. From the studies of Chistovich and her colleagues it appears that the concept of formant is of limited utility in the low-frequency region of the spectrum (< 1 kHz) where the harmonic spacing exceeds the critical bandwidth. Consequently, it is likely that the spectrum maximum in the first formant region corresponds to a single spectral component. However, this dominant component need not necessarily be the one with the largest amplitude. For higher formants, where individual harmonics are unlikely to be fully resolved by the auditory system, some form of "frequency-averaging" as proposed by Carlson *et al* (1975), is a distinct possibility.

An alternative method of investigating the auditory basis of vowel quality is found in a study by Houtgast (1974a). Houtgast has been concerned with developing a psychophysical procedure to determine the "internal" spectrum of a signal as "seen" by the auditory system. Toward this end, he has developed a means of inferring the amplitude level of each internal component of a complex sound by determining its "pulsation-threshold." In a typical experiment, a masker tone, set at a fixed level, is alternated with a test tone, whose level is under subject control. When the test tone is presented at a sufficiently low amplitude, only the masker tone is audible. As the intensity of the test tone is increased, it appears to merge with the masker, and is perceived as a continuous tone. At higher levels, the test tone is perceived as a pulsating tone, distinct from the masker. The pulsation threshold is defined as the highest level at which the tone is still perceived as a continuous tone, before pulsating. Houtgast (1973) has shown, using a single frequency test tone, that this method provides an estimate of auditory frequency selectivity similar to that obtained in single unit studies of the cochlear nerve. Houtgast (1974a) has also shown that the auditory representation of a complex signal (consisting of the first 10 harmonics of a 250 Hz fundamental), as determined by the pulsation threshold, is consistent with other psychophysical estimates of the ear's resolving power. In view of the excellent agreement between pulsation threshold and physiological estimates of

frequency selectivity, Houtgast (1974b) determined the pulsation threshold measurements for the first 32 harmonics of the vowels [a], [E], and [i]. From these data, it is possible to reconstruct an "internal" spectral representation for each vowel. The spectrum envelopes of these "internal" spectra are similar to those of the acoustic spectra. However, there are a number of interesting differences. For example, the amplitude of the harmonics just below the second formant appears to be suppressed in the vowels [i] and [E], resulting in a "sharpening" of the second formant peak. When F1 and F2 are relatively close together, as in the vowel [a], the second formant peak is reduced in amplitude. Possible physiological correlates of this spectral suppression are discussed in Chapter 1.64.

Another method of determining the auditory spectrum of vowel sounds has been developed by Sachs and Zurek (1979). A vowel is presented to one ear while a pure tone, equal in frequency to one of the vowel's harmonics, is presented to the other. The amplitude (and phase for frequencies below 1200 Hz) of the tone is adjusted until the tonal image appears to emanate from the center of the head. Using this procedure, it is possible to obtain an estimate of the level for each harmonic of the vowel. Auditory representations of vocalic spectra obtained with this "contralateral-probe" procedure are similar, in certain respects, to those obtained with the pulsation threshold method. For example, the auditory and physical spectra correspond closely in the low-frequency region of the spectrum ($< 1\text{--}1.5$ kHz) for both methods. Yet, there are some very striking differences as well. First, the amplitude of high frequency components (> 1.5 kHz) is often 20-30 dB below the level of the acoustic spectrum, even in the region of spectral maxima. Second, the peaks in the internal spectrum are more pronounced than those in the acoustic spectrum. Harmonics on the high-frequency side of a spectral peak usually suffer a larger reduction in amplitude than those on the low frequency side of the first formant. This amplitude reduction can be as much as 20-40 dB.

The pulsation threshold and contralateral-probe techniques offer somewhat contrasting estimates of the auditory spectrum of vowel sounds. The pulsation threshold procedure suggests that the internal spectrum is similar to the acoustic spectrum, except for the enhancement of the second formant peak. The contralateral-probe estimate of the auditory spectrum deviates considerably from the acoustic spectrum. A significant amount of spectral sharpening in the region of the first formant is evident. Moreover, the amplitude of high-frequency harmonics is considerably smaller than would be predicted from Houtgast's study. These contrasting views of vocalic spectra may very well reflect genuine differences in the manner in which complex spectra are encoded. The pulsation threshold probably reflects the spatial excitation pattern of peripheral (cochlear nerve) units. The contralateral-probe procedure, on the other hand, is based on more central auditory mechanisms, and may reflect, at least in part, the temporal patterns of auditory neurons.

1.63 "PLACE" REPRESENTATION OF VOWEL SPECTRA

The perception of vowel quality is ultimately based on physiological mechanisms. In this chapter, and the one that follows, two possible means are considered by which vowel quality is encoded in the auditory system. The first possibility is that the vocalic spectrum is coded in terms of the spatial excitation pattern of auditory neurons. Each harmonic in the spectrum governs the discharge activity of a specific population of fibers. The amplitude of each spectral component is represented in terms of the unit discharge rate. Fibers tuned to the formant peaks fire at a faster rate than fibers tuned to the spectral troughs. The resulting auditory representation is a two-dimensional profile, in which the tonotopic affiliation of a fiber is given by the x-coordinate and the discharge rate is expressed by the y-coordinate. The location of the formant peaks could be determined by finding the maximum points in the y-dimension.

This "place" model provides a simple, straight-forward physiological framework for encoding the spectra of broadband, complex signals such as speech. As a consequence, this model forms the basis of the initial processing stage of most automatic speech recognition systems (Pols, 1971; Alinat, 1975; Dolmazon *et al.*, 1977; Karnickaya *et al.*, 1975; Searle *et al.*, 1979; Weinstein *et al.*, 1975; Zwicker *et al.*, 1979). However, the model makes two important assumptions, neither of which are empirically supportable. The first concerns the linearity of the auditory system's response to stimulus amplitude. The model assumes that a 10 dB increment in signal level results in a simple multiplication of the discharge rate of all fibers. As discussed below, this is simply not the case. The second assumption involves the system's dynamic range. Speech maintains its intelligibility over a range of approximately 100 dB (Pollack and Pickett, 1958). Any physiological model must incorporate a dynamic range of similar magnitude. Unfortunately, any model based on discharge rate will have difficulty meeting this criterion because of the relatively restricted dynamic range of auditory neurons (Chapter 1.42).

These assumptions have been put to a direct test by Sachs and Young (1979). They recorded the discharge activity of a large, tonotopically-distributed population of (cat) cochlear nerve fibers in response to four different synthetic vowels. At low intensities (40 dB SPL), the spectrum envelope of the vowels are reflected in the discharge rate profile. However, as signal level increases, a large number of units become rate-saturated. As a consequence, the contrast, in terms of discharge rate, between the peaks and valleys of the spectrum becomes progressively compressed. Above 60-65 dB SPL, the discharge rate of almost all cochlear nerve fibers tuned to frequencies below 10 kHz have reached saturation level. At this point, there is effectively no contrast between the peaks and valleys of the spectrum.

Although Sachs and Young's (1979) study makes it unlikely that vocalic spectra are encoded in terms of the spatial excitation pattern of the cochlear nerve as a whole, it does not entirely rule out the possibility of place encoding. For there are a small proportion of cochlear nerve fibers (approximately 5%) whose dynamic range is considerably larger (approximately 70 dB) than the majority of units. The spontaneous discharge rates of these large dynamic range units are extremely low (< 1 spike/sec), about 50-100 spikes/sec lower than most fibers. Sachs and Young (1979) find that the vocalic spectrum is indeed preserved in the spatial excitation pattern of these fibers over a large dynamic range.

It is also conceivable that the representation of vocalic spectra observed in the cochlear nerve is not entirely representative of the excitation pattern at higher levels of the auditory system. For example, there is evidence of lateral inhibitory networks functioning in the dorsal cochlear nucleus (Evans, 1975). These may operate to enhance the contrast between peaks and valleys in a broadband spectrum by selectively depressing the discharge rate of DCN units tuned to the spectral valleys and it is possible that the large dynamic range fibers of the cochlear nerve may play a role in this inhibition. However, lateral inhibitory mechanisms are yet to be systematically studied, and as a consequence, their role in the processing of speech sounds remains obscure.

The evidence reviewed in this chapter suggests that vocalic spectra are probably not represented in terms of the spatial excitation of auditory neurons, at least at the level of the cochlear nerve. Although there is a small subpopulation of units with a sufficiently large dynamic range to preserve a place representation of the vocalic spectra at high signal levels, it is unlikely that the auditory representation of vowel sounds would be based on such a small population of nerve fibers. The section that follows focusses on an alternative means of encoding the spectra of vowels, one that is not constrained by the limited dynamic range of peripheral auditory neurons.

1.64 THE ROLE OF NEURAL PERIODICITY IN VOWEL PERCEPTION

The intelligibility of speech is preserved over a range of at least 100 dB SPL, far exceeding the dynamic range of most cochlear nerve fibers. Although the possibility that lateral inhibitory mechanisms "sharpen" the place representation of vocalic spectra at higher levels of the auditory system cannot be excluded, the available evidence suggests that the representation of vocalic spectra is probably based on some other parameter than the discharge rate of cochlear nerve fibers. Young and Sachs (1979) and Sachs and Young (1980) suggest that this alternative parameter is the temporal discharge pattern of auditory neurons.

Their conclusion is based on the response of a large population of cochlear nerve fibers to vowels presented at intensities spanning a 60 dB range. At low intensities (<40 dB SPL), units tend to synchronize to spectral components lying in the center of their response areas. Place and temporal representations of the vocalic spectrum would, at these signal levels, be equivalent. As sound pressure level increases, the discharge rate of fibers tuned to the first formant saturates. With further increases in signal level, units originally synchronized to frequencies adjacent to the formants become phase-locked to the formant frequencies. At moderate-to-high intensities (>60 dB SPL), most of the fibers tuned to frequencies between the first and second, and second and third formants synchronize their discharge patterns to the first and second formants, respectively. In this fashion, the spectral peaks of a vowel are preserved in terms of the temporal discharge pattern of cochlear nerve fibers up to at least 80 dB SPL (the highest signal level used in Young and Sachs' (1979) study).

Although comparable population studies have not been performed in other auditory regions, several studies suggest that neural periodicity continues to play an important role at higher levels of the auditory pathway. Moore and Cashin (1974; 1976; 1977), Rupert *et al* (1977), and Caspary *et al* (1977) have found that a significant proportion of units in the ventral cochlear nucleus phase-lock in response to vocalic stimuli. At low signal levels (10-20 dB above a unit's rate threshold), units tuned to components in the formant region synchronize their discharge pattern to the formant frequency. Other units synchronize to the pitch period of the vowel (Caspary *et al*, 1977). At higher signal levels, a unit's discharge pattern can be significantly modified by energy present in the unit's inhibitory region. A moderate amount of energy in the inhibitory sideband of a unit may actually enhance the temporal response of the neuron to vocalic stimuli (Moore and Cashin, 1974).

It is likely that the spectrum of low-frequency speech sounds is preserved in the temporal discharge pattern of auditory neurons up to at least the level of the superior olivary complex. Tuschitani and Boudreau (1964) report transmitting intelligible speech through a macroelectrode located in the the medial superior olive (of the cat). Above the level of the superior olivary complex, the brainstem response to speech sounds has been little studied. Watanabe and Sakai (1978) report that many neurons in the inferior colliculus (of the cat) synchronize to the pitch period of natural speech. However, it is not clear from this study whether collicular units actually synchronize to the first or second formants. Chapter 7 of the present study addresses this question in more detail.

1.7 PHYSIOLOGICAL BASIS OF THE FREQUENCY-FOLLOWING RESPONSE

Most physiological studies focus on the response characteristics of single nerve cells. The single unit approach provides detailed information concerning the discharge patterns of a small sample of neural elements evoked by acoustic stimuli. However, the picture supplied by such studies is extremely selective given the large number of neurons in the auditory system. In a relatively homogeneous neural population, such as the cochlear nerve, this drawback may be ameliorated to a certain degree through population studies (Pfeiffer and Kim, 1975; Kim and Molnar, 1979; Sachs and Young, 1979; 1980; Young and Sachs, 1979). However, this procedure (involving the recording of many hundreds of units) is extremely time-consuming, and is consequently inappropriate for exploratory studies.

The selective view provided by single unit studies is particularly troublesome for stimuli producing complex auditory sensations whose neural manifestation probably involves large scale integration of many different populations of nerve cells at various levels of the auditory pathway.

Given these drawbacks of single unit recording techniques, is there an alternative technique that provides similar information about the phase-locking activity of auditory neurons? In this section the history of research on periodic auditory macropotentials is reviewed, as is evidence relating these potentials to single unit activity.

1.71 THE RELATIONSHIP OF DEPTH-RECORDED PERIODIC BRAINSTEM POTENTIALS (FREQUENCY-FOLLOWING RESPONSE) TO THE TEMPORAL DISCHARGE PATTERNS OF SINGLE UNITS

Fifty years ago, Wever and Bray (1930a) recorded periodic macropotentials from the cochlear nerve of the cat in response to pure tones and natural speech (Chapter 1.43). Although subsequent studies were to suggest that the cochlear microphonic contributed to their recording (Davis *et al*, 1933; Derbyshire and Davis, 1935), there is little doubt that a significant portion of the response was of neural origin. Their basic findings were extended to auditory brainstem nuclei by Kemp *et al* (1937), who recorded synchronized responses (using concentric electrodes) up to 2500 Hz in the trapezoid body and 1000-1500 Hz at the lateral lemniscus.

These initial studies were not followed up until the 1960's, when Tsuchitani and Boudreau (1964) recorded (using signal averaging techniques) periodic "wave" activity from the medial superior olive in response to frequencies up to 4 kHz. They also found that speech spoken into the ear of a cat could be understood by an observer listening to the amplified signal recorded from the MSO. Warnick and Starr (1968) showed that binaural beats, produced by presentation of slightly different frequencies to the two ears of a cat, are reflected in the macropotential recorded from the MSO. The response to binaural

stimulation resembled an amplitude-modulated sinusoid with an envelope periodicity corresponding to the beat frequency.

Worden and Marsh (1968) compared the periodic potential recorded from a number of brainstem nuclei in the cat with the cochlear microphonic (CM) recorded from the round window of the same animals. They observed that the potential recorded from the brainstem electrodes - the "frequency-following response" (FFR) - differs from the CM in a number of respects. (1) While the CM response to a pure tone is relatively sinusoidal, the FFR spectrum contains significant energy at harmonics of the tonal frequency. This is consistent with the observation that auditory units are activated by only one polarity of the stimulus waveform. (2) The onset latency of the CM is approximately 1-2 msec. The latency of FFR is significantly longer (3-6 msec), lengthening at auditory nuclei progressively more rostral to the cochlear nuclei. (3) FFR disappears prior to CM in terminal anoxia, indicating the dependence of FFR on normal metabolism of neural tissue. (4) The onset of FFR is often considerably larger than the subsequent portion of the response, suggestive of neural adaptation. (5) The upper limit of FFR is approximately 5 kHz, consistent with the upper limit of peripheral auditory phase-locking for single units (Rose *et al*, 1967). In contrast, the CM reproduces frequencies up to the limit of hearing (approximately 40 kHz in the cat). In a subsequent study (Marsh *et al*, 1970) demonstrated that FFR recorded from ventral cochlear nucleus or medial superior olive is abolished upon sectioning of the cochlear nerve or cryogenic cooling of the VCN. In contrast, the CM is unaffected by either procedure.

Marsh *et al* (1974) studied the distribution of FFR throughout the auditory brainstem pathway. Their study indicated that there are two discrete pathways for transmission of FFR through the brainstem nuclei. One pathway transmits FFR from CN to ipsilateral DNLL and IC via the MSO. The other path leads from CN directly to contralateral VNLL, DNLL and IC, bypassing the superior olivary complex altogether. Their observations also indicate that FFR is recordable from all nuclei in which a substantial number of phase-locking neurons have been found, and that the upper limit of FFR at each nucleus is consistent with data published for single unit synchrony. These limits are approximately 5 kHz for the cochlear nucleus (Starr and Hellerstein, 1971), 3-4 kHz for trapezoid body and medial superior olive, and 1.5-2.0 kHz for lateral lemniscus and inferior colliculus. FFR has not been recorded from levels above the inferior colliculus, consistent with the very small proportion of units observed to phase-lock in the MGB and auditory cortex.

A more direct study of the relationship between single unit activity and FFR was published by Holstein *et al* (1969). They found that the envelope of (CN) unit spike activity, when integrated over time, corresponds closely to FFR recorded from a small population of neurons in the same region of the CN.

Further evidence relating FFR to single unit activity is found in a study by Marsh et al (1972). They simultaneously recorded the response of single units and FFR in the cochlear nucleus, as well as the CM, to tones presented with a variable amount of masking noise. At noise levels sufficient to attenuate or abolish the FFR, the synchrony of the single units was also disrupted (although the discharge rate was often unaffected). The CM, on the other hand, displayed relatively little degradation with increasing noise levels.

In summary, a substantial body of evidence indicates that the frequency-following response recorded from the auditory brainstem pathway of the cat is a true neural potential and that it represents the summation of single neural elements phase-locked to the frequency of the stimulating signal. This conclusion is based on the following observations: (1) The FFR is a periodic potential whose fundamental frequency is equal to that of the stimulus frequency. (2) Its behavior is distinct from the cochlear microphonic in many ways, most notably in terms of its latency, upper frequency limits, and behavior under terminal anoxia. (3) The latency of FFR is appropriate to the level from which it is recorded. For CN this latency is about 2 msec, and increases appropriately at progressively higher levels of the auditory brainstem pathway. (4) The upper frequency limit of FFR in various nuclei corresponds to the upper limits of synchronized single unit activity. (5) The FFR deteriorates at masking levels similar to those which disrupt the phase-locking behavior of single units

1.72 THE RELATIONSHIP OF SCALP-RECORDED FREQUENCY-FOLLOWING RESPONSE TO DEPTH-RECORDED ACTIVITY

Ten years ago, Jewett et al (1970) and Jewett and Williston (1971) reported recording a potential from the human scalp, which they suggested was of brainstem origin. This potential, recorded in response to trains of clicks, consisted of a series of short-latency peaks. These peaks, they hypothesized, represented the onset activity of neurons at successive levels of the auditory brainstem pathway. The brainstem origins of this potential were confirmed by Buchwald and Huang (1975) in a series of lesion studies in the cat. They identified the cochlear nerve as the generator of the first peak (Wave I) and showed that successive peaks, spaced at approximately 1 msec intervals, have their origins in the cochlear nucleus, superior olivary complex, ventral nucleus of the lateral lemniscus, and the inferior colliculus. The latency of this last peak, Wave V, is approximately 6 msec.

It is known from gross action potential studies of the cochlear nerve (Teas et al, 1962; Eldredge, 1978) that the largest contribution to the click-evoked response comes from neurons tuned to frequencies above 2 kHz (Note 1.10). Consequently, the brainstem-evoked response (BSER) recorded by Jewett et al (1970) probably does not significantly reflect the activity of low-frequency, phase-locking units.

To determine if the activity of low-frequency brainstem units could be recorded from the human scalp, Moushegian et al (1973) recorded (differentially between vertex and linked earlobes) the scalp response to low-frequency tone bursts, using signal-averaging techniques. These scalp potentials were periodic, following the frequency of the tone up to 2 kHz, and were abolished in the presence of masking noise. Their onset latency was approximately 6 msec.

The latency of these scalp-recorded FFR suggested an upper brainstem source. Smith et al (1975) tested this hypothesis by comparing FFR recorded from the scalp of cats with FFR recorded from the brainstem nuclei of the same animals. The latency of FFR recorded from the IC ($x=5.3$ msec) was found to be practically identical to the latency of the scalp potential ($x=5.6$ msec). This correspondence suggests that the IC is the source of scalp FFR. To test this hypothesis, FFR was recorded bilaterally from IC and MSO, as well as from the scalp. Cryogenic cooling of the IC resulted in significant attenuation of both IC and scalp FFR, while FFR recorded from MSO remained at the control amplitude. This indicated that brainstem nuclei caudal to the IC do not significantly contribute to FFR recorded from the scalp, at least in the cat. Although it is not possible to determine the source of scalp-recorded FFR in human subjects so directly, several pieces of evidence suggest an upper brainstem source for human FFR. First, the onset latency of FFR recorded from the scalp of cat and human is practically identical (Figure 1.16; Marsh et al, 1975). Second, the spectra of cat and human responses to pure tone stimuli are very similar. Third, the upper frequency limit of FFR recorded from the human scalp (2 kHz) is the same as that recorded from the IC of the cat.

A number of investigators have reported recording scalp FFR of significantly shorter latency than 6 msec (Sohmer and Pratt, 1976; 1977; Stillman et al, 1978; Scherg and Brinkmann, 1979). For example, Sohmer and Pratt (1977) found two distinct scalp signals when recording differentially between vertex and ipsilateral earlobe. One, with a latency of 6 msec, is almost certainly of upper brainstem origin, for this potential is not recordable in patients with lesions confined to the upper medulla and midbrain (Sohmer et al, 1977). The other potential has a latency of 1 msec and can be cancelled by flipping the polarity of the stimulus (equivalent to a phase shift of 180 degrees) and adding the resultant potential to the original scalp signal. The behavior of this potential is consistent with the cochlear microphonic.

Stillman et al (1978) also distinguish two different scalp potentials. However, both are of considerably longer latency than the CM potential recorded by Sohmer and colleagues. One potential (FFP2), with a latency of 6 msec, appears to be of the same origin as FFR recorded by Moushegian et al (1973) and Smith et al (1975). The other potential (FFP1) has a latency between 1.4 and 1.8 msec shorter than FFP2 and is particularly prominent at high intensities (>65 dB SL). Stillman et al (1978) suggest this signal is generated by the same

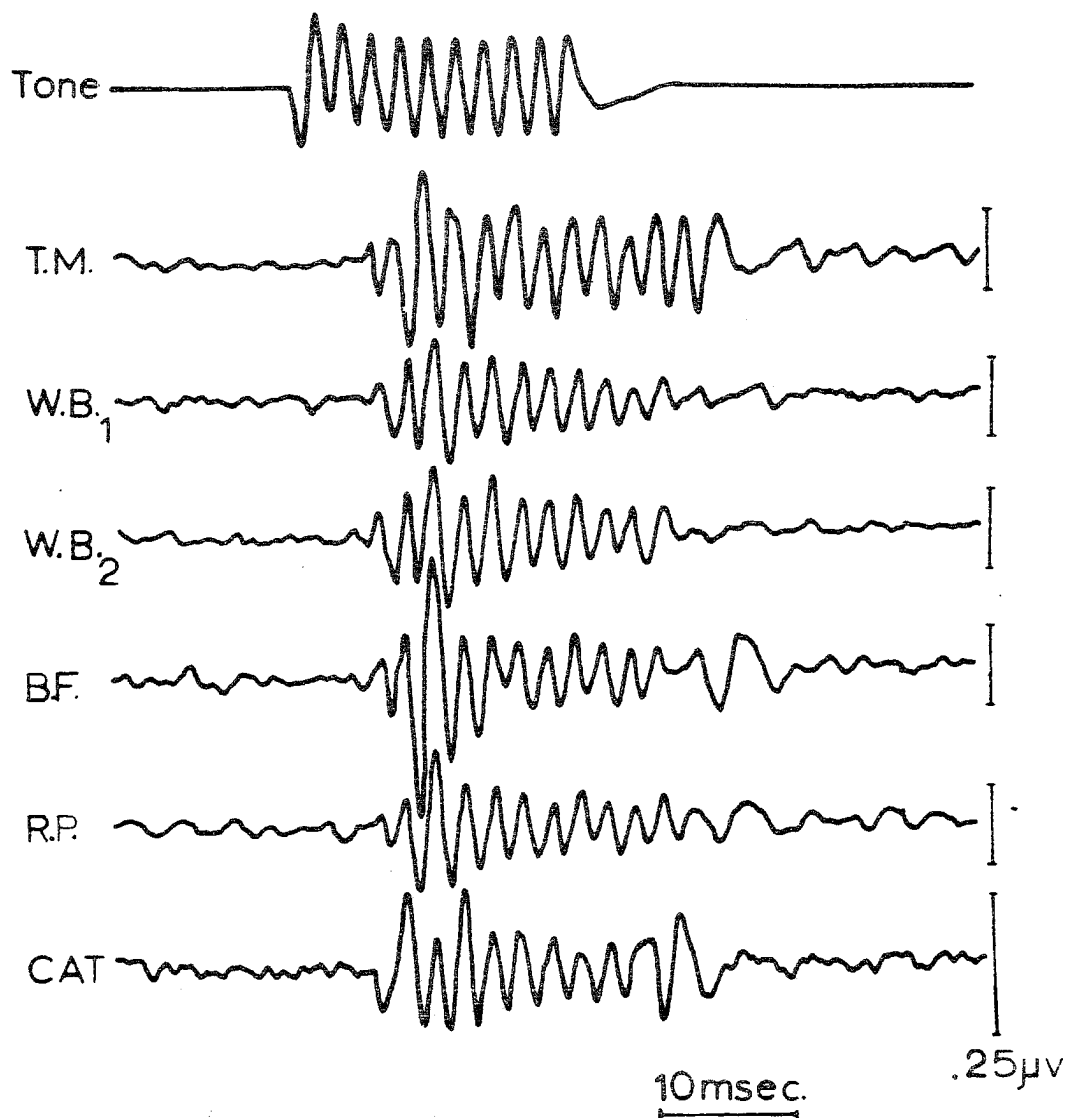


Figure 1.16 Frequency-following responses recorded from the vertex of one feline and four human subjects in response to a 500 Hz signal. Top trace is the averaged microphone response to the 500 Hz stimulus. Note the consistency of the averaged response across subjects, both in terms of the waveform periodicity and response latency. Averages consist of 7000 responses. Stimulus intensity = 85 dB SPL.

source as wave III of the BSER (probably the medial superior olive). Scherg and Brinkman (1979) have differentiated these two components on the basis of their latency and amplitude behavior. The latency of FFPI ranges between 1.4 msec (70 dB HL) and 3.5 msec (40 dB HL), while FFP2 latency varies between 6.8 msec (70 dB HL) and 12.1 msec (30 dB HL). Significantly, FFPI is present when the scalp potential is differentially recorded between vertex and ipsilateral earlobe and is absent when the potential is recorded between vertex and contralateral earlobe. This suggests that the generator of FFPI is located close to the ipsilateral ear and below the level of binaural interaction (e.g. cochlear nucleus or cochlear nerve).

Huis in't Veld *et al* (1977a) report recording FFR with a latency of approximately 3 msec (90 dB SPL). This potential is observed when FFR is recorded differentially between mastoid and ipsilateral neck. However, it is not clear whether the generator of this potential is the same as FFPI. Because of its nonlinear behavior as a function of frequency, Huis in't Veld *et al* (1977a) suggest that this FFR may have several sources, including the CM, CN, and cochlear nerve. They also report an FFR of even earlier latency (approximately 2 msec) when the response is recorded differentially between vertex and ipsilateral mastoid. Because of these complications, Huis in't Veld and colleagues recommend recording FFR only between vertex and a site of demonstrated electrophysiological neutrality, such as the (contralateral) neck.

1.73 TONOTOPIC SPECIFICITY OF SCALP-RECORDED FREQUENCY-FOLLOWING RESPONSE

At present, there is no generally accepted electrophysiological technique to assess hearing function below 2 kHz. Onset-type responses such as the BSER and the gross action potential primarily reflect the activity of units tuned to frequencies above 2 kHz. As a consequence, there is considerable interest in determining whether FFR provides an accurate method of monitoring the activity of low-frequency neurons.

Because of the asymmetrical pattern of cochlear excitation, a significant number of high-frequency units will respond to high-intensity, low-frequency stimulation. Only at relatively low signal levels (<40 dB SL) can a tonotopically-discrete pattern of activity be assured. Unfortunately, because of the low amplitude of the far-field recorded brainstem response (approximately 0.25 uv) relative to the cortical EEG, FFR cannot be recorded from the human scalp at very low intensities. Its threshold is approximately 35-40 dB SL. Consequently, there is some doubt about the degree to which the response reflects the activity of low-frequency units.

A number of investigators have attempted to infer the tonotopic specificity of scalp-recorded FFR by means of masking studies or recordings from patients with frequency-selective hearing losses.

(a) Masking Studies

At low signal levels, there is little doubt that scalp-recorded FFR is generated by low-frequency units. De Boer et al (1977) presented 500 Hz tone bursts at 50-60 dB SPL in the presence of broadband noise of variable bandwidth (250-4000 Hz or 800-4000 Hz). Masking of the FFR occurred only when the noise band encompassed the signal frequency. Yamada et al (1979) obtained comparable results for low-intensity tones.

As the intensity of the tone increases, the masking pattern becomes more complex. Huis in't Veld et al (1977b) presented a 500 Hz tone burst at 83 dB p.e. SPL in the presence of noise bands (300 Hz wide) with center frequencies ranging from 250-4000 Hz. Noise bands centered below 2 kHz generally reduced the amplitude of the FFR by a significant amount. Noise centered at 3 kHz and 4 kHz had a negligible effect on the FFR. They conclude that FFR to high-intensity, 500 Hz tones is largely the product of neurons tuned to frequencies below 2 kHz. Studies using broadband noise have generally reached similar conclusions (Moushegian et al, 1978; Gardi and Merzenich, 1979).

At certain levels and cutoff frequencies, masking noise has paradoxical effects. For example, several investigators (Davis and Hirsh, 1976; Huis in't Veld et al, 1977b; de Boer et al, 1977) have observed that FFR is often enhanced in amplitude and "smoother" in the presence of high-pass noise (low-frequency cutoff=1.5 or 2.0 kHz). This is particularly evident for signal levels of 60 dB SL or greater. The origin of this paradoxical pattern is not clear. De Boer et al (1977) and Davis and Hirsh (1976) suggest these effects are indicative of multiple cochlear sources for FFR, each with its own phase vector. Under appropriate conditions, two or more sources can occur in counter phase and thereby cancel each other out. When one of these sources is masked, the result is an enhancement of the unmasked FFR. Consistent with this hypothesis is Davis and Hirsh's observation that the latency of FFR usually increases by about 1 msec when the cutoff frequency of the noise is around 2 kHz. Huis in't veld et al (1977b), however, suggest the locus of such effects is at a higher level, in view of the fact that such paradoxical masking is not observed with a mastoid derivation.

(b) Hearing Loss Studies

A number of studies indicate that FFR to low-frequency tones is recordable despite significant high-frequency hearing loss (> 2 kHz), (Sohmer et al, 1977; Huis in't Veld et al, 1977b; Moushegian et al, 1978; Yamada et al, 1978). For example, Yamada et al (1978) recorded FFR from patients suffering from the following types of hearing loss: (1) pure high-frequency loss; (2) severe high-frequency loss mixed with moderate low-frequency loss; (3) severe hearing loss across the spectrum; and (4) low-frequency loss. FFR threshold was significantly elevated for all but the high-frequency loss subjects. Moreover, for each group, the elevation in FFR threshold was correlated with the degree of low-frequency loss.

In summary, the tonotopic specificity of scalp-recorded FFR is highly dependent on signal intensity. At low-to-moderate levels (<55 dB SL), scalp-recorded FFR is generated primarily by a relatively small population of low-frequency neurons. At higher intensities, the tonotopic bandwidth of the activated population expands to encompass a considerably larger number of units. At high intensities (>70 dB SL), the tonotopic affiliation of contributing units may extend to two octaves above the frequency of the stimulus. As a consequence, the tonotopic specificity of FFR can only be assured when stimulus levels are below 55 dB SL per component.

1.74 AMPLITUDE BEHAVIOR OF FREQUENCY-FOLLOWING RESPONSE AS A FUNCTION OF SIGNAL LEVEL AND SIGNAL-TO-NOISE RATIO

The threshold of scalp-recorded frequency following response is approximately 45-50 dB SPL. In contrast, FFR can be recorded from the auditory brainstem nuclei of the cat at significantly lower signal levels (approximately 20 dB SPL for a 500 Hz tone). In part, this disparity in FFR threshold may be attributable to the signal amplitude loss characteristic of far-field recordings. There can be as much as a thousand-fold decrease in signal size from depth to scalp responses in simultaneous recording of FFR in cat (the depth-recorded response is often of sufficient magnitude as to preclude the necessity of signal-averaging). An additional factor is the increased noise from cortical neural activity and muscle potentials.

Above this threshold level, the amplitude of scalp FFR grows linearly with increasing signal intensity up to levels approaching 65-75 dB SPL (Marsh *et al*, 1975: Figures 3 and 4). Beyond this level, FFR amplitude asymptotes or diminishes due to middle ear muscle activation and intracochlear distortion.

The growth of FFR amplitude with increasing signal level is probably due to two factors. One involves the increase in discharge rate of auditory neurons synchronized to the signal. The second reflects the fact that the number of units synchronized to a signal increases as a function of intensity. Because of the unidirectional course of the traveling wave, high-frequency units will discharge prior in time to low-frequency units. Consequently, growth in FFR amplitude due to the recruitment of high frequency neurons should be reflected in the onset latency of the response. Scalp-recorded FFR to a 500 Hz tone burst presented over a range of stimulus intensities is illustrated in Figure 1.17. Notice that the onset latency does not appreciably change over a 20 dB range and that the phase of the FFR is uniform over this same range (Marsh *et al*, 1975). Huis in't Veld *et al* (1977b) have also shown that the latency of FFR to simple tones does not appreciably change over a fairly broad range of intensities. This suggests that FFR is generated primarily (though perhaps not exclusively) by units tuned to the signal frequency. Interestingly, Anderson *et al* (1971) observed that the phase of single unit synchronization remains constant as a function of intensity only for those units tuned to the stimulus frequency.

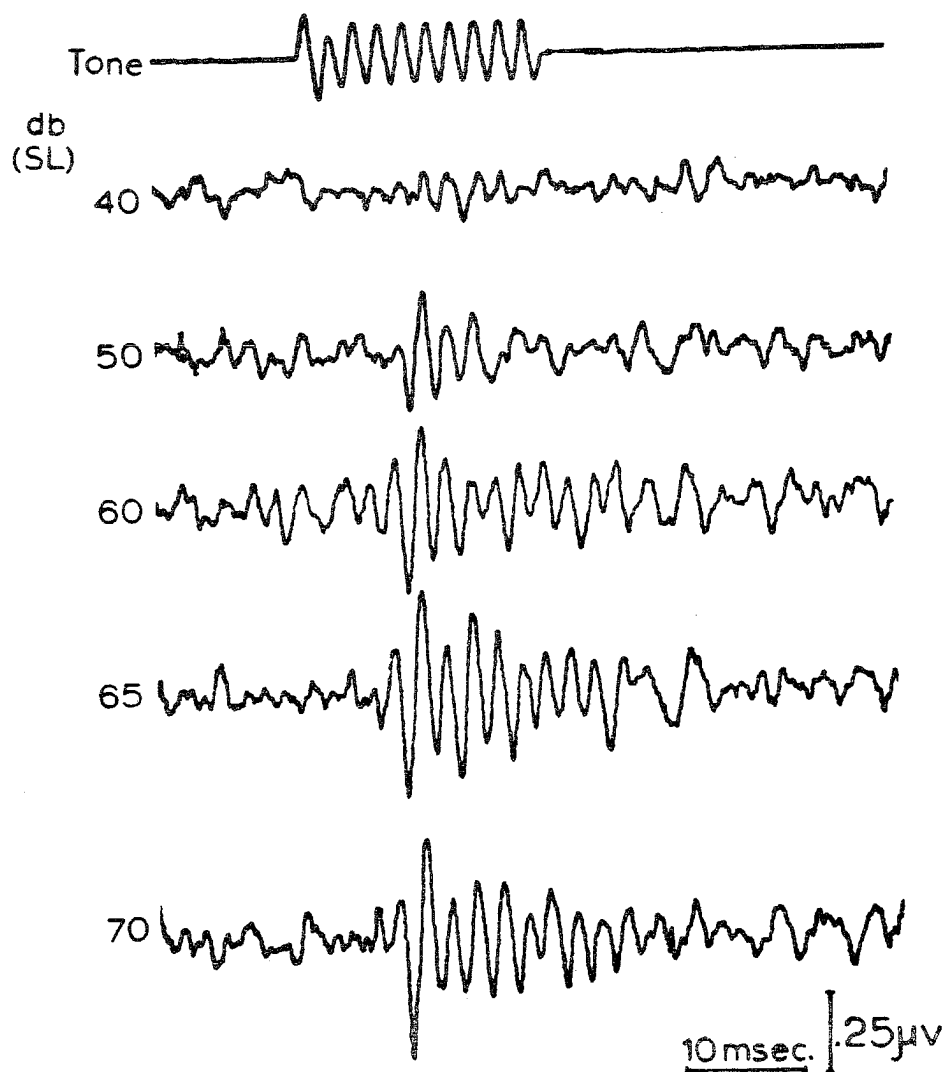


Figure 1.17 FFR for a single subject in response to a 500 Hz tone presented over a range of stimulus intensities. Stimulus level is indicated in terms of sound pressure level above subject threshold (SL). Notice that the latency of the response does not appreciably change over a 20 dB range. Averages consist of 2000 responses.

The amplitude of scalp-recorded FFR is significantly reduced in the presence of bandpass noise (100-900 Hz) of sufficiently high intensity. Marsh *et al* (1975) have shown that FFR to a 500 Hz tone undergoes progressive degradation for S/N ratios of less than 10 dB. Such degradation includes changes in onset latency, amplitude and phase. For S/N ratios below detection threshold (approximately -5 dB) FFR is hardly recognizable and its amplitude is diminished by 15-25 dB.

1.75 SUMMARY: BASIC PROPERTIES OF THE FREQUENCY-FOLLOWING RESPONSE

The frequency-following response is a neurogenic potential representing the integrated activity of auditory neurons phase-locked to the stimulus spectrum. Its period, when recorded in response to a pure tone, matches the period of the stimulus. It is recordable from auditory brainstem nuclei in the cat as well as from the scalp of human subjects. The response has been recorded to frequencies up to 5 kHz in the CN (Worden and Marsh, 1968; Marsh *et al*, 1972) and up to 2 kHz in the IC (Marsh *et al*, 1974). Scalp FFR has been recorded to frequencies up to 2 kHz (Moushegian *et al*, 1973; Greenberg *et al*, 1978; Greenberg and Marsh, 1979). The onset latency of scalp FFR (when recorded differentially between vertex and contralateral neck) is approximately 6-10 msec (depending on signal level), consistent with an upper brainstem origin (IC or LL). Masking studies indicate that the tonotopic affiliation of units contributing to the scalp-recorded response is confined largely to neurons tuned to the signal spectrum when stimulus level is near FFR threshold (~35-50 dB SL). At higher stimulus levels, the bandwidth of the tonotopic population contributing to the scalp response broadens. At high intensities (>70 dB SL), scalp FFR is generated by a broad population of units, whose tonotopic bandwidth may span two octaves or more.

The amplitude of scalp FFR increases proportional to signal level over a range of approximately 25-35 dB. At higher intensities, its amplitude asymptotes due to activation of the acoustic reflex and the presence of intracochlear distortion. The amplitude, phase, and onset latency of the scalp response change in the presence of high levels of masking noise centered around the stimulus frequency. The FFR is significantly reduced in amplitude at noise levels sufficient to mask the presence of the tone.

Because the frequency-following response provides a method of monitoring the periodic activity of upper brainstem neurons in human subjects, it offers an opportunity of studying the role of neural periodicity in the perception of such primary auditory attributes as pitch, timbre and loudness. In this section certain evidence was reviewed which suggests that the pitch of low-frequency tones is based on the temporal discharge pattern of auditory neurons and that this phase-locked activity is reflected in the periodic nature of the frequency-following response. So far, the frequency-following response has been discussed only in relation to single frequency stimuli.

However, as discussed in Chapters 1.21-1.29, the pitch of complex, biological signals is based on more than a single frequency. To what extent, then, does the frequency-following response reflect the pitch of complex tones?

CHAPTER II

GENERAL EXPERIMENTAL METHODS

The experimental methods described in this chapter apply to most of the experiments discussed in Chapters 3 - 7. Certain experimental variables, such as stimulus intensity and stimulus duration, are specified separately for each experiment.

2.1 SIGNAL GENERATION

Signals for all experiments were digitally-generated on an Interdata 74 minicomputer (Figure 2.1). Each frequency component was synthesized independently and its intensity digitally-adjusted to a standard reference level (77 dB SPL) using a Bruel and Kjaer 1/2" microphone coupled to an earphone with a 2 c.c. coupler. Individual frequency components were subsequently digitally summed and the composite signal's onset and offset smoothed with a 5 msec linear rise/decay function. The composite signal was then played through the acoustic system, recorded by the B & K microphone, amplified and redigitized. The calibration signal (an average of 50 or 100 presentations) was then spectrally analyzed (Figure 2.2) to insure that the sound pressure level of each frequency component was accurate.

2.2 SIGNAL PRESENTATION

The stimulus presentation system is also illustrated in Figure 2.1. The system consisted of the Interdata computer (sample rate=15,623 samples/sec), which retrieved a designated stimulus file off disc and outputted the signal through a D/A converter. The stimulus was subsequently bandpassed-filtered between 150 - 2,800 Hz (Krohn-Hite #330N; attenuation characteristic=24 dB/octave) to minimize any onset transient, amplified and presented through a pair of TDH 39 headphones (flat response between 500-5000 kHz; -6 dB/octave rolloff above and below these limits) encased within an mu-metal shielded container to eliminate electromagnetic radiation as a potential source of recording artifact. Sound pressure level was digitally controlled. Stimuli were presented at a rate of 9 or 15 per sec in blocks of 1000 (Note 2.1).

2.3 FREQUENCY-FOLLOWING RESPONSE RECORDING SYSTEM

The subject was seated in a comfortable, reclining chair within a shielded, sound-attenuated (approximately 50 dB attenuation at 1 kHz) chamber. The brainstem evoked response was recorded differentially between an active electrode placed on the vertex and a reference electrode located on either the side or the back of the neck. One side

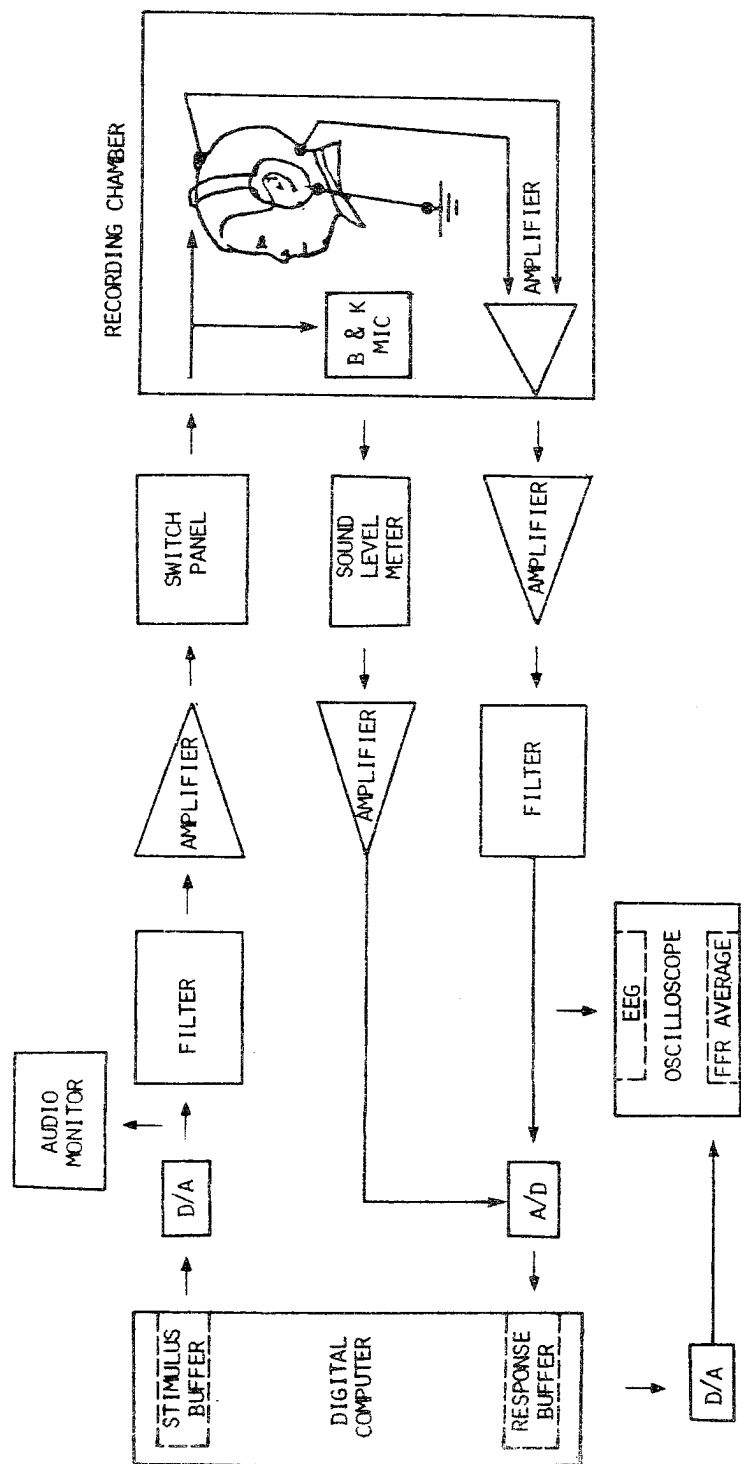


Figure 2.1 Stimulus Generation and FFR recording system.

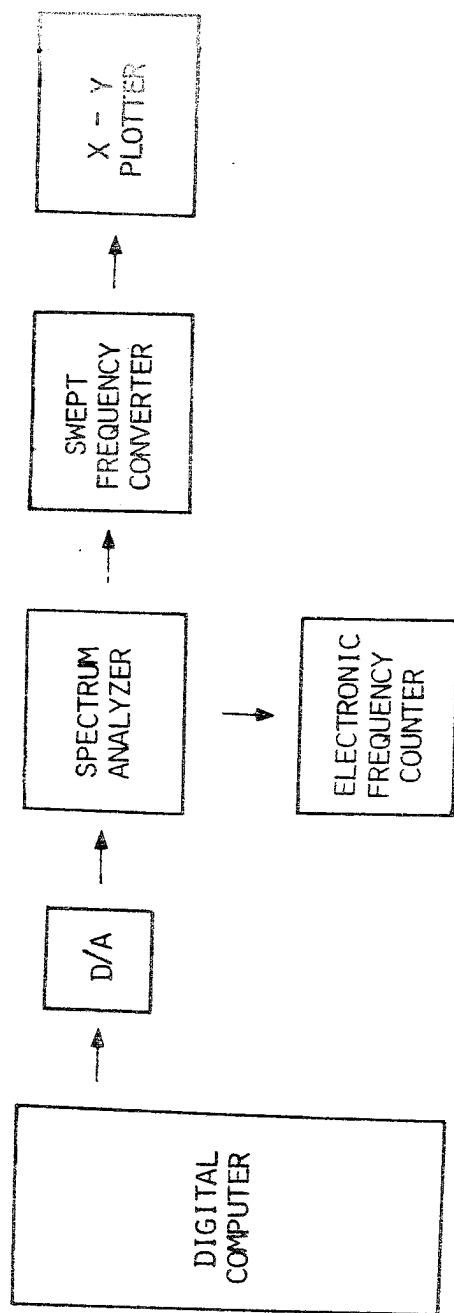


Figure 2.2 Analog spectrum analysis system.

of the neck always served as ground. Grass gold-plated electrodes and Grass electrode cream were used in all experiments. The quality of the scalp-electrode contact was monitored with an impedance meter using a 1 kHz signal. The electrode impedance was always under 5 kOhms (and usually under 2 kOhms). The brainstem signal was amplified in two stages. The first stage amplifier (1000x) was located inside the chamber close to the subject. The second stage amplifier (125x) was located outside the booth. Total gain was 125,000. The response was then bandpass-filtered (200-3,000 Hz) and sampled through the analog-to-digital (A/D) converter into the computer (Figure 2.1). Sampling of the vertex potential by the A/D converter was intercalated with the stimulus waveform outputted through the D/A converter, on a point-by-point basis for the entire 62.5 msec analysis epoch. This insured that stimulus onset was coincident with the beginning of the analysis epoch and that the entire stimulus was phase- and time-locked to the vertex potential within that epoch. Each response was summed with preceding responses in the 62.5 msec buffer until an average of 1000 responses was accumulated. The average was then normalized to compensate for DC drift and stored on disc for subsequent spectral analysis.

The background EEG and the averaged FFR were continuously monitored on an oscilloscope throughout the course of the experiment. The background level of the on-going EEG must be exceptionally low for optimum recording of the FFR. Consequently, superior averages were usually obtained while the subject slept (Note 2.2). Any indication of subject movement in the EEG resulted in temporary suspension of stimulus presentation and signal averaging. Recording resumed only after the EEG returned to its original background level. Recording sessions usually lasted between one and two hours.

2.4 SPECTRAL ANALYSIS OF THE FREQUENCY-FOLLOWING RESPONSE

FFR averages were analyzed in one of two ways. Usually, the response spectrum was measured using a series of digital narrow bandpass filters (bandwidth: approximately 30 Hz). These filters were centered around frequencies ranging from 244 Hz to 976 Hz, in 122 Hz steps. For other frequencies, the FFR spectrum was determined with a (Tektronics 1L5) analog spectrum analyzer and the analysis plotted on a calibrated x-y plotter (Figure 2.2). The analysis bandwidth of the spectrum analyzer was usually 20 Hz.

2.5 SUBJECTS

Subjects consisted of a total of sixteen volunteers (9 males and 7 females) between the ages of 8 and 57 years, with no known history of hearing loss. Subjects with elevated FFR or detection thresholds were excluded from further study.

CHAPTER III

PRELIMINARY STUDIES OF THE FREQUENCY-FOLLOWING RESPONSE TO THE "MISSING" FUNDAMENTAL

3.1 INTRODUCTION

In the introductory chapter it was shown that the low pitch of complex tones cannot be explained in terms of classical place or periodicity theories of pitch perception. The classical place theory of von Helmholtz (1863), which proposed that the low pitch of a tonal complex is based on the activity of neurons excited by the fundamental component, has proven untenable in light of Schouten's (1938) elegant cancellation study and subsequent confirmation of his basic finding by Licklider (1954) and Thurlow and Small (1955). Classical periodicity theory has fared no better. Schouten (1940c) demonstrated that the perception of low pitch could not possibly be based on the temporal pattern of neurons synchronized to the amplitude modulation pattern of the signal's waveform frequency. This he accomplished by showing that the pitch of a harmonic complex changes when its frequency components are shifted upwards by a small amount, despite the fact that the waveform modulation frequency remains unchanged. Schouten, having determined that the low pitch ("residue") of complex tones is based on neither the fundamental nor the waveform envelope suggested that the pitch of the complex is derived from the activity of auditory neurons synchronized to the temporal fine-structure of narrowly-spaced, aurally-unresolved harmonics of the signal (Chapter 1.22).

Schouten's model was ultimately challenged as a consequence of studies indicating that the most salient sensation of low pitch stems from widely-spaced, aurally-resolved components, rather than from the high-order harmonics (Plomp, 1967; Ritsma, 1967; Houtsma and Goldstein, 1972; Bilsen, 1973). These studies also raised some doubt about the role of neural periodicity in the perception of pitch, for it was assumed that there would be no significant degree of neural interaction between contiguous harmonics that are perceptually resolved by the ear. However, this assumption has proven incorrect. Rose et al (1969; 1974) and Evans (1978) have shown that the temporal discharge activity of peripheral auditory neurons tuned to frequencies between contiguous, aurally-resolved components does in fact reveal a significant degree of cochlear interaction and that this interaction provides a means by which the fundamental frequency of the stimulus complex could be represented in terms of the periodic discharge pattern of peripheral units.

However, it is not known whether this sensitivity to the fundamental period of the signal is preserved in the periodic discharge activity of neurons located at higher levels of the auditory pathway. We sought to investigate this issue by recording the frequency-

following response - which represents the aggregate periodic response of neurons located in the upper auditory brainstem pathway (Chapter 1.72) - during presentation of complex signals known to produce a salient percept of low pitch.

Frequency-following responses were recorded to three different stimulus pairs. Each pair consisted of a pure tone and a complex tone whose constituent frequencies were equivalent to the second through fifth harmonics of the pure tone frequency. Consequently, both signals would be expected to produce the same pitch despite that fact that they have no spectral component in common. Is the spectrum of FFR generated by the complex tone similar to the spectrum of FFR produced by the pure tone? Or does it primarily reflect the frequencies of the stimulus spectrum?

3.2 EXPERIMENTAL METHODS

Stimuli were (a) four-component complex tones, consisting of the second through fifth harmonics of a common fundamental (1. $f_0=244$ Hz - 488+732+976+1220 Hz; 2. $f_0=366$ Hz - 732+1098+1464+1830 Hz; 3. $f_0=488$ Hz - 976+1464+1952+2440 Hz) and (b) pure tones equal in frequency to these same fundamentals (244, 366, and 488 Hz). Each frequency component began in sine phase. Signal duration was 45 msec. Signal level was 70 dB SPL/component, presented binaurally at a rate of 9 per sec. FFR averages consisted of 2000 or 3000 responses, which were recorded differentially between vertex and the left side of the neck. All other aspects of the experimental protocol and recording procedure were as specified in Chapter 2.

The spectrum of the averaged responses was determined by passing the accumulated average through a digital-to-analog converter into a low-frequency analog spectrum analyzer and plotting the output of the analyzer on an x-y plotter (Figure 2.2).

3.3 EXPERIMENTAL RESULTS

A series of illustrative frequency-following responses to pure and complex tones is shown in Figure 3.1. Each pure tone signal generates a brainstem response whose period matches the stimulus period and whose primary spectral component is the same frequency as the stimulus. Thus, the 244 Hz tone generates a response with an energy maximum at 244 Hz, the 366 Hz signal generates a response whose spectral maximum is 366 Hz, etc. Each complex stimulus produces a response with a period equal to the fundamental frequency of the signal (i.e. equivalent to the period of the pure tone) and whose energy is concentrated at a frequency band one octave below the lowest stimulus component. This frequency corresponds to the low pitch of the stimulus and its amplitude is considerably greater than the amplitude of the spectral components associated with the stimulus components.

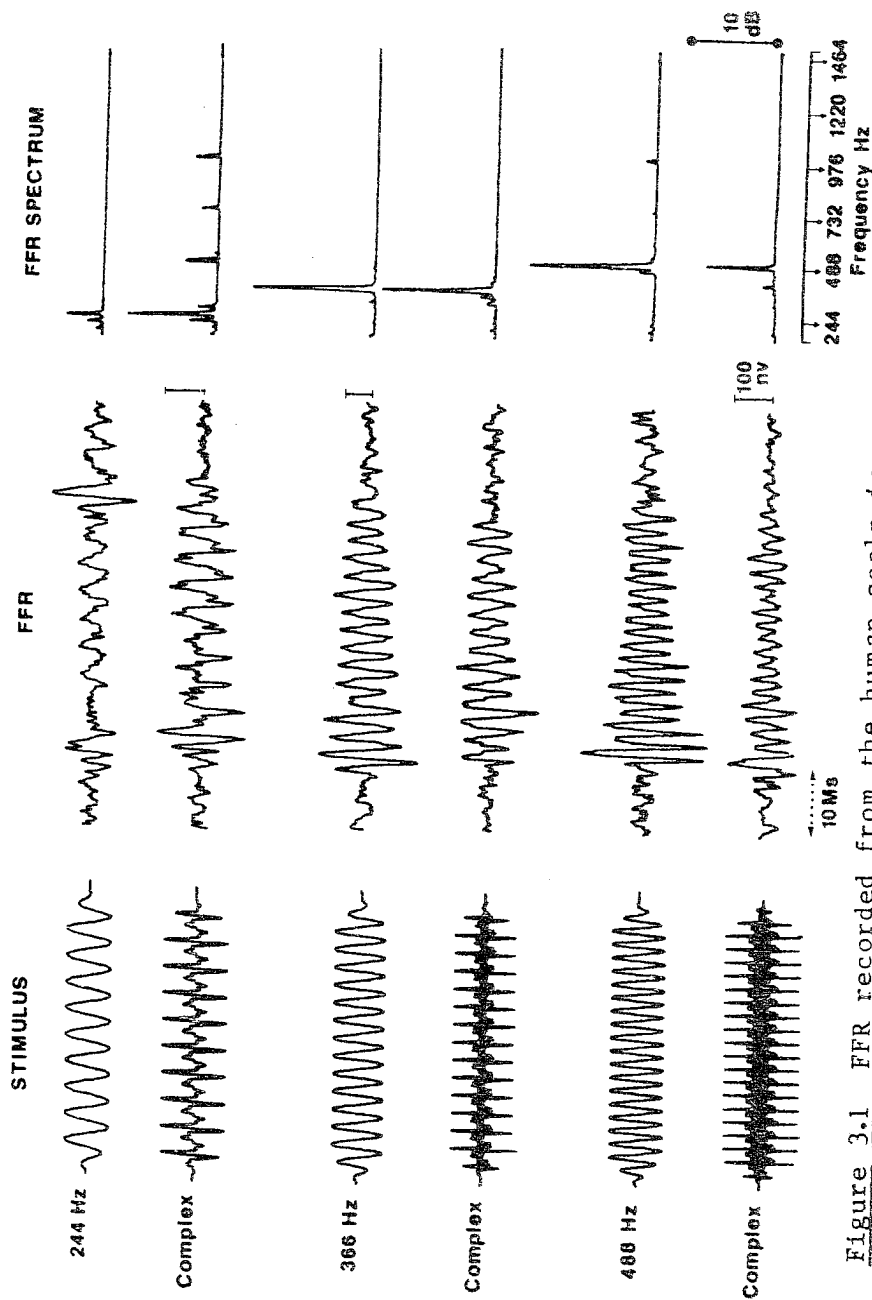


Figure 3.1 FFR recorded from the human scalp in response to pure and complex tones of variable frequency. Each stimulus pair consists of a pure tone (244, 366, 488 Hz) and a complex tone equivalent to the second through fifth harmonics of the pure tone frequency. The stimulus waveforms are shown in column 1. A representative FFR generated by each stimulus is shown in column 2 (analysis epoch = 63 msec). Column 3 shows the analog spectral analysis for each response. Note the similarity in spectra of FFR recorded to the pure tone and the complex. Stimulus level was 70 dB SPL for the pure tone stimuli and 82 dB SPL (70 dB SPL/component) for the complex signals. Averages consist of 2000 responses for the bottom and top set of FFR. The middle averages consist of 3000 responses.

3.4 DISCUSSION

Complex tones generate FFR whose predominant energy lies one octave below the lowest frequency component of the stimulus and which corresponds to the fundamental frequency of the signal. It is tempting to conclude from this simple demonstration that FFR recorded from the scalp of human subjects does indeed reflect periodic neural activity related to the perception of low pitch. However, as will become evident in the chapters that follow, it may be necessary to qualify this conclusion in several important respects.

For example, it is conceivable that the spectrum of the scalp response to a complex tone reflects the presence of a non-linear distortion product equal to the frequency of the "missing" fundamental. In principle, it should be possible to eliminate combination tones as a possible source of scalp FFR by recording at signal levels below 40 dB SPL. However, because the threshold of scalp is approximately 45-60 dB SPL/component, it is usually necessary to record responses at signal levels (70 dB SPL/component) where combination tones of substantial amplitude may be generated (Goldstein, 1967).

3.5 FREQUENCY-FOLLOWING RESPONSE TO THE "MISSING" FUNDAMENTAL IN THE PRESENCE OF LOW-FREQUENCY MASKING NOISE

If FFR to complex tones is an artifact of intracochlear distortion the amplitude of the scalp response should be significantly reduced in the presence of bandpass noise centered at the frequency of the (missing) fundamental. This is because any combination tone product generated through non-linear intracochlear interaction of two or more stimulus components propagates to the appropriate tonotopic location along the cochlear partition (Goldstein and Kiang, 1968; Greenwood *et al*, 1976; Smoorenburg *et al* 1976; Buunen and Rhode, 1978). Thus, auditory nerve fibers respond to a combination tone as if it were an externally applied stimulus. Consequently, band-limited noise of sufficient intensity will perceptually and physiologically "mask" the combination tone. Licklider (1954) and Thurlow and Small (1955) used low-frequency band-pass noise to demonstrate perceptually that the "missing" fundamental is not an artifact of intracochlear distortion. (Chapter 1.29 (b)). Using a similar experimental procedure, we sought (in collaboration with J.C. Smith, J.T. Marsh, and W.S. Brown [Smith *et al*, 1978; Smith, 1976]) to demonstrate that the FFR to the "missing" fundamental is likewise not an artifact of non-linear distortion.

This hypothesis was tested by comparing the scalp FFR recorded to a pure tone (366 Hz) (Note 3.1) and a complex stimulus whose components were equivalent to the second through fifth harmonics of a 366 Hz fundamental (732+1098+1464+1830 Hz) to FFR recorded to the same stimuli in the presence of bandpassed noise (approximately 200 Hz wide) centered at 366 Hz. Signal level was 81 dB SPL for the simple tone and 69 dB SPL/component (81 dB SPL overall sound pressure level) for the

complex. Overall SPL of the noise was 86 dB SPL (spectrum level=40 dB SPL). Brainstem responses were recorded (differentially between vertex and one earlobe) from seven subjects, all of whom reported no prior history of hearing loss.

To confirm that the complex stimulus produced a salient low pitch corresponding to the missing fundamental, each subject listened to the complex tone in one ear and to a pure tone generated by a separate audio oscillator in the other, while adjusting the oscillator to match the low pitch of the complex signal. All subjects adjusted the oscillator to within 2 Hz of the 366 Hz missing fundamental.

In the absence of masking noise, FFR generated by the complex tone is similar in many respects (latency, spectrum, amplitude) to FFR produced by the simple tone (Figure 3.2: rows 1 and 2). FFR generated by the same stimuli recorded in the presence of the noise shows a different pattern (Figure 3.2: rows 3 and 4). FFR to the pure tone is reduced in amplitude by approximately 8 dB. If FFR recorded to the complex tone were the product of a distortion product, the noise should reduce the amplitude of the response by an amount comparable to the pure tone response. However, this is not the case. The amplitude of the complex tone FFR is reduced only by about 1 dB - an amount that is likely to be caused by the noise masking the spectral components of the stimulus complex. Comparison of these mean masking decrements shows the FFR to the pure tone to be significantly masked [$t(1,6) = 2.26, p < .05$] while FFR to the complex tone is not [$t(1,6) = 0.44$]. This comparison indicates that the FFR to the missing fundamental is not the result of non-linear distortion, but rather reflects the activity of auditory neurons tuned to frequencies above the band of masking noise.

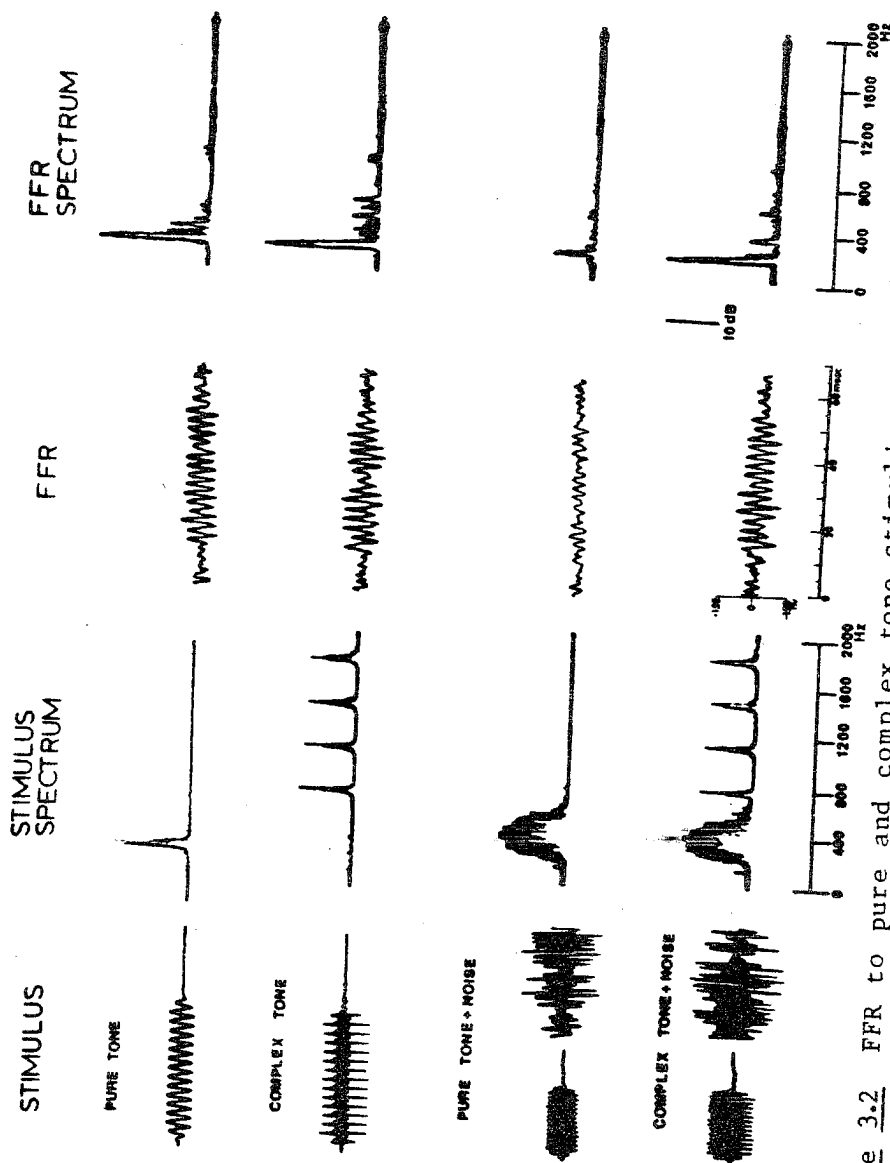


Figure 3.2 FFR to pure and complex tone stimuli presented in silence and in the presence of masking noise. Column 1 shows the stimulus waveforms for the pure and complex tone stimuli. The pure tone frequency was 366 Hz. The complex tone consisted of the second through fifth harmonics of a 366 Hz fundamental. The bottom two rows of this column show pure and complex tones, plus a 366 Hz-centered masking noise (bandwidth=200 Hz). Column 2 shows the analog spectrum analysis of the stimulus waveforms. Column 3 shows the composite averaged FFR of seven subjects. Column 4 is the analog spectrum analysis of the adjacent, composite FFR. Stimulus level was 81 dB SPL for both pure and complex stimuli.

CHAPTER IV

SPECTRAL BASIS OF THE FREQUENCY-FOLLOWING RESPONSE TO THE "MISSING" FUNDAMENTAL

4.1 INTRODUCTION

The low pitch of complex signals stems from a relatively narrow region of the spectrum spanning the first eight to ten harmonics of the fundamental (Smoorenburg, 1970; Bilsen 1973; Houtsma and Goldstein, 1972). The limits of this "existence" region is often attributed to a constraint on the operation of the central pitch processor which limits analysis to those components spectrally resolved in the auditory periphery. However, this interpretation is subject to question in view of the fact that a similar upper limit obtains when the stimulus components are presented to separate ears (Houtsma and Goldstein, 1972; Bilsen, 1973). This finding implies that the minimum separation of components capable of generating a low pitch reflects some auditory property other than peripheral frequency analysis. One possibility is that the frequency resolution of the central auditory pathway is similar to that of the auditory periphery (Scharf et al, 1976). Another possibility, by no means incompatible with the first, is that the existence region represents that region of the spectrum which can be encoded in terms of the temporal discharge patterns of auditory neurons (Ritsma, 1962).

Related to the existence region is the concept of "dominance." The pitch of a broadband signal is determined principally by the value of spectral components lying within a narrow frequency range, largely irrespective of the relative amplitude or bandwidth of the portion of the signal lying outside this region (Ritsma, 1967; Ritsma and Bilsen, 1970). The actual limits of this "region of dominance" varies among subjects and is also constrained by fundamental frequency, as shown in Table 4.1 (adapted from Ritsma, 1967).

Fundamental Frequency (Hz)	Spectral Regions (4 Observers)			
	01	02	03	04
100	300 - 400	400 - 500	300 - 500	400 - 500
200	600 - 800	800 - 1000	800 - 1000	1000 - 1200
400	800 - 1200	800 - 1200	800 - 2000	1600 - 2000

TABLE 4.1 Region of Dominance for Low Pitch
(Three Observers) [Data from Ritsma, 1967]

Table I indicates that the dominance region encompasses an area between two and five times the fundamental frequency of the signal and that there is a slight negative correlation between fundamental frequency and the harmonic rank of the dominant components (i.e. the

higher the fundamental, the lower the rank of the dominant harmonics). The dominance principle has been verified for a number of different experimental stimuli, including (1) complex tones (monotic) - Plomp, 1967; Ritsma, 1967; Bilsen, 1973; Houtsma and Goldstein, 1972; Patterson and Wightman, 1976; (2) complex tones (dichotic) - Houtsma and Goldstein, 1971; 1972; Bilsen, 1973; (3) comb-filtered noise (monotic) - Ritsma and Bilsen, 1970; Wilson, 1974; Bilsen, 1977; Yost and Hill, 1978; comb-filtered noise (dichotic) - Bilsen and Goldstein, 1974; Bilsen, 1977; (4) jittered pulse trains - Rosenberg, 1966). The principle of dominance has also been verified using a number of different experimental procedures. These include (1) pitch discrimination - e.g. Plomp, 1967; Ritsma, 1967; Ritsma and Bilsen, 1970; Yost and Hill, 1978; (2) pitch matching - Patterson and Wightman, 1976; and (3) musical interval recognition (monotic and dichotic) - Houtsma and Goldstein, 1971; 1972). The region of dominance, as delineated by these studies, encompasses the spectral region between 300 and 2000 Hz. The region's upper boundary implies that as the fundamental frequency increases, the rank of the dominant harmonics decreases, consistent with the trend in Table 4.1. It also implies that the upper limit of low pitch is between 700 and 1000 Hz (consistent with Ritsma's, 1962 estimate of 800 Hz).

Explanations of the region of dominance fall into two basic categories. Accounts based on place mechanisms emphasize the importance of wide spacing between components (Terhardt, 1970; Wightman, 1973a) and the fact that the pitch of the stimulus complex is based on the pitch of the individual spectral components (Terhardt, 1975; van den Brink, 1970; 1975). However, place models make a number of predictions that are incompatible with the experimental evidence. One corollary of the place model is that the strength of a signal's pitch increases as the spacing between spectral peaks widens. This corollary predicts that, ceteris paribus, the dominant harmonics will be the lowest components in the spectrum. This prediction is incorrect. The lowest components are dominant only for relatively high fundamentals (Chapter 1.29 (d)). Another prediction implied by this corollary is that the strength of low pitch increases proportional to the fundamental frequency. However, the pitch generated by signals with high fundamentals actually declines in strength above approximately 250 Hz (Ritsma and Bilsen, 1970) to 500 Hz (Yost and Hill, 1978; 1979).

Another proposed explanation of the dominance region is based on the ability of auditory units to encode the signal spectrum in terms of the temporal pattern of neural discharge (Goldstein, 1978). The precision of neural phase-locking is greatest for those frequencies (600-2000 Hz) which lie in the region of dominance (Rose et al, 1967: Figure 1). If one assumes that the frequency of the individual spectral components is analyzed in terms of the temporal firing pattern of auditory units (Goldstein and Srulovicz, 1977), the basis of the dominance principle lies in the precision with which the signal spectrum is measured (see Chapters 1.32 and 4.4 for further details concerning Goldstein's model).

Goldstein's explanation is consistent with many features of the dominance principle, including the dominance of low components for dichotically-presented signals (Houtsma and Goldstein, 1972; Bilson, 1973), the positive correlation between pitch ambiguity and rank of the stimulus harmonics, and the intensity independence of the dominance principle. However, Goldstein's model fails to answer a number of key questions. What is the physiological basis for the operation of the central pitch processor? If the central processor operates on the temporal firing pattern of unit activity, is the output of the pitch extractor also encoded in terms of neural periodicity? Why is the low pitch of complex tones restricted to fundamentals below 1 kHz? And why is the most salient pitch generated by signals with fundamentals between 200-500 Hz?

The FFR represents the aggregate periodic activity of units located in the upper auditory brainstem pathway and as such may serve as an index of neural periodicity with which to examine the neural mechanisms underlying the perception of low pitch. The frequency-following response was recorded to complex tones of variable harmonic rank and fundamental frequency in order to determine whether the existence region and region of dominance for low pitch might be correlated with some dimension of the scalp response.

4.2 EXPERIMENTAL METHODS

Stimuli consisted of two-component, 40 msec tone bursts, equivalent to successive harmonics of a common fundamental. The common fundamental ranged between 244 and 488 Hz. Stimulus specification varied as a function of the fundamental frequency in the following way:

Fundamental	Harmonic Rank	Stimulus Frequencies
244 Hz	1-2,...,7-8	244+488,...,1952+2196 Hz
366 Hz	1-2,...,5-6	366+732,...,1830+2196 Hz
488 Hz	1-2,...,4-5	488+976,...,1952+2440 Hz

Stimuli were presented binaurally at a rate of 9 per second. Stimulus level was 70-80 dB SPL/component. All harmonics were presented at the same SPL for any given subject. Others aspect of stimulus generation and presentation were as specified in Chapters 2.1 and 2.2.

Responses were recorded differentially between vertex and either the side or the back of the neck. Ground was placed on one side of the neck. Averages consisted of 2000-3000 responses for each stimulus.

Subjects consisted of 7 individuals ranging between the ages of 12 and 57. None had any previous history of hearing loss. The results of the 244 Hz series is based on 5 subjects. The 366 Hz series is based on 4 subjects, and the 488 Hz data is based on 3 subjects.

4.3 EXPERIMENTAL RESULTS

Figure 4.1 illustrates a set of responses to a series of two-component stimuli of variable harmonic rank for a single subject. Displayed are responses for each of eight pairs of two-harmonic stimuli (fundamental frequency = 244 Hz) as well as the relative amplitude of each response. Several features of these data are of interest:

- (a) The primary spectral component of each response is equal to 244 Hz, a frequency which corresponds to both the period of the stimulus waveform and the pitch of the tonal complex. Each response, except that generated by the eighth and ninth harmonics, exhibits a marked periodicity equal to the fundamental frequency of the stimulus.
- (b) The amplitude of the responses varies as a function of the harmonic rank of the stimulus components. The largest response is generated by the third and fourth harmonics, rather than by the fundamental and second harmonic. The relatively small response to the lowest stimulus makes it unlikely that responses to the other stimuli are generated by combination tones (Chapter 3). For responses to stimuli higher than the third and fourth harmonics, FFR amplitude decreases with progressively increasing harmonic rank. The periodic quality of the response all but disappears for the FFR generated by harmonics eight and nine.

The amplitude behavior of these responses is similar in a number of respects to psychophysical measures of pitch strength. First, the "existence region" of periodic FFR encompasses the first eight harmonics - essentially the same as the psychophysical "existence region" (exclusive of combination tones) for low pitch (Smootenburg, 1970; Houtsma and Goldstein, 1972; Bilsen, 1973; Buunen *et al.*, 1974). Moreover, the fundamental and second harmonic generate a substantially smaller response than other pairs of low harmonics. Thus the major component of FFR periodicity is unlikely to be produced by combination tones. Consequently, the periodicity evident in the FFR reflects a true "missing" fundamental. The low harmonics, exclusive of the fundamental, generate a substantial response. Response amplitude falls off sharply above the sixth harmonic. Thus the amplitude pattern appears to parallel in large measure, the limits of aural frequency resolution. However response amplitude is not entirely predictable by the harmonic spacing of the stimuli. The largest response is generated by the third and fourth harmonics, not by the second and third (or the fundamental and second) as one would expect if aural frequency resolution were the sole determining factor in the generation of FFR periodicity.

- (c) The onset latency of the response progressively shortens as stimulus frequency increases. This suggests that each response is generated by a slightly different population of neurons and that these neurons are tuned to progressively higher frequencies. Consistent with this interpretation is the fact that the latency difference between the responses to the highest and lowest stimuli

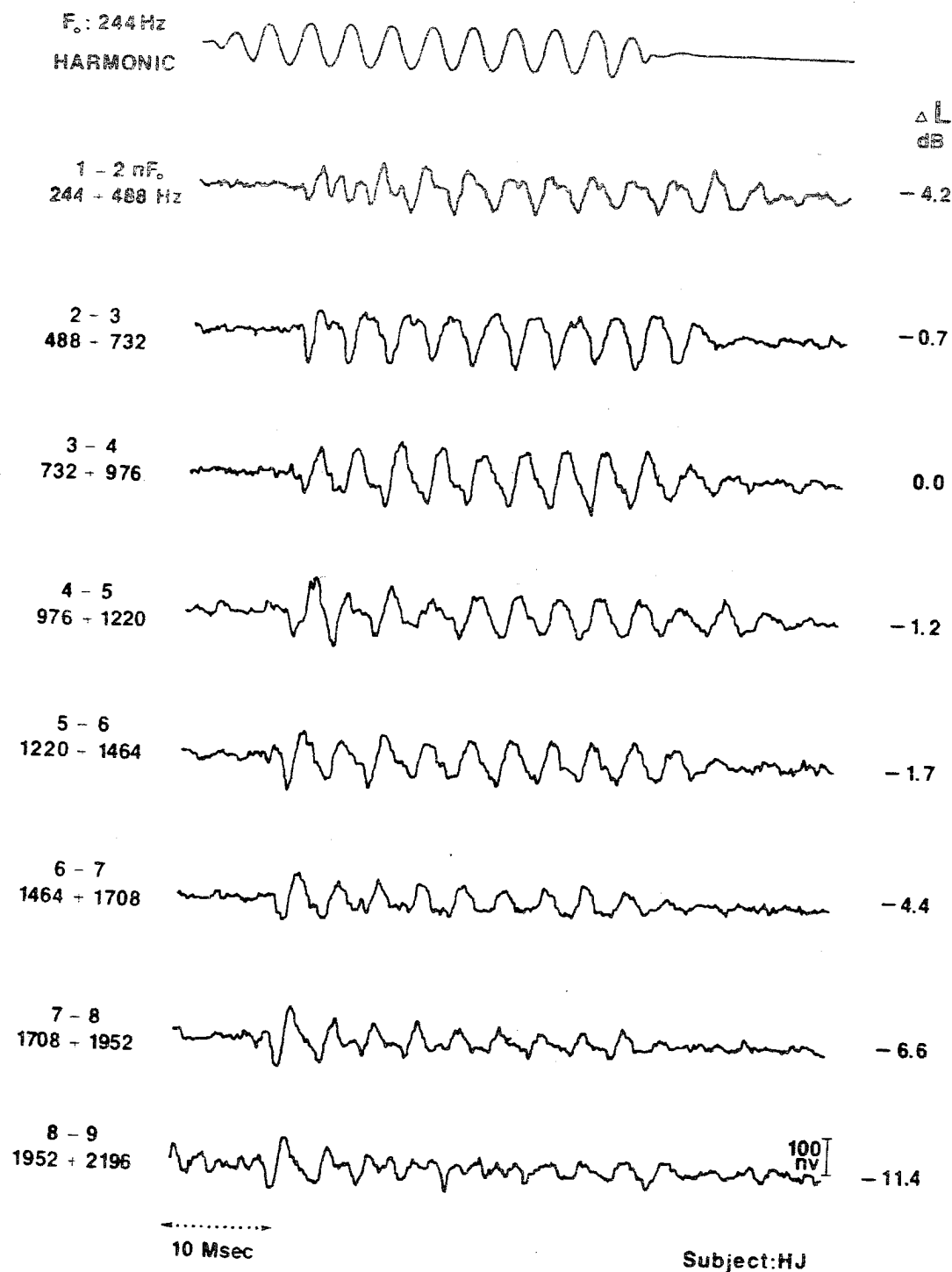


Figure 4.1 FFR to a series of two-component signals of variable harmonic rank. The fundamental is 244 Hz. The harmonic ranks and stimulus frequencies are indicated in the left-hand column. The relative level of the fundamental component (244 Hz) in the FFR spectrum is indicated in the right-hand column for each response. The spectrum level of the response generated by the third and fourth harmonics serves as the reference. Stimulus intensity was 73 dB SPL/component. Each average consists of 2000 responses.

is approximately 1.25 msec. This interval corresponds closely to empirical travel time functions for excitation of single units with corresponding characteristic frequencies in both the cochlear nerve (of the squirrel monkey: Greenwood, 1977) and the AVCN (of the cat: Gibson *et al*, 1977). There is no reason why the travel time excitation functions for units in the upper auditory brainstem should differ appreciably from those in more peripheral nuclei.

The amplitude of the FFR evoked by each harmonic pair is summarized for five subjects in Figure 4.2. The amplitude pattern is similar to that illustrated in Figure 4.1. However, in contrast to the data shown in Figure 4.1, the amplitude of the fundamental component evoked by each of the two components (filled circles) is compared with the FFR generated by a four-component tone consisting of the second through fifth harmonics of the same fundamental. The amplitude of the fundamental component for the four-component response serves as the reference level for each subject in order to show that the amplitude of responses to the most potent two-component stimuli are not appreciably smaller than the response to a larger number of harmonics. This is important, because certain investigators (e.g. Evans, 1978) have suggested that the low pitch produced by a pair of harmonics is qualitatively different from (or at least, less salient than) tones composed of three or more harmonics. However, Bilsen (1973) has shown that in terms of pitch discrimination ability, there is no significant difference between a pair of "dominant" harmonics and a signal composed of twenty components. Consequently, it may be of considerable significance that the largest two-component responses (generated by the third and fourth harmonics) are often as large as (if not larger than) those evoked by the four-component signal.

As with the responses shown in Figure 4.1, the largest amplitude FFR is produced by the third and fourth harmonics, followed by the responses generated by the second and third, first and second, and fourth and fifth harmonics. There is a progressive decline in FFR amplitude for responses generated by harmonics above the third and fourth. For purposes of comparison, Bilsen's (1973) psychophysical measure of pitch strength for a comparable series ($F_0=200$ Hz) of two-component tones is also shown. It is apparent that Bilsen's psychophysical measure of pitch strength matches the FFR amplitude data very closely.

Figure 4.3 illustrates, for a single subject, responses to a series of tones with a fundamental frequency of 366 Hz. As with the 244 Hz stimulus series (Figure 4.1), the largest responses are generated by the lower harmonics. Also evident is the progressive decline in FFR amplitude with increasing harmonic rank. However, the largest response is generated by the second and third harmonics rather than by the third and fourth harmonics. Furthermore, periodic FFR is evident only for the first five harmonics.

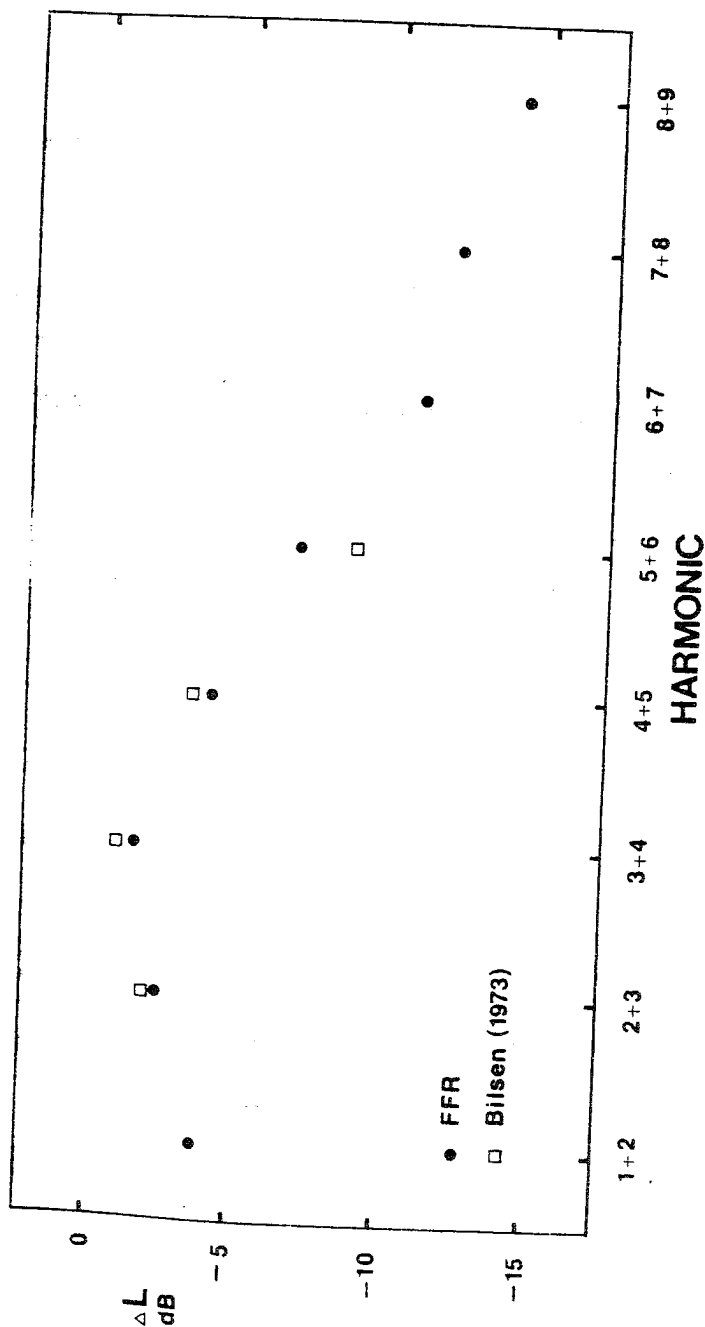


Figure 4.2 FFR amplitude as a function of the harmonic rank of the stimulus components. Plotted is the level of the fundamental component (244 Hz) in the spectra of FFR generated by two-component tones of variable harmonic rank. Reference level is the amplitude of the fundamental component in the FFR generated by a four-component tone consisting of the second through fifth harmonics of the same fundamental. Each data point (filled circles) represents the mean relative spectrum level of FFR recorded from four subjects. Psychoacoustic estimates of the strength of the pitch produced by analogous stimuli ($f_0=200$ Hz) are from Bilsen (1973).

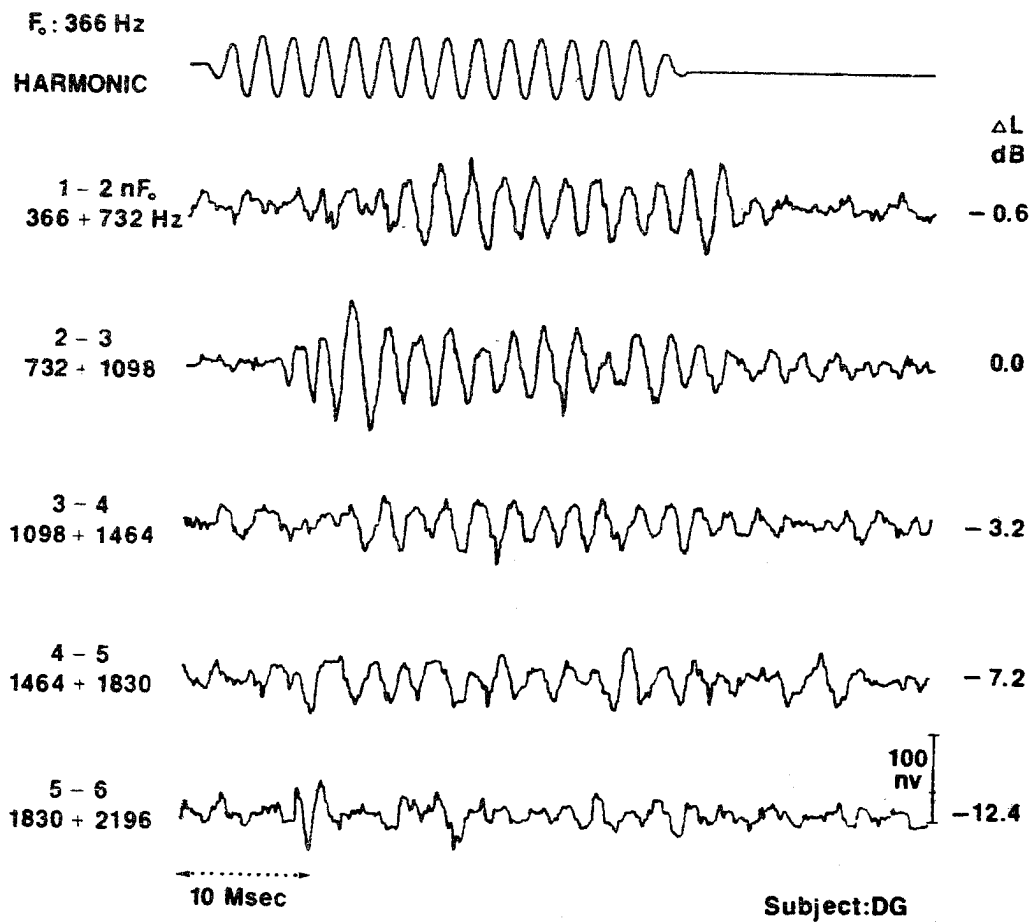


Figure 4.3 FFR to a series of two-component signals of variable harmonic rank. The fundamental is 366 Hz. Stimulus intensity = 70 dB SPL/component. Each average consists of 3000 responses.

Figure 4.4 shows FFR to a series of tones with a 488 Hz fundamental. This series of responses differs from the 244 Hz series to an even greater extent. The largest response is generated by the lowest harmonic pair and the periodicity of the FFR diminishes considerably above the third harmonic.

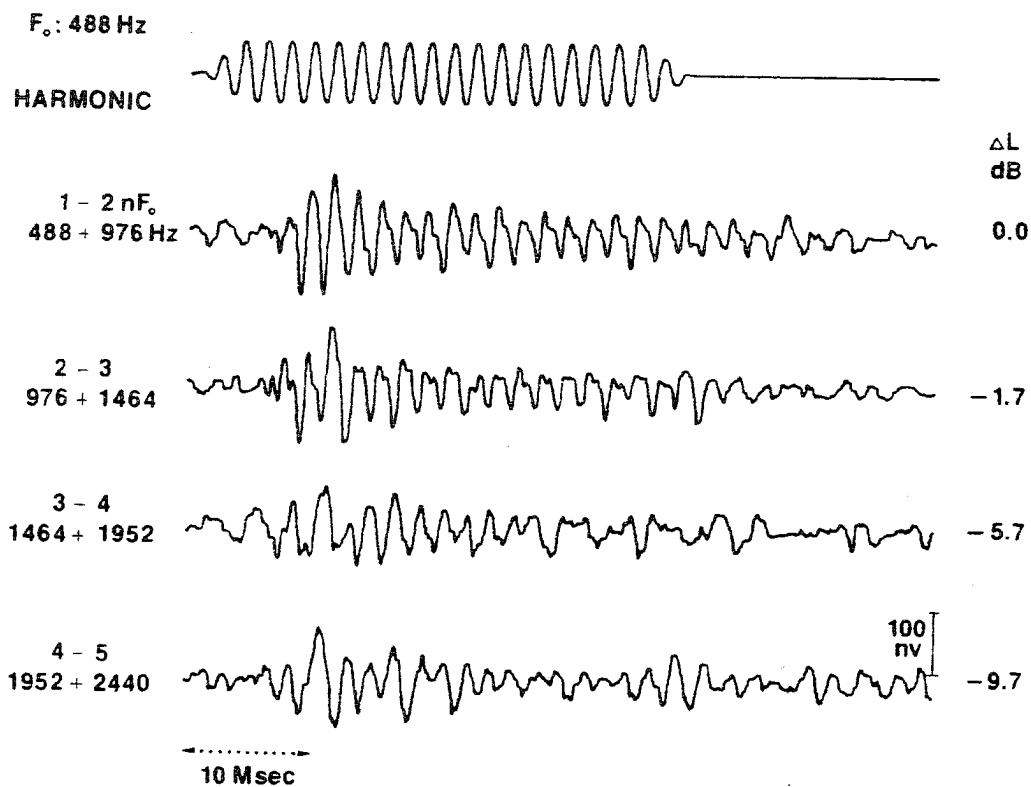
In Figure 4.5, the amplitude of each response (relative to the level of the response generated by the fundamental and second harmonic) is plotted separately for each fundamental frequency. Data points represent the mean for a group of subjects. It is evident from these data that the largest responses are generated by the lower harmonics, regardless of the fundamental frequency. However, the harmonic rank of the components generating the largest FFR is not the same for each fundamental. For tones with a 244 Hz fundamental, the largest response is evoked by the third and fourth harmonics. For the 366 Hz tones, it is the second and third harmonics which produce the biggest response. And for the 488 Hz series, the largest FFR is generated by the fundamental and second harmonic. There is clearly a negative correlation between fundamental frequency and the harmonic rank of the "dominant" components.

The pattern illustrated in Figure 4.5 suggests that the concept of dominance for the FFR applies to a spectral region rather than to a particular set of harmonics. In Figure 4.6 the data shown in Figure 4.5 is replotted in terms of the frequencies of the stimulus components. The data points represent the lower frequency of each two-component tone. When FFR amplitude is plotted in such a fashion, the curves for each fundamental series show a very similar pattern. The largest responses are generated by frequencies below 1 kHz. The region of dominance for the FFR lies between 500 and 1000 Hz. Above this region, there is a progressive decline in response amplitude that is relatively independent of the harmonic rank of the stimulus components.

4.4 DISCUSSION

These data suggest that the primary factor determining FFR amplitude is the frequency of the stimulus components rather than their harmonic rank. In particular, it appears that the largest FFR is generated by components with frequencies between 500 and 1000 Hz. There is reason to believe that this pattern has some significance for models of pitch perception. For the strongest pitch is produced by stimuli whose spectral components lie within this same region (Rosenberg, 1965; Ritsma, 1967; Plomp, 1967; Ritsma and Bilsen, 1970). This correspondence suggests that the mechanisms underlying the generation of FFR and the perception of pitch may be related.

Four different explanations of the dominance region for low pitch have been proposed. Terhardt (1970) has suggested that the lower harmonics are dominant for low pitch because the central pitch processor operates best on fully-resolved spectral information. Thus, the strongest pitch will be generated by components which are farthest



Subject: TM

Figure 4.4 FFR to a series of two-component signals of variable harmonic rank. The fundamental is 488 Hz. Stimulus intensity was 75 dB SPL/component. FFR waveforms represent the sum of 3000 responses.

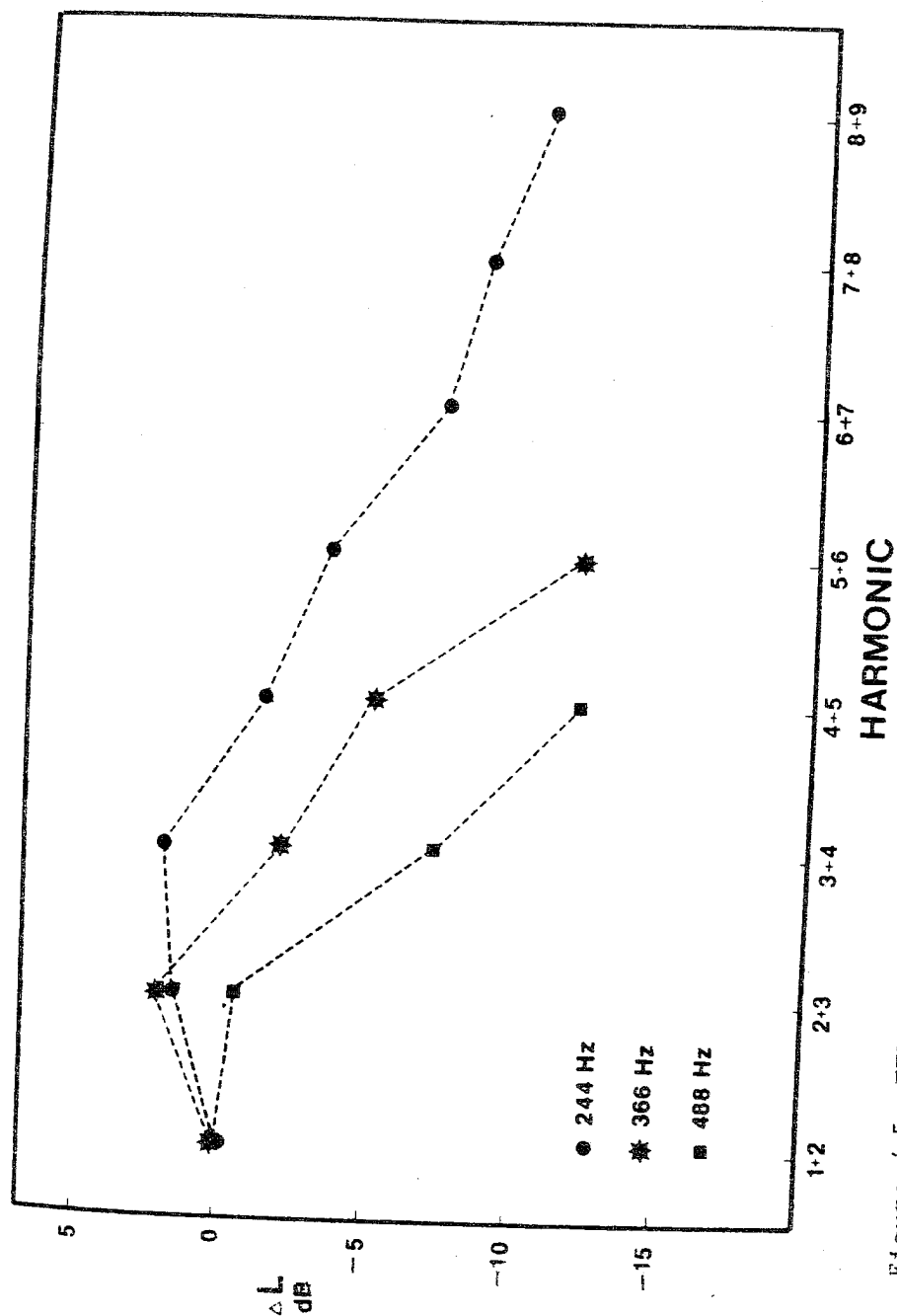


Figure 4.5 FFR amplitude as a function of harmonic rank for three different series of two-component stimuli. The level of the fundamental component of the FFR spectrum (relative to the amplitude of the FFR produced by the fundamental and second harmonic) is plotted for three different fundamental frequencies. Data for the 244 Hz series is based on 5 subjects, while data for the 366 Hz and 488 Hz series are based on 4 and 3 subjects, respectively.

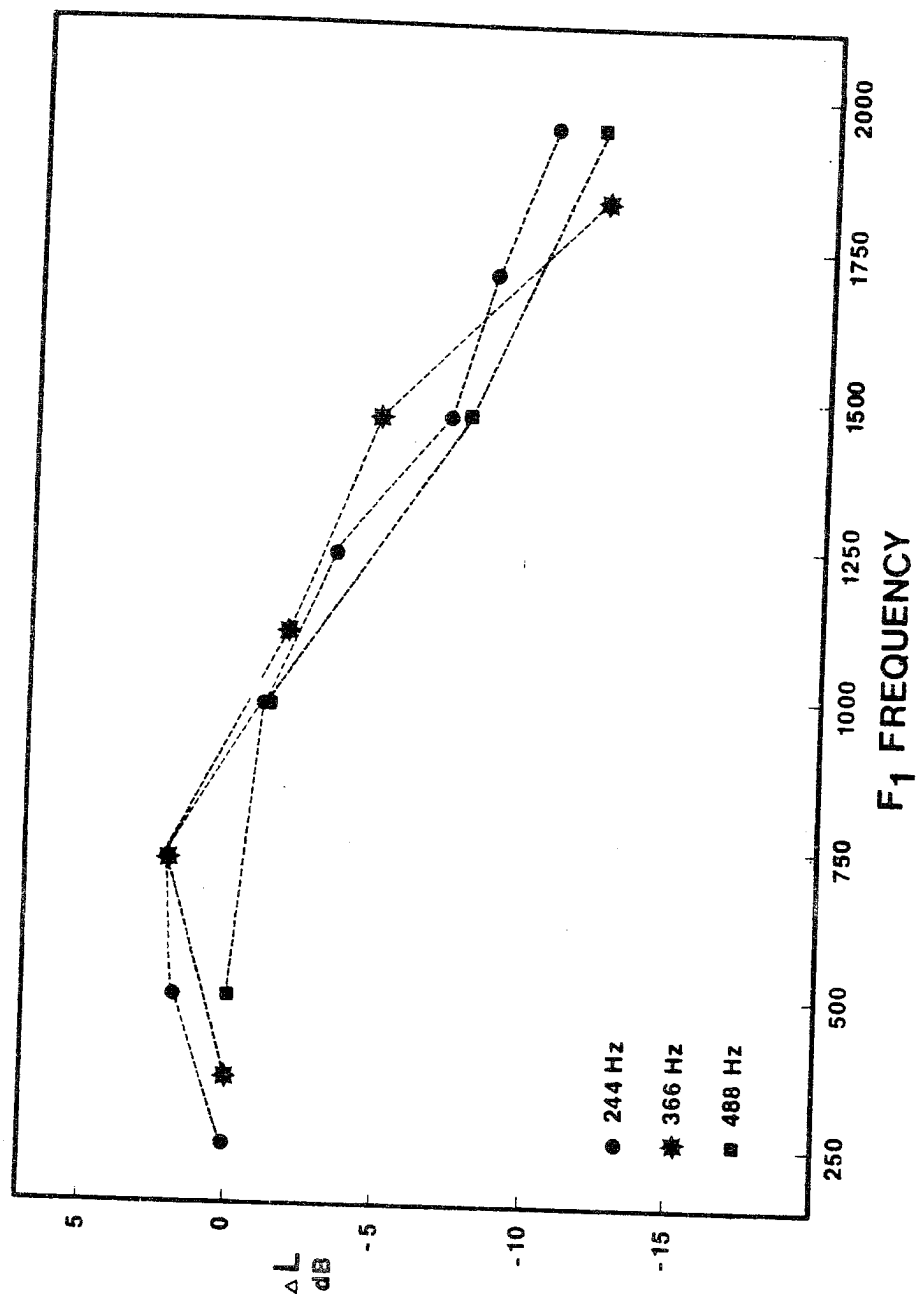


Figure 4.6 Figure 4.5 is replotted in terms of the frequency of the lower stimulus component.

apart and which aurally interact the least. Terhardt's explanation, however, does not take into account the fact that the strongest pitch rarely stems from the most widely separated (and hence most fully-resolved) components. Rather, for signals with fundamentals in the speech range (100 - 300 Hz), it is the third, fourth and occasionally fifth harmonics which are dominant for low pitch. Another reason why (cochlear) frequency resolution probably does not play a determining role in the dominance of low harmonics is the fact that the lower harmonics produce the most salient pitch even when only a single frequency component is presented to each ear (Houtsma and Goldstein, 1972; Bilsen, 1973).

Bilsen *et al* (1975) suggest that the dominance principle may be based on some form of non-linear suppression and/or lateral inhibition. They have recorded from cells in the dorsal cochlear nucleus during presentation of comb-filtered noise. This signal consists of wide-band noise containing many spectral peaks. The peaks are equally-spaced in terms of linear units (i.e. Hz). Each spectral peak of the CFN is set, in turn, to coincide with the unit's characteristic frequency. Then, the corresponding spectral valley of each peak is set at the unit's CF. The unit's ability to distinguish between a spectral peak and the adjacent valley in terms of its discharge rate provides a measure of the cell's spectral resolving power. One would expect the greatest difference in discharge rate to occur between the lowest peak and valley. And this is, in fact, the case when the difference in peak-valley spectrum level is large (20 dB - Evans, 1977 [DCN]; 35 dB - Narins and Evans, 1980 [cochlear nerve]). However, when the peak-valley level difference is about 11 dB the largest modulation in the discharge rate is produced by the peak corresponding to the second (Bilsen *et al*, 1975: Figures 2 and 3) or third (ibid: Figure 1) harmonic and the adjacent valley. This result implies the operation of some form of non-linear suppression or neural inhibition that is most effective over a limited frequency range. The functional result of this mechanism would be to enhance the neural excitation pattern of units tuned closely to the spectral peaks relative to the activity of cells tuned to frequencies lying between contiguous peaks. This selective enhancement of the neural excitation pattern tuned to the region around the third harmonic would presumably bias the central pitch processor in favor of this portion of the spectrum.

Goldstein (1978) has suggested an alternative explanation of the dominance region. He observes that the temporal variance of neural phase-locking is smallest for frequencies between 600 and 2000 Hz. If one assumes that the auditory system analyzes the spectrum of a signal on the basis of neural timing information (Goldstein and Srulovicz, 1977), then it follows that the precision of aural frequency measurement will be proportional to the variance of the timing information conveyed in the synchronized discharge activity of single auditory units. Consequently, one would expect that the ability to discriminate single frequencies would be most acute between 600 and 2000 Hz. This is, in fact, the case. But what does the precision of aural frequency measurement have to do with the strength of low pitch?

Goldstein proposes that the pitch of complex signals is based on a pattern recognition mechanism which finds a harmonic template that most closely matches the spectrum of the external signal. The more precise the determination of the external spectrum, the smaller the probability that significant errors in this template matching process will occur. Thus, those components which are measured with the greatest degree of precision are least likely to produce a significant pitch ambiguity. And according to Goldstein's model, the smaller the pitch ambiguity, the stronger the low pitch of the signal will be.

Thus, Goldstein's model posits that the source of the dominance region lies in the phase-locking behavior of single auditory neurons. Peripheral frequency resolution is a factor only in so far as it impedes the process of aural frequency measurement. Any set of frequency components separated by more than approximately 10% will be analyzed. However, this does not necessarily imply that all harmonics which are located in the dominant part of the spectrum (e.g. the third through tenth harmonics of a 200 Hz fundamental) will generate a strong pitch. According to Goldstein's model, this is because the probability of a template matching error is roughly proportional to the harmonic rank of the stimulus components. The harmonic rank of the stimulus components is important only because of the probabilistic basis of central pitch extraction. This would explain why the low harmonics maintain their dominance under dichotic stimulation.

Yet another explanation of the dominance region comes from the research group at Bell Laboratories (Flanagan; 1962; Rosenberg, 1965; 1966). They observe that the size of the traveling wave along the basilar membrane is largest for frequencies between 500 and 1000 Hz. Consequently, information concerning both the stimulus spectrum and stimulus periodicity will be signalled to higher auditory centers by a larger number of neurons than analogous information in other regions of the spectrum. The strength of a tone's pitch would consequently be proportional to the number of units phase-locked to the cochlear waveform and would be independent of the harmonic rank of the stimulus components (except in so far as it affects the number of synchronized units). This explanation would also be compatible with the dominance of low harmonics under dichotic stimulation since the size of the neural population encoding the signal's pitch would presumably be determined by the number of peripheral units synchronized to the stimulus components.

The present results suggest that the dominance region is probably a consequence of the size of the neural population synchronized to the cochlear stimulus. The data are thus qualitatively consistent with certain features of Bilson *et al's* (1975), Rosenberg's (1966) and Goldstein's (1978) proposals. However, it is not possible to favor any single explanation over the others of the basis of the present data because the magnitude of the FFR is a function of at least three factors. The first concerns the fact that the amplitude of the FFR is proportional to the number of neurons synchronized to the cochlear waveform. The second is that the greater the precision with which the

activated neural population is synchronized to the cochlear waveform or signal spectrum, the larger the overall response. The third factor involves the phase dispersion of the synchronized activity. A large degree of phase dispersion will result in a certain amount of response attenuation due to partial phase cancellation of the aggregate response caused by the interference of multiple sources with different phase vectors. By reducing the phase dispersion of the contributing neural elements, it is possible, in principle, to enhance the magnitude of the overall response. Thus, the magnitude of the traveling wave will not necessarily produce a large response unless the phase dispersion of the synchronized neural population is relatively small and the precision of phase-locking relatively precise. It is likely, that all three factors are relevant to the physiological basis of the dominance principle.

The most reliable predictor of FFR amplitude is the frequency of the stimulus components. The largest responses are generated by signals centered around the region of 750 Hz, regardless of the fundamental frequency or harmonic rank of the stimulus components. This finding suggests that frequency resolution per se is probably not the most important factor determining the amplitude of the FFR and, in this sense, is consistent with both Goldstein's and Rosenberg's explanations. Goldstein's model implies that the origin of the dominance principle lies in the superior precision of neural phase-locking. It is difficult to evaluate this hypothesis with the FFR because of the low-pass characteristic of the scalp response which precludes recording FFR of sizable magnitude to frequencies above 500 Hz. Consequently, it is not possible to directly measure the precision of neural synchronization to the component frequencies in the region of dominance. However, the waveforms of FFR recorded to harmonics in the dominance region are often characterized by having sharper peaks than FFR produced by non-dominant harmonics. This feature is particularly evident when the fundamental frequency of the stimuli is 244 Hz or lower. This aspect of the FFR may be an indication of tighter phase-locking. It may also be significant that this feature is usually most apparent in younger subjects, particularly those with unusually low low-frequency thresholds.

CHAPTER V

PHASE-SENSITIVITY OF THE FREQUENCY-FOLLOWING RESPONSE TO THE "MISSING" FUNDAMENTAL

5.1 INTRODUCTION

Periodicity models of pitch perception have traditionally focussed on the temporal properties of the stimulus waveform. The classical periodicity model of the late nineteenth century (Wundt, 1880; Rutherford, 1886) emphasized the importance of the amplitude modulation pattern ("envelope") of the stimulus waveform (Chapter 1.21). Around 1940 (Schouten, 1940a,c), the periodicity model was revised in order to take into account the operation of cochlear frequency analysis. Because the lowest ten harmonics of a complex signal are analyzed as separate frequency components in the cochlea, the periodic properties of the waveform are only preserved through the interaction of the higher, aurally unresolved harmonics in the inner ear (Chapter 1.22).

If the sensation of pitch is derived from auditory neurons synchronized to the envelope of the cochlear waveform (resulting from the interaction of contiguous components), would one not expect the salience of the pitch to be a function of the "peakedness" (i.e. depth of amplitude modulation) of the effective stimulus waveform? Not necessarily. However, for the present discussion, it will be assumed that the answer is in the affirmative, in order to follow the logic of some historically important experiments. These experiments sought to test the periodicity model by modifying the peakedness of the stimulus waveform envelope and observing any change that might occur in either the salience or value of the signal's pitch. The simplest way to modify the waveform envelope without affecting the spectrum is through manipulation of the relative phase configuration of the stimulus components. The most distinct waveform periodicity occurs when all components begin in the same phase and the flattest waveform results when the starting phase of each component is randomly assigned (Figure 1.5). If the pitch of the signal is based on some temporal property of the stimulus waveform, the pitch of the random phase stimulus should sound considerably less distinct than the pitch of the phase-coherent signal. Licklider (1955), using broadband stimuli, found this to be the case. So did Mathes and Miller (1947) using three-component signals composed of narrowly-spaced harmonics. Ritsma and Engel (1964) observed that the pitch value of three-component tones (consisting of harmonics separated by less than 10%) is also sensitive to the phase configuration of the signal. These studies, all designed to test Schouten's model of the residue, appeared to support the importance of the temporal properties of the stimulus waveform in the perception of pitch.

However, beginning in the mid-1960's, it became increasingly clear that the high, unresolved harmonics do not make a significant contribution to the pitch of complex tones (Chapter 1.25; Ritsma, 1962; 1963; Rosenberg, 1965; 1966; Ritsma, 1967; Plomp, 1967; Ritsma and Bilsen, 1970; Terhardt, 1970; Houtsma and Goldstein, 1972; Bilsen, 1973). The perceptual quality called the "residue" by Schouten, and assumed to constitute the basis of low pitch, has been relabelled "roughness" (Terhardt, 1970) and is considered to be a sensation quite separate from low pitch. The conclusions reached by Licklider (1955), Mathes and Miller (1947) and Ritsma and Engel (1964) appear to apply to the perception of roughness rather than to low pitch. Buunen (1975) has shown that the sensation of roughness is indeed based on the amplitude modulation properties of the waveform created by the cochlear interaction of unresolved harmonics (separated by less than 12%) and is consequently quite sensitive to the relative phase configuration of the signal. It is distinguished from low pitch in that it has a sharp, "buzz"-like quality and is most salient when the fundamental is lower than 100 Hz. Above 100 Hz, the roughness sensation grows weaker, disappearing altogether for fundamentals above 300 Hz (Terhardt, 1970). In contrast, low pitch is a product of more widely-spaced harmonics and whose salience does not appear to be appreciably affected by the phase configuration of the signal (Chapter 1.27; Bilsen, 1973; Buunen *et al.*, 1974; Buunen, 1975), except when the pitch is based on combination tones generated by marginally resolved components (i.e. separated by approximately 8-12% - Chapter 1.26; Buunen *et al.*, 1974; Buunen, 1975).

The phase insensitivity of low pitch is hardly surprising in view of the fact that the pitch of complex tones is based exclusively on those components resolved by the ear (Chapter 1.25). The manifestation of phase effects requires significant interaction among three or more components. This will occur only when components are separated by less than 10%. The absence of phase effects for low pitch is generally interpreted as evidence against any model of pitch perception based on the temporal discharge pattern of auditory neurons. This conclusion is invalid, because it ignores the fact that a significant degree of cochlear interaction exists for any pair of stimulus components separated by an octave or less (Brugge *et al.*, 1969). This interaction is reflected in the temporal activity of cochlear nerve fibers (Rose *et al.*, 1969; Brugge *et al.*, 1969; Javel, 1980) and AVCN units (Rose *et al.*, 1974) tuned to frequencies between the components. Consequently, information concerning the fundamental frequency of the stimulus is available to higher auditory centers in the periodic activity of peripheral units.

In Chapter 4 it was shown that the fundamental frequency of two-component tones is represented in the scalp-recorded frequency following response as long as the harmonics are separated by more than approximately 15%. These results suggest that the fundamental is encoded in the periodic activity of upper auditory brainstem neurons. However, it has been suggested by Hall (1979) that scalp FFR does not represent neural activity directly involved in the processing of low

pitch, but reflects, rather, the amplitude modulation pattern ("envelope") of the stimulus waveform. From the studies of Schouten (1940c), de Boer (1956a,b) and others, it is quite clear that the pitch of complex tones is not a product of the stimulus waveform envelope (Chapters 1.22 and 1.23). Thus, if the periodicity of the scalp-recorded FFR can be shown to be governed by the modulation pattern of the stimulus waveform, this would imply that the FFR does not represent neural processes involved in the perception of low pitch. Hall's hypothesis is subject to test in one of two ways. First, one can record FFR to inharmonic signals and compare the responses with those obtained using harmonic complexes. The results of such a comparison are discussed in Chapter 6. A second means of testing Hall's hypothesis is to record FFR to stimuli in which the modulation depth of the signal is systematically controlled through manipulation of the stimulus phase configuration. From the studies of de Boer (1956a), Bilsen (1973) and Buunen *et al* (1974) it is known that the salience of low pitch is not appreciably affected by such phase manipulations, as long as the components are separated by more than 12-15%. If the FFR is based on the envelope modulation pattern of the waveform, the largest responses should be generated by phase-coherent stimuli with very peaked envelopes. The smallest responses should be produced by waveforms with a low degree of amplitude modulation.

5.2 COMPARISON OF PHASE-SENSITIVITY OF FFR TO TONES OF LOW AND HIGH HARMONIC RANK

Hall's hypothesis was tested using two separate series of three-component signals. In one series the stimuli consisted of the third, fourth, and fifth harmonics of a 244 fundamental. This set of stimuli were similar to two stimulus complexes employed in Hall's (1979) study (Note 5.1). Stimuli in the other set were composed of the seventh, eighth, and ninth harmonics of the same fundamental. If the frequency following response to the missing fundamental parallels the psychophysically-defined properties of low pitch, one would expect that the FFR will be insensitive to the phase configuration of the low harmonic complexes in which the stimulus components are aurally resolved. However, the FFR should reflect the waveform envelope characteristic of the high harmonic signals due to the fact that the components are separated by less than 15% (i.e. the critical bandwidth).

5.21 EXPERIMENTAL METHODS

(a) Stimulus Generation and Presentation

Stimuli consisted of two series of three-component, 40 msec tone bursts. In one series, the components were equivalent to the third, fourth, and fifth harmonics of a 244 Hz fundamental (732+976+1220 Hz). In the other, components were equivalent to the seventh, eighth, and ninth harmonics of the same fundamental (1708+1952+2196 Hz). Signals in each series were identical in all respects except the starting phase of the tone bursts' middle component (f_2). The lower (f_1) and upper (f_3)

components of each tone always began in sine phase (0 degrees). The starting phase of the middle component (f_2) varied between 0 and 180 degrees, in 45 degree steps. The amplitude of the three tonal components were set at the same intensity for all stimuli.

Stimulus intensity was set relative to each individual's FFR threshold. FFR threshold was defined as the lowest intensity level at which the stimulus fundamental (244 Hz) dominates the FFR spectrum. FFR threshold varied between 50 and 65 dB SPL (40-55 dB SPL/component). Stimulus level was 5 dB SPL above FFR threshold. Stimuli were presented binaurally at a rate of 15 per sec. Other stimulus parameters were as specified in Chapter 2.2.

(b) FFR Analysis

Brainstem responses were recorded differentially between vertex and the back of the neck. The left side of the neck served as ground. Each average usually consisted of 3000 responses. Occasionally, more than 3000 responses were collected for a response average in order to enhance the signal-to-noise ratio of the FFR to stimuli of exceptionally low stimulus intensity. Other recording parameters were as specified in Chapter 2.3.

The amplitude of the fundamental component was computed for each average using a digitally-implemented spectrum analysis. Analysis was restricted to the portion of the FFR following the "on" component of the response (approximately 8-10 msec following stimulus onset).

(c) Subjects

Three females (ages 12-28) and four males (ages 25-57) with no previous history of hearing loss, served as subjects.

5.22 EXPERIMENTAL RESULTS

(a) Phase Effects: Low Harmonics

By shifting the starting phase of f_2 relative to f_1 and f_3 , it is possible to systematically vary the modulation depth of the stimulus waveform envelope, as shown in the left-hand column of Figure 5.1. The depth of waveform amplitude modulation is greatest when all three tonal components begin in the same phase (phase-coherent) or when f_2 is in counterphase to f_1 and f_3 ($f_2=180$ degrees). The modulation depth is smallest when the phase of f_2 leads the phase of f_1 and f_3 by 90 degrees (quadrature-phase). An intermediate depth of modulation results when the phase of f_2 leads f_1 and f_3 by 45 or 135 degrees.

Figure 5.1 illustrates the relationship between the stimulus envelope modulation depth and amplitude of the frequency following response for a tonal complex composed of low harmonics. Stimulus intensity was 65 dB SPL (5 dB above FFR threshold). In the left-hand column are the acoustic waveforms for each stimulus phase configuration. Frequency following responses for each stimulus are shown on the right. Each stimulus produces a response whose major

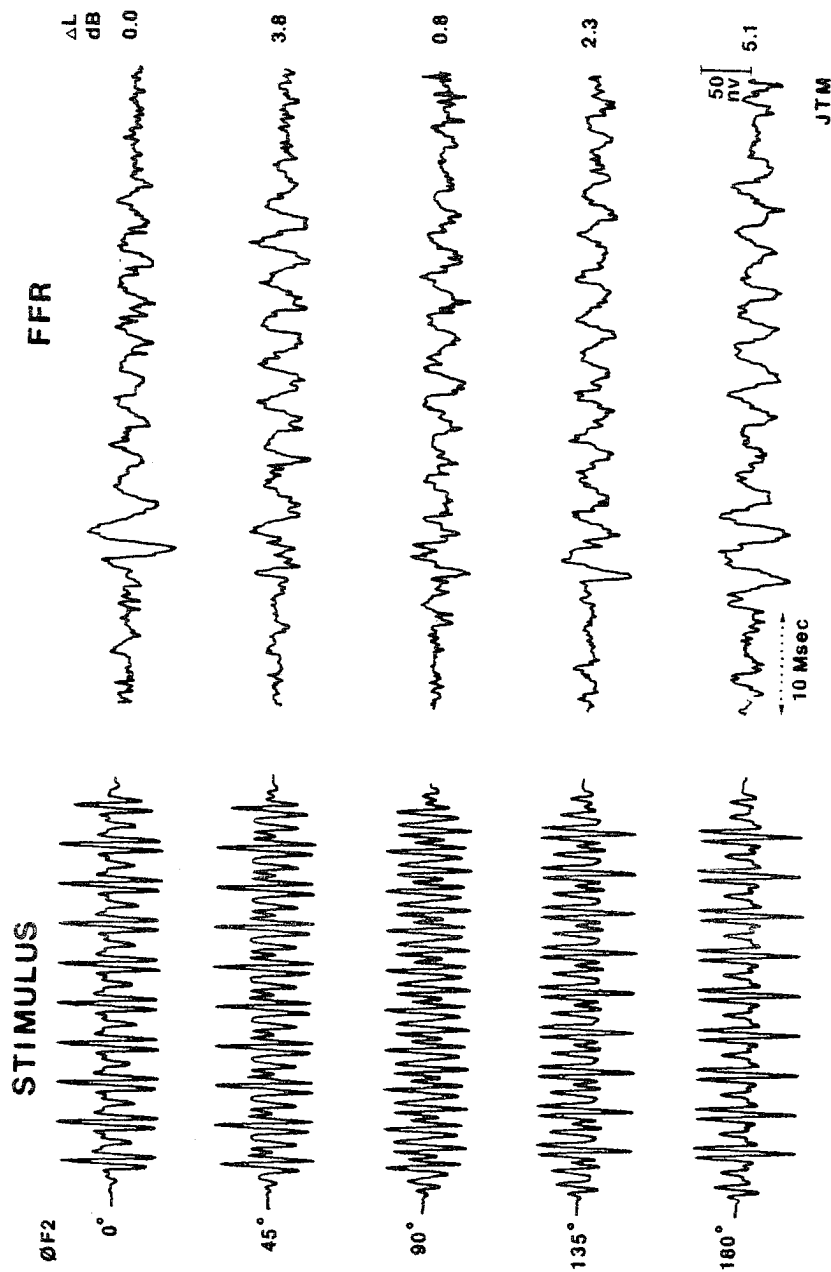


Figure 5.1 FFR recorded from a single subject to three-component tones of variable phase configuration and low harmonic rank. Stimuli consisted of the third, fourth, and fifth harmonics of a 244 Hz fundamental. The lower (f1) and upper (f3) components began in sine phase for all stimulus conditions. The starting phase of the middle component (f2) varied from 0 to 180 degrees, in 45 degree steps. On the left-hand side are the stimulus waveforms produced by phase manipulation of f2. On the right are response waveforms generated by these stimuli. The right-most column indicates the relative level of the fundamental component in the FFR spectrum for each response. Stimulus intensity = 70 dB SPL (5 dB above FFR threshold). FFR waveforms are averages of 3000 responses.

period corresponds to both the period of the stimulus waveform and the pitch of the tonal complex. The amplitude of this 244 Hz (fundamental) component is shown to the right of each response. The amplitude of the fundamental is expressed in dB relative to the amplitude of the fundamental evoked by the phase-coherent ($f_2=0$ degrees) stimulus.

If there were a systematic relationship between stimulus envelope modulation depth and FFR amplitude (of the fundamental component), one would expect the largest responses to be produced by the phase-coherent and counterphase stimuli, and the smallest response to be produced by the quadrature-phase stimulus. However, for this subject, the largest responses are, in fact, produced by stimuli whose middle component starting phase is either 45 or 180 degrees. The smallest amplitude response is produced by the phase-coherent stimulus. Thus, for tones composed of widely-spaced harmonics, there does not appear to be a systematic relationship between the modulation depth of the stimulus waveform and FFR amplitude (Note 5.2).

(b) Phase Effects: High Harmonics

Figure 5.2 illustrates the relationship between stimulus waveform modulation depth and FFR amplitude for a tonal series composed of narrowly-spaced harmonics. Stimulus level was 63 dB SPL (5 dB above FFR threshold). In contrast to the pattern exhibited in Figure 5.1, it is apparent that the amplitude of FFR evoked by tone bursts composed of narrowly-spaced harmonics is related to the depth of stimulus waveform modulation. The largest responses are produced by the phase-coherent and counterphase stimuli. The smallest response is evoked by the quadrature-phase stimulus. There is a five-fold difference in FFR magnitude between the largest and smallest responses. However, this amplitude disparity is probably not attributable to threshold changes, as all responses exhibit a large "on" response.

(c) Comparison of Phase Effects for Low and High Harmonics
A comparison between FFR evoked by phase-coherent and quadrature phase stimuli of different harmonic rank for a single subject is shown in Figure 5.3. For this subject, the quadrature-phase response is slightly larger (1.6 dB) than the phase-coherent response for stimuli composed of widely-spaced harmonics. In contrast, the quadrature-phase response is considerably smaller than the phase-coherent response for narrowly-spaced harmonics.

The relative amplitude of the phase-coherent and quadrature-phase responses was computed in this manner for 7 subjects. Stimulus intensity was always 5 dB above FFR threshold for any individual subject (50-65 dB SPL). Figure 5.4 shows that the amplitude of the quadrature-phase and phase-coherent responses produced by the low harmonic stimuli are almost identical. The quadrature-phase response was, on the average, only 0.6 dB larger. On the other hand, there is a significant disparity between the amplitude of the phase-coherent and quadrature-phase responses for the high harmonic stimuli.

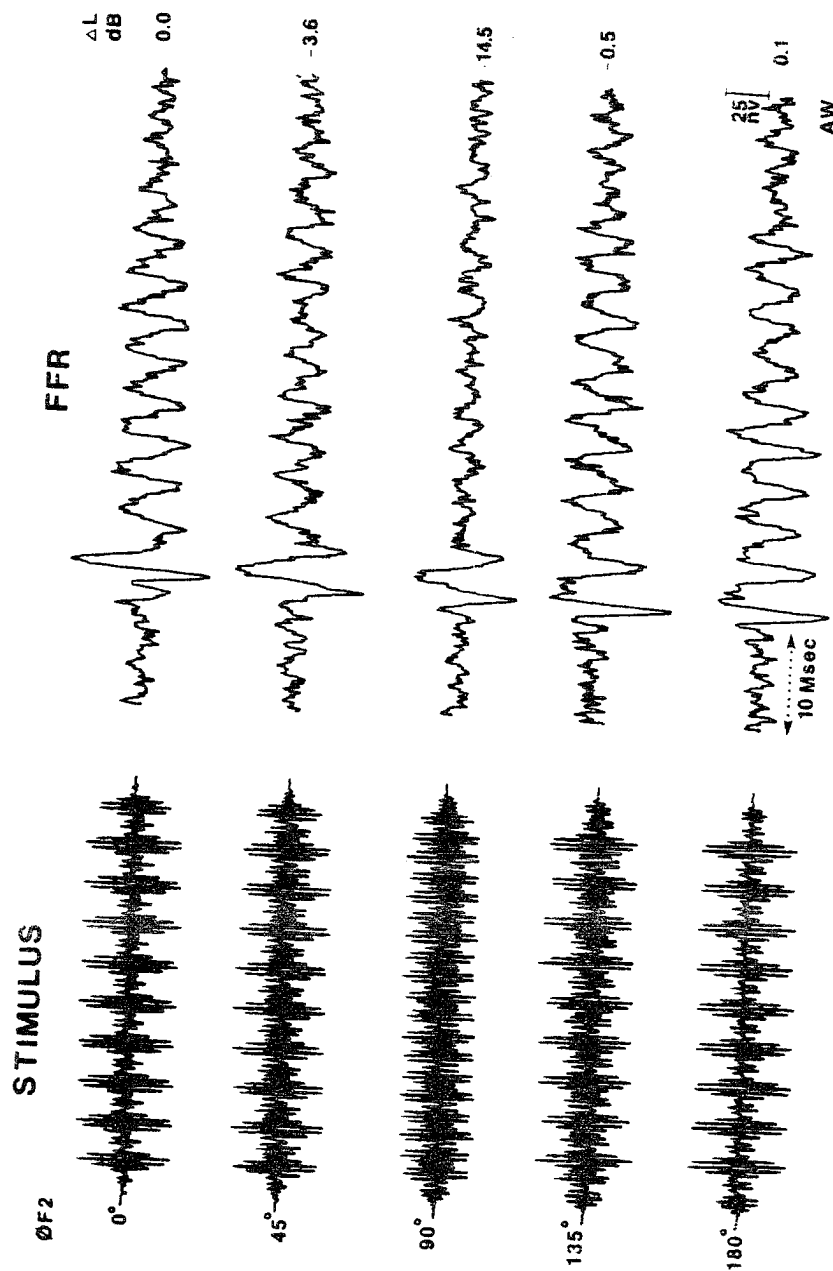


Figure 5.2 FFR recorded to three-component stimuli of variable phase configuration and high harmonic rank. Stimuli consisted of the seventh, eighth, and ninth harmonics of a 244 Hz fundamental. Stimulus intensity was 68 dB SPL (5 dB above FFR threshold). Averages consist of 3000-5000 responses (Response waveforms and relative spectrum levels are normalized to compensate for unequal response numbers in each average).

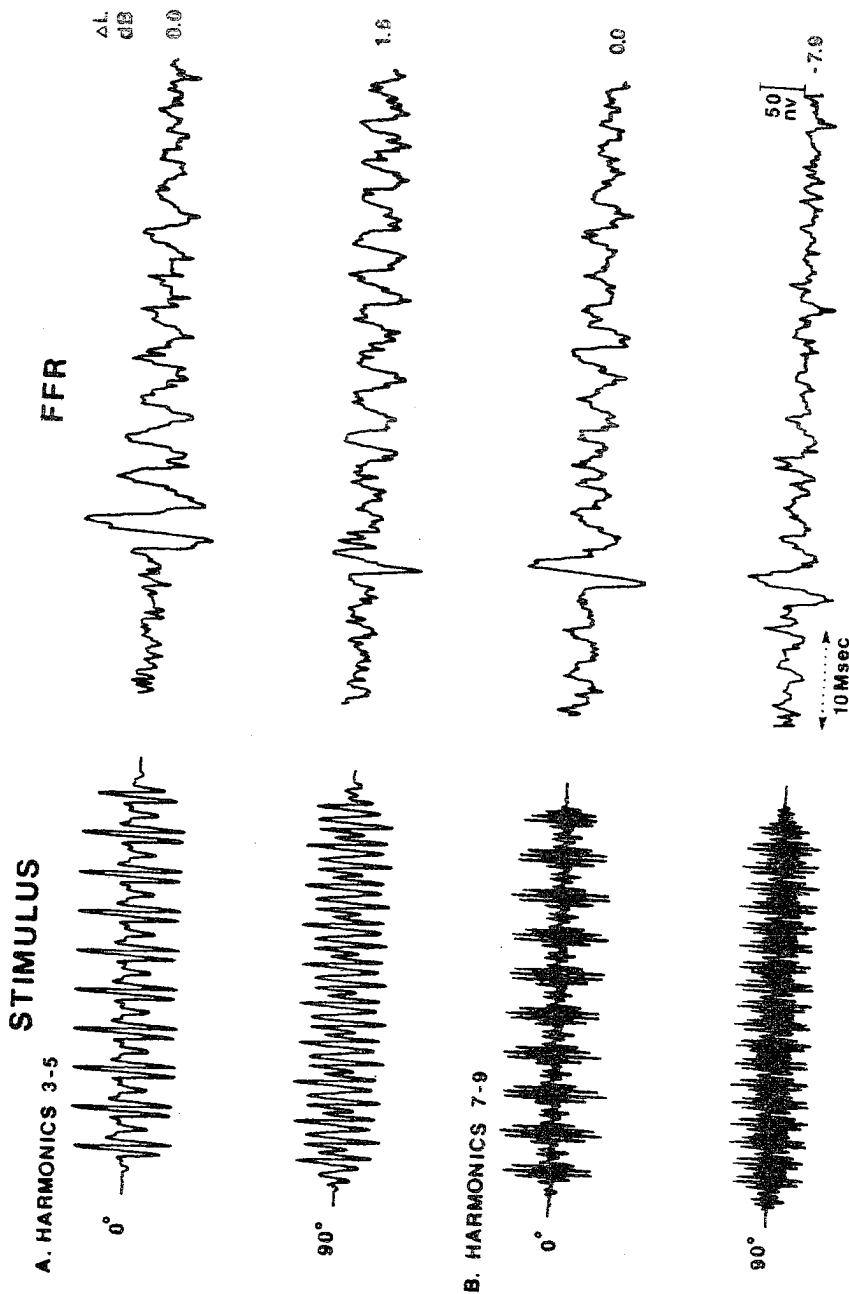


Figure 5.3 Comparison of FFR recorded from a single subject to phase-coherent (f2) starting phase = 0 degrees) and quadrature phase (f2) starting phase = 90 degrees) stimuli. Stimulus intensity was 70 dB SPL (5 dB above FFR threshold). Averages consist of 5000 responses.

JTM

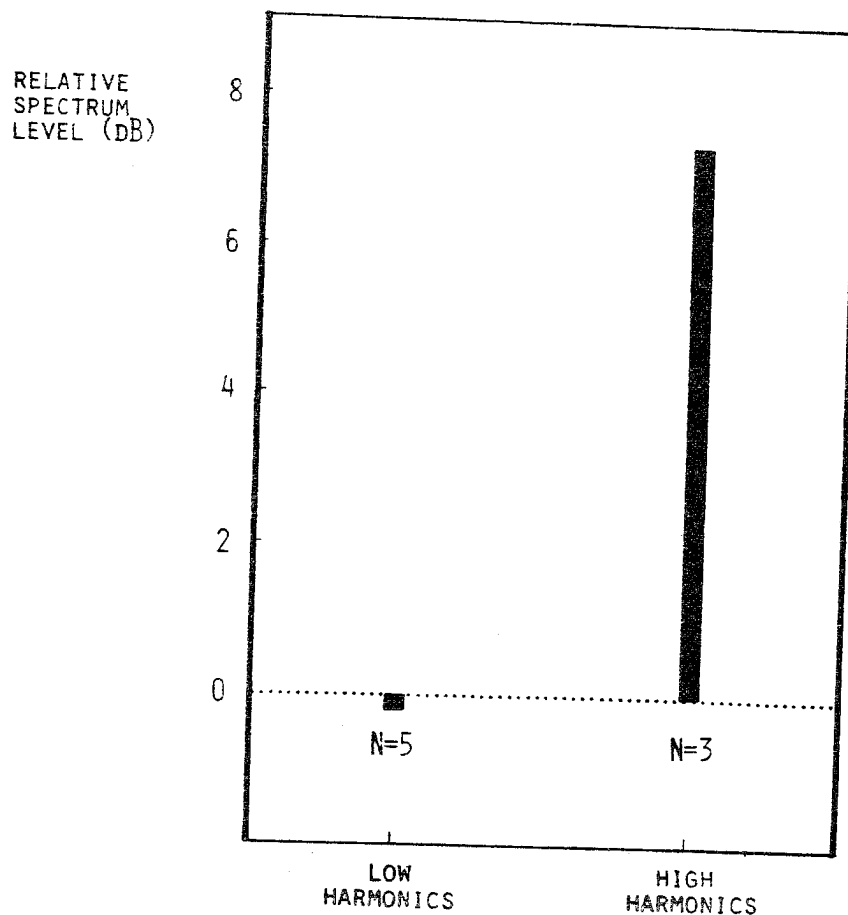


Figure 5.4 The spectrum level difference (of the fundamental component) between FFR produced by phase-coherent and quadrature phase stimuli for two series of three-component tones (reference = spectrum level of FFR generated by the quadrature phase signal). Low harmonic stimuli consisted of the third, fourth, and fifth harmonics of a 244 Hz fundamental. High harmonic stimuli consisted of the seventh, eighth, and ninth harmonics of the same fundamental. Means for the low harmonic responses are based on the analysis of FFR collected from 5 subjects. Means for the high harmonic responses are based on 3 subjects. Averages consist of 3000-5000 responses. Stimulus intensity was 5 dB above FFR threshold (55-75 dB SPL).

5.23 DISCUSSION

The results of the present experiment indicate that scalp-recorded FFR does not follow the envelope of the stimulus waveform when the tone's components are of low harmonic rank (i.e. separated by at least 20%). In contrast, the scalp response to stimuli composed of narrowly-spaced harmonics, at the very margin of the existence region for low pitch, does appear to be sensitive to the amplitude modulation characteristic of the stimulus waveform. These data suggest that FFR to tones of low harmonic rank is indeed generated by a signal that has undergone some form of cochlear frequency analysis. Consequently, it should be possible to infer the frequency selectivity of individual subjects by presenting a series of quadrature-phase signals of variable harmonic rank.

The present results also suggest that the FFR is generated primarily by units tuned to relatively low frequencies in the general vicinity of the stimulus components. This is because the frequency resolution of the cochlea broadens (in absolute terms) as one approaches the basal region. Thus, a set of low frequency components, analyzed as separate components in the apical portion of the cochlea, will be unresolved in the basal region. If the FFR was primarily the product of high-frequency units, one should observe a significant amplitude decrement in response to the low harmonic quadrature-phase stimulus.

However, this does not mean that scalp FFR never reflects the activity of high-frequency units. It will be recalled from the discussion in Chapter 1.74 that the FFR to pure tone stimuli may be based on the activity of a broad population at moderate-to-high intensities. In the present study, signal level was maintained at a relatively low level (45-60 dB SPL/component). However, most FFR studies have been conducted at considerably higher stimulus intensities. For example, the spectrum level of the stimuli in Hall's (1979) study was 70-74 dB SPL/component. At these signal levels, the FFR is generated by a considerably larger population of units. It is possible that at intensities comparable to those used in Hall's (1979) study, the FFR does become sensitive to the amplitude modulation pattern of the stimulus waveform.

5.3 PHASE-SENSITIVITY OF THE FFR AS A FUNCTION OF INTENSITY

The frequency selectivity of the auditory system diminishes at high intensities (> 70 dB SPL). For example, Pick (1977) has shown, using a comb-filtered noise (CFN) signal, that the precision with which the spectral peaks can be psychophysically resolved diminishes for intensities above 70-80 dB SPL. Evans (1977), and Narins and Evans (1980) have shown that the ability of single cochlear nerve fibers to resolve the spectral peaks of a CFN signal also declines with level for intensities above 70 dB SPL. If the FFR is indeed sensitive to the frequency resolving power of the auditory system (Chapter 5.2), this

decline in frequency selectivity might be observed in the scalp response to low harmonic stimuli at high signal levels. This hypothesis was tested by recording FFR to phase-coherent and quadrature-phase signals of low harmonic rank at levels up to 15 dB above the those used in Chapter 5.2.

5.31 EXPERIMENTAL METHODS

Stimuli were three-component tones, composed of the third, fourth and fifth harmonics of a 244 Hz fundamental. These signals were similar to those specified in Chapter 5.2 except that the signal level ranged between 5 and 20 dB SPL above FFR threshold (i.e. from 45-60 dB SPL/component up to 75 dB SPL/component). Subjects were 3 males and 3 females, all of whom had participated in experiment 5.2. All other experimental parameters were identical to those specified in Chapter 5.21.

5.32 EXPERIMENTAL RESULTS

In Chapter 5.22 it was shown that the amplitude of the FFR is not significantly affected by the modulation depth of the stimulus waveform when the signal level is 5 dB above FFR threshold. However, as signal intensity increases above this level, the FFR does become sensitive to the waveform envelope. This pattern is illustrated for a single subject in Figure 5.5. At 5 dB above FFR threshold (65 dB SPL), FFR to the phase-coherent and quadrature-phase signals are of approximately equal magnitude. However, the growth functions of FFR to the two signals are very different. The amplitude of the phase-coherent response grows with increasing stimulus level. The quadrature-phase response, in contrast, actually declines in amplitude over the same intensity range. Consequently, there is a large disparity in FFR amplitude between the two signals at higher signal levels. The onset latency of the responses (particularly the phase-coherent phase responses) decreases with increasing stimulus intensity. This decrement in onset latency is consistent with the hypothesis that, at high intensities, an increasing number of high-frequency units are contributing to the scalp response.

A comparison of the phase-coherent and quadrature-phase responses for a number of different subjects is shown in Figure 5.6. These data follow the same pattern as illustrated in Figure 5.5. At the lowest intensity level (5 dB SPL above FFR threshold), there is no significant difference between phase-coherent and quadrature-phase responses. However, at higher signal intensities, a considerable disparity between the two types of responses can be observed. Moreover, this amplitude difference increases with increasing signal level.

This pattern is true of all but one subject. The response pattern for this subject is illustrated in Figure 5.7. For this subject, the quadrature response is slightly larger than the phase-coherent response at each intensity level. As in Figure 5.5, the onset latency of the

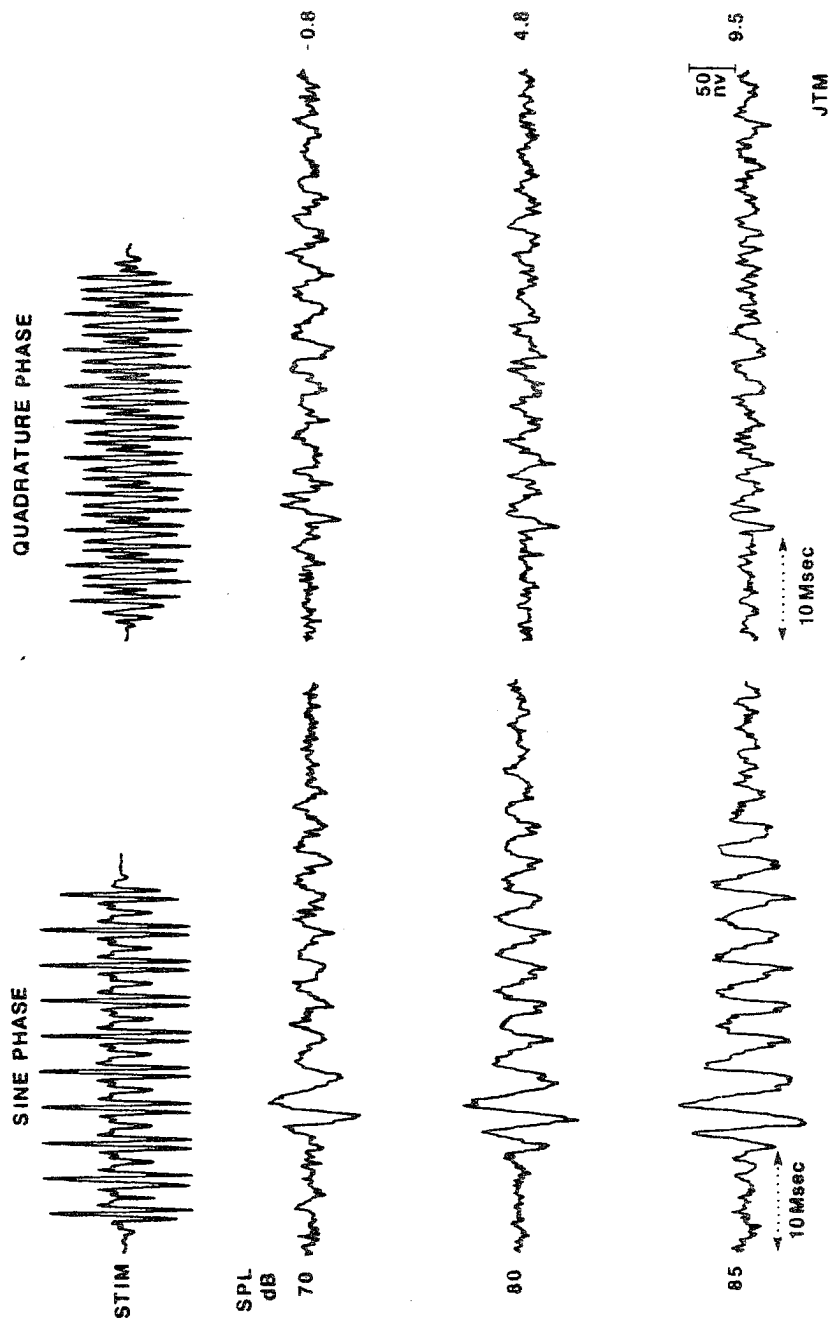


Figure 5.5 FFR recorded from a single subject to phase-coherent and quadrature phase stimuli over a range of intensities. FFR threshold was 65 dB SPL. The right-hand column indicates the relative spectrum level (reference = spectrum level of FFR produced by the quadrature phase signal). Averages consist of 3000 responses.

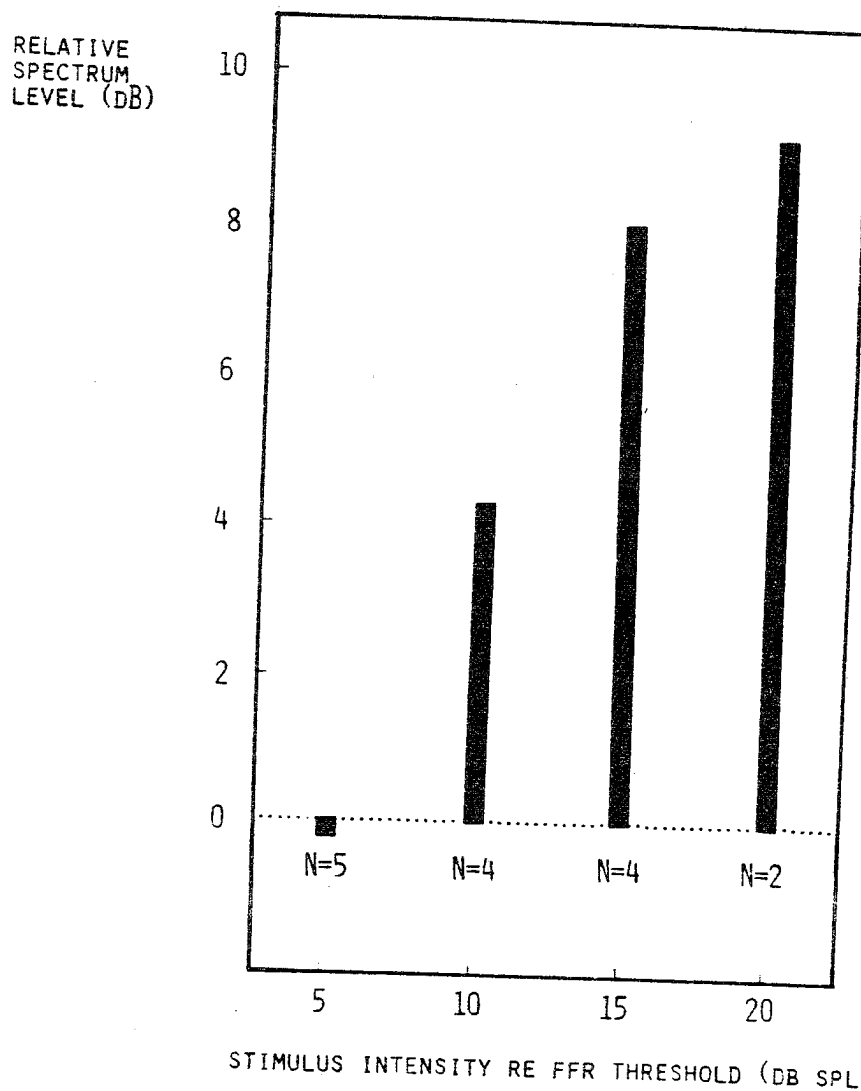


Figure 5.6 The spectrum level difference (reference = quadrature phase response) between FFR produced by phase-coherent and quadrature phase stimuli over a range of intensity levels (expressed in terms of FFR threshold). Stimuli were three-component tones composed of the third, fourth, and fifth harmonics of a 244 Hz fundamental. Mean spectrum levels indicated are means based on 2 to 5 subjects.

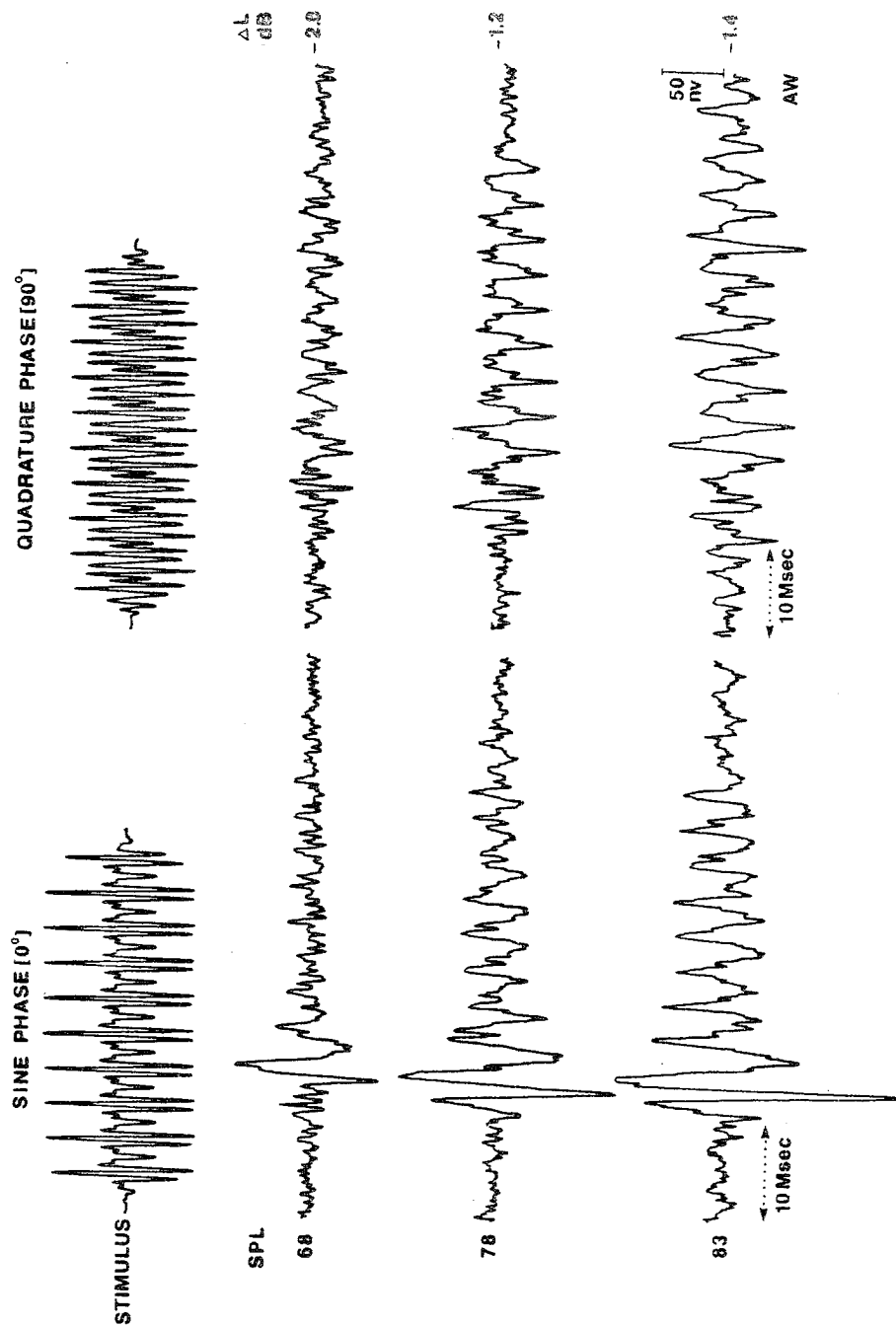


Figure 5.7 Atypical pattern of FFR recorded to phase-coherent and quadrature phase stimuli as a function of intensity. FFR threshold = 63 dB SPL. Averages consist of 5000 responses.

response diminishes with increasing signal intensity, suggestive of a contribution to the response by high-frequency units. It is not clear why the responses of this subject should differ from the prevailing pattern. One possibility is that the frequency resolving power of this subject is superior to that of the other subjects.

5.33 DISCUSSION

At moderate-to-high signal levels, the FFR does indeed appear to follow the waveform envelope of the stimulus. This phenomenon is clearly evident for signal intensities 10 dB SPL or more above FFR threshold (Figure 5.6). Why is the amplitude of the FFR affected by the stimulus waveform envelope at these levels when it appears to be relatively insensitive to the same parameter at lower intensities? One possibility was discussed in Chapter 5.23, namely the progressive recruitment of higher frequency, more basal units with increases in signal level. This hypothesis is consistent with the progressively shorter onset latencies observed in the responses recorded at higher intensities. However, if this were the only factor involved, one would expect to observe at least some response decrement at higher intensities for subject AW in Figure 5.7.

Another potentially relevant factor is the diminishing frequency selectivity observed at high signal intensities (Chapter 5.3). This possibility has significant implications for models of pitch perception, for it implies that the salience of low pitch may diminish at high signal levels. Unfortunately, this prospect has not been systematically explored. Most psychophysical studies of pitch have been conducted at intensities below 50 dB SPL. At these levels, no effect of stimulus phase configuration on the salience of low pitch has been reported (for harmonics lower than the eighth). The effect of stimulus phase configuration has not been explored for intensities above 70 dB. If FFR recorded at high signal levels does reflect neural processes related to the processing of low pitch, the present data would imply that low pitch does become sensitive to the stimulus phase configuration at levels above 70 dB SPL. However, it is also possible that FFR recorded at high signal intensities unduly reflects the activity of units not directly involved in the processing of low pitch (see Chapter 6).

5.4 PHASE-SENSITIVITY OF THE FFR AS A FUNCTION OF RELATIVE SPECTRUM LEVEL OF THE STIMULUS COMPONENTS

If the FFR recorded in response to the quadrature-phase stimulus reflects, at high signal levels, an actual decrement in frequency resolving power, it should be possible to boost the amplitude of the response by impeding the interaction of the stimulus components. In principle, this may be accomplished by raising the amplitude of the signal's middle component relative to the level of the sidebands in order to prevent the lower sideband from interacting with the other components.

If, on the other hand, FFR recorded at high intensities reflects the synchronization of high-frequency units to the waveform envelope, there should be a further decrement in the response to the quadrature phase signal when the level of the middle component is raised above the level of the sidebands because the waveform envelope becomes even flatter under this condition.

5.41 EXPERIMENTAL METHODS

Stimuli were similar to those illustrated in Figure 5.1 They consisted of a series of tones whose components were equivalent to the third, fourth, and fifth harmonics of a 244 Hz fundamental. The starting phase of the middle component varied between 0 and 180 degrees in 45 degree steps. The starting phase of the sidebands was always 0 degrees. The tones differed from those of Figure 5.1 only with respect to the relative amplitude of the stimulus components. The intensity of the upper and lower components was 70 dB SPL, while the level of the middle component was 76 dB SPL (equivalent to a modulation index of 1.00). Overall stimulus intensity was 82 dB SPL.

5.42 EXPERIMENTAL RESULTS

Figure 5.8 shows that the amplitude of the FFR is virtually insensitive to the stimulus phase configuration when the level of the middle component is 6 dB higher than the sidebands. There is practically no difference in the amplitude of the responses. Each signal produces a clearly periodic response which in no way reflects the amplitude modulation pattern of the stimulus waveform.

In Figure 5.9, the effect of raising the level of the middle component is even more evident. The figure compares the phase-coherent and quadrature-phase responses for two sets of signals. The top two rows display the responses produced when all three components are of equal intensity (modulation index = 2.00). There is a nearly 4 dB disparity in their amplitude levels. At this intensity level, the FFR is clearly responding to the interaction of all three components. When the level of the middle component is raised by 6 dB, a different pattern emerges, as shown in the bottom two rows of the same figure. There is practically no difference in the amplitude of the phase-coherent and quadrature-phase responses, despite significant differences in the modulation pattern of the stimulus waveforms of the two stimuli. This response pattern would not be expected to occur if the FFR were merely following the envelope of the stimulus waveform.

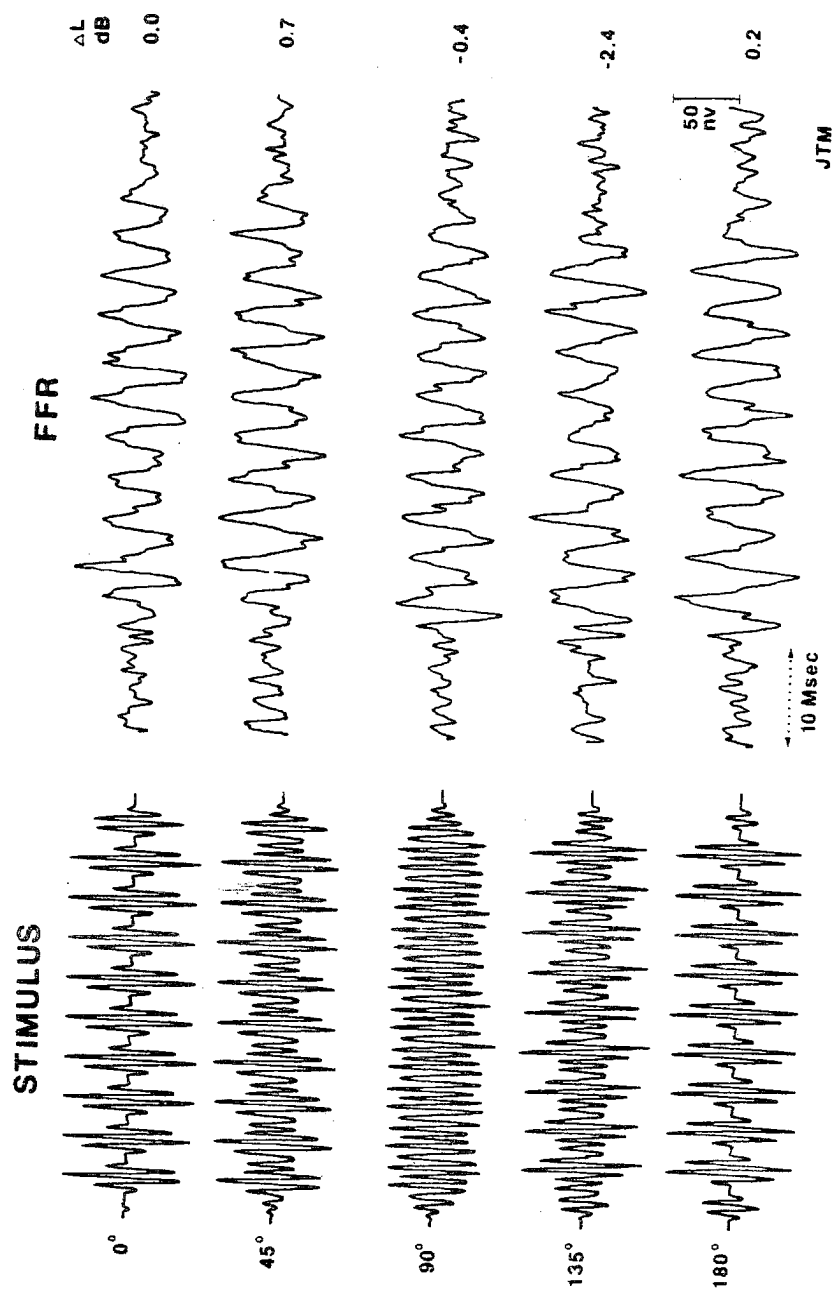


Figure 5.8 FFR recorded from a single subject to three-component tones of variable phase configuration and low harmonic rank. Stimuli consisted of the third, fourth, and fifth harmonics of a 244 Hz fundamental. Modulation index = 1.00. Stimulus level = 80 dB SPL (70 dB SPL/component for f1 and f3 and 76 dB SPL for f2). FFR waveforms are averages of 2000 responses.

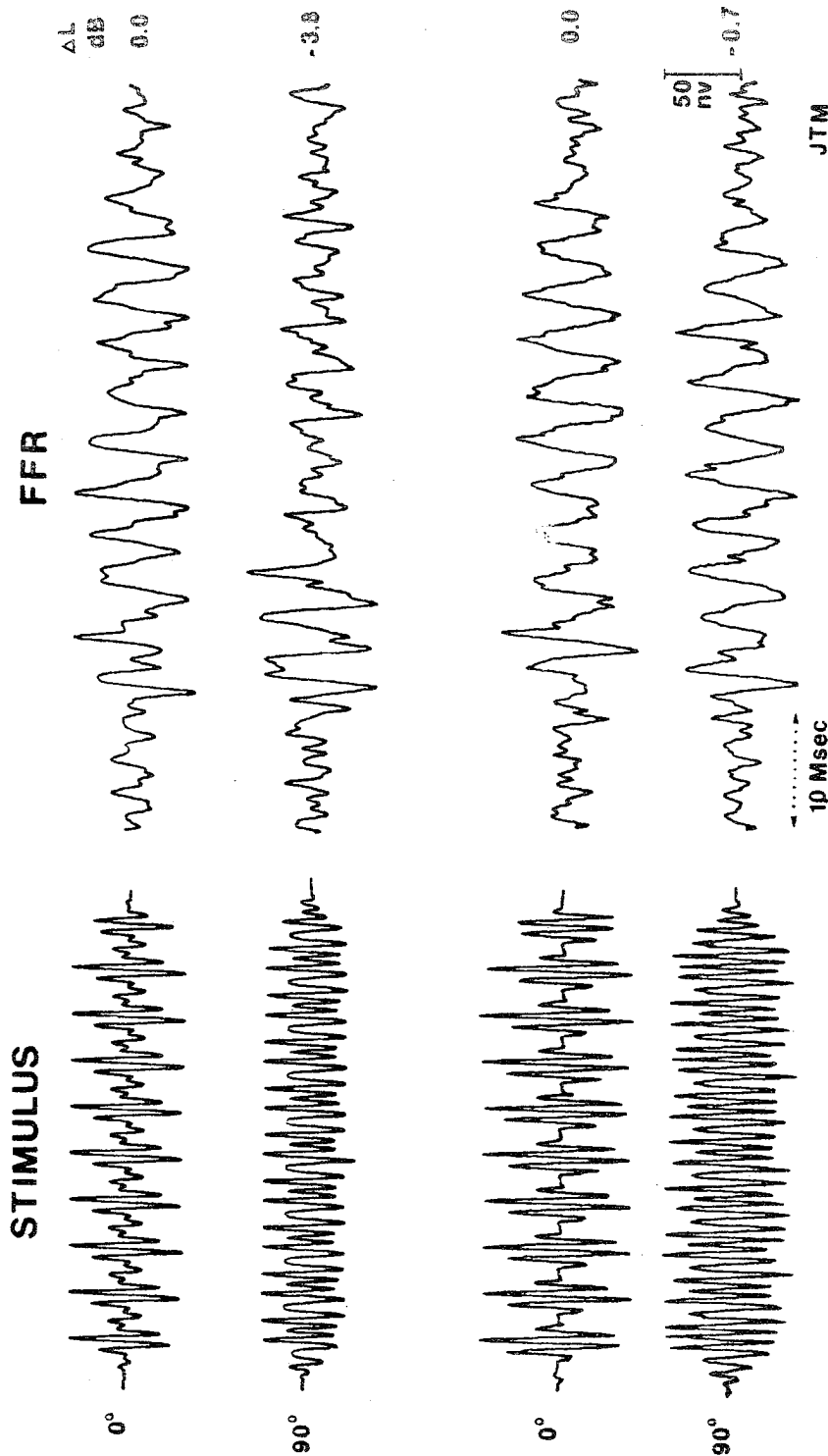


Figure 5.9 Comparison of FFR recorded to phase-coherent and quadrature phase signals for two different modulation indices. At top are the stimulus waveforms and FFR produced by signals with a modulation index of 2.00. Bottom rows represent same for stimuli with a modulation index of 1.00. Stimulus level = 80 dB SPL for the stimuli illustrated in the top rows (70 dB SPL/component) and 82 dB SPL for the signals shown in the bottom rows (70 dB SPL/component for f1 and f3 and 76 dB SPL for f3). Averages consist of 2000 responses.

5.43 DISCUSSION

This demonstration suggests that FFR recorded at relatively high intensities is not generated by units tuned to very high frequencies. If the FFR were, one would expect the FFR to reflect the modulation pattern of waveform, since the stimulus components would presumably be unresolved in the basal portion of the cochlea. Although it is unlikely the the FFR at high intensities reflects only the activity of units tuned to the vicinity of the stimulus frequencies, it is also unlikely that the scalp response is generated by an extremely broad range of units. Rather, the generating population probably spans an intermediate tonotopic range of approximately two octaves.

It may be of potential significance that many psychophysical studies of low pitch have been conducted with stimuli of modulation index equal to 1.00 (e.g. Schouten *et al*, 1962; Ritsma, 1962; de Boer, 1956a,b). This amplitude configuration may be optimal for preserving the internal spectrum of the signal at high intensities and for rendering the low pitch insensitive to the stimulus phase configuration. Indeed, Ritsma (1962) found that the salience of low pitch was in fact most pronounced when the modulation index was equal to 1.15 (i.e. the level of the middle component was 5 dB greater than the level of the sidebands).

5.5 SUMMARY

In Chapter 5.2 it was shown that the spectrum and amplitude of the scalp-recorded FFR is insensitive to the phase configuration of the stimulus when the harmonic components are separated by at least 20% and presented at low-to-moderate intensity levels. If the stimulus components are separated by less than 15%, FFR amplitude is significantly diminished in response to a quadrature-phase signal. This response to high harmonics suggests that the FFR is responding to the interaction of all three spectral components. This pattern suggests that the FFR is sensitive to the waveform envelope only when the components are not aurally resolved. Thus, it should be possible, in principle, to provide a rough estimate of an individual's frequency selectivity on the basis of the response to quadrature-phase stimuli of variable harmonic rank.

At high stimulus intensities, the FFR becomes sensitive to the phase configuration of even low harmonic stimuli (Chapter 5.3). In view of the results presented in Chapter 5.2, it is possible that this phase-sensitivity reflects a decrease in frequency selectivity at higher intensities, an explanation consistent with both psychophysical (Pick, 1977) and physiological (Evans, 1977; Narins and Evans, 1980) studies of frequency resolution at high signal levels. This hypothesis was tested in chapter 5.4 using stimuli of variable phase configuration and modulation indices. If the original decrement observed in the quadrature-phase response was due to increased component interaction as a consequence of diminishing frequency selectivity, it should be

possible to enhance the quadrature-phase response by impeding aural interaction among the components. This was done by decreasing the modulation index from 2.00 to 1.00 (i.e. increasing the level of the middle component to be 6 dB above the level of the sidebands). This procedure successfully restored the amplitude of the quadrature-phase FFR and rendered the responses to other phase configurations equally insensitive to the waveform envelope modulation pattern.

CHAPTER VI

FREQUENCY-FOLLOWING RESPONSE TO INHARMONIC SIGNALS

6.1 INTRODUCTION

Inharmonic signals have played a key role in the development of contemporary pitch theory. They were first used by Schouten (1940c) to demonstrate that the pitch of a complex tone is based neither on the period of the waveform envelope nor on the lowest frequency of the stimulus complex (Chapter 1.23). Subsequently, it was shown by de Boer (1956a,b) that the pitch produced by these signals can be predicted (to a first approximation) with the following equation:

$$(1) \quad p = f_c / \underline{n}$$

where f_c is the frequency of the signal's middle component, and $\underline{n}$ is the effective harmonic rank of that component. The harmonic rank, $\underline{n}$, can be found by determining the closest integer value that solves the following equation:

$$(2) \quad \underline{n} = f_c / (f_{n+1} - f_n)$$

where $f_{n+1} - f_n$ is the difference frequency (the reciprocal of the envelope period). When $f_c / \underline{n}$ is an integral value, as in the case of an harmonic series, p is equivalent to the difference frequency ($f_{n+1} - f_n$). However, when $f_c / \underline{n}$ is unequal to an integral value, the resulting pitch will deviate from the difference frequency by an amount that is approximately equal to $\underline{d}$ (delta) $f / \underline{n}$. For example, if the frequency of each component in an harmonic series $(n-1)f_o + nf_o + (n+1)f_o$ is shifted upwards by $\underline{df}$ Hz, the resulting pitch shifts upward from the pitch of the harmonic series by an amount that is equal to $\underline{df} / \underline{n}$ (for $\underline{df} < f_o / 2$). This proportional relation between the pitch and center frequency of the signal, the "first effect of the pitch shift", was shown by de Boer to be compatible with two different mechanisms of pitch extraction. The pitch of the signal composed of high, aurally-unresolved components, de Boer attributed to the timing pattern of auditory fibers synchronized to the fine-structure of the cochlear waveform. De Boer reasoned that the pitch generated by low, aurally-resolved components, must be based on some other mechanism because the components would not be expected to interact in the cochlea sufficiently to produce a waveform from which the low pitch of the signal could be derived. He suggested that the pitch generated by low harmonics was the product of a template-matching operation in which the auditory system attempts to find a harmonic series which most closely matches the frequencies of the inharmonic signal. De Boer demonstrated that these two mechanisms were mathematically equivalent and suggested that the extraction of pitch was probably based on both types of operation.

Equation (1) implies that the magnitude of the pitch shift is largest for widely-spaced components and decreases in proportion to the harmonic rank of the stimulus frequencies. However, de Boer (1956a, b; and subsequently Schouten *et al.*, 1962; Smoorenberg, 1970) observed that for narrowly-spaced components, the actual pitch shift is somewhat greater than would be predicted by equation (1). Using equation (1), it is possible to calculate the effective harmonic rank ($\underline{n}'$) generating $\underline{p}$:

$$(3) \quad \underline{n}' = f_c' / \underline{p}$$

where f_c' is the effective center frequency and $\underline{n}'$ is an integer. By this method, Smoorenberg was able to determine that the highest effective rank for any inharmonic signal was 8, regardless of the actual rank of the stimulus components. This implied that the pitch of inharmonic signals stems from components separated by at least 12% even when the separation between the physical stimulus components is considerably less. Smoorenberg accounted for this discrepancy by showing that the pitch of narrowly spaced components is derived from combination tones, which effectively extend the signal spectrum down several harmonics (Chapter 1.26). This explained why the pitch of narrowly spaced components was always higher than predicted by equation (1).

Smoorenberg's study implied that the low pitch of complex signals always stems from the resolved portion of the spectrum. This implication appeared incompatible with fine-structure theories of pitch perception (e.g. Schouten *et al.*, 1962), because aurally-resolved components should not interact sufficiently to enable the period of the signal to be encoded in terms of the temporal discharge pattern of auditory units. This was indeed the conclusion drawn by Whitfield (1970a,b) who recorded from units in the cochlear nucleus during presentation of harmonic and inharmonic signals. He found that both types of stimuli produced similar period histograms, and concluded from this that the pitch of complex signals was not likely to be based on the timing behavior of single auditory units.

However, Evans (1978) has shown that Whitfield's conclusion is based on an erroneous assumption. The period histogram provides only a very rough estimate of the degree to which a unit is synchronized to the period of a particular frequency. This is because the index of synchronization is based on the proportion of cell discharge occurring within a time interval equal to one half the period of the stimulating frequency (Chapter 1.43; Figure 1.11). Consequently, the period histogram does not provide the temporal resolution necessary to discriminate between the temporal discharge patterns synchronized to slightly different periods.

Evans (1978) and Javel (1972; 1980) have found that inharmonic stimuli are distinguishable from their harmonic counterparts when the temporal activity of single cochlear nerve fibers is plotted in terms of interspike intervals (Chapter 1.43; Figure 1.12). The interspike

pattern resembles a half-wave rectified version of the signal's temporal fine-structure. This is particularly the case for units with CFs in the vicinity of the stimulus components. Moreover, the fine structure is preserved in the temporal discharge activity even when the stimulus components are separated by 20% or more. In fact, the temporal fine-structure tends not to be preserved as well in the firing pattern when the components are separated by smaller frequency intervals (Javel, 1972). Javel (1980) also found that the fine-structure is preserved less well in the firing pattern of units tuned to frequencies considerably above the components.

The results of Evans (1978) and Javel (1972; 1980) suggest that information concerning the pitch of harmonic and inharmonic signals can be represented in terms of the temporal discharge activity of single cochlear nerve fibers. This conclusion is contrary to what would have been predicted on the basis of psychophysical studies of frequency resolution. Apparently, the seeming absence of perceptual interaction between contiguous frequency components should not be taken to imply that these components do not physiologically interact in the cochlea. It would, therefore, be wise to maintain a distinction between perceptual and physiological indices of cochlear interaction. Thus, a major conceptual premise of contemporary pitch theories - that the low pitch of complex tones cannot possibly be attributed to cochlear interaction of contiguous components - needs to be re-examined in light of Evans' and Javel's findings.

The importance of Evans' and Javel's results lie in their demonstration that the discharge activity of cochlear nerve fibers contain sufficient information for the extraction of low pitch by a central pitch processor operating on the temporal pattern of neural discharge. However, it is not known whether this temporal information is preserved at higher levels of the auditory pathway and, if so, whether the central processing of pitch is actually based on the temporal representation of the cochlear waveform.

To shed light on these questions, and to explore in general the role of temporal neural mechanisms in the processing of inharmonic signals, frequency-following responses were recorded to inharmonic complexes composed of widely-spaced components. These responses were compared with those recorded to harmonic signals of analogous harmonic rank to determine whether the two types of signal could be distinguished on the basis of some property of the FFR.

6.2 EXPERIMENTAL METHODS

Stimuli consisted of five three-component tones. Three were inharmonic signals whose frequency components were non-integral multiples of the difference frequency ($f_{n+1} - f_n = 244$ Hz). Stimulus components were equivalent to $1.5 + 2.5 + 3.5 [f_{n+1} - f_n]$ ($366 + 610 + 854$ Hz); $2.5 + 3.5 + 4.5 [f_{n+1} - f_n]$ ($610 + 854 + 1098$ Hz); or $3.5 + 4.5 + 5.5 [f_{n+1} - f_n]$ ($854 + 1098 + 1342$ Hz). The other stimuli were

harmonic counterparts of the inharmonic signals, whose frequency components were equivalent to either the second, third, and fourth harmonics ($488 + 732 + 976$ Hz) or third, fourth, and fifth harmonics ($732 + 976 + 1220$ Hz) of a 244 Hz fundamental. Thus, the period of the stimulus waveform envelope was the same for both types of stimuli (4.1 msec). However, both the qualitative and quantitative nature of the pitch produced by the inharmonic stimuli is very different from that of the harmonic complexes. The harmonic signals produce a salient low pitch that corresponds closely to the fundamental frequency (244 Hz). The inharmonic tones, on the other hand, produce a much less salient, more "wooden" sounding sensation of pitch. The inharmonic stimuli used in the present experiment can produce, under the appropriate experimental conditions, at least three separate low pitches. Two of these pitches lie either slightly below or slightly above the difference frequency. The third reflects the actual period of the stimulus waveform (8.2 msec) and corresponds to the "true" fundamental of the signal (122 Hz). This last pitch is reported by Gerson and Goldstein (1978) to be difficult to perceive (not all observers can hear it) and is present only when the stimulus components are of relatively low harmonic rank.

FFR was recorded from the scalp of two subjects. For subject LL, overall stimulus intensity was 68 dB SPL (58 dB SPL/component). For subject TM, overall stimulus level was approximately 70 dB SPL ($f_1 = 55$ dB SPL, $f_2 = 60$ dB SPL, $f_3 = 65$ dB SPL). Stimuli were binaurally presented at a rate of nine per second. All other stimulus presentation parameters were as specified in Chapter 2.2.

FFR were recorded differentially between vertex and the left side the neck. The ground electrode was placed on the right neck. FFR averages consisted of between 2500 and 4000 responses for subject LL and 3000 responses for subject TM. All other recording parameters were as specified in Chapter 2.3.

6.3 EXPERIMENTAL RESULTS

Frequency-following responses for a set of harmonic and inharmonic signals are illustrated for a single subject in Figure 6.1. Stimulus waveforms are shown directly above each response. Responses to inharmonic stimuli are illustrated in the left-hand column. FFR to their harmonic counterparts are shown on the right. Several features of these responses are of interest. The harmonic stimuli evoke responses whose periodicity is equal to the fundamental of the stimulus, namely 244 Hz. In contrast, the periodicity of the inharmonic responses is considerably less uniform. This is particularly evident for response 1 (and for response 3, to a lesser degree) where the period appears to be about twice as long as the period of the harmonic responses. This longer period corresponds to the actual repetition rate of the stimulus waveform. Gerson and Goldstein (1978) have shown that under the appropriate experimental conditions a pitch corresponding to this "true" fundamental is perceptible and that this pitch is most salient

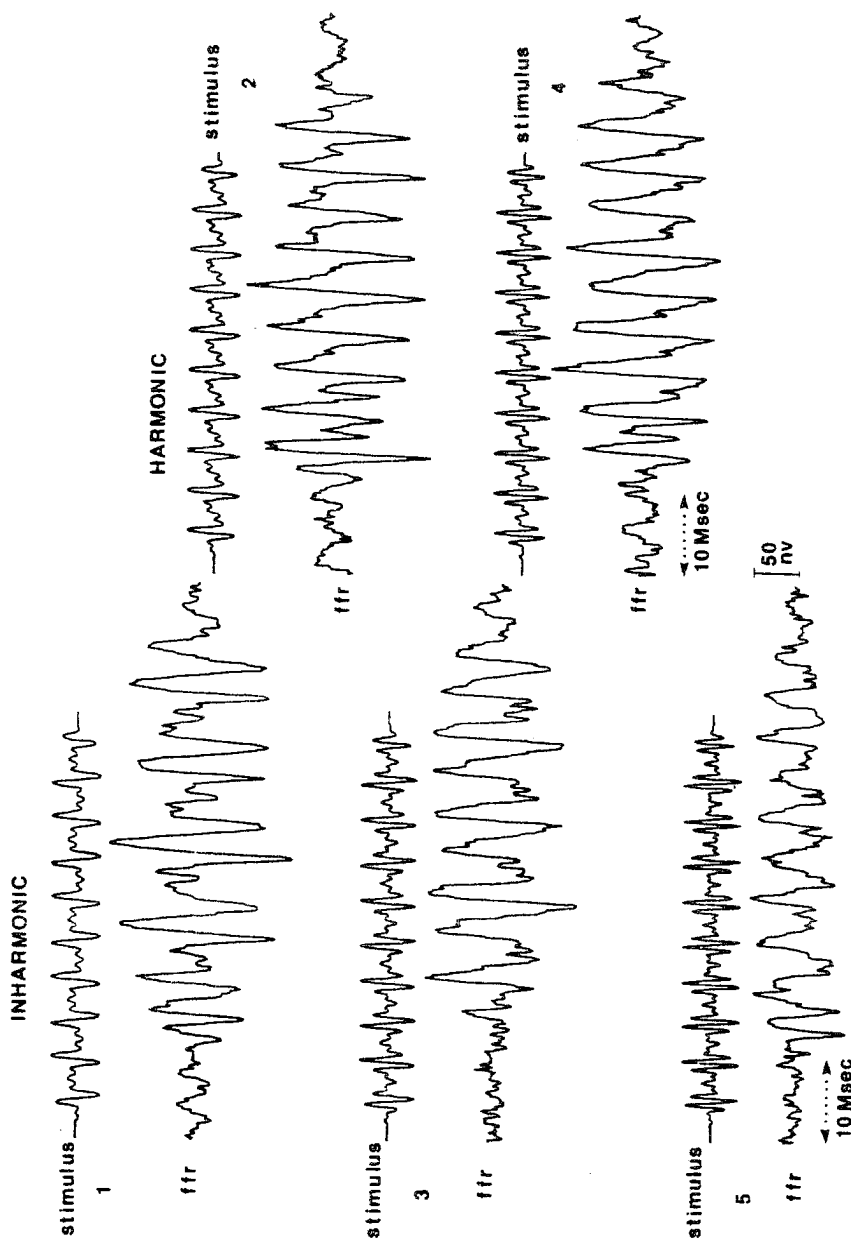


Figure 6.1 FFR recorded to three-component harmonic and inharmonic tones. Responses to the inharmonic stimuli are shown on the left. On the right are FFR generated by their harmonic counterparts. Stimulus waveforms are illustrated above each response. Stimulus spectra of stimuli were as follows: Stimulus 1 = 366+610+854 Hz; Stimulus 2 = 488+732+976 Hz; Stimulus 3 = 610+854+1098 Hz; Stimulus 4 = 732+976+1220 Hz; Stimulus 5 = 854+1098+1342 Hz. Stimulus intensity = 58 dB SPL/component. Averages consist of 4000 responses.

for stimuli composed of the very lowest harmonic ranks. This pattern is similar to that observed in the present set of waveform responses, where the true period is most evident for responses to the lowest inharmonic signals.

Related to this feature of the inharmonic responses is a striking asymmetry in the morphology of the response waveforms. The periodicity of the waveform is more regular for the negative polarity of response 1 and the positive polarity of responses 3 and 5. The opposite polarity is characterized by higher frequency components that may possibly be related to the waveform fine structure. Responses to the harmonic stimuli display no such waveform fine-structure composition or morphological asymmetry.

Fourier analysis of the FFR waveforms indicated that responses to both the harmonic and inharmonic stimuli contain a significant amount of energy at the difference frequency (244 Hz). This fact is not entirely surprising in view of the fact that stimulus levels used in the present experiment were 10-15 dB above FFR threshold. It was shown in Chapter 5 that FFR recorded at these levels reflects the interaction of three contiguous stimulus components and that such interaction produces a large component at the envelope repetition frequency. However, in view of the waveform differences observed between harmonic and inharmonic responses, we sought to determine whether the pitch associated with inharmonic signals might be reflected in some other parameter of the FFR.

Toward this end, the temporal interval between waveform peaks was digitally computed for each response. The response waveform was displayed on a CRT and a cursor was used to mark the position of each peak in the negative polarity of the response waveform. The position of each peak was determined to within one sample point of the actual peak. Consequently, the temporal resolution of this analysis is limited only by the sampling rate of the computer (15,623 samples/sec). The temporal resolution is approximately 64 microseconds, indicating a frequency resolution of about 4 Hz in the relevant spectral region (244 Hz).

The results of this analysis (performed on the negative polarity of each response) is shown for two subjects in Table 6.1. Each interval value represents the mean of from seven to twelve measurements, and is expressed as a proportion of the interval corresponding to the difference frequency (4.1 msec). The standard deviation is indicated to the right of each mean.

STIMULUS Harmonic	SUBJECT: LL		SUBJECT: TM		THEORETICAL	
	Interval	S.D.	Interval	S.D.	F1+F2	F2+F3
1. 488+732+976	1.00	.03	1.01	.03	1.00	1.00
2. 732+976+1220	1.01	.04	1.00	.02	1.00	1.00
<u>Inharmonic</u>						
3. 366+610+854						
Low Interval	0.77	.01	0.76	.08	0.79	0.85
High Interval	1.25	.02	1.24	.05	1.37	1.21
4. 610+854+1098						
Low Interval	0.92	.02	0.89	.04	0.85	0.89
High Interval	1.12	.02	1.13	.04	1.21	1.15
5. 854+1098+1342						
Low Interval	0.89	.03	0.90	.02	0.89	0.91
High Interval	1.14	.03	1.11	.02	1.15	1.11

TABLE 6.1 Interval Analysis for Responses to Harmonic and Inharmonic Stimuli.

Intervals in the right-hand columns represent theoretical estimates of the predicted intervals based on the assumption that the responses were generated by the interaction of either the lower two (f_1+f_2) or upper two (f_2+f_3) frequency components. For the harmonic signals, there will, of course, be no difference in the period of the waveform produced by these pairs of components. The predicted interval will always correspond to the period of the waveform envelope. This is consistent with the fact that stimulus components of different harmonic rank produce (approximately) the same pitch. Comparison of the computed with the predicted intervals shows that the responses to the harmonic signals conforms to the difference frequency interval with a very small degree of variance.

The situation for the inharmonic signals is very different. In these instances, the predicted intervals correspond not to the envelope periodicity, but to the pseudo-periods of the waveform (Figure 1.4). When the component frequencies are midway between harmonics of the difference frequency, two pseudo-periods describe the waveform periodicity equally well. However, the duration of these pseudo-periods depends on the harmonic rank of the stimulus components. For components of low harmonic rank, the pseudo-periods deviate considerably from the period of a harmonic stimulus (Stimulus 3). Thus, the intervals measured for the FFR to stimulus 3 deviate by as much as 25% from the intervals measured for responses to stimuli 1 and 2. For components of higher harmonic rank, the pseudo-periods will be much closer to the period of the harmonic waveform. Consistent with this pattern, one

observes that the intervals computed for FFR to stimuli 4 and 5 are much closer to the period of the harmonic FFR. It is of interest that the interval analyses for these stimuli are practically identical, for both subjects. This suggests that the FFR may be generated by the same pair of frequency components in both instances ($854+1098$ Hz). For stimulus 4 these would be f_2 and f_3 , while for stimulus 5, f_1 and f_2 would be the generating components. This analysis is consistent with psychophysical studies of pitch generated by analogous inharmonic comb-filtered noise signals (Ritsma and Bilsen, 1970; Yost and Hill, 1978). These studies indicate that the pitches generated by this type of a broadband inharmonic signal are equal to about .89 and 1.14 of the difference frequency regardless of the harmonic rank of the lowest spectral peak, as long as at least two of the peaks are in the dominance region for low pitch. In the present experiment, the analogous signals were stimuli 4 and 5, both of which contain components in the dominance region. The results of this interval analysis are largely consistent with psychophysical studies of inharmonic pitch. Those stimuli lying within the dominance region generate responses whose interpeak intervals are very close to those obtained in perceptual experiments with spectrally similar types of signals.

In view of the close correspondence between the interval analysis and psychophysical measures of inharmonic pitch, why are the same results not obtainable through a conventional Fourier analysis? The present analysis computed the intervals for only a single polarity of the response waveform. An similar analysis for the opposite polarity did, in fact, yield results comparable to the original Fourier analyses, namely a predominance of intervals equal to the period of the waveform envelope. This fact suggests that information relevant to the pitch of the signal may be reflected in only a very restricted portion of the FFR waveform.

Another means by which to isolate FFR activity of more direct relevance to the processing of pitch is an adaptation of the compound histogram originally developed by Gobleck and Pfeiffer (1969) and Arthur et al (1971) for the temporal analysis of single unit discharge patterns. To isolate neural activity that is temporally synchronized to a particular stimulus frequency, the stimulus is presented twice, once in its original polarity, and a second time with its polarity inverted. The temporal record of discharge activity obtained under this second condition is then itself inverted in polarity and added to the activity collected with the initial stimulus polarity (Huis in't Veld et al, 1977a,b). Thus, activity synchronized to a restricted portion of the stimulus cycle will summate, while background activity not phase-locked to the stimulus frequency will tend to cancel or at least be significantly attenuated. This procedure is illustrated in Figure 6.2 for FFR recorded in response to a pure tone stimulus. On the left are responses obtained with opposite stimulus polarity conditions. Notice that the onsets of the two responses occur at approximately the same time and are consequently independent of stimulus polarity. The remainder of the response is, however, synchronized to the stimulus

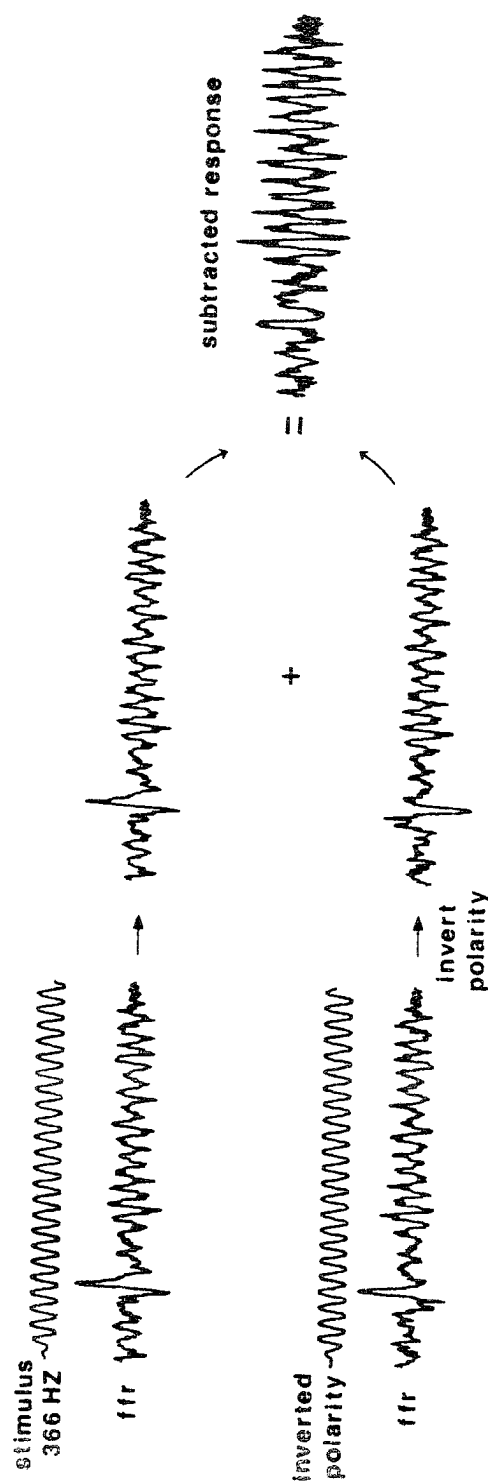


Figure 6.2 An illustration of the response subtraction procedure for a pure tone stimulus (366 Hz). The left-hand column shows the FFR generated by the signal presented in opposing stimulus polarities. The polarity of one of the responses is subsequently inverted, as shown in the middle trace (bottom). The two responses are then added to form the composite, subtracted response. This procedure is equivalent to subtracting one response from the other. The resulting waveform is an FFR which, in principle, reflects only activity phase-locked to the stimulus frequency.

waveform. In the middle column, the polarity of the bottom response is inverted and then added to the other response (equivalent to subtraction of one waveform from the other). The result is a waveform whose synchronized component is approximately twice the magnitude of the original response. Other portions of the response, such as the onset component are reduced somewhat in magnitude. Thus, this "response subtraction" has the effect of enhancing those aspects of the FFR reflecting neural activity phase-locked to the stimulus frequencies.

In Figure 6.3, the response subtraction procedure is applied to the analysis of FFR generated by a three-component inharmonic signal and the results compared with an alternative procedure of "response addition." The right-hand column illustrates the response subtraction procedure. The top trace is the response generated by the signal in its original polarity. The middle trace is the FFR produced by the same stimulus with its polarity inverted. The bottom trace is the subtracted waveform. This last waveform is distinguished from the original responses in that it contains less energy in the low frequency region and emphasizes the part of the spectrum containing the component frequencies. Although its low-frequency spectrum is reduced in magnitude, it contains peaks at locations that are consistent with both the interval analysis presented in Table 6.1 and the results of numerous psychophysical experiments. In contrast, the response addition procedure yields a very different result. The low-frequency component of the FFR is enhanced, while the higher frequencies associated with the spectral components are greatly diminished. The fact that the low-frequency periodic component of the FFR does not shift when the stimulus polarity is inverted indicates that this aspect of the FFR is the result of activity synchronized to the stimulus envelope rather than to the stimulus frequencies. The spectrum of this added response is dominated by a large peak at the waveform envelope frequency, 244 Hz.

6.4 DISCUSSION

Inharmonic signals have served an important function in the development of recent pitch theory. Because the pitch produced by these signals is very different from the frequency of the waveform envelope and the difference tone, they serve, in principle, to distinguish low pitch from such contaminating sensations as roughness and combination tones. However, the very nature of the pitch they produce is extremely ambiguous and highly ephemeral. This is particularly true for signals such as those used in the present experiment in which the components are equivalent to odd harmonics of a common fundamental. It is not entirely clear why the pitch of inharmonic signals sounds so different from that produced by harmonic complexes (Chapter 1.20 (B)) nor it is clear whether this pitch is actually based on precisely the same mechanisms that mediate the pitch of harmonic complexes. It is possible that the greater salience of harmonic pitch is due, at least in part, to contributions made by combination tones and roughness.

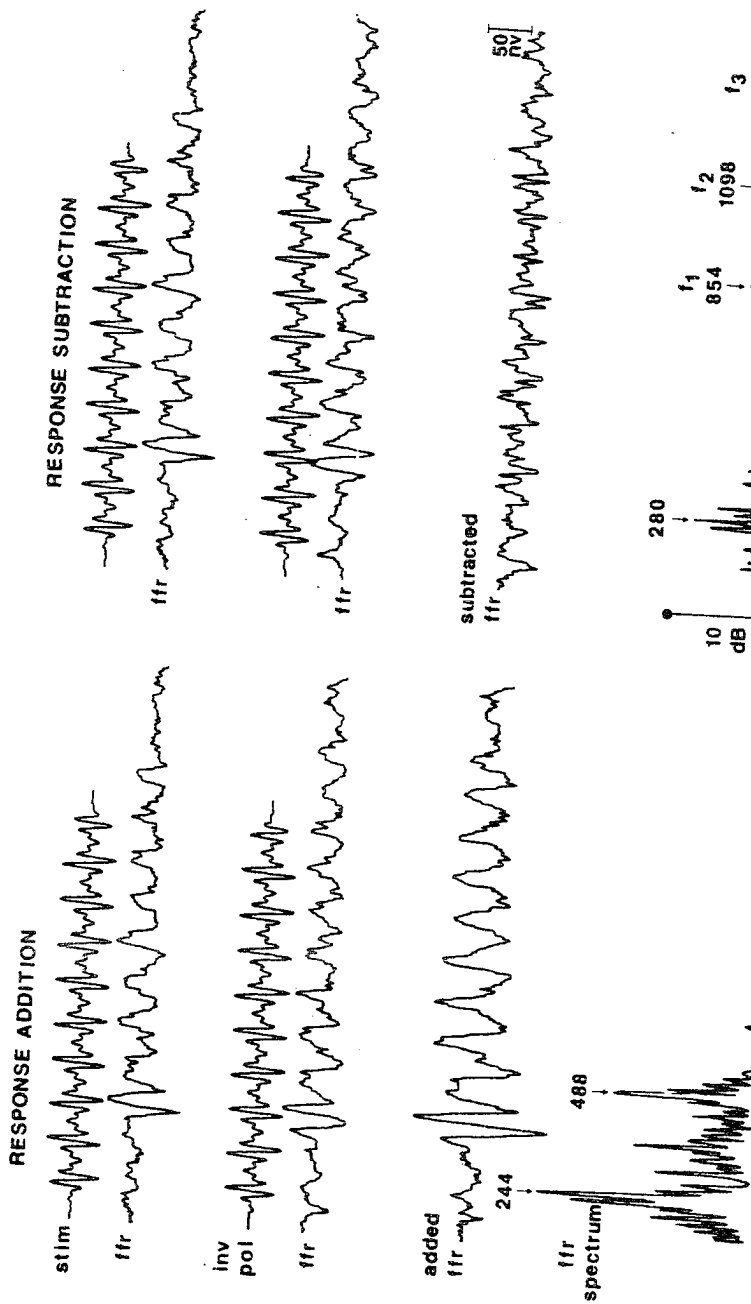


Figure 6.3 The response subtraction procedure illustrated for a complex, inharmonic tone and compared with the waveform and spectra of the composite response obtained through response addition. On the right, the response subtraction procedure is illustrated for a signal consisting of frequencies 854, 1098 and 1342 Hz. The spectrum of the subtracted FFR contains a primary peak at 280 Hz and smaller peaks at the stimulus frequencies. The primary peak corresponds to the frequency of one of the pitches produced by the stimulus. Addition of FFR generated by opposite polarity stimuli results in a composite response whose major spectral component (244 Hz) corresponds to the envelope periodicity. The component at 488 Hz probably represents a rectification distortion product of the primary spectral peak. Stimulus intensity was 70 dB SPL/component. Each average consists of 2000 responses. Composite averages consist of 4000 responses.

In light of the unusual nature of inharmonic pitch, it is perhaps not so surprising that the present results are themselves somewhat ambiguous. Spectral analysis of the full FFR waveform indicates a substantial component at the difference frequency for the inharmonic stimuli. Only when the spectrum of a single polarity of the waveform is analyzed does the analysis suggest a possible relationship between the FFR spectrum and the pitch of the inharmonic signals.

Why should the full waveform of the FFR reflect the envelope of the stimulus waveform? One possibility is that it does indeed reflect the pitch sensation produced by these stimuli. Patterson (1973) and Patterson and Wightman (1976) have found that pitch matches to analogous stimuli often correspond to the difference frequency. Although this result was interpreted by them as a match to the octave above the so-called "true" fundamental, it is also possible that for certain inharmonic signals, the difference frequency does produce the most salient pitch. Another possibility is that the 244 Hz component in the present series of responses reflects the rectification product of a 122 Hz (true fundamental) component filtered out of the FFR. A third possibility relates to the observation made in Chapter 5 that the FFR can reflect the waveform stimulus amplitude modulation characteristic when stimulus intensity is sufficiently high. At this higher stimulus level, it is likely that a substantial contribution is made to the scalp response by units tuned to frequencies above the stimulus components. Javel (1980) has shown that such units tend to follow the stimulus envelope rather than the fine-structure of the stimulus waveform. If this is the case, the difference frequency component of the scalp response may very well reflect the contribution of higher frequency units.

However, as was shown by the interval analysis and spectrum analysis of a single polarity of the scalp response, there does appear to be certain information contained in the FFR that does correspond to the pitches produced by the inharmonic signals. In an attempt to isolate this potentially pitch-relevant information, FFR was recorded to inharmonic signals of opposite polarity and the response generated by one polarity subtracted from the other response. This technique effectively cancels most of the envelope-following component of the scalp response, leaving behind activity that is synchronized primarily to the stimulus components. When the responses generated by opposite polarity stimuli are added, instead of subtracted, the result is a response synchronized to the waveform envelope. It is tempting to infer from this pattern that the envelope-following component of the FFR reflects the activity of units tuned to frequencies substantially above the component frequencies, while the component synchronized to the stimulus frequencies reflects the activity of units tuned to frequencies in the general vicinity of the stimulus spectrum. Consistent with this interpretation is Huis in't Veld *et al's* (1977a) observation that the latency of FFR obtained with the response subtraction procedure does not substantially change with increasing signal level. This would imply that the procedure eliminates the contribution made by the more basal units to the scalp response.

Evans (1978) and Javel (1980) suggest that information concerning the fine-structure of the cochlear waveform related to the pitch of inharmonic signals is reflected principally in the temporal discharge pattern of those cochlear nerve fibers tuned to frequencies in the vicinity of the stimulus components. Thus, any recording technique that monitors a substantial number of higher frequency units will be unable to "pickup" neural activity synchronized to the fine-structure of the waveform. Rather, the activity will primarily reflect the envelope of the stimulus waveform (Javel, 1980). The present results are generally consistent with the conclusions of Evans (1978) and Javel (1980). When the scalp response reflects the activity of a relatively broad range of brainstem units, the FFR spectrum is dominated by energy at the envelope frequency (244 Hz). Elimination of the envelope-following component through the response subtraction procedure reduces the FFR to a waveform with energy at both the stimulus components and at frequencies relevant to the pitch produced by these signals.

However, the present results are far from conclusive. There is no clear correspondence between the pitch produced by inharmonic signals and the spectrum of the FFR. In part, this lack of direct correspondence may reflect the fact that scalp FFR records the activity of a larger unit population than is directly relevant for the processing of inharmonic pitch. Cancellation of the high-frequency component of the FFR leaves a waveform that appears to bear a closer correspondence to the pitches produced by the signal. Yet the results are hardly overpowering and are not always replicable. Another reason for the lack of conclusive results may have to do with the very nature of inharmonic pitch. The sensation produced by inharmonic signals is highly ephemeral. Sometimes the observer has difficulty matching any sort of pitch at all. It is conceivable that the neural information upon which such pitch matches are based is itself so subtle as to elude detection by conventional recording techniques. The data of Evans (1978) and Javel (1980) lend support to this view. For in neither instance do the interval histograms show unambiguous peaks at the predicted values. Rather, in the case of Evans' study, the histogram makes it clear that the cochlear nerve fiber is primarily synchronized to one of the component frequencies. However, some of the peaks are slightly larger than the others. These intervals correspond to the approximate pitches produced by the signal. The interval histograms presented in Javel's study show an even less direct correspondence between neural timing behavior and the pitches produced by the stimuli. If inharmonic pitch is based on such activity, it must surely involve some very complex neural processing at higher levels to extricate the pitch-relevant information from the mass of neural timing information available in the auditory periphery.

CHAPTER VII

FREQUENCY-FOLLOWING RESPONSE TO SYNTHETIC VOWELS

7.1 INTRODUCTION

The intelligibility of speech is preserved over a range of at least 100 dB SPL. In contrast, the dynamic range of most cochlear nerve fibers is limited to approximately 20-40 dB. The discharge rate of single cochlear fibers increases with intensity up to a certain level and then saturates. At moderate-to-high signal levels, most cochlear nerve fibers will have reached their saturation level in response to a broadband stimulus. As a consequence, the spectrum of two or more broadband sounds cannot be distinguished on the basis of the spatial excitation pattern of the cochlear nerve. Although it is possible that lateral inhibitory mechanisms in the cochlear nucleus and beyond serve to enhance the dynamic range of more central neurons, their relevance for neural encoding of speech remains untested and consequently remains in doubt. Until such evidence is forthcoming, it should probably be assumed that the auditory representation of speech sounds is based on some parameter other than the spatial excitation pattern of neural elements.

One possible alternative means by which the spectra of low-frequency speech sounds (such as vowels) may be encoded, is in terms of the temporal discharge pattern of auditory neurons. Young and Sachs (1979) and Sachs and Young (1980) have shown that the spectra of vocalic stimuli are preserved in the synchronized discharge patterns of single cochlear fibers over an intensity range of at least 70 dB. The extended dynamic range of the temporal representation is based on the suppression of synchronized activity to frequencies outside the formant regions. This "synchrony suppression" becomes apparent at moderate-to-high intensities (>60 dB SPL) when the discharge rate of most cochlear nerve fibers has reached saturation level. At these levels, fibers which were synchronized, at lower intensities, to frequencies in the center of their response area, synchronize instead to frequencies in the region of the first formant. Thus, the number of fibers synchronized to F1 increases with level. At high stimulus intensities, therefore, a considerable proportion of low-frequency fibers are synchronized to the first formant. In contrast, the number of fibers synchronized to F2 or F3 does not appreciably increase as a function of signal intensity. Rather, the number of units locked to these formants (particularly F2) remains relatively constant over a large intensity range. This fact is somewhat surprising in view of the behavior of units tuned to frequencies below F2. One would expect a similar pattern of synchrony recruitment for F2 and F3. Instead, these high-frequency fibers synchronize to F1 at levels above 50-60 dB SPL.

At the highest levels used in Young and Sachs' study (77-78 dB SPL)

one observes an almost total domination of the synchronized response of low-frequency cochlear fibers by the first formant. Exceptions to this domination include units synchronized to F2 and to frequencies lower than F1. The synchronized response to F3 diminishes at the highest intensity level, apparently suppressed by the response to either F1 or F2. The synchronized response to F2 may also become suppressed at this level when it is relatively close to F1 (e.g. the vowel [a]).

Another parameter of the synchronized response that changes with level is the representation of fundamental frequency. At low-to-moderate intensities (<50-60 dB SPL), a substantial number of higher-frequency fibers synchronize to the fundamental. This synchronization does not, however, represent a response to the lowest spectral component. Rather, these units are phase-locked to the envelope of the cochlear waveform created by the interaction of unresolved components. This is evident from the fact that synchronization to the fundamental occurs primarily for fibers tuned to frequencies above the eighth harmonic. Hashimoto et al (1975) and Kiang et al (1979) have also observed this "pitch-synchronous" response for the activity of high-frequency fibers. One also observes, at this low intensity level, a synchronized response in the temporal patterns of fibers tuned to frequencies below F1. As stimulus level increases, the number of high-frequency units exhibiting this pitch-synchronous activity diminishes, although there is still a substantial number of such fibers synchronized to the fundamental. At the same time, the synchronized response to the fundamental increases in the fibers tuned below F1. The activity of the low-frequency fibers may very well reflect the sort of cochlear interaction between contiguous components observed by Rose et al (1969) and Javel (1972; 1980) [Chapter 6]. At the highest intensities, there is a further decrement in the synchronized response of high-frequency units and a slight decrease in the response of low-frequency fibers as well.

Young and Sachs' study suggests that at the level of the cochlear nerve, the spectrum of low-frequency speech sounds is likely to be encoded in terms of the temporal discharge pattern of cochlear fibers. The first (and to a lesser degree), second, and third formants maintain their auditory representation in terms of the periodic discharge activity of single fibers over a dynamic range spanning at least 70 dB. Over this same range, synchronized activity to most other portions of the spectrum is effectively suppressed. This portrait of cochlear nerve activity is significant because it provides a possible physiological basis for many features of vowel perception. Among those aspects of vowel coding that are consistent with their data are the following:

- (a) Vowel identity is relatively independent of relative formant amplitude (Ainsworth and Millar, 1972; Chistovich et al, 1979). Thus, the level of F2 can be as low as 28-45 dB below F1 before a significant change in vowel quality occurs. This is consistent with the fact that the synchronized response of cochlear fibers to F2 is not appreciably suppressed by F1, even when the amplitude of F2 is 20 dB lower than F1 (for the vowel [i]).

- (b) It is difficult to distinguish single formant approximations to two-formant synthetic vowels when the frequency of F2 is relatively close to F1 (Carlson et al., 1970;1975; Chistovich et al., 1979). This is possibly related to the synchrony suppression observed for F2 in the vowel [a] at intensities above 60 dB SPL.
- (c) Pitch discrimination of synthetic vowels does not deteriorate upon filtering out the lowest frequency components (Klatt, 1973). This is consistent with the fact that a large synchronized response to the fundamental frequency is observed in fibers tuned to harmonics above the fundamental component.

These parallels between psychophysical and physiological studies suggest that at least certain features of vowel perception may be based on the temporal pattern of neural discharge. In particular, Young and Sachs' (1979) study provides a compelling case for the auditory analysis of vocalic spectra being based on the synchronized activity of low-frequency fibers. However, it is not clear whether the vocalic spectrum continues to be represented in terms of synchronized activity at higher levels of the auditory pathway. No study comparable to Young and Sachs' has been performed for any of the auditory brainstem nuclei. Nevertheless, several studies do suggest that neural periodicity continues to play a role in the encoding of vocalic spectra at higher levels of the pathway. Moore and Cashin (1974; 1976; 1977), Rupert et al. (1977), and Caspary et al. (1977) have each demonstrated that units in the anteroventral cochlear nucleus exhibit phase-locked behavior in response to vowel sounds. In particular, many low-frequency units synchronize their discharge pattern to the frequency of the first formant at low intensities. At higher intensities, the discharge pattern of a unit is affected by the amount of energy falling into its inhibitory sidebands. However, relatively little quantitative data describing this phenomenon are available at present. Many units also exhibit a pitch-synchronous response to vowel sounds. Again, there is little data available, but it appears that this form of synchronized activity is present at both lower and upper levels of the brainstem pathway. Caspary et al. (1977) found that most units tuned to frequencies outside the formant region exhibited this response in the AVCN. Watanabe and Sakai (1978) have observed pitch-synchronous activity for many units in the inferior colliculus.

The present study sought to determine whether the spectrum of vowel sounds might be temporally coded at the level of the upper auditory brainstem pathway. Toward this end, frequency-following responses were recorded from the scalp of human subjects during presentation of synthetic vowels. Of particular interest was whether the first and possibly higher formant would be discernable in the spectrum of the scalp response.

7.2 EXPERIMENTAL METHODS

(a) Stimulus Generation and Presentation

Stimuli consisted of two sets of digitally-synthesized signals corresponding to the vowels [I] (hid), [E] (head), [ae] (had), and [a] (hot). These stimuli were synthesized on a digital parallel resonance synthesizer (Note 7.1), which enabled independent control of formant frequency, formant level, and formant bandwidth. The synthesizer calculated the digital waveform values for a single glottal pulse and subsequently concatenated a series of these pulses to produce the stimulus output. Stimuli were recorded on audio tape and redigitized and stored on the Interdata computer for on-line signal presentation.

Formant frequencies for each vowel were as follows:

Vowel	F1	F2	F3
[I]	400	2400	3000
[E]	600	1800	2400
[ae]	800	1600	2400
[a]	800	1200	2400

Fundamental frequency was 100 Hz for one set of stimuli, and 200 Hz for the other set. Formant bandwidth was 50 Hz for F1, 75 Hz for F2, and 150 Hz for F3. F2 amplitude was usually 5 dB to 10 dB below the level of F2, even for vowels such as [I] and [E] where the F1-F2 level disparity is usually 15-20 dB in natural speech (Peterson and Barney, 1952). These levels were designed to compensate for the low-pass characteristics of the scalp-recorded FFR (Chapter 1.72). Stimulus duration was 60 msec. Stimuli were bandpass-filtered between 200 and 3500 Hz (Note 7.2) and presented binaurally at a rate of nine per second (Note 2.1). Stimulus intensity was 70-75 dB SPL. Other stimulus presentation parameters were as specified in Chapter 2.1.

(b) Recording Procedure

FFR was recorded differentially between vertex and one side of the neck. The other side of the neck served as ground. Averages usually consisted of 2000-4000 responses. Other recording parameters were as specified in Chapter 2.2.

(c) Subjects

Subjects consisted of four, normal hearing subjects (three male, one female) between the ages of 11 and 56.

(d) Spectral Analysis

Responses were analyzed using an analog spectrum analyzer (Figure 2.2) with an approximate spectral resolution of 50 Hz.

7.3 EXPERIMENTAL RESULTS

Frequency-following responses to each set of vocalic stimuli are illustrated for a single subject in Figure 7.1. The stimulus waveform (microphone response) is shown above each response. Several features of these responses are of interest. First, the response latency of the FFR (6-8 msec) to these synthetic vowels is similar to that of FFR recorded to low-frequency pure tones (Figure 1.16) and is consistent with an upper brainstem source for the FFR. Second, most of the FFR waveforms have a well-defined periodic quality that corresponds to the periodicity of the stimulus waveforms. Because of the sharp rolloff characteristic of the headphones, frequencies below 200 Hz are greatly attenuated. Consequently, the signal level for the 100 Hz component was at least 40 dB below the level of F1. Therefore the periodic responses recorded to these stimuli reflects a "missing fundamental" (Chapter 3). Waveform periodicity is most apparent for those responses generated by stimuli whose first formant lies within the region of dominance (Chapter 4). Thus, the vowels [I] ($F_1=400$) and [E] ($F_1=600$) generate the most periodic responses when the fundamental is 100 Hz, but produce a relatively small periodic response when the fundamental is equal to 200 Hz. The vowels [a] and [ae] ($F_1 = 800$ Hz) exhibit a complementary pattern, producing a strong periodic response when the fundamental is 200 Hz, but generating a considerably less periodic waveform when the fundamental is 100 Hz.

Spectral analysis of FFR to the vowel [I] ($F_0 = 200$ Hz) is shown in Figure 7.2 for a group of four different subjects. These responses have certain features in common. First, each response contains a significant amount of energy at the fundamental frequency of the stimulus (200 Hz). This component can be as large as the spectral component corresponding to the first formant (e.g. subject TM). Second, the energy maximum in each response corresponds to the frequency of the first formant, 400 Hz. However, the degree to which this spectral component dominates the response varies from subject to subject. Sometimes there is a considerable amount of energy at other low-frequency components, as seen in the traces for subjects WB and TM. More often, though, one observes a suppression of spectral energy to frequencies above the first formant. This phenomenon is not merely a reflection of the relative amplitude of the stimulus components, but represents rather a form of response suppression analogous to that observed in single cochlear nerve fibers by Young and Sachs (1979).

Figures 7.3 - 7.6 illustrate the relationship between the stimulus spectrum and spectrum of the FFR for eight different vowels. This relationship is complex and depends, to a large extent, on the particular vowel. Several aspects of these responses are of particular interest:

(a) Each response has a considerable amount of energy in the low frequencies, regardless of the amplitude of the corresponding spectral components in the stimulus. This pattern is particularly evident for the vowels [a] and [ae], where the level of the lowest harmonics in the

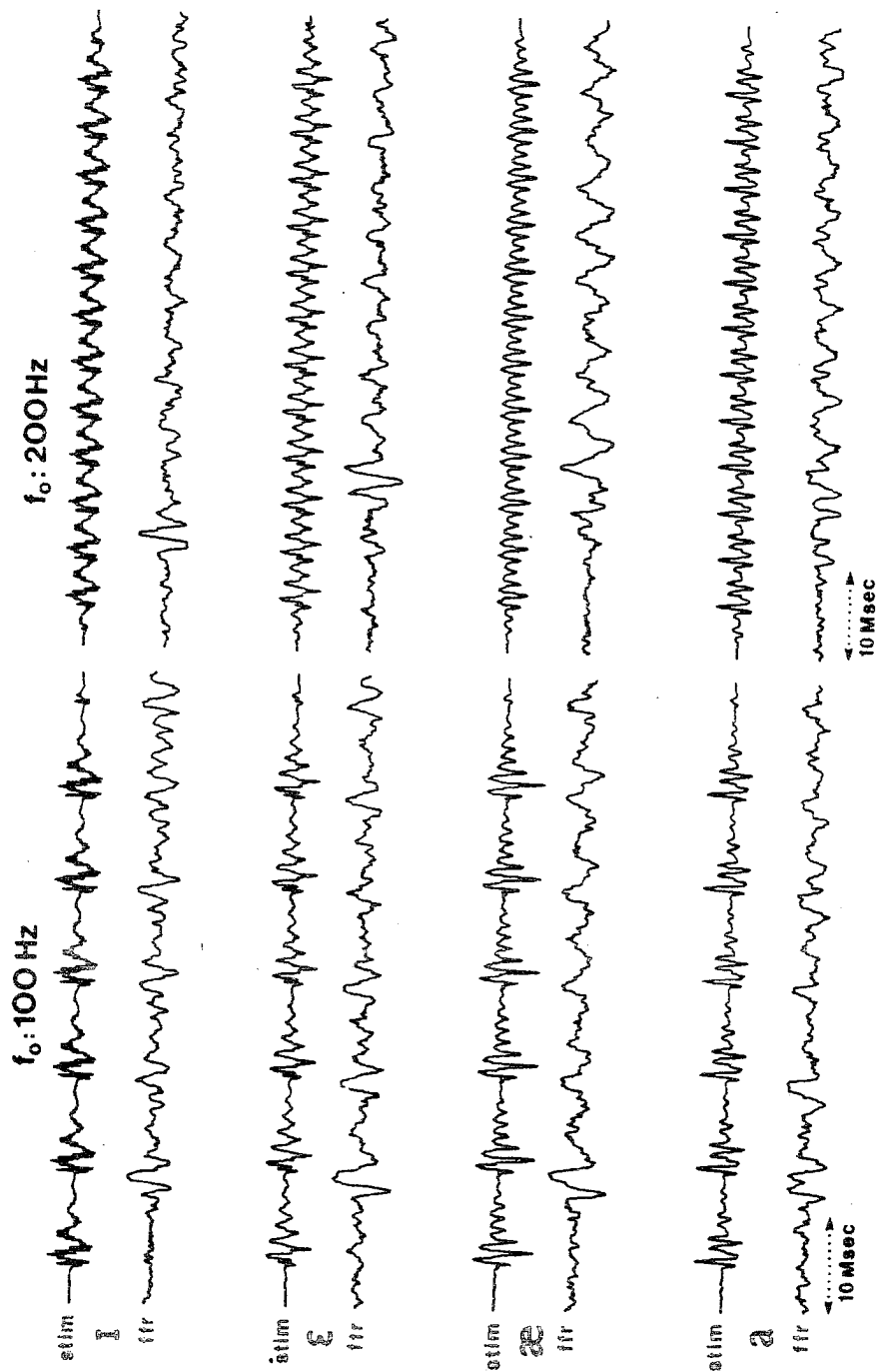


Figure 7.1 FFR recorded in response to synthetic vowels. The stimulus waveform is shown above each FFR waveform. On the left are the responses generated by stimuli whose fundamental frequency = 100 Hz. On the right are responses to stimuli whose fundamental = 200 Hz. Stimulus intensity = 70 dB SPL. Averages consist of 2000 responses.

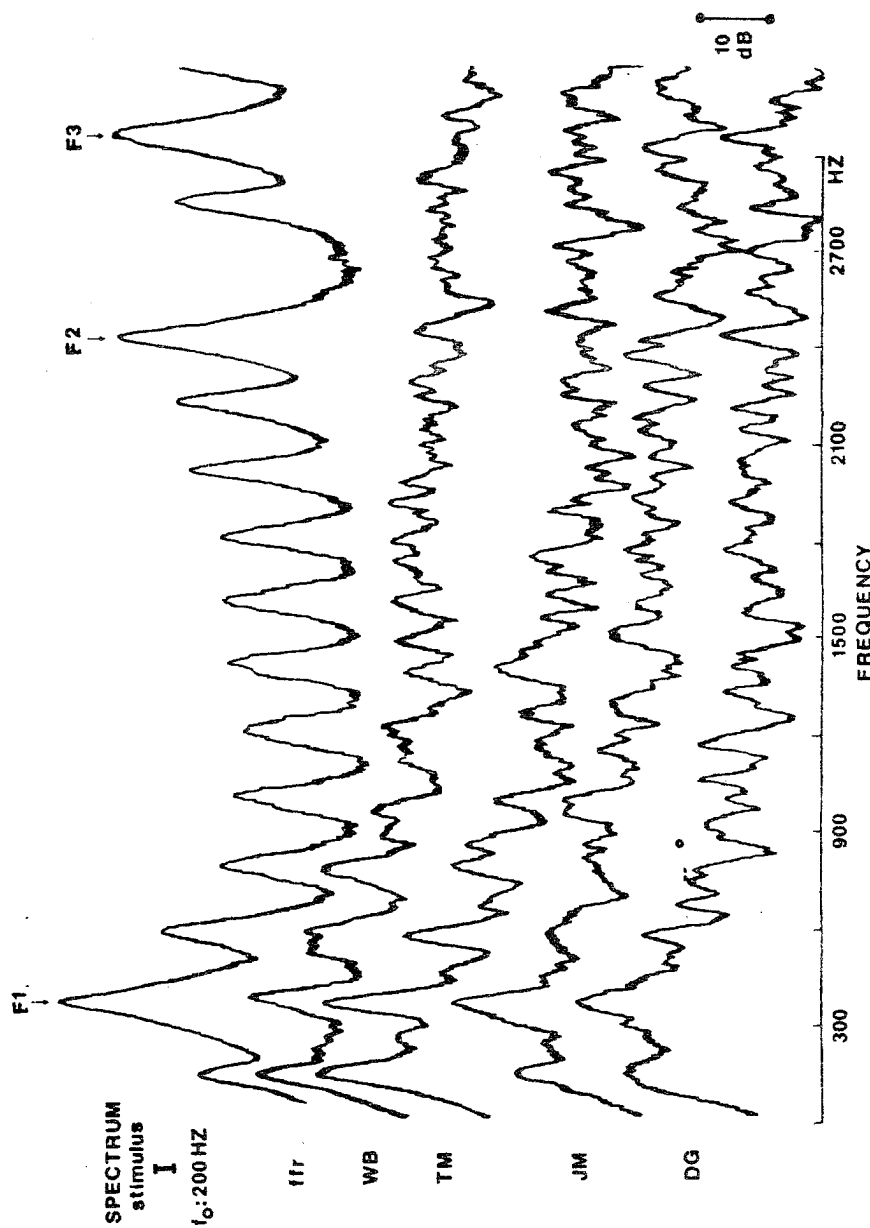


Figure 7.2 Spectra of FFR recorded to the vowel [I] for four different subjects (WB-DG). Top trace is the stimulus spectrum. Spectral peaks corresponding to the first three formants are indicated by arrows. Bandpass of the spectral analysis = 200-3200 Hz. Spectral analysis performed using an analogue spectrum analyzer (Chapter 2.4). Spectral resolution was approximately 50 Hz.

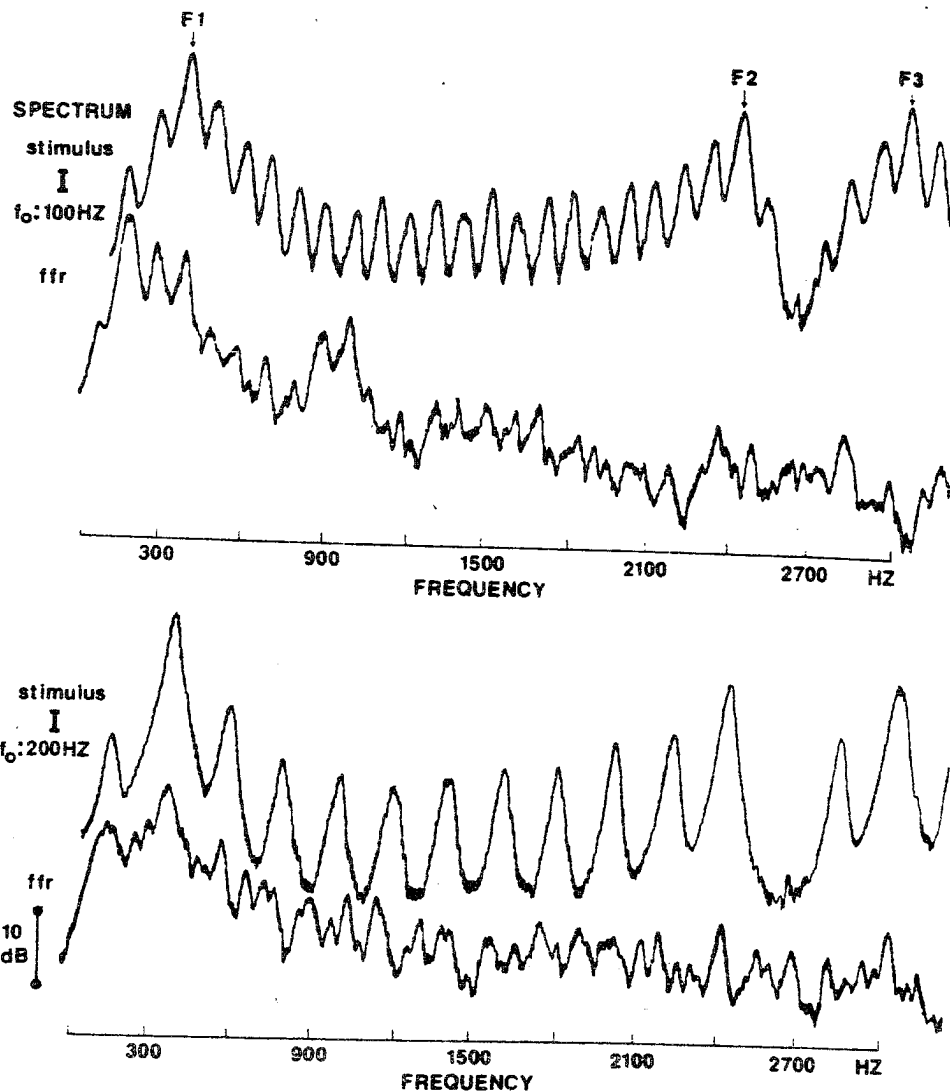


Figure 7.3 Spectral analysis of FFR recorded to the vowel [I] for a single subject. Stimulus and FFR spectra are shown for stimuli of two different fundamental frequencies.

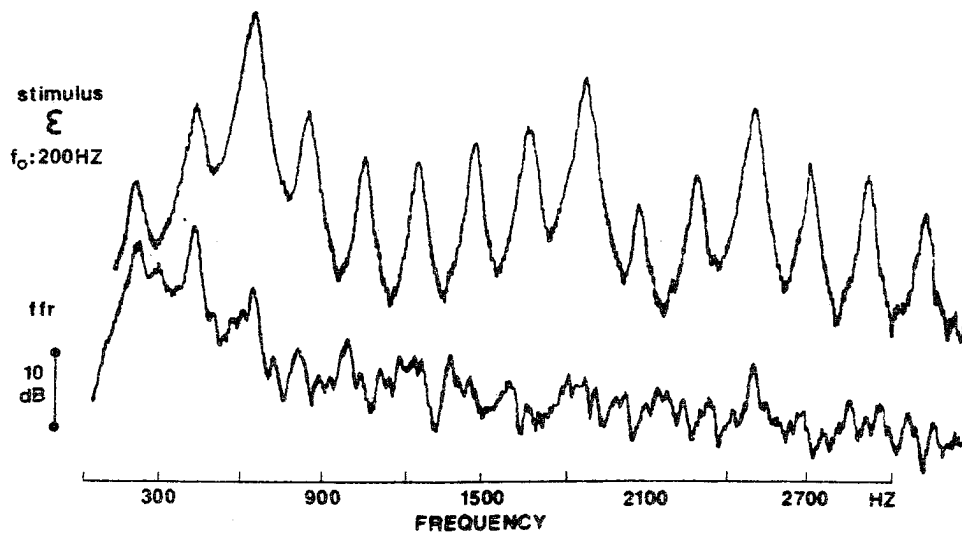
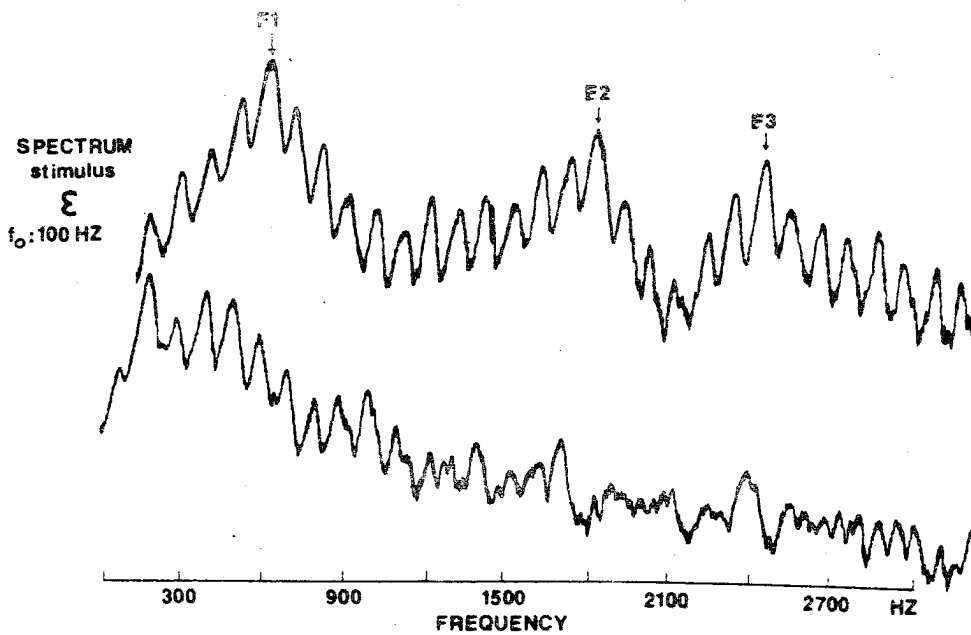


Figure 7.4 Spectral analysis of FFR recorded to the vowel [E] for a single subject.

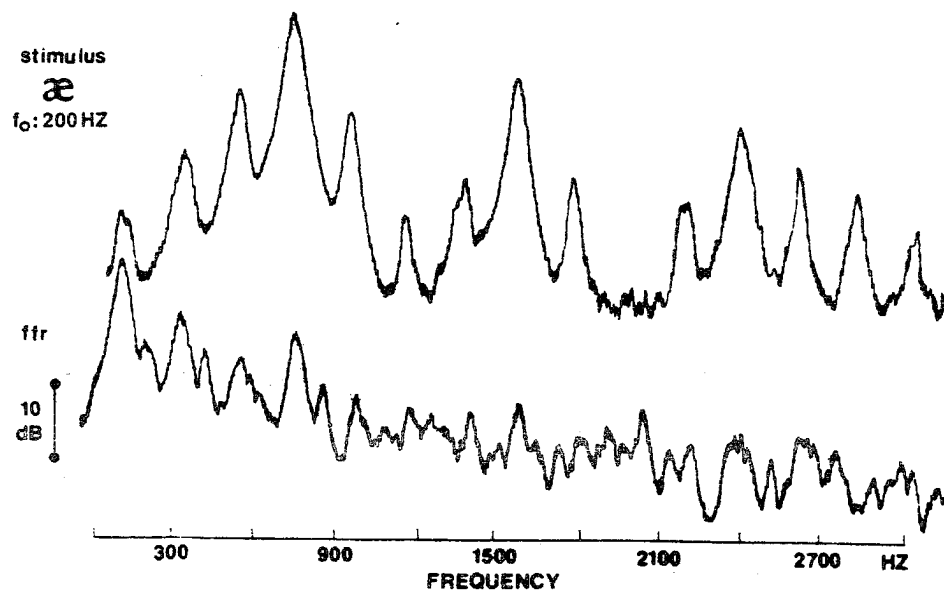
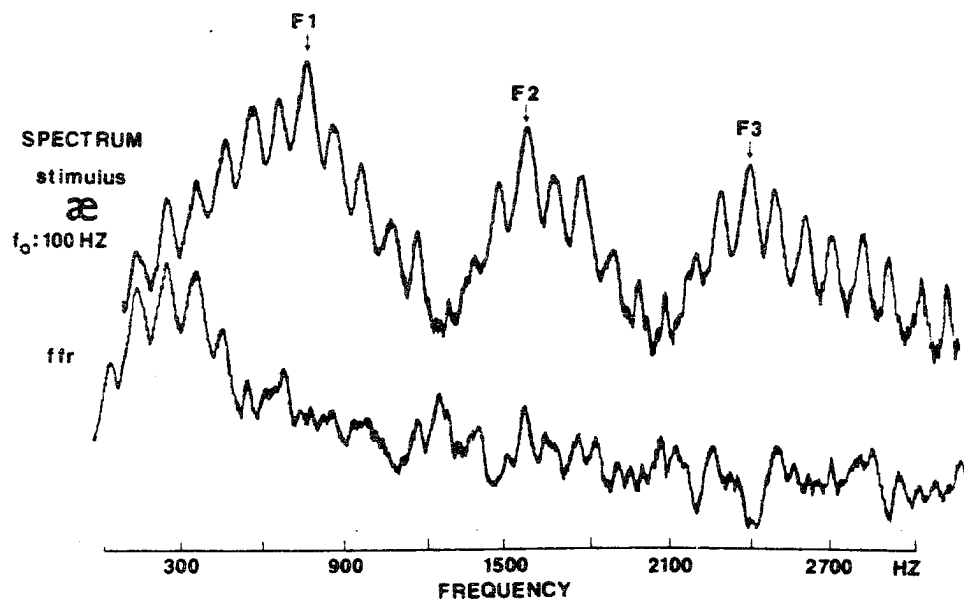


Figure 7.5 Spectral analysis of FFR recorded to the vowel [æ] for a single subject.

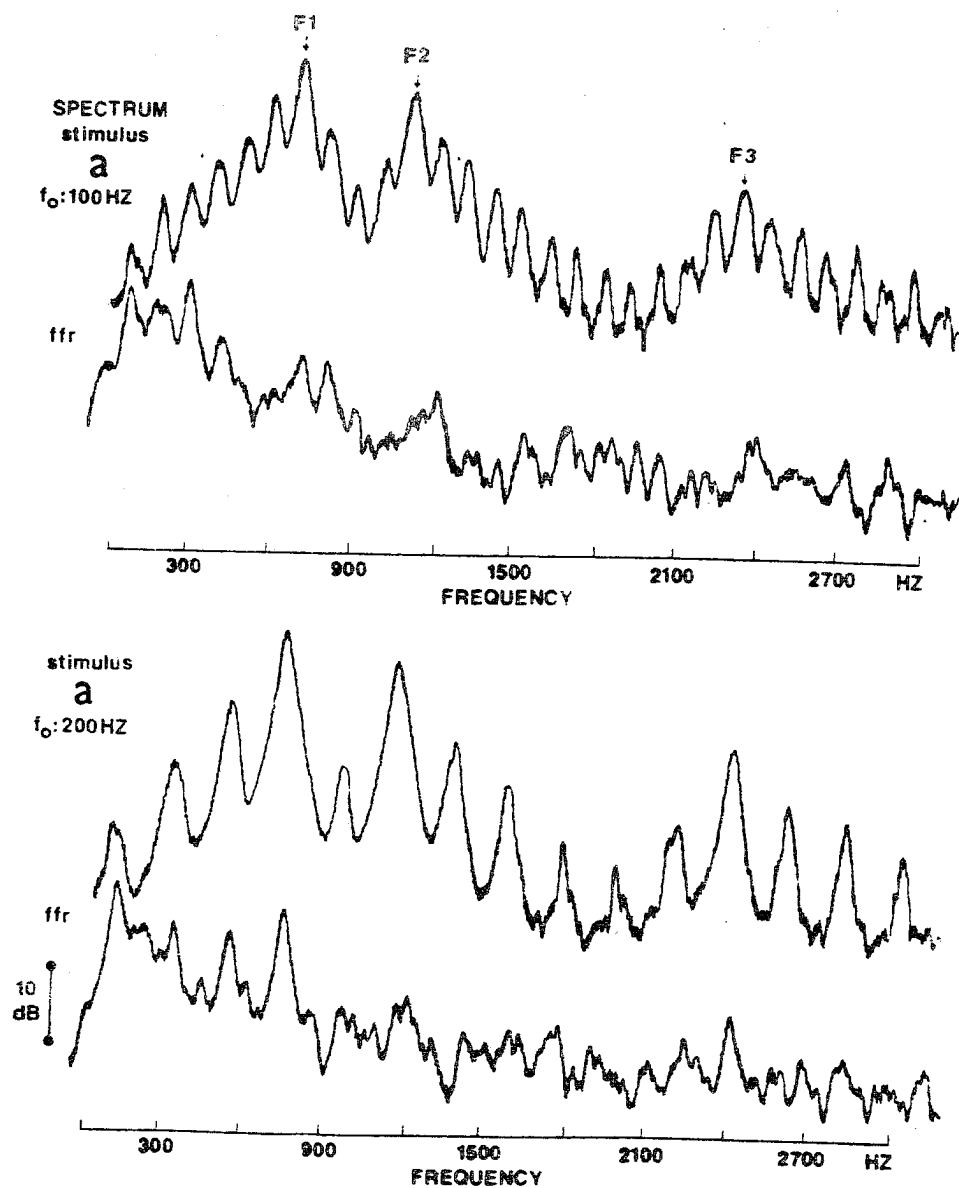


Figure 7.6 Spectral analysis of FFR recorded to the vowel [a] for a single subject.

stimuli are approximately 20-30 dB below the intensity of the harmonics in the region of the first formant. The spectrum level of these components is consequently as low as 30-40 dB SPL (approximately 10-15 dB below FFR threshold). This fact implies that the low harmonics are not the primary source of this low-frequency component of the FFR.

(b) In contrast to Figure 7.1, the response component corresponding to the first formant is rarely the largest component of the FFR. This reflects the fact that the amplitude of the low frequency components in the FFR are of considerable magnitude regardless of the stimulus spectrum. If this aspect of the vocalic response is taken into account, one often observes an abrupt attenuation of spectral energy for frequencies directly above the first formant. This attenuation (or suppression; see Chapter 7.3 (g)) is particularly evident in Figures 7.3 (a), 7.5 (b) and 7.6 (b).

(c) Despite the fact that the spectrum level of the second and third formants are within 5 to 10 dB of the first formant level, the amount of energy in the FFR at these frequencies is much below that of the low frequencies. This pattern reflects, at least in part, the fact that the amplitude of scalp FFR diminishes with increasing frequency above 500 Hz (Chapter 1.72). However, despite the absence of obvious spectral peaks in the region of F2 and F3, there are occasionally regions of apparent suppression on either the low- or high-frequency side of the formant (e.g. F2: Figures 7.3(a), 7.6(a); F3: 7.3(b), 7.4(a), 7.4(b)).

(d) There is always a significant amount of energy in the FFR corresponding to the fundamental frequency of the 200 Hz series of vowels. This reflects the periodic quality of the responses shown in Figure 7.1. Moreover, the amplitude of the 200 Hz component is largest for those vowels ([ae] and [a]) whose first formant lies in the region of dominance for low pitch. Because the FFR was bandpass-filtered between 200 and 3000 Hz, it is not possible to determine the amount of energy in the FFR corresponding to the fundamental of the 100 Hz stimulus series. However, it is likely, judging from the periodic quality of the response waveforms (Figure 7.1) that a large 100 Hz component would be evident in unfiltered versions of the FFR.

(e) One occasionally observes a "spurious" peak in the FFR spectrum. An example of such is illustrated in Figure 7.3 (a), where a local spectral maximum occurs around 900-1000 Hz in the absence of any corresponding energy peak in the stimulus spectrum. The origin of this energy maximum is unclear. However, it is likely to be more than a mere response artifact, for it has been observed in several subjects. Moreover, this peak has been observed by other investigators using other experimental methods. For example, Voight *et al* (1980) have observed an analogous peak in the synchronized component of cochlear nerve fibers responses to the vowel [E]. A similar peak shows up in the contralateral-probe measurements of Sachs and Zurek (1979). It is possible that this peak represents some nonlinear response to widely-separated energy maxima.

(f) When the first and second formants are relatively close, as in the case of the vowel [a], an apparent suppression of the second formant component of the response is evident. This suppression is particularly evident in Figure 7.6 (b), but can also be observed in Figure 7.6 (a). One manifestation of suppression is illustrated in this latter example. Normally, the response to the 1300 Hz harmonic would be suppressed by the second formant. However, in this case, it appears that the suppressive potential of the 1200 Hz component has been itself suppressed by the first formant component. The suppression of the second formant in the FFR is consistent with the observations of Young and Sachs (1979) for the temporal response of single cochlear nerve fibers. It is also consistent with the studies of Carlson et al (1970; 1975), Chistovich and Lublinskaja (1979), and Chistovich et al (1979), both of whom find that for the vowel [a], F1 and F2 can be replaced by a single formant without significant loss of intelligibility.

(g) The correspondence between stimulus spectrum and response spectrum is usually superior for those vowels in which the harmonics are widely-spaced (i.e. $F_0=200$ Hz). Although characteristic of all responses, the pattern is particularly evident for the vowels [I] (Figure 7.3) and [a] (Figure 7.6). The closer correspondence between stimulus and response of the spectra is probably due to different factors. One reflects the fact that there is less low-frequency energy in the responses to the 200 Hz stimulus series (e.g. Figure 7.3). Second, the amplitude of the spectral peak corresponding to the first formant is larger when the harmonics are widely separated. This suggests that spectral components below a spectral maximum may occasionally act to suppress the response to the spectral peak, if they are closely-spaced. This hypothesis is consistent with the results of an experiment in the spectrum of a two-component signal (366+610 Hz) was compared to that of a spectrally similar three-component signal (366+488+610 Hz). The only difference between the two signals was the presence of the middle component (488 Hz). Figure 7.7 shows that the spectrum level of the FFR component corresponding to the upper frequency (610 Hz) is effectively suppressed when the 488 Hz component is added to the signal. The amount of suppression increases even further when the level of the middle component is raised 10 dB above the level of sidebands. This pattern suggests that the ability of a component to suppress the response to a neighboring harmonic depends, at least in part, on the spacing between the two energy bands. Furthermore, the present results imply that components below 600 Hz are more effective in generating FFR than components of higher frequency. Thus, even though there is more energy in the 600 Hz region for the vowel [E], the lower components tend to dominate the FFR spectrum. The suppressive potential of the lower harmonics is particularly striking when harmonic spacing is relatively close.

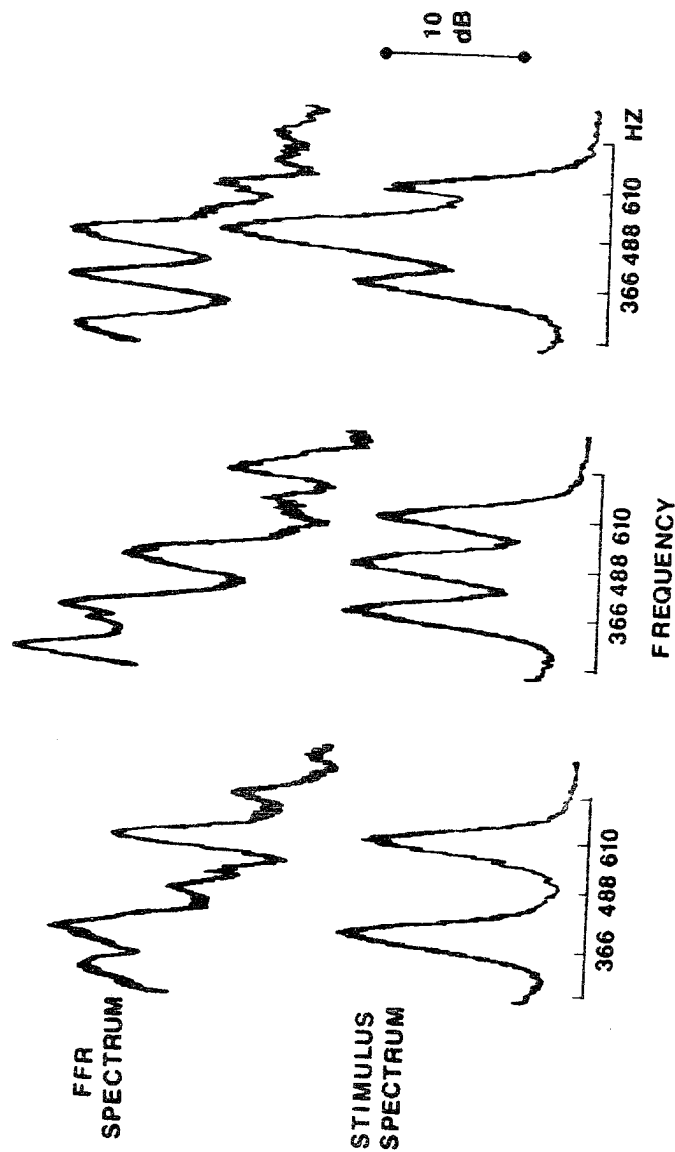


Figure 7.7 A possible FFR correlate of synchrony suppression. FFR recorded from a single subject to two- and three-component stimuli. At bottom are the stimulus spectra. The spectra for the associated FFR are shown above.

7.4 DISCUSSION

The dynamic range of most peripheral auditory fibers (as measured in terms of discharge rate) is limited to approximately 20-40 dB SPL. In contrast, speech is intelligible over a range of 100 dB SPL. This disparity suggests that the spectrum of speech sounds is probably encoded in terms of some neural parameter other than the spatial excitation pattern of auditory neurons. Young and Sachs (1979) suggest that the spectrum of low-frequency (< 5 kHz) speech sounds (such as vowels) is represented in the auditory periphery in terms of the temporal discharge pattern of individual neurons. In the present study, the periodic activity of neurons in the upper auditory brainstem pathway were recorded to determine whether the spectrum of synthetic vowels might be similarly encoded at a higher level of the auditory pathway.

Young and Sachs (1979) suggest that the spectrum of vocalic signals is preserved over a large dynamic range as a consequence of the synchrony suppression operative at high intensities. At low-to-moderate intensities, units synchronize to components lying in the center of their response areas. However, as stimulus level increases above 60 dB SPL, the temporal discharge patterns of these same neurons tend to synchronize instead to the frequency of the most intense harmonic either in the F1 or F2 region. As a consequence, synchrony to harmonics outside the formant regions is suppressed and is replaced by a phase-locked response to the first or second formant frequencies.

In this set of experiments it has been observed that a noticeable dip in the spectrum level of the FFR often occurs on the high-frequency side of the formant region. This phenomenon is particularly evident in Figure 7.6(b) where the energy above the first formant component is significantly attenuated. A similar form of spectral attenuation is evident in Figure 7.5 (b). However, it is not clear from these examples whether the spectral component is actually being suppressed by the presence of the formant or if the apparent attenuation merely reflects the lower spectrum level of the associated stimulus component. The experiment illustrated in Figure 7.7 strongly implies that the spectral attenuation does indeed reflect synchrony suppression. In this experiment, the response to the upper frequency component was markedly affected by the addition of a lower frequency component. Consistent with the findings of Young and Sachs (1979) and others (e.g. Sachs and Kiang, 1968; Abbas and Sachs, 1976; Sachs and Abbas, 1976; Javel *et al.*, 1978), frequencies higher than the suppressor are attenuated considerably more than lower frequencies (when the disparity between suppressor and suppressee is less than 15-20 dB). Thus, it is likely that the spectral dips observed on the high-frequency side of the formant region are the consequence of some form of synchrony suppression.

The present results also suggest that if the perception of vowel quality is based on temporal parameters of neural discharge, it is unlikely that the auditory system estimates the location of the first

formant peak on the basis on some form of interpolation or frequency averaging. The low-frequency region of the FFR spectrum is characterized by peaks corresponding to the component frequencies. There is no evidence of any sort of frequency averaging. Furthermore, it is difficult to see how the formant peak could be inferred from the relative amplitude of the components in the formant region. There is often a marked suppression of synchronized activity on the high frequency side of a formant region. Moreover, the relationship between the amplitude of a frequency component and the magnitude of the corresponding spectral component in the FFR is hardly linear. In addition, if some form of formant interpolation were involved, one would expect that the definition of the response spectrum would be superior for the FFR to the 100 Hz fundamental stimulus series. However, the FFR spectrum is, in fact, more clearly delineated when the fundamental is 200 Hz. This suggests that for the resolved portion of the spectrum, the temporal activity of brainstem units synchronizes mainly to individual frequency components.

For the unresolved part of the spectrum, however, it is certainly conceivable that some form of frequency averaging may occur. No single component dominates the response pattern of any sizeable population of neurons and, as a consequence, units in this region may synchronize to two or more contiguous components (e.g. Evans, 1978: Figure 13). Evidence of frequency averaging is occasionally seen in the region of the FFR spectrum corresponding to the higher formants, where the local peak lies midway between adjacent components of the stimulus spectrum (e.g. F3 in Figure 7.4(a)).

Young and Sachs (1979) studied the population response of the cochlear nerve to different vowels by recording the activity of a large number of single fibers distributed over a broad tonotopic range. A major advantage of this technique is that it enables the response of a tonotopically-restricted portion of the cochlear nerve to be specifically-determined, while at the same time providing a means to study the response of the cochlear nerve as a whole.

In contrast, the scalp-recorded frequency-following response reflects the activity of a large, tonotopically-indeterminate population of neurons. This fact makes it difficult to compare the FFR directly with the discharge patterns of single auditory units. Thus, it is not clear whether the predominance of low-frequency energy in the scalp-recorded response reflects the activity of units tuned to the low-frequency region of the spectrum or rather reflects the activity of high-frequency units. This is a significant issue because it offers a potential explanation of the large disparity between the vocalic spectrum (particularly for [ae] and [a]) and the spectrum of the brainstem response. If this energy is derived from low-frequency units, it would imply that the disparity is a result of a bias towards the low-frequencies in the temporal discharge pattern of upper brainstem units. However, if high-frequency units constitute the primary source of the low-frequency energy, then the disparity between the cochlear nerve representation of vocalic spectra and that provided by the FFR

would be due to the difference in recording technique.

In Chapter 6, it was shown that units tuned to the higher frequencies make a disproportionate contribution to the scalp response, particularly at moderate to high intensities. Moreover, it was shown that the envelope-following component of the scalp response is probably the result of this high-frequency activity. In the present study, the overall sound pressure level was 70-75 dB. At this level, it is likely that a substantial number of high-frequency units are activated by the vocalic stimuli. Young and Sachs (1979), as well as Kiang *et al* (1979) and Hashimoto *et al* (1975), have shown that at this intensity level a substantial proportion of cochlear nerve fibers tuned to the high-frequencies are synchronized to the pitch period of the stimulus. Watanabe and Sakai (1978) have observed a similar pattern for units in the inferior colliculus. Most of the scalp responses illustrated in Figure 7.1 contain a prominent "pitch-synchronous" component equal to the period of the vocalic waveform. It therefore seems likely that this "pitch-synchronous" component of the FFR reflects the activity of high-frequency units synchronized to the envelope of the vocalic waveform.

Young and Sachs (1979) have shown that as the intensity of the vocalic signal increases, the proportion of units synchronized to the formant frequencies (particularly F1) also increases. At levels above 60 dB SPL many high-frequency units, which were synchronized to the pitch period at lower intensities, phase-lock to the frequencies in the region of the first formant. Consequently, if a similar process occurred at the upper brainstem level, one would expect that as the signal level rises above 60 dB, the pitch-synchronous component of the FFR would diminish and be replaced by synchronization to the first formant.

7.5 FREQUENCY-FOLLOWING RESPONSE TO SYNTHETIC VOWELS AS A FUNCTION OF STIMULUS LEVEL

This hypothesis was tested by recording FFR to the vowel [a] over an intensity range between 50 and 70 dB SPL. Experimental methods were similar to those specified in Chapter 7.2. Results for a single subject are shown in Figure 7.8. At the lower and intermediate intensity levels, one observes a large pitch-synchronous component in the FFR. However, when the signal level increases above 60 dB SPL, this component of the scalp response begins to diminish in amplitude and is replaced by a following to frequencies in the first formant region.

In a separate experiment, FFR was recorded to a filtered version of the vowel [E] at two different intensity levels, 55 and 65 dB SPL. The vowel was bandpass-filtered between 150 and 800 Hz (attenuation characteristic = 24 dB/ octave) and presented at an overall intensity level of either 55 (Figure 7.9: left-hand column) or 65 dB SPL (Figure 7.10: left-hand column). At the lower intensity level, the stimulus produces a response that follows both the pitch period and the frequencies in the region of the first formant. Because the response

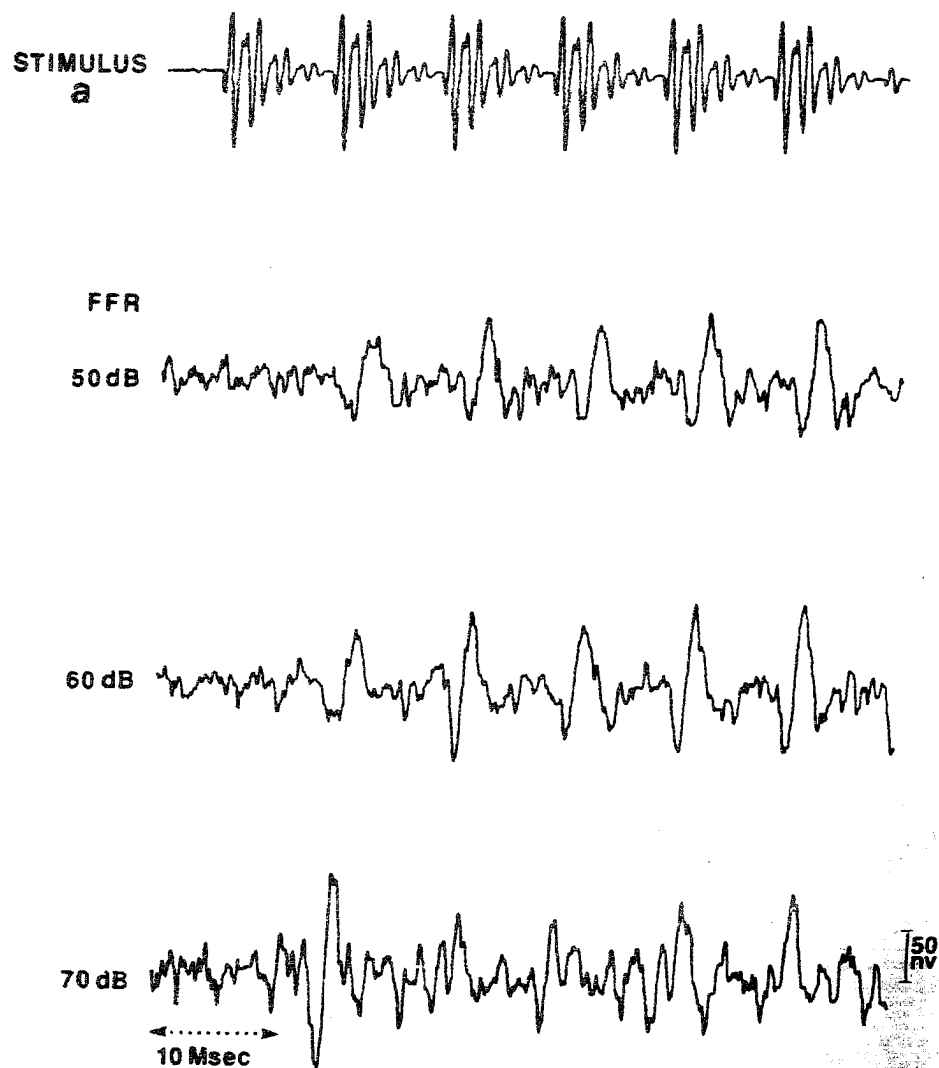


Figure 7.8 FFR recorded for a single subject to the vowel [a] as a function of intensity. Top trace shows the stimulus waveform. Averages consist of 4000 responses.

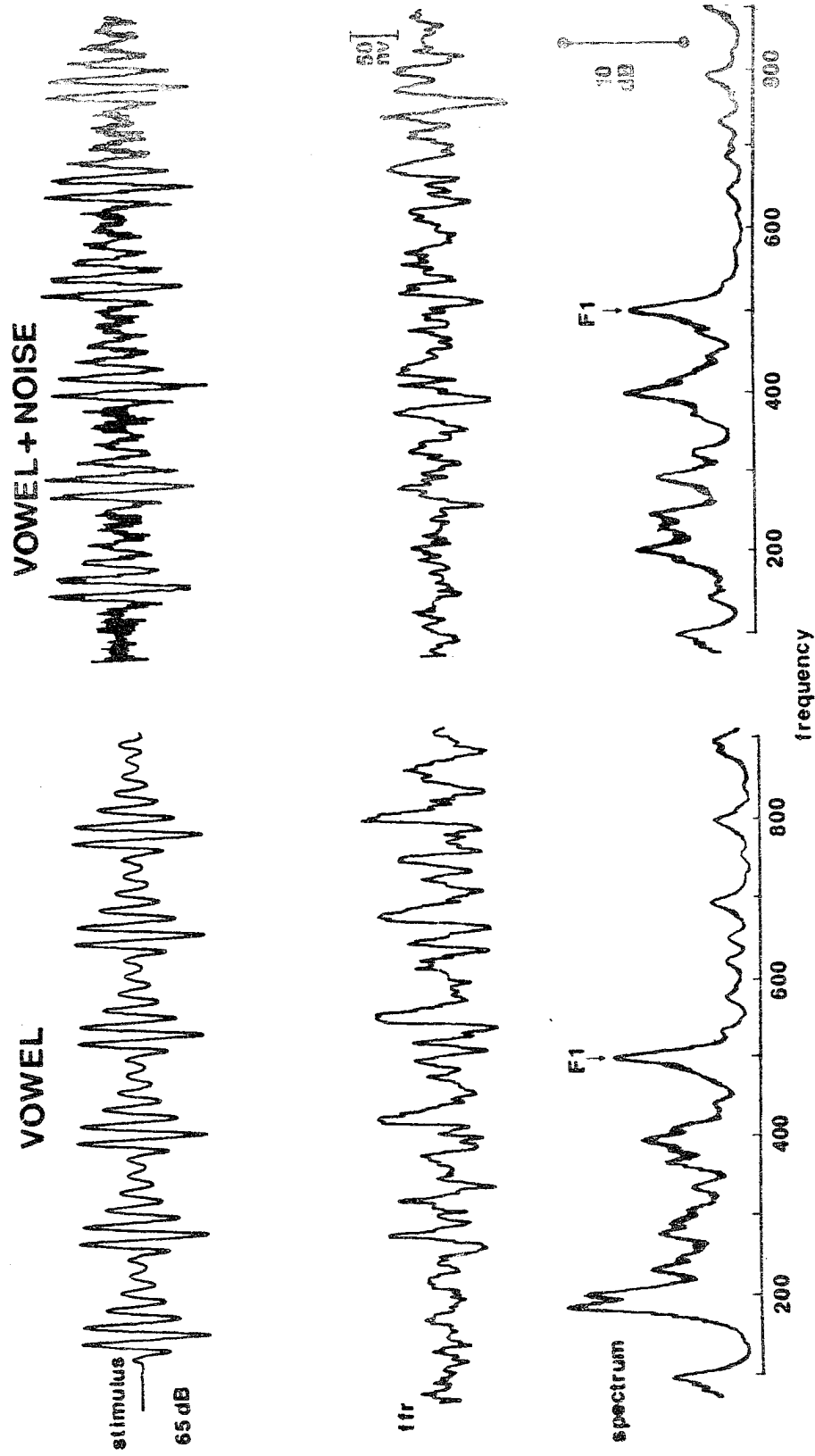


Figure 7.9 FFR recorded for a single subject to the vowel [E] filtered between 150 and 800 Hz, and presented either alone, or in the presence of masking noise (2.5-5.0 kHz). Stimulus level for the filtered vowel = 55 dB SPL. Noise level = 70 dB SPL. Averages consist of 5000 responses.

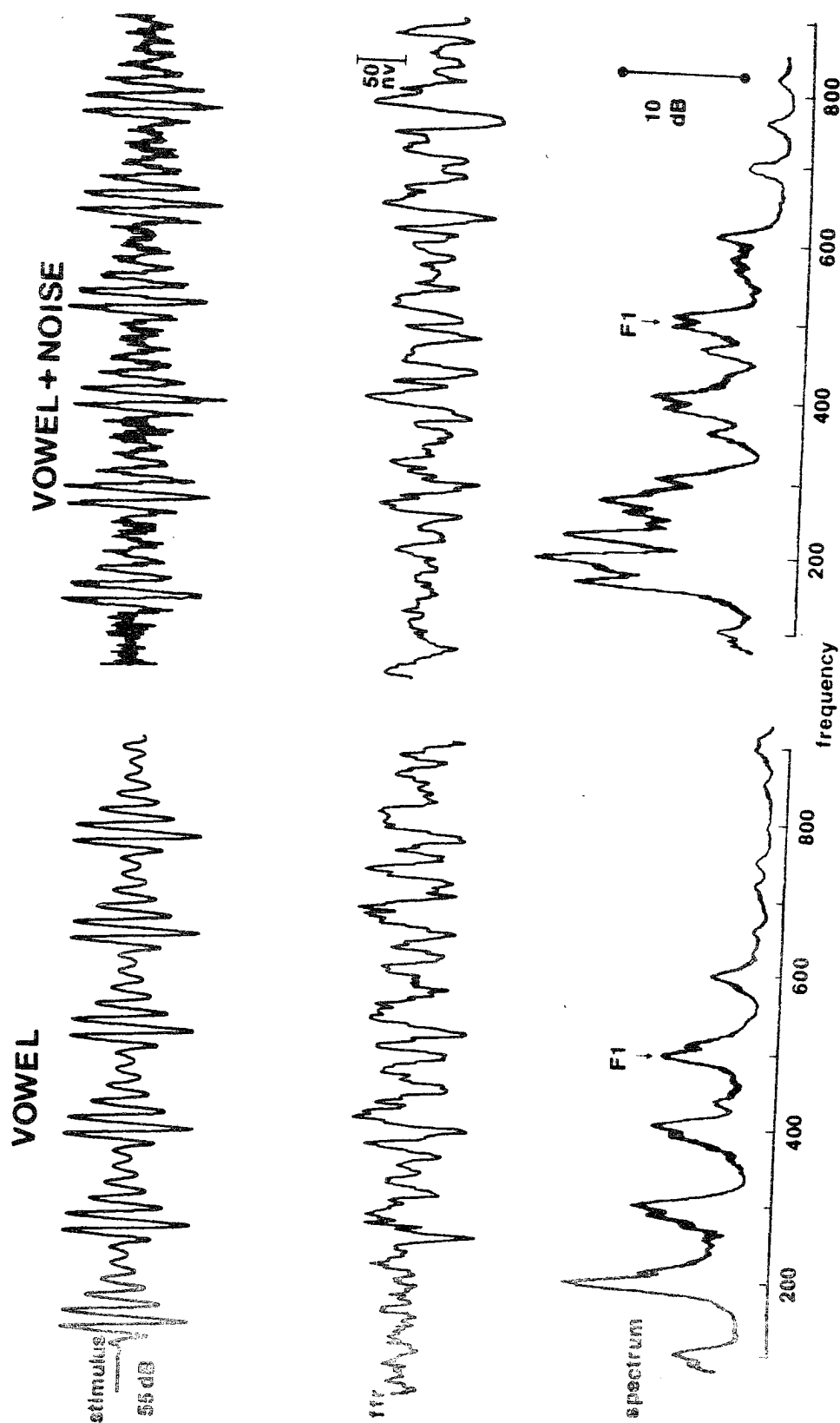


Figure 7.10 FFR recorded to the same stimuli as illustrated in Figure 7.9. Stimulus level = 65 dB SPL. Noise level = 70 dB SPL. FFR waveforms are averages of 5000 responses.

was bandpass-filtered between 200 and 3000 Hz, a relatively small amount of energy is present at 100 Hz. However, the spectral components at 200, 300 and possibly 400 Hz are neural rectifier distortion products of the 100 Hz component (Note 8.2). Thus the magnitude of the pitch synchronous component of the FFR can be indirectly inferred from the magnitude of these lowest harmonics of the fundamental. At an intensity 10 dB above this level (Figure 7.10: left-hand column), one observes an increase in the formant-related component. The reduction of energy in the region immediately below F1 may reflect a slight reduction in the pitch-synchronous component of the response. This pattern is consistent with Young and Sachs' (1979) study of the cochlear nerve response. At intensities above 60 dB SPL they observed that the pitch-synchronous response of high frequency units diminished slightly and was replaced by a synchronization to the frequency of the first formant.

To determine if the energy in the low-frequency portion of the FFR spectrum is due to the recruitment of high-frequency units synchronized to the pitch period, FFR was recorded to the same stimuli in the presence of masking noise bandpassed-filtered (attenuation characteristic=24 dB/octave) between 2.5 and 5 kHz and presented at an overall level of 70 dB SPL. If FFR recorded at 65 dB SPL is largely generated by high frequency units, one should observe an attenuation of the pitch-synchronous component of the response in the presence of high-frequency masking noise. In contrast, the noise should have little effect on the FFR recorded at lower intensities, because the response is generated primarily by units tuned to frequencies below the noise. In other words, the noise should attenuate the 65 dB response more than the 55 dB response. This is in fact the case, as illustrated in Figure 7.10 (right-hand column). The pitch-synchronous component of the higher intensity response is particularly affected by the addition of masking noise. In contrast, the low-frequency portion of the spectrum for the 55 dB response (Figure 7.9; right-hand column) is not attenuated in the presence of the same level of noise although there appears to a slight change in the synchronized pattern.

These demonstrations, although preliminary in nature, suggest that the FFR recorded to vowel sounds largely reflects the activity of units tuned to frequencies above the first formant. These units are largely synchronized to the waveform envelope. Consequently, the FFR spectrum will be heavily weighted towards the low frequencies. Only at the highest intensities (> 70 dB SPL) should the majority of units synchronize to the first formant. Unfortunately, at these levels, the FFR often deteriorates in quality, in part due to distortion of probable cochlear origin. Moreover, the background EEG often increases in amplitude due to the irritating effects of high signal levels. In addition, response attenuation occasionally occurs as a result of activation of the acoustic reflex. Consequently, it is not practical to record FFR to vocalic stimuli at levels sufficiently high to observe the domination of units' synchronized activity by the first formant. FFR recorded at lower intensities reflects multiple sources of activity that make it difficult to distinguish activity phase-locked to the

formants from activity synchronized to the envelope. This is particularly troublesome for vowels such as [I], where the two types of information overlap in frequency. For vowels with relatively high first formants (e.g. [ae] and [a]), the problem is less serious but still tends to obscure the formant-synchronized component of the FFR.

7.6 FREQUENCY-FOLLOWING RESPONSE TO SYNTHETIC VOWELS USING THE RESPONSE-SUBTRACTION PROCEDURE

In Chapter 6.3 it was observed that the envelope-following component of the FFR could be cancelled by the response subtraction procedure (Figure 6.5) and that the resultant waveform probably reflects the activity of units tuned to the component frequencies. This procedure might therefore provide a means to monitor that population of neurons most directly synchronized to the formant frequencies. In Figure 7.11 the spectrum of FFR recorded to the vowel [a] is illustrated for two different stimulus polarities. Both spectra contain a significant amount of low-frequency energy. Although peaks corresponding to F1 and F2 are evident in the FFR spectra, their level is considerably below that of the low frequencies. However, when the response to the inverted polarity stimulus is subtracted from the original response, the low-frequency component of the FFR is largely cancelled, without significantly reducing the amplitude of the first and second formant components. In addition, there is a small set of peaks in the region of 2100 - 2400 Hz that may possibly reflect periodic activity synchronized to F3.

The FFR provides a means to monitor the activity of upper auditory brainstem in response to vocalic stimuli and to compare the spectra so obtained with the results of single unit studies performed with comparable signals. However, the spectrum of FFR recorded through conventional means is biased in favor of frequencies below 500 Hz. As a consequence, the neural representation so obtained may not accurately reflect the activity of the upper brainstem as a whole. In an effort to minimize this low-frequency bias, the FFR was recorded to the same vowel presented in opposite stimulus polarity. This procedure effectively cancels much of the low-frequency energy without attenuating the portion of the spectrum corresponding to the spectral maxima of the stimuli. The result is a spectrum that is dominated by the first (and occasionally second) formant and which is qualitatively similar to results obtained for single cochlear nerve fibers.

To determine if the response-subtraction procedure yields comparable results for a larger set of stimuli, FFR was recorded to four different vowels and the responses subtracted from responses obtained to the same stimuli presented in the opposite stimulus polarity.

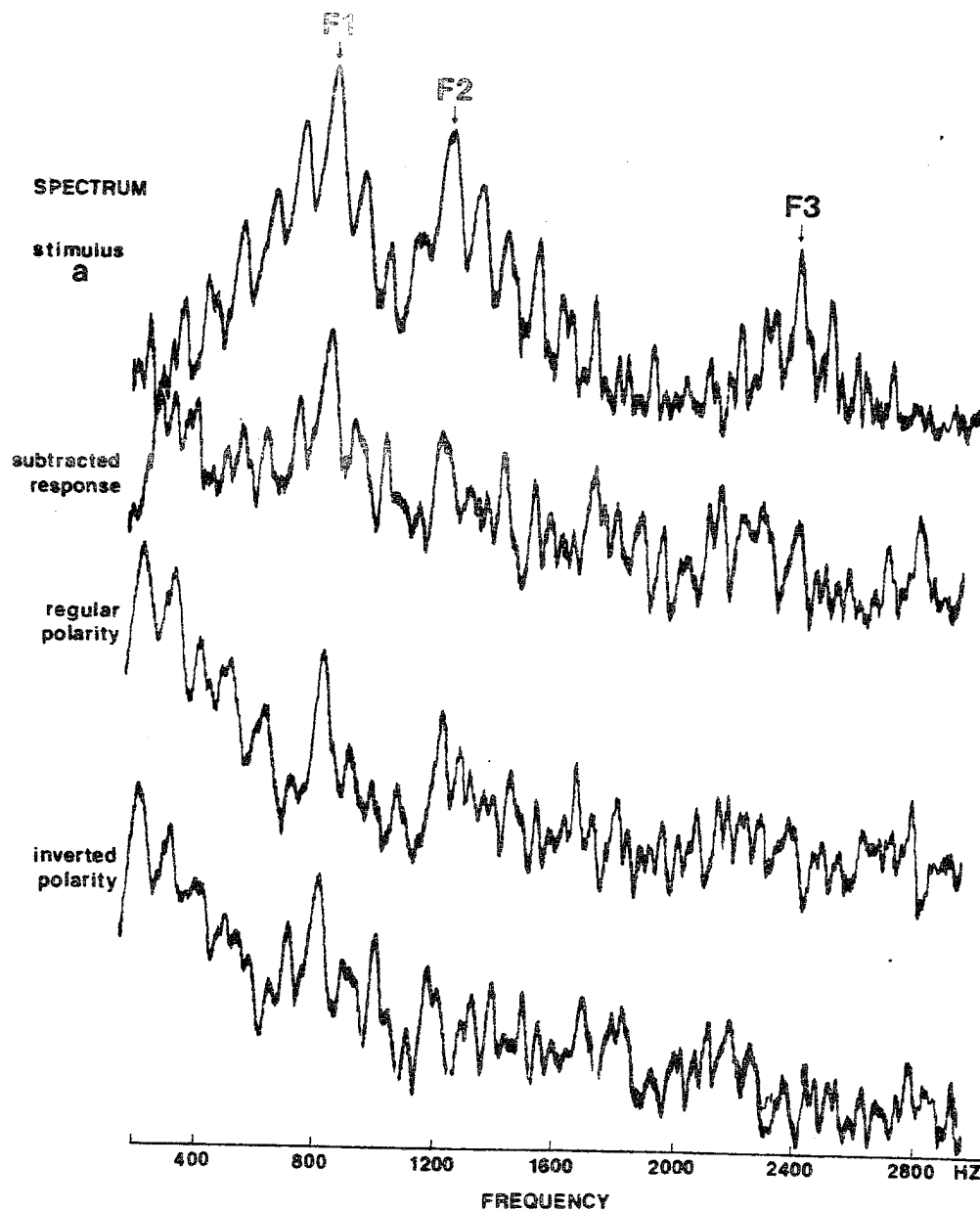


Figure 7.11 Spectral analysis of FFR and subtracted FFR recorded to the vowel [a]. The top trace is the stimulus spectrum. The bottom traces are the spectra of FFR recorded to the same stimulus presented in opposite stimulus polarity. The remaining trace is the spectrum of the subtracted response.

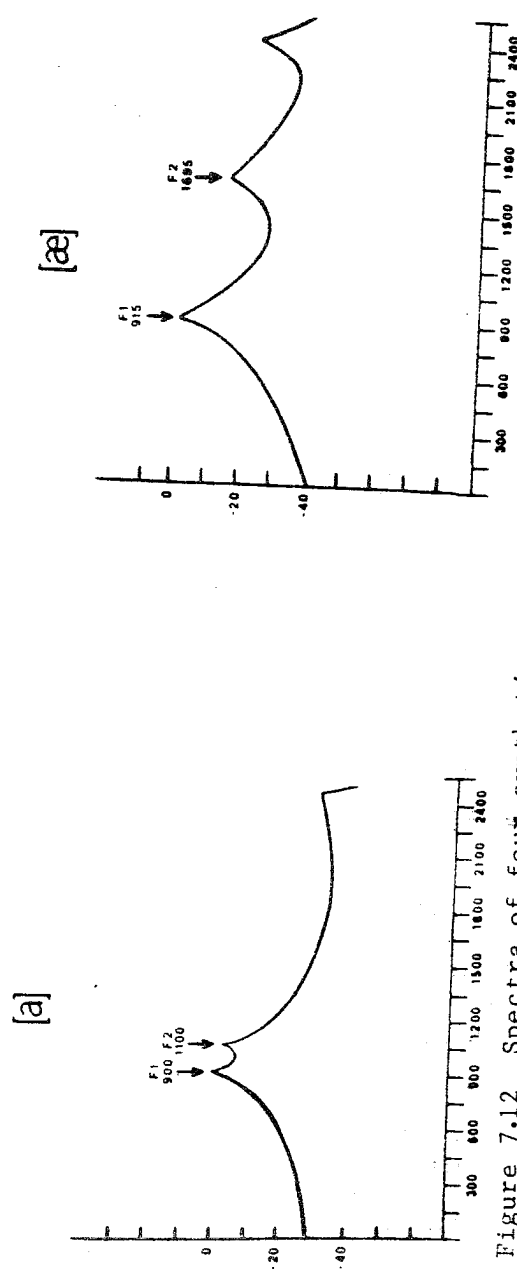
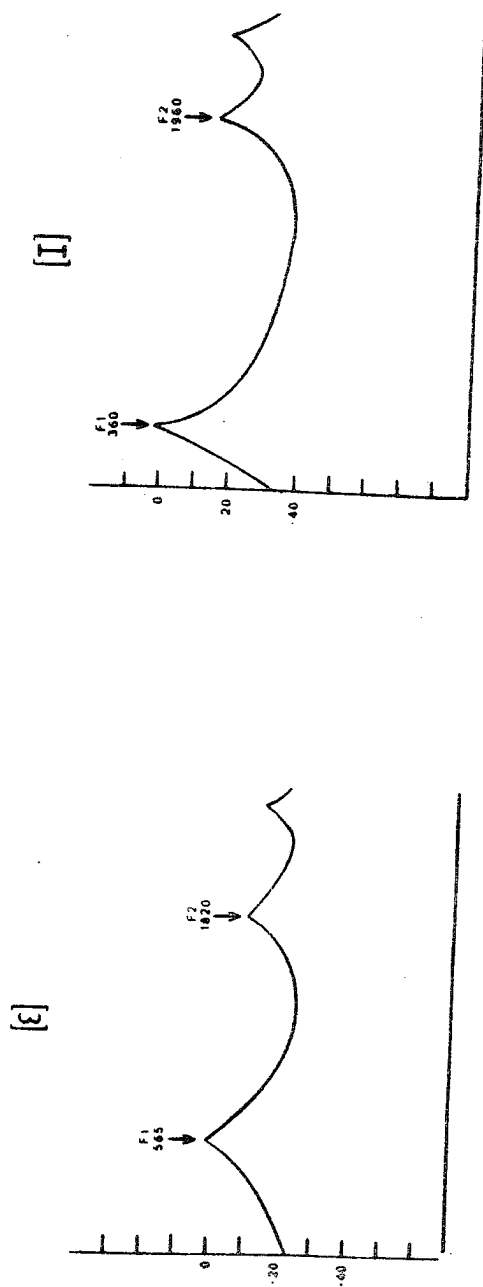


Figure 7.12 Spectra of four synthetic vowels used in the response-subtraction experiment. The location of the first and second formants are indicated by arrows.

7.61 EXPERIMENTAL METHODS

Stimuli were synthetic vowels corresponding to the vowels [I], [E], [ae], and [a]. Spectral analyses of the microphone responses to the four stimuli are shown in Figure 7.12 F3, F4 and F5 were set to 2440, 3400, and 4200 Hz for all vowels. In contrast to stimuli used in the previous experiments, these vowels were synthesized on a digital serial formant synthesizer in which the level of the individual formants was computed according to a model of the vocal tract (Gold and Rabiner, 1968). Formant bandwidth was 50 Hz for F1, 100 Hz for F2, 150 Hz for F3, and 200 Hz for F4 and F5. Fundamental frequency was 150 Hz and stimulus duration was approximately 53 msec. Stimuli were recorded on audio tape and redigitized by the Interdata computer for on-line stimulus presentation. Stimulus intensity was 75 dB SPL. Stimuli were presented monaurally at a rate of 9 per second. Stimulus polarity was inverted midway through each average and the response to the inverted polarity signal subtracted from the average obtained using the initial uninverted stimulus polarity. The "subtracted" averages consisted of between 4000 and 6000 responses.

The electrode configuration placed the active electrode at vertex and the reference electrode on the earlobe contralateral to the ear of stimulation. The opposite earlobe served as ground. Responses were filtered between 200 and 3000 Hz. Offline digital Fourier analysis was performed on the subtracted FFR. Each spectral band was approximately 30 Hz wide. Responses were obtained for three subjects (2 males, 1 female, between the ages of 8 and 34). All other experimental parameters were as specified in Chapter 2.

7.62 EXPERIMENTAL RESULTS

Results for a single subject are illustrated in Figure 7.13. In contrast to the FFR waveforms shown in Figure 7.1, these subtracted responses do not display a large waveform envelope-following component. Rather, they appear to contain a substantial amount of higher-frequency energy. Thus, even though the stimuli were presented at a fairly high intensity level, the response-subtraction procedure appears to have cancelled much of the pitch-synchronous component of the original response.

A comparison of the stimulus and response spectra for each vowel indicates that the FFR contains a significant amount of energy in the region of the first formant. This is the case for each vowel including [ae] and [a], where the formant component is 900 Hz. Moreover, the spectral peak in the FFR corresponding to the first formant dominates the spectrum in every instance. Occasionally, the largest peak in the FFR spectrum corresponds to the second highest peak in the formant region, as shown for the vowel [I]. This may reflect the fact that the more intense component of the formant is in a less sensitive auditory region. In addition, there is some energy in the low frequencies (e.g. [E] and [ae] which may be indicative of incomplete cancellation of the

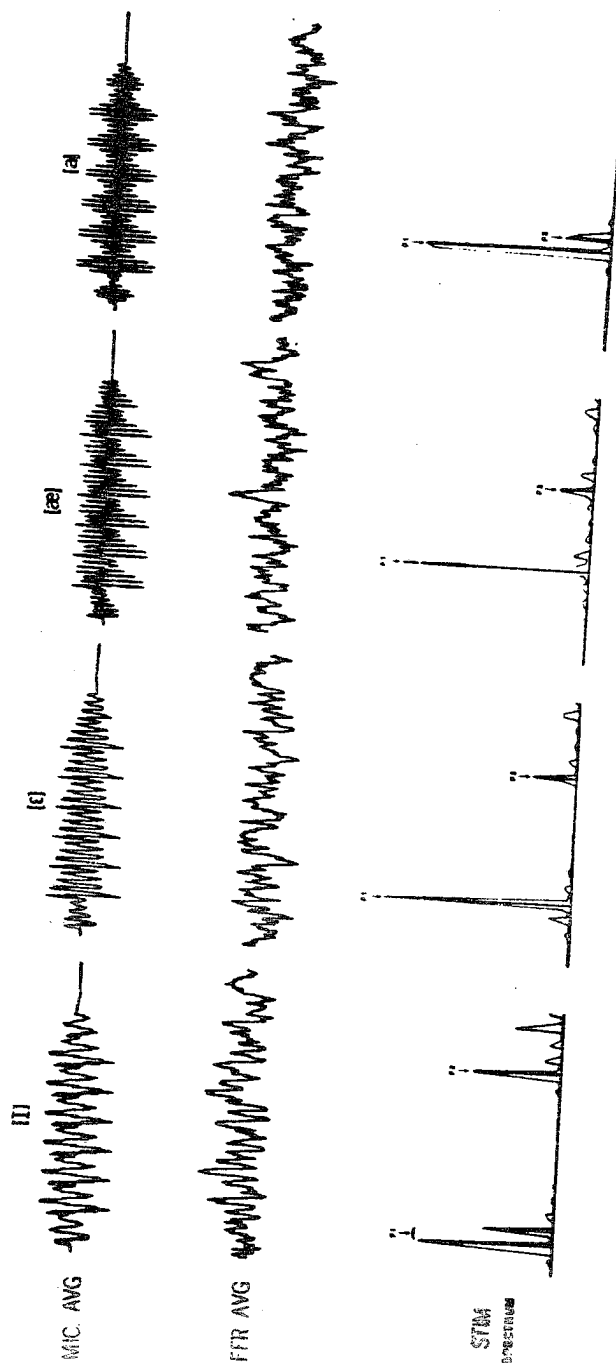


Figure 7.13 Subtracted FFR recorded from a single subject in response to four synthetic vowels. The top row illustrates the microphone average of the stimulus waveform. The next row shows the composite, subtracted FFR recorded to these stimuli. The stimulus spectra are shown in a linear plot on the row below. The location of the first and second formants are indicated by arrows. The bottom row shows the linear spectrum analysis of the subtracted FFR. Stimulus intensity was 75 dB SPL. Averages consisted of 4000 responses.

pitch-synchronous response. Consistent with FFR averages shown in Figures 7.3-7.6, there is relatively little energy in the FFR spectrum corresponding to the second and third formants.

The results for all three subjects are illustrated in Figure 7.14. Although, there is some residual energy in the region below F1, the spectrum of the FFR is dominated by energy in the region of the first formant for most of the responses. In contrast to the spectra of responses obtained in the conventional manner, the present set of FFR emphasizes the portion of the FFR spectrum associated with the first formant. The results are particularly striking for the vowels [ae] and [a]. However, there are certain exceptions to these generalizations (responses to the vowel [E] for REN and KAT; responses to [ae] and [a] for subject KAT) and these tend to be of two varieties. One involves incomplete cancellation of the low-frequency energy associated with the pitch-synchronous response (the responses to [ae] for all three subjects, and the response to [a] for subject KAT), which is probably due to a certain degree of asymmetry in the responses obtained to each stimulus polarity. The other type of exception involves responses to the vowel [E]. The spectrum peak of this vowel is 600 Hz. Davis (1976) has observed that the FFR to a 600 Hz signal is considerably smaller than that obtained to a 500 Hz signal. We have observed a similar attenuation of responses to signals in this frequency range in most (but not all subjects, e.g. the response to [E] for WSB) subjects. It is not entirely clear what mechanism is responsible for this response attenuation (see Chapter 7.3(7) for one possible explanation).

7.63 DISCUSSION

The FFR provides a means by which to study auditory brainstem mechanisms involved in the processing of vowel quality. When recorded in the conventional manner, the scalp response reflects the periodic activity of (probably high frequency) units synchronized to the envelope of the stimulus waveform. This "pitch-synchronous" component of the response is reflected in the predominance of low-frequency energy in the FFR spectrum. The spectrum of FFR recorded in this manner provides the misleading impression that the temporal activity of brainstem units is heavily biased towards frequencies below 500 Hz, independent of the spectrum level of individual stimulus harmonics. When the response subtraction procedure is applied to scalp FFR a very different picture emerges. The location of the first formant is often discernable in the FFR spectrum. It corresponds to the largest energy peak in the spectrum. In those few instances where this is not the case, the first formant is usually discernable as a substantial peak with a considerable degree of spectral suppression on the high-frequency side. These exceptions could probably be eliminated by insuring that the responses recorded to opposite polarities of the stimulus are roughly symmetrical and of approximately equal amplitude.

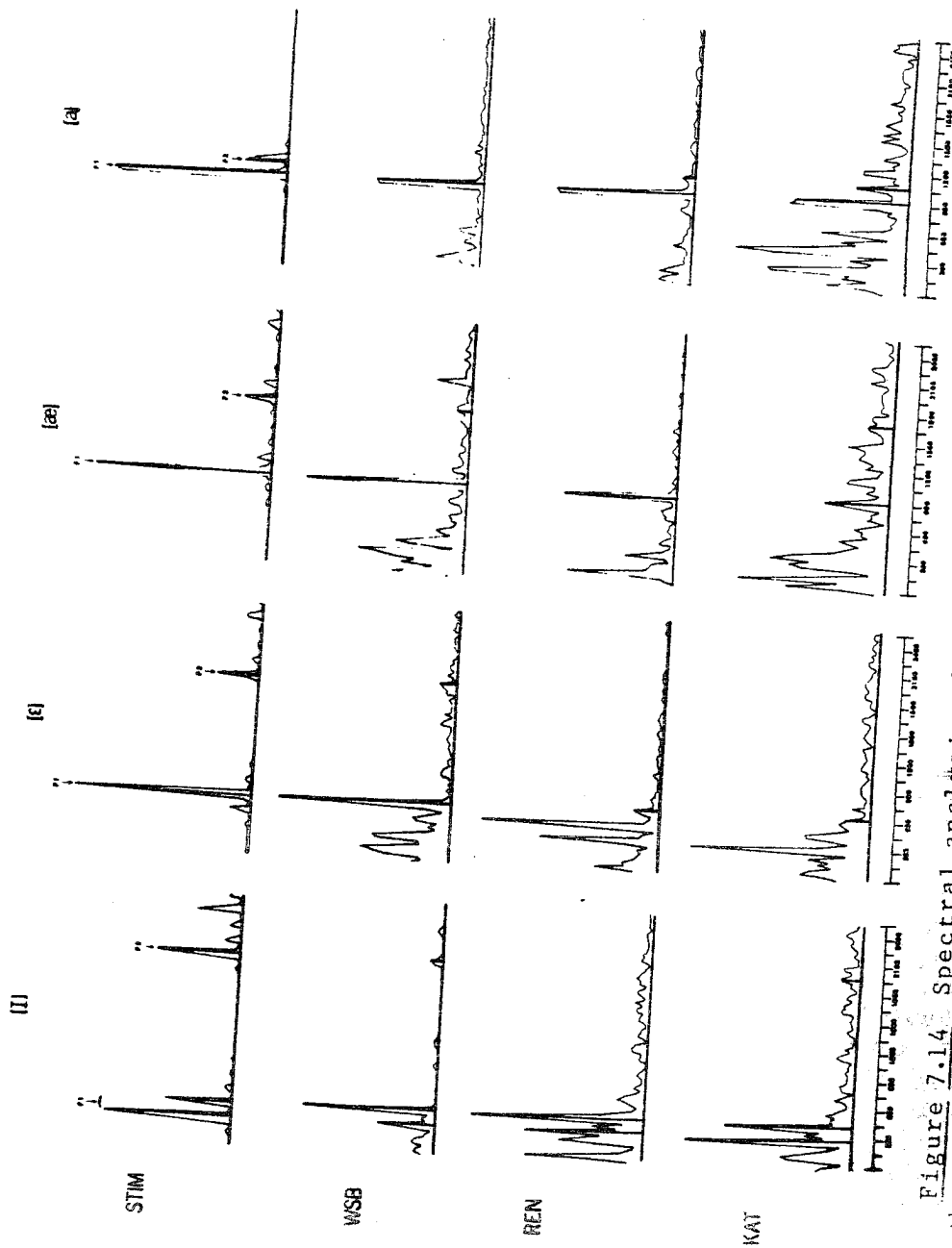


Figure 7.14 Spectral analysis of composite, subtracted FFR recorded from three subjects. The top row contains the spectra of the stimuli. Stimulus level = 75 dB SPL. Averages consist of 4000-6000 responses.

Thus, the results of the present experiment are generally consistent with the conclusions reached by Young and Sachs (1979). The first formant appears to be represented in terms of the temporal discharge patterns of auditory neurons. Response synchrony to components outside the formant region appear to be suppressed or at least "drowned out" by the predominant synchronization to frequencies in the formant region. The major discrepancy between the results of the present study and those obtained by Young and Sachs for single cochlear nerve fibers involves the representation of the second and third formants. In the present study, the spectrum level of FFR in the F2 and F3 regions was considerably lower than F1. In fact, FFR to these higher formants was hardly distinguishable, except for certain responses to the vowel [a]. This exception may provide a clue as to the basis of the disparity. The upper frequency limit of phase-locking in the cochlear nerve is approximately 5 kHz (Rose et al, 1967). In contrast, the upper limit in the inferior colliculus is roughly 2-3 kHz (Rose et al, 1966). Moreover, relatively few neurons in the IC actually phase-lock to frequencies above 1-1.5 kHz. Consistent with this is the fact that the amplitude of FFR recorded from the IC of the cat diminishes sharply for frequencies above 1.5 kHz (Marsh et al, 1974). Thus, the attenuation of the FFR spectrum observed in the F2 and F3 regions may very well reflect a constraint on the phase-locking capabilities of upper brainstem units.

The attenuation of F2 and F3 is also consistent with the contralateral-probe study of Sachs and Zurek (1979). They found that the internal level of individual components above 1-1.2 kHz is considerably below the external level of the spectrum. This is in marked contrast to the pattern observed for frequencies below and including the first formant. The contralateral probe technique probably reflects the synchronized activity of neurons in the central auditory pathway (Chapter 1.62). Consequently, it is of interest that the estimate of the internal spectrum obtained with the response-subtraction procedure is qualitatively similar to that derived from the contralateral-probe technique.

The relatively small amount of energy observed in the F2 and F3 regions of the spectrum may also reflect the fact that even at the level of the cochlear nerve, the number of fibers synchronized to these higher formants is relatively few. However, this population of fibers continues to synchronize to F2 and F3 when most other fibers have become synchronized to F1. This suggests that F2 and F3 may be represented in terms of the temporal discharge pattern of a relatively small, but highly synchronized group of neurons. If this were the case, one would not expect this population to make a major contribution to the scalp response. Even so, one occasionally observes in the FFR spectrum a small, but distinct peak (or set of peaks) in the high formant region (e.g. Figures 7.4(a), 7.6(b)).

CHAPTER VIII

SUMMARY AND CONCLUSIONS

8.1 INTRODUCTION

The auditory analysis of complex sounds plays an important role in speech communication and musical appreciation. Yet, despite over a century of intensive investigation, the auditory mechanisms underlying the perception of complex signals, such as speech, remain obscure. For many years, it was assumed that the frequency analysis of such signals is based primarily on the spatial distribution of neural excitation. However, recent studies suggest that the representation of complex spectra is unlikely to be based on "place" mechanisms of frequency analysis (Chapters 1.42 and 1.62).

An alternative means of frequency coding, based on the temporal discharge pattern of auditory neurons, was shown in Chapter 1.63 to be far more compatible with psychophysical and physiological studies of complex signal analysis. The studies reviewed suggest that the perception of pitch and timbre are largely based on temporal properties of auditory activity.

To test this hypothesis experimentally, the periodic activity of brainstem units was recorded from the scalp of human subjects during presentation of complex signals relevant to the perception of pitch and vowel quality. In previous studies (Worden and Marsh, 1968; Marsh and Worden, 1968; Marsh *et al.*, 1974), it has been shown that the period of this brainstem potential, the frequency-following response, matches the period of the stimulating waveform and that the response reflects the activity of units in the upper auditory brainstem nuclei which are synchronized to the stimulus frequency.

The present study investigated whether the spectrum and low pitch of complex signals also are reflected in this brainstem potential. The study was divided into two parts. The first part (Chapters 3-6) was concerned with determining the extent to which the behavior of the scalp-recorded FFR corresponds to the perception of pitch under a variety of stimulus conditions known to influence the magnitude and salience of the pitch sensation. The second part (Chapter 7) examined the degree to which scalp-recorded FFR contains spectral information concerning the formants of vocalic signals. In the present chapter, the major results of each part of the overall study are summarized and discussed.

8.2 FREQUENCY-FOLLOWING RESPONSE CORRELATES OF LOW PITCH

In Chapter 1, the major features of low pitch were enumerated (Chapter 1.29) and shown, in principle, to be compatible with a model based on the periodic discharge behavior of auditory neurons (Chapter 1.5). This hypothesis was then put to the experimental test by determining the degree to which these properties of low pitch are mirrored in the amplitude and spectral behavior of the scalp-recorded frequency-following response. In this section, the experimental results are considered for each feature in turn.

(a) The Pitch of Complex Tones does not Depend on the Fundamental Component

The classical "place" theory of pitch (Helmholtz, 1863) hypothesized that the pitch of a complex tone is based on the lowest frequency component of the signal. This hypothesis predicts that when the fundamental is absent or significantly attenuated, the pitch of the complex should jump an octave. Seebeck (1841) showed that this is not the case. The pitch of a complex signal remains basically unchanged when the amplitude of the fundamental is reduced to negligible proportions.

In previous studies of the frequency-following response, it was observed that FFR to single frequency tones produces a periodic scalp response equal in frequency to the stimulus. The question explored in Chapter 3 was whether FFR recorded in response to complex tones lacking the fundamental component, would reflect primarily the stimulus components or the "missing" fundamental. It was shown that the FFR to such stimuli is dominated by spectral energy at the frequency of the missing fundamental and contains a relatively small amount of energy at frequencies corresponding to the stimulus components.

(b) The "Missing" Fundamental is not Attributable to Combination Tones Produced by Intracochlear Distortion

Seebeck's (1841) original demonstration of the "missing" fundamental was ascribed by Helmholtz (1863) and others to the presence of some form of aural distortion. This distortion, it was hypothesized, would serve to re-introduce the fundamental in either the middle ear or the cochlea. Schouten (1938) showed, however, that the sensation of low pitch is not based on a combination tone when the signal is presented at low intensities.

Because the FFR is recorded at signal levels considerably higher than those used in Schouten's demonstration, it was necessary to rule out combination tones as the source of FFR recorded in response to the complex tones. The combination tone hypothesis was shown to be untenable in light of the fact that FFR to the tonal complex is not appreciably attenuated in the presence of masking noise centered at the fundamental frequency. At the same time, it was shown that FFR to a pure tone equal in frequency to the fundamental was significantly reduced in amplitude by the same noise.

(c) Spectral Limits of Low Pitch

If the pitch of complex tones does not stem from the fundamental, from where does it come? In the 1930's Schouten proposed that a tone's low pitch (the "residue") is derived from neural synchronization to the cochlear waveform produced through the interaction of unresolved, closely-spaced harmonics. His hypothesis was shown to be untenable in light of the fact that low pitch stems exclusively from aurally-resolved components (Smoorenburg, 1970; Houtsma and Goldstein, 1972; Bilsen, 1973; Buunen *et al*, 1974).

In Chapter 4, FFR was recorded to a series of two-component signals of variable fundamental frequency and harmonic rank to determine the spectral limits of signals capable of generating FFR. It was found that FFR corresponding to the fundamental is generated only by the lowest six to eight harmonics.

(d) The Dominance of Low Harmonics

The strongest sensation of pitch is generated by low harmonics (particularly ranks three and four) lying in the region between 500 and 1500 Hz (Ritsma, 1967; Plomp, 1967; Houtsma and Goldstein, 1972; Bilsen, 1973).

In Chapter 4 it was shown that the largest brainstem responses are produced by stimuli whose components lie in the region between 500 and 1000 Hz, irrespective of the fundamental frequency.

(e) Salience of Low Pitch: Fundamental Frequency

The most salient sensation of low pitch is produced by signals whose fundamentals lie between 250 and 500 Hz (Bilsen and Ritsma, 1970; Plomp, 1975; Yost and Hill, 1978). The strength of low pitch declines sharply for fundamentals above 500 Hz and below 100 Hz.

In the present study, the largest brainstem responses were consistently generated by signals with a fundamental of 244 Hz or 366 Hz. The amplitude of the FFR falls off very sharply for frequencies above 488 Hz.

(f) Salience of Low Pitch: Number of Components

The salience of pitch produced by just two frequency components situated in the "region of dominance" is of greater magnitude than the pitch generated by a far larger number of components located outside this favored region (Ritsma, 1967; Bilsen, 1973).

In chapter 4 it was shown that the amplitude of FFR generated by the most potent two-component signal was equal to or larger than that produced by stimuli composed of four harmonics.

(g) Pitch Salience as a Function of Modulation Depth

Ritsma (1962) observed that the modulation depth of the signal influences the salience of low pitch. A modulation depth of approximately 115% (i.e. the level of f_2 is 5 dB greater than f_1 and f_3) produces the optimum pitch salience.

In Chapter 5, it was shown that a modulation depth of 100% produced a brainstem response effectively insensitive to the stimulus phase-configuration regardless of signal level (Figure 5.8). Moreover, the amplitude of FFR generated by a quadrature phase signal is increased when the modulation depth is lowered from 200% (all components equal amplitude) to 100% (f_2 is 6 dB above the level of f_1 and f_3).

(h) Phase-insensitivity of Low Pitch

The pitch of complex tones is insensitive to the phase configuration of the stimulus components (Bilsen, 1973; Wightman, 1973a) except for the influence of combination tones generated by closely-spaced harmonics (Smootenburg, 1970; Buunen et al., 1974).

In Chapter 5, FFR was recorded to three-component tones of variable phase-configuration and waveform modulation depth. At low-to-moderate intensities (5 dB above FFR threshold), no systematic relationship between FFR amplitude and stimulus waveform modulation depth was observed for components of low harmonic rank (3+4+5). However, FFR recorded to a set of high harmonics, at the very margin of the "existence" region (7+8+9) for low pitch show a rather different pattern. For these stimuli, FFR amplitude is, indeed sensitive to the modulation pattern of the waveform envelope.

When the intensity level of these stimuli is increased to 10 dB above FFR threshold or higher, a different pattern emerges. At these levels, the amplitude of the FFR is influenced by the modulation depth of the stimulus waveform, even for low harmonics separated by more than a critical band. Two possible interpretations of this change in pattern were offered. One possibility is that at higher intensities the scalp response is generated mostly by units whose characteristic frequencies lie considerably above that of the stimulus components. This hypothesis implies that harmonics which would be resolved by units tuned to lower frequencies, remain unresolved by these more basal units. A second possibility, not entirely incompatible with the first, is that the changing pattern of the FFR reflects the deterioration of frequency resolution at high intensities. The second alternative is presently favored in view of the fact that raising the level of the middle stimulus component effectively renders the FFR insensitive to the stimulus phase-configuration at high intensities. This result implies that the sensitivity of the FFR to the waveform modulation pattern is a product of neural interaction among units tuned to the portion of the spectrum encompassing the stimulus components.

(i) The Pitch of a Complex is not Based on the Waveform Modulation Frequency

Inharmonic signals have played a key role in the development of contemporary pitch theory. The pitch produced by such stimuli deviates from the waveform envelope repetition frequency by an amount that is roughly proportional to the harmonic rank of the stimulus components. Two different, albeit mathematically equivalent, mechanisms have been proposed to explain the pitch generated by inharmonic

signals. One proposal, related to Schouten's original model of the residue, suggests that the pitches reflect the synchronized activity of auditory neurons to the temporal fine-structure of the cochlear waveform. This model was considered by de Boer (1956a) and others to apply only to the high, unresolved components since the low, resolved harmonics would not be expected to interact sufficiently in the cochlea to produce a waveform capable of generating the necessary fine-structure. This implied that the pitch of low harmonics must be based on some form of harmonic template-matching mechanism.

A number of physiological studies suggest that this assumption is inaccurate. Any single pair of contiguous components separated by an octave or less will interact in some region of the cochlea. This interaction produces a waveform from which the fine-structure can be potentially derived (Rose *et al*, 1969; Evans, 1978; Javel, 1980). Thus a temporal mechanism could, in principle, underlie the pitch generated by inharmonic signals. This hypothesis was tested by recording FFR to inharmonic signals of low harmonic rank and comparing the responses to harmonic signals with similar spectra. It was observed that the FFR to these inharmonic signals differed from FFR to their harmonic counterparts in a number of subtle ways. The interpeak intervals of the responses to the inharmonic signals were shown to differ from those generated by harmonic signals and to correspond to the pitches generated by these signals. However, spectral analyses of the entire waveform usually indicate a large component at the envelope repetition frequency. The discrepancy between these two methods of analysis is due to an asymmetry in the waveform of the inharmonic response. The spectrum of the positive polarity contains a large component at the difference frequency, while the spectrum of the negative polarity contains considerable energy at frequencies corresponding to the pitches associated with the signal. Thus, it appears that at moderate-to-high intensities, FFR to a complex signal may consist of two components, one reflecting synchronized activity to the waveform envelope, the other reflecting activity relevant to the processing of pitch.

In a separate experiment, the envelope-following component of the response was cancelled by subtracting the response recorded to the signal presented in one stimulus polarity from the response obtained to the same stimulus presented in the opposite polarity. The spectrum of the subtracted response contains energy at both the stimulus frequencies and at frequencies corresponding to the pitch shift of inharmonic signals. This spectral pattern suggests that the response-subtraction procedure cancels the contribution of high-frequency units while retaining information derived from neurons synchronized to the components themselves. The spectral analysis of the subtracted response indicates a significant amount of energy in the frequency range corresponding to the pitch produced by the signal.

(j) The Low Pitch of Complex Tones is Ambiguous

Complex tones are capable of generating more than a single low pitch. Although this pitch ambiguity is most apparent for inharmonic signals, even harmonic complexes may generate multiple pitches, particularly when the stimulus components are of relatively high harmonic rank (Schouten et al., 1962).

In Chapter 6, it was observed that the spectrum of a response generated by an inharmonic signal may, on occasion, have two separate peaks in the pitch-relevant region (Figure 6.3). Moreover, in informal observations, the spectrum of FFR recorded to harmonic stimuli of high harmonic rank (ranks 7-9) was found to contain multiple peaks in the region of the "missing" fundamental.

(k) Intensity-independence of Dominance Principle

Frequency components lying inside the region of dominance may determine the pitch of a broadband signal despite the fact that its level is 40 dB or more below the level of the other components. The level of these components need be only 10 dB SL in order to exert their dominance over the remainder of the spectrum.

The threshold of scalp recorded FFR is 45-60 dB SPL/component, some 40 dB or more above behavioral threshold of low pitch. This fact precludes a direct determination of the intensity-independence of the dominance effect for FFR in human subjects.

(l) Low Pitch in the Absence of Peripheral Interaction

A sensation of low pitch may be heard when only a single frequency is presented to each ear (Houtsma and Goldstein, 1972; Bilsen, 1973). This pitch sensation is faint and requires extensive listening experience to hear reliably. The existence of a dichotically-mediated pitch suggests a central origin for at least some aspects of low pitch. However, this finding does not necessarily mean that central mechanisms play a determining role, to the exclusion of more peripheral mechanisms. It is likely that low pitch is based on both peripheral and central mechanisms. Under typical listening conditions, the pitch percept reflects the operation of both mechanisms. Under dichotic stimulation, the percept is based on neural activity confined to central nuclei. This would account for why the salience of dichotic pitch is so ephemeral compared to the pitch produced by monotic stimulation.

We have attempted to record FFR to three-component dichotic signals, in which the signal to one ear consisted of the third and fifth (or fourth and sixth) harmonics of a 244 Hz fundamental, while the signal presented to the other ear consisted of the fourth (or fifth) harmonic of the same fundamental. Due to limitations in our signal generation equipment, it was not possible to generate two independent, phase-locked signals. The results obtained under these conditions are somewhat inconclusive (Note 8.1).

In summary then, we have shown that many of the major features of low pitch are reflected in either the spectrum or amplitude of the frequency-following response recorded from the scalp of human subjects. This pattern of correspondence implies that many aspects of pitch perception are based on the temporal activity of neurons in the auditory periphery and brainstem nuclei. For, by the level of the upper brainstem region, the general pattern of temporal neural activity is of a form that bears, in many respects, a striking correspondence to many of the parameters involved in the perception of low pitch (Note 8.2).

8.3 THEORETICAL IMPLICATIONS OF THE PRESENT STUDY FOR MODELS OF PITCH PERCEPTION

The neural mechanisms underlying the perception of pitch are, as yet, poorly understood. For this reason, contemporary models of pitch perception are based primarily on psychoacoustic studies. Although much can be learned from perceptual experimentation, the results of such studies are often physiologically ambiguous. The findings may be explicable in terms of more than one physiological mechanism.

Pitch research is a field in which the physiological ambiguity of psychoacoustic experimentation is particularly acute. For example, the fact that low pitch stems only from the lowest eight to ten harmonics has been interpreted by a number of theorists as support for a "place" model of pitch (Terhardt, 1974). However, in recent years it has become clear that the temporal properties of neural activity are also sensitive to the aural resolution of frequency components. Consequently, experiments which have traditionally served to distinguish between a "place" and a "periodicity" code for pitch have turned out to be largely irrelevant. The fact that low pitch is based on the spectrum of the signal does not preclude the relevance of neural periodicity for the processing of pitch. In principle, the spectral input to the central pitch processor is just as likely to be encoded in terms of periodic neural activity as it is in terms of the spatial distribution of auditory excitation.

Another issue of considerable theoretical significance is the physiological mechanism by which the percept of pitch is "synthesized" from the auditory representation of the signal spectrum. To address this question one must know how the spectrum is encoded in the central processor and the form in which pitch information is represented at this level. It is within this context that the results of the present study may have certain implications for models of pitch perception.

The frequency-following response reflects the periodic activity of auditory neurons. For this reason it can serve as a measure of neural periodicity with which to address the following question: to what extent is the pitch of complex tones based on the temporal discharge properties of auditory neurons?

The results of the present study suggest that neural periodicity plays a significant role in the processing of low pitch, though not necessarily in the form originally envisioned by early proponents of periodicity theory. Early periodicity models ignored the operation of cochlear filtering and, largely for this reason, the relationship between temporal neural coding and aural resolution remained unclear for many years. For example, Wever's (1949) the volley theory was illustrated only in terms of the pitch of single frequency signals. It was not entirely clear how the volley principle might be extended to complex signals. Most researchers assumed, like Schouten, that volley theory could only be of relevance to pitch coding in the absence of aural resolution. Thus, it seemed logical that the pitch of complex signals would stem from the unresolved portion of the spectrum. When it was shown that low pitch stems from the lower, rather than the higher harmonics, neural periodicity lost favor as a plausible mechanism of pitch extraction. The rejection of periodicity mechanisms in the early 1970's was based on two assumptions. The first assumption was that frequency components that are perceptually resolved do not physiologically interact in the cochlea. The second was that the auditory representation of the signal spectrum is based on the spatial excitation pattern of neural activity. The first assumption has been proven incorrect as a result of the studies of Rose *et al* (1969; 1974), Javel (1972; 1980) and Evans (1978). The validity of the second assumption for frequencies below 5 kHz is now being questioned as a result of the cochlear nerve studies of Sachs and Young (1979; 1980) and Young and Sachs (1979) and the theoretical work of Goldstein and Sruлович (1977). At the same time, the relevance of "place" mechanisms for the encoding of low frequencies is now being questioned as a consequence of the limited dynamic range of peripheral auditory neurons.

In view of these recent developments, it is likely that temporal neural mechanisms play a central role both in the peripheral coding of the spectrum and in the pitch extraction operations of more central auditory regions. This conclusion is consistent with the results of the present study, in so far as it was shown that many perceptual characteristics of low pitch are mirrored in the periodic activity of upper auditory brainstem neurons. This correspondence between low pitch and scalp-recorded FFR suggests that most of the relevant neural processing occurs at or below the level of the inferior colliculus.

By what means might the pitch of complex signals be encoded in terms of the temporal discharge pattern of auditory units? In my view, there are two likely possibilities. One possibility is that the neural processing upon which low pitch is based occurs at a fairly peripheral level of the auditory system. For example, this information might be obtained through the interaction between contiguous components observed by Rose *et al* (1969), Javel (1972; 1980) and Evans (1972) in the temporal discharge activity of cochlear nerve fibers. The other possibility is that the pitch of a complex signal is neurally "synthesized" through some form of temporal cross-correlation performed in a central region of the auditory pathway. To achieve this cross

correlation, all that would be required, in principle, is a population of neural "coincidence detectors" that receive as their input temporal information concerning two (or more) contiguous frequency components.

The existence of low pitch in the absence of peripheral interaction (Houtsma and Goldstein, 1971; 1972; Bilsen, 1973; van den Brink, 1975) implies that at least certain aspects of neural pitch extraction are performed in the central auditory pathway. However, this need not imply that peripheral processes do not participate in the formation of low pitch. For, the pitch produced by dichotic stimulation is considerably less salient than that generated by monaural stimulation, suggesting that at least certain components of the pitch sensation are derived from peripheral processing.

These considerations suggest that future auditory research might fruitfully explore the relative contributions of peripheral and central processes to the formation of low pitch and address questions such as the following: Are there two separate pitch processors, operating in parallel, and if so, what are their respective roles in the processing of pitch? To what extent is the information obtained through the two mechanisms redundant? And does the degree of redundancy influence the salience of the perceived pitch? These are but a few of the questions that remain to be explored by future research.

8.4 FREQUENCY-FOLLOWING RESPONSE CORRELATES OF VOWEL QUALITY

We turn our attention finally to the auditory representation of vowel quality. In Chapter 1.62 it was observed that vowel quality is based on the distribution of energy in the low-frequency region (< 5 kHz) of the signal spectrum. Consequently, the auditory representation of vowel quality is likely to be based on mechanisms responsible for the analysis of the frequency spectrum. The issue addressed in Chapter 7 concerned the extent to which the vocalic spectrum is encoded in terms of the temporal discharge pattern of auditory neurons. Toward this end, frequency-following responses were recorded in response to a series of synthetic vowels in order to determine whether certain parameters of the signal spectrum, such as the formant frequencies and the fundamental, are reflected in the scalp-recorded response.

The results of experiments reported in Chapter 7 suggest that certain parameters of the vocalic spectrum are, in fact, temporally coded at the level of the upper auditory brainstem. These include the first (and possibly second and third formants) as well as the fundamental frequency. Each parameter is discussed in turn.

(a) Fundamental Frequency

FFR recorded in response to synthetic vowels exhibit a marked periodicity equal to the period of the stimulating waveform. This is evident regardless of whether the fundamental frequency of the stimulus is 100 or 200 Hz. The periodicity is particularly marked when the first formant lies in the "region of dominance."

It was hypothesized that this periodic component of the FFR may reflect, in part, the activity of relatively high-frequency neurons synchronized to the envelope of the waveform created through the interaction of the high, unresolved harmonics. This hypothesis is consistent with the studies of Young and Sachs (1979) and Kiang *et al* (1979) who show that single cochlear nerve fibers tuned to frequencies above the lower formants synchronize their discharge to the fundamental frequency. It is also possible that the periodic component of the FFR reflects the response of units to harmonics lying in the region of dominance for low pitch.

(b) First Formant

The spectrum of the FFR often contains local energy maxima corresponding to the frequency of the first formant. This correspondence is most evident for the vowel [I], where F1 lies below 500 Hz. For other vowels, in which F1 is located at 600 or 800 Hz, the correspondence between the FFR and signal spectrum is less clear. This is due to the predominance of low-frequency energy (< 500 Hz) in the FFR spectrum. The predominance of this low-frequency energy was interpreted as reflecting the neural rectification products of units synchronized to the waveform envelope. Despite the presence of this low-frequency energy, local energy maxima in the vicinity of F1 are often observed. These peaks were usually characterized by a rather abrupt reduction of energy on the high frequency side. It is possible that this abrupt attenuation is indicative of some form of "synchrony suppression", similar to that observed by Young and Sachs (1979) for the cochlear nerve.

In a separate series of experiments, it was shown that the low-frequency energy observed in the FFR can be effectively cancelled, along with the "pitch-synchronous" component of the response, by means of the response-subtraction procedure. The resulting FFR spectrum contains, in most instances, a large peak at the frequency corresponding to the first formant.

(c) Second and Third Formants

Although the spectra of FFR often contain local energy maxima at frequencies corresponding to F2 and F3, the correlation between FFR and signal spectrum is much less evident than is the case for F1. In both normal and subtracted responses the spectrum level of these peaks is considerably lower than that of the peaks associated with F1. This attenuation of the response spectrum for frequencies above 1 - 1.5 kHz probably reflects two separate factors. One concerns the fact that relatively few units in the upper auditory brainstem pathway phase-lock to frequencies above 1.5 kHz (Rose *et al*, 1966; Marsh *et al*, 1974). The second factor is that, even at the level of the cochlear nerve, only a relatively small proportion of fibers synchronize to F2 and F3 (Young and Sachs, 1979). Thus, it would be difficult to observe periodic activity synchronized to these high frequencies without the use of depth recordings made directly from the brainstem pathway.

In summary, the results of the present study indicate that the fundamental frequency and first formant of vocalic signals are reflected in the temporal discharge pattern of upper auditory brainstem neurons as recorded from the human scalp by means of far-field techniques. The "pitch-synchronous" activity of these brainstem units is particularly evident when the FFR is recorded to the signal presented in a uniform polarity. The first formant component of the response becomes evident when the pitch-synchronous component of the FFR is cancelled. Formants above the first are not well represented in the scalp-response.

8.5 THEORETICAL IMPLICATIONS FOR MODELS OF VOWEL PERCEPTION

Current models of speech perception generally assume that the auditory analysis of the signal spectrum is based on the spatial distribution of neural excitation. This assumption is implicit in various attempts to model the peripheral auditory response to speech through computation of the "loudness density function" of the spectrum (Zwicker *et al.*, 1979; Bladon and Lindblom, 1979; Karnickaya *et al.*, 1975; Alinat, 1975). Recent physiological studies of the cochlear nerve response to synthetic vowels cast grave doubts on the validity of this approach. At signal levels corresponding to conversational level (60-70 dB SPL), most cochlear nerve fibers are discharging at their maximum rate, thus precluding the possibility of distinguishing spectral peaks on the basis of the spatial excitation pattern. Over this same intensity range, however, formant frequencies are distinguishable in terms of the temporal discharge pattern of cochlear fibers.

Young and Sachs (1979) suggest that the vocalic spectrum is preserved in the discharge pattern of the cochlear nerve over a large dynamic range through the operation of "synchrony suppression." At moderate-to-high signal levels, when most of the cochlear nerve has reached rate saturation, this mechanism serves to maintain the neural contrast between spectral peaks and valleys by reducing the amount of synchronized activity to frequencies outside the formant regions. Fibers tuned to a formant frequency phase-lock to this frequency over a broad range of intensities. Fibers tuned to non-formant frequencies phase-lock to frequencies lying in the center of the response area only at low-to-moderate intensities. At higher signal levels, they, too, synchronize to a formant frequency (usually F1).

It is evident from this pattern of neural activity that information concerning the location of formant frequencies can be represented in terms of the temporal discharge pattern of the cochlear nerve. Over a broad range of intensities, a contrast between spectral peaks and the remainder of the spectrum is preserved in terms of the amount of synchronized activity. At low-to-moderate intensities, when relatively few fibers are synchronized to the formants, this contrast can be signalled primarily in terms of the difference in synchronized rate. Once fibers tuned to the formant frequencies reach rate saturation, any further increase in intensity would result in a loss of neural

contrast, if synchronized rate were the only relevant parameter. However, at moderate-to-high intensities, a large number of fibers, spanning a broad tonotopic range, become synchronized to (primarily) the first formant. Thus, the neural contrast between the formants and the remainder of the spectrum would be preserved in terms of the difference in the total number of fibers synchronized to these portions of the spectrum.

Despite the force of their data, Young and Sachs (1979) argue for retaining a "place" model of auditory spectrum analysis. In their view, the auditory representation of the vocalic spectrum is based on the contrast in the synchronized rate of units tuned to the formant frequencies and the amount of synchronized activity displayed by other units to frequencies lying in the center of their response area. Thus, frequency information would be signalled by the tonotopic affiliation of individual fibers, while the amplitude of each frequency component would be coded in terms of the synchronized rate.

In my opinion, Young and Sachs' proposal is not very plausible, for it would require that the auditory system keep track of the appropriate synchronized frequency for each fiber. Otherwise, it would not be possible to determine whether the synchronized activity exhibited by a particular fiber was the result of a purely excitatory response to a component located in the middle of its response area or was the product of "synchrony suppression." Furthermore, it is not entirely clear that the behavior of cochlear nerve fibers at high intensities is appropriately characterized by the term "synchrony suppression." The term implies a reduction in the amount of synchronized activity exhibited by a nerve fiber. However, it is quite clear from Young and Sachs' data that an actual reduction in synchronized rate rarely occurs. Rather, the frequency to which the fiber synchronizes, changes with level. The process involves suppression only from the standpoint of the fiber's tonotopic identity. From the perspective of the system as a whole, a change in a fiber's synchronization pattern involves a process of "synchrony recruitment" by the first formant.

With this change in perspective, it is possible to envision how the vocalic spectrum may be temporally encoded at the level of the cochlear nerve. It is clear from Young and Sachs' study that, at moderate-to-high intensities, only certain portions of the spectrum are temporally encoded to any significant degree. These temporally-encoded regions include the first three formants and those frequencies below F1. What is encoded at the level of the cochlear nerve is information concerning the frequency of the spectrum peaks, and this information is transmitted to higher centers in terms of the temporal pattern of neural discharge. Information about the signal's pitch is probably extracted from the region of the spectrum just below F1. The frequency composition of the remainder of the spectrum is largely irrelevant.

This view of vowel encoding implies that the actual configuration of the spectrum envelope is largely irrelevant for the auditory representation of vowel quality. What matters is the location of the

prominent spectrum peaks. Consistent with this view are studies by Mushnikov and Chistovich (1969), Ainsworth and Millar (1972), and Chistovich and Lublinskaja (1979) which show that vowel quality is remarkably insensitive to changes in the relative amplitude level of the formants (Chapter 1.62).

In natural speech, the spectrum is changing continuously over time. Under these conditions it is probably neither necessary nor efficient for the auditory system to encode the entire spectrum. Detailed representation of the spectrum is not required because this information can be largely reconstructed from knowledge of the fundamental and the formants. Under the circumstances, what is required is a relatively rapid and precise means to track the location of the spectral peaks. Temporal coding is ideally suited for the task as it provides a significantly more precise means to measure the spectrum than place coding (Siebert, 1970), particularly over brief durations of time (Goldstein and Srulovicz, 1977).

The discussion, so far, has focussed on the neural representation of vocalic signals at the level of the cochlear nerve. How might the vocalic spectrum be encoded at higher levels of the auditory pathway? Although there are no systematic single unit studies which specifically address this question, there is reason to believe that most of the significant parameters of the vocalic spectrum (e.g. F1 and the fundamental frequency) are temporally encoded up to at least the level of the superior olivary complex (Chapter 1.64). The results of the present study suggest that even at the level of the upper auditory brainstem pathway, the first formant and the fundamental continue to be temporally encoded. At higher levels, the picture is much less clear. Only a very small proportion of units in the medial geniculate body appear to be capable of phase-locking (Rouiller *et al*, 1979). And true phase-locking has yet to be observed in the auditory cortex. It therefore seems likely that the temporal code is transformed into another form at these highest levels of the auditory pathway. The nature of this neural transformation will undoubtedly form the subject of much future research.

NOTES

- 1.1 Ohm's original formulation proposed that the ear preserves the phase of each frequency component (de Boer, 1976: 492). Helmholtz (1863) modified Ohm's law in accordance with his belief that the ear was, within limits, essentially "phase-deaf."
- 1.2 In the first edition of Helmholtz's book, the resonators were associated with the arches of Corti.
- 1.3 These special listening conditions involved sounding the frequency of the partial independently before listening for it in a complex. He also used a resonator to enhance the partial's perceptibility. Plomp (1964), without the use of resonators, found that the ability to isolate individual partials declines significantly between the fifth and ninth harmonic (see Chapter 1.25).
- 1.4 The pitch ambiguity of signal C is explicable in terms of the dominance of low harmonics (Chapter 1.25). The lower pitch probably stemmed from the fourth, fifth, and sixth harmonics, while the pitch an octave above was most likely a product of the second and fourth harmonics. The presence of this second, higher pitch was probably the result of the relatively small amount of energy at the third harmonic.
- 1.5 The level of odd-order combination tones ($f_1 - n(f_2 - f_1)$) relative to the primaries (f_1, f_2) is a function of f_2/f_1 . For $f_2/f_1 = 1.1$, the cubic difference tone ($CDT = 2f_1 - f_2$) is as great as 10-15 dB down from f_1 ; $3f_1 - 2f_2$ is about 30 dB down; $4f_1 - 3f_2$ is about 40 dB down. When $f_2/f_1 = 1.33$, the CDT is about 40 dB down (Goldstein, 1970: Figure 1).
- 1.6 A subsequent revision of the optimum processor model specifies those conditions under which this restriction need not apply (Gerson and Goldstein, 1978).
- 1.7 These predicted pitch values were computed by assuming that the central processor had erred in assigning the harmonic ranks of the stimulus components. In Goldstein's model, the probability of such a template mismatch increases with increasing harmonic rank. Thus, tones composed of low harmonics would not be expected to generate ambiguous pitches as often as signals composed of high harmonics.
- 1.8 This is true as long as the frequency components are of roughly equal intensity. When the relative level of contiguous components exceeds 5-10 dB, the temporal discharge pattern of neurons tuned to the less intense component becomes synchronized to the higher amplitude component (Rose et al, 1969; 1971; 1974; Brugge et al 1969).

- 1.9 Under appropriate listening conditions, certain observers also match the pitch of the signal to the "true" fundamental (Gerson and Goldstein, 1978).
- 1.10 The gross action potential reflects primarily high-frequency units due to the differential velocity of the traveling wave. Its velocity is greatest for the basal portion of the basilar membrane. A broadband signal, such as a click, will activate a large number of nerve fibers, tuned to frequencies spanning the entire spectrum. However, the activity of low-frequency fibers will be obscured due to the progressive lag in their activation. At more basal portions of the cochlea, a large number of fibers will be activated more or less simultaneously. It is this simultaneous activation of a large number of cochlear nerve fibers that is reflected in the gross AP.
- 2.1 These repetition rates are considerably higher than would be expected to occur in a natural communication context and might thus appear to compromise the relevance of the experimental results for the processing of speech and musical sounds. However, the present series of experiments were designed to monitor the activity of brainstem units relevant to the processing of auditory attributes based directly on the signal spectrum. As such, it is unlikely that appreciably longer stimulus durations would have significantly affected the results obtained. Pilot experiments with stimuli of 100 msec duration indicated that the amplitude of the FFR is largest during the initial 60 msec of the response. The decline of FFR amplitude after this initial interval probably reflects neural adaptation. The effect of stimulus repetition rate on FFR amplitude depends on the bandwidth of the signal. For pure tones, Davis and Hirsh (1979) have found no significant amplitude decrement for rates up to 35 per second. The click-evoked BSER (where signal bandwidth is usually about 10 kHz) is sensitive to repetition rate for rates above 10 sec (Davis, 1976). We have found that both pure tone and complex tone stimuli (comprised of two to four components) can be presented up to 15 per second with no decrement in response amplitude. In fact, there is some augmentation of the scalp response for the faster rates. However, with broadband stimuli, such as synthetic vowels, the maximum rate, without amplitude reduction, is approximately 9 per second (the maximum rate used in this study).
- 2.2 This statement may seem somewhat surprising in view of the popular conception of hearing as an exclusively "active" process. However, the neural activity of the brainstem is not significantly affected by the state of the cortex. Consequently, the subject's state of consciousness does not affect the FFR, except in so far as the background EEG, of cortical origin, interferes with the brainstem activity being averaged.

- 3.1 In a previous publication (Smith et al., 1978), the stimulus frequency was reported to be 365 Hz. This value has been revised in light of subsequent frequency measurements.
- 5.1 Two of the stimuli used in Hall's (1979) study were three-component tones consisting of the third, fourth, and fifth harmonics of a 200 Hz fundamental (600+800+1000 Hz). The modulation depth was either 1.0 or 2.0.
- 5.2 It is not entirely clear why the largest FFR is produced when the middle component is in counterphase to f_1 and f_3 . One possibility is that, due to the effective cochlear rectification of the signal, one polarity of the stimulus waveform is a slightly more effective stimulus than the other polarity. However, this does not explain why the 45 degree stimulus produces a response almost as large as that generated by the counterphase signal.
- 7.1 The stimuli were generated in the Phonology Laboratory, Department of Linguistics, University of California, Berkeley. I thank Dr. John Ohala for permission to use the facilities of the Phonology Laboratory.
- 7.2 Although stimuli were filtered above 200 Hz, this did not appreciably affect the stimulus spectrum below 200 Hz, because of the poor frequency response of the headphones to frequencies below this point.
- 7.3 Due to the unilateral excitation properties of the hair cells, the effective stimulus is essentially a rectified (and filtered) version of the original waveform. The attendant neural rectification produces an asymmetric waveform which, when spectrally analyzed, contains considerable energy at harmonics of the fundamental frequency. These spectral harmonics are rectifier distortion products (Young and Sachs, 1979).
- 8.1 FFR recorded to these signals were compared to monaural control signals composed of the same three frequency components as presented dichotically. Usually, the spectrum level of the 244 Hz component of the brainstem response was at least several dB lower than the fundamental component of the FFR generated by the monaural response. Also, it was considerably lower in amplitude than the 488 Hz component (the difference frequency for the two-component signal). However, in at least one instance the 244 Hz component of the dichotic response was approximately the same level as the 488 Hz component and dominated the spectrum of the FFR in the region below 488 Hz. However, it is not clear whether this isolated result reflects neural activity related to the synthesis of information presented to the opposing ears or rather some subharmonic of the two-component signal's difference frequency (488 Hz).

- 8.2 Certain attributes of low pitch were not reflected in some parameter of the FFR (such as the pitch of dichotic signals). It is possible that these attributes may require signal levels below the effective threshold of scalp-recorded FFR for their expression. It is of interest, in this regard, that the great majority of psychophysical studies of low pitch have used stimulus levels below 50 dB SL. There has been no systematic study of the effect of stimulus level on the salience of low pitch. This is an issue of considerable theoretical interest, for at high intensities a large unit population with a very broad tonotopic distribution is responding to the signal. If the pitch-relevant information is being transmitted through the discharge activity of a selected proportion of these units, how does the central pitch processor identify the correct population? It is possible that the dominance for low pitch of frequencies over a very restricted region of the spectrum is related to this selection process.

Another factor of potential relevance is the use of far-field recording techniques. The use of scalp electrodes increases the probability of picking up extraneous neural activity due to the large size of the unit population contributing to the scalp response. This risk can be minimized by the use of depth electrodes which record the activity of relatively small neural populations of (approximately) determinate tonotopic affiliation. This is a potentially significant issue because of the likelihood that the information of most direct relevance to the processing of low pitch would be found in the discharge activity of units tuned most closely to the stimulus components, since this is the only population that is continuously active over the entire dynamic range of low pitch. For this reason, it will ultimately be necessary to validate the present results with recordings obtained directly from the auditory brainstem nuclei (of e.g. the cat). For only by direct recording methods will it be possible to observe the phase-locked activity of auditory units at intensities sufficiently low as to adequately restrict the tonotopic distribution of activated units.

In the present study, we have suggested that the response-subtraction procedure may reflect the activity of units tuned to the region of the spectrum encompassing the stimulus components. In view of the significance of tonotopic specificity discussed above, it would be of interest to determine whether the activity obtained from the scalp through this procedure is comparable to that obtained from depth electrodes at low stimulus intensities and using conventional recording procedures.

REFERENCES

- Abbas, P.J. and Sachs, M.B. (1976) Two-tone suppression in auditory nerve fibers: Extension of a stimulus-response relationship. J. Acoust. Soc. Am. 59: 112-122.
- Ainsworth, W.A. and Millar, J.B. (1972) The effect of relative formant amplitude on the perceived identity of synthetic vowels. Language and Speech 15: 328-341.
- Aitkin, L.M., Anderson, D.J. and Brugge, J.F. (1970) Tonotopic organization and discharge characteristics of single neurons in nuclei of the lateral lemniscus of the cat. J. Neurophysiol. 33: 421-440.
- Alinat, P. (1975) A study of the French language's phonemes by means of an artificial cochlea: Application to speech recognition. Rev. Tech. Thompson CSF 7: 91-123.
- Allanson, J.T. and Whitfield, I.C. (1956) The cochlear nucleus and its relation to theories of hearing. In Third London Symposium on Information Theory, C. Cherry (ed.). London: Butterworth, pp. 269-286.
- American National Standards Institute (1973) Psychoacoustical Terminology. ANSI S3.20.
- Anderson, D.J., Rose, J.E. and Brugge, J.F. (1971) Temporal position of discharges in single auditory fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects. J. Acoust. Soc. Am. 49: 1131-1139.
- Arthur, R.M., Pfeiffer, R.R., and Suga, N. (1971) Properties of "two-tone inhibition" in primary auditory neurons. J. Physiol. 212: 593-609.
- Bekesy, G. von (1928) Zur Theorie des Horens: die Schwingungsform der Basilarmembran. Physik. Zeits. 29: 793-810.
- Bekesy, G. von (1934) Über die nichtlinearen Verzerrungen des Ohres. Ann. Phys. 20: 809-827.
- Bekesy, G. von (1960) Experiments in Hearing. New York: McGraw-Hill.
- Bekesy, G. von (1972) The missing fundamental and periodicity detection in hearing. J. Acoust. Soc. Am. 51: 631-637
- Bernstein, J. (1871) Untersuchungen über den Erregungsvorgang im Nerven- und Muskelsysteme. Heidelberg: Carl Winter's Universitätsbuchhandlung.

- Bilsen, F.A. (1973) On the influence of the number and phase of harmonics on the perceptibility of the pitch of complex signals. Acustica 28: 60-65.
- Bilsen, F.A. (1977) Pitch of noise signals: Evidence for a "central spectrum." J. Acoust. Soc. Am. 61: 150-161.
- Bilsen, F.A. and Goldstein, J.L. (1974) Pitch of dichotically delayed noise and its possible spectral basis. J. Acoust. Soc. Am. 55: 292-296.
- Bilsen, F.A. and Ritsma, R.J. (1970) Some parameters influencing the perceptibility of pitch. J. Acoust. Soc. Am. 47: 469-475.
- Bilsen, F.A., ten Kate, J.H., Buunen, T.J.F. and Raatgever, J. (1975) Responses of single units in the cochlear nucleus of the cat to cosine noise. J. Acoust. Soc. Am. 58: 858-866.
- Bladon, R.A.W. and Lindblom, B. (1979) Auditory modeling of vowels. J. Acoust. Soc. Am. 65: S5(A).
- Boer, E. de (1956a) On The Residue in Hearing. Academic Thesis, University of Amsterdam.
- Boer, E. de (1956b) Pitch of inharmonic signals. Nature 178: 535-536.
- Boer, E. de (1976) On the "residue" and auditory pitch perception. In Handbook of Sensory Physiology V/3, W.D. Keidel and W.D. Neff (eds.). Berlin: Springer Verlag, pp. 479-583.
- Boer, E. de, Machiels, M.B. and Krudenier, C. (1977) Low level frequency following response. Audiology 16: 229-240.
- Boudreau, J.C. (1965) Neural volleying: Upper frequency limits detectable in the auditory system. Nature 208: 1237-1238.
- Brink, G. van den (1970) Two experiments on pitch perception: Diplacusis of harmonic AM signals and pitch of inharmonic AM signals. J. Acoust. Soc. Am. 48: 1355-1365.
- Brink, G. van den (1975) Monaural frequency-pitch relations as the origin of binaural diplacusis for pure tones and residue sounds. Acustica 32: 166-174.
- Brugge, J., Anderson, D., Hind, J. and Rose, J. (1969) Time structure discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. J. Neurophysiol. 32: 386-401.
- Buchwald, J. and Huang, C-M. (1975) Far-field acoustic response: Origins in the cat. Science 189: 382-384.

- Burns, E.M. and Viemeister, N.F. (1976) Nonspectral pitch. J. Acoust. Soc. Am. 60: 863-869.
- Buunen, T.J.F. (1975) Two hypotheses on monaural phase effects. Acustica 34: 98-105.
- Buunen, T.J.F., Bilsen, F.A. and Brink, G. van den (1974) Phase effects in a three component signal. J. Acoust. Soc. Am. 55: 297-303.
- Buunen, T.J.F. and Rhode, W. (1978) Responses of fibers in the cat's auditory nerve to the cubic difference tone. J. Acoust. Soc. Am. 64: 772-781.
- Carlson, R., Granstrom, B. and Fant, G. (1970) Some studies concerning perception of isolated vowels. STL-QPSR 2-3: 19-35.
- Carlson, R., Fant, G., and Granstrom, B. (1975) Two-formant models, pitch, and vowel perception. In Auditory Analysis and Perception of Speech, G. Fant and M. Tatham (eds.). London: Academic Press, pp. 55-82.
- Caspary, D.M., Rupert, A.L., and Moushegian, G. (1977) Neuronal coding of vowel sounds in the cochlear nuclei. Exp. Neurol. 54: 414-431.
- Chistovich, L.A. (1971) Auditory processing of speech stimuli - evidences from psychoacoustics and neurophysiology. Proc. 7th Int. Congr. Acoust., pp. 27-41.
- Chistovich, L.A. and Lublinskaya, V.V. (1979) The "center of gravity" in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. Hearing Research 1: 185-195.
- Chistovich, L.A., Sheikin, R.L., and Lublinskaja, V.V. (1979) "Centers of gravity" and spectral peaks as determinants of vowel quality. In Frontiers of Speech Communication Research, B. Lindblom and S. Ohman (eds.). London: Academic Press, pp. 143-157.
- Cramer, E.M. and Huggins, W.H. (1958) Creation of pitch through binaural interaction. J. Acoust. Soc. Am. 30: 413-417.
- Davis, H. (1976) Principles of electric response audiometry. Ann. Otol. Rhinol. and Laryngol 85: Supplement 28.
- Davis, H., Derbyshire, A.J., Lurie, M.H., and Saul, L.J. (1933) Further analysis of cochlear activity and auditory action currents. Trans. Am. Otol. Soc. 23: 106-116.
- Davis, H. and Hirsh, S.K. (1976) The audiometric utility of brainstem responses to low-frequency sounds. Audiology 15: 181-195.

- Davis, H. and Hirsh, S. (1979) A slow brain stem response for low-frequency audiometry. Audiology 18: 445-461.
- Derbyshire, A.J. and Davis, H. (1935) The action potentials of the auditory nerve. Am. J. Physiol. 113: 475-504.
- Dolmazon, J., Bastet, L., and Shupljakov, V. (1977) A functional model of peripheral auditory system in speech processing. IEEE Trans. ASSP 25: 261-264.
- Dye, R.H. Jr. and Hafter, E.R. (1980) Just-noticeable differences of frequency for masked tones. J. Acoust. Soc. Am. 67: 1746-1753.
- Eldredge, D.H. (1978) Interpretation of evoked neural responses: Facts and illusions related to methods for recording whole-nerve action potentials. In Evoked Electrical Activity in the Auditory Nervous System, R.F. Naunton and C. Fernandez (eds.). New York: Academic Press, pp. 151-164.
- Erulkar, S.D. (1972) Comparative aspects of sound localization. Physiol. Rev. 52: 237-359.
- Evans, E.F. (1975) The cochlear nerve and cochlear nucleus. In Handbook of Sensory Physiology V/2, W.D. Keidel and W.D. Neff (eds.). Berlin: Springer Verlag.
- Evans, E.F. (1977) Frequency selectivity at high signal levels of single units in cochlear nerve and nucleus. In Psychophysics and Physiology of Hearing, E.F. Evans and J.P. Wilson (eds.). London: Academic Press, pp. 185-192.
- Evans, E.F. (1978) Place and time coding of frequency in the peripheral auditory system: Some physiological pros and cons. Audiology 17: 369-420.
- Evans, E.F. and Palmer, A.R. (1975) Responses of units in the cochlear nerve and nucleus of the cat to signals in the presence of bandstop noise. J. Physiol. 252: 60-62P.
- Fant, G. (1960) Acoustic Theory of Speech Production. The Hague: Mouton.
- Fastl, H. and Stoll, G. (1979) Scaling of pitch strength. Hearing Research 1: 293-302.
- Flanagan, J.L. (1962) Computational model for basilar membrane displacement. J. Acoust. Soc. Am. 34: 1370-1376.
- Fletcher, H. (1924) The physical criterion for determining the pitch of a musical tone. Phys. Rev. 23: 427-437.
- Fletcher, H. (1929) Speech and Hearing. New York: Van Nostrand.

- Galambos, R. and Davis, H. (1943) The response of single auditory-nerve fibers to acoustic stimulation. J. Neurophysiol. 6: 39-58.
- Gardi, J. and Merzenich, M. (1979) The effect of high-pass noise on the scalp-recorded frequency-following response (FFR) in humans and cats. J. Acoust. Soc. Am. 65: 1491-1500.
- Gerson, A. and Goldstein, J.L. (1978) Evidence for a general template in central optimum processing for pitch of complex tones. J. Acoust. Soc. Am. 63: 498-510.
- Gibson, M.M., Hind, J.E., Kitzes, L.M., and Rose, J.E. (1977) Estimation of traveling wave parameters from the response properties of cat AVCN neurons. In Psychophysics and Physiology of Hearing, E.F. Evans and J.P. Wilson (eds.). London: Academic Press, pp. 57-68.
- Goblick, T.J. and Pfeiffer, R.R. (1969) Time-domain measurements of the cochlear nonlinearities using combination clicks stimuli. J. Acoust. Soc. Am. 46: 924-938.
- Gold, B. and Rabiner, L.R. (1968) Analysis of digital and analog formant synthesizers. IEEE Trans. Audio-Electroacoust. AU-16: 81-94.
- Goldstein, J. (1967) Auditory nonlinearity. J. Acoust. Soc. Am. 41: 458-479.
- Goldstein, J.L. (1970) Aural combination tones. In Frequency Analysis and Periodicity Detection in Hearing, R. Plomp and G. Smoorenburg (eds.). Leiden: Sijthoff., pp. 230-245.
- Goldstein, J.L. (1973) An optimum processor theory for the central formation of the pitch of complex tones. J. Acoust. Soc. Am. 54: 1496-1516.
- Goldstein, J.L. (1978) Mechanisms of signal analysis and pattern perception in periodicity pitch. Audiology 17: 421-445.
- Goldstein, J.L. and Kiang, N.Y.-S. (1968) Neural correlates of the aural combination tone 2F1-F2. Proc. IEEE 56: 981-992.
- Goldstein, J.L. and Srulovicz, P. (1977) Auditory-nerve spike intervals as an adequate basis for aural spectrum analysis. In Psychophysics and Physiology of Hearing, E.F. Evans and J.P. Wilson (eds.). London: Academic Press, pp. 337-346.
- Green, D.M. (1976) An Introduction to Hearing. Hillsdale, N.J.: Lawrence Erlbaum Associates.

- Greenberg, S., Smith, J.C., Marsh, J.T., and Brown, W.S. (1978) Human frequency-following response to synthetic vowels. J. Acoust. Soc. Am. 63: S76(A).
- Greenberg, S. and Marsh, J.T. (1979) Human frequency-following responses to synthesized vowels. Paper presented at the 59th Annual Meeting of the Western Psychological Society, San Diego.
- Greenwood, D.D. (1977) Empirical travel time functions on the basilar membrane. In Psychophysics and Physiology of Hearing, E.F. Evans and J.P. Wilson (eds.). London: Academic Press, pp. 43-53.
- Greenwood, D.D., Merzenich, M.M., and Roth, G.L. (1976) Some preliminary observations on the interrelations between two-tone suppression and combination tone driving in the anteroventral cochlear nucleus of the cat. J. Acoust. Soc. Am. 59: 607-633.
- Hall, J.W. III (1979) Auditory brainstem frequency following responses to waveform envelope periodicity. Science 205: 1297-1299.
- Hashimoto, T., Katayama, Y., Murata, K. and Taniguchi, I. (1975) Pitch synchronous response of cat cochlear nerve fibers to speech sounds. Jap. J. Physiol. 25: 633-644.
- Helmholtz, H.L.F. von (1863) Die Lehre von Tonempfindungen als Physiologie Grundlage der Theorie der Musik. Braunschweig: F. Vieweg und Sohn. English translation by A.J. Ellis of the 4th edition (1877), On the Sensations of Tone as a Physiological Basis for the Theory of Music, (1897). Reprinted by Dover Publications, New York, 1954.
- Helmholtz, H.L.F. von (1868) Über den Muskelton. Verhandl. Naturhist.-Medicin. Vereins Heidelberg 4: 88-90.
- Holstein, S.B., Buchwald, J.S., and Schwafel, J.A. (1969) Tone response patterns of the auditory nuclei during normal wakefulness, paralysis, and anesthesia. Brain Research 15: 483-499.
- Hoogland, G.A. (1953). The Missing Fundamental: A Place Theory of Frequency Analysis in Hearing. Academic Thesis, University of Utrecht.
- Houtgast, T. (1973) Psychophysical experiments on "tuning curves" and "two-tone inhibition." Acustica 29: 168-179.
- Houtgast, T. (1974a) Lateral Suppression in Hearing. Doctoral Dissertation, Free University, Amsterdam.
- Houtgast, T. (1974b) Auditory analysis of vowel-like sounds. Acustica 31: 320-324.

- Houtsma, A.J.M. and Goldstein, J.L. (1971) Perception of musical intervals: Evidence for the central origin of the pitch of complex tones. MIT-RLE Technical Report 484.
- Houtsma, A.J.M. and Goldstein, J.L. (1972) The central origin of the pitch of complex tones: Evidence from musical interval recognition. J. Acoust. Soc. Am. 51: 520-529.
- Huis in't Veld, F., Osterhammel, P. and Terkildsen, K. (1977a) Frequency following auditory brainstem responses in man. Scand. Audiol. 6: 27-34.
- Huis in't Veld, F., Osterhammel, P. and Terkildsen, K. (1977b) Frequency selectivity of the 500 Hz frequency following response. Scand. Audiol. 6: 35-42.
- Javel, E. (1972) A neurophysiological investigation of single unit responses to some periodic complex auditory stimuli. Doctoral Dissertation, University of Pittsburgh.
- Javel, E. (1980) Coding of AM tones in the chinchilla auditory nerve: Implications for the pitch of complex tones. J. Acoust. Soc. Am. 68: 133-146.
- Javel, E., Geisler, C.D., and Ravindran, A. (1978) Two-tone suppression in auditory nerve of the cat: Rate-intensity and temporal analyses. J. Acoust. Soc. Am. 63: 1093-1104.
- Jewett, D.L., Romano, M.N., and Williston, J.S. (1970) Human auditory evoked potentials: Possible brain stem components detected on the scalp. Science 167: 1517-1518.
- Jewett, D.L. and Williston, J.S. (1971) Auditory-evoked far fields averaged from the scalp of humans. Brain 94: 681-696.
- Karnickaya, E., Mushnikov, V., Slepokurova, N., and Zhukov, S. (1975) Auditory processing of steady-state vowels. In Auditory Analysis and Perception of Speech, G. Fant and M. Tatham (eds.). London: Academic Press, pp. 37-54.
- Kemp, E.H., Coppee, G.E., and Robinson, E.H. (1937) Electrical responses of the brain stem to unilateral auditory stimulation. Am. J. Physiol. 120: 304-315.
- Kiang, N.Y-S., Watanabe, T., Thomas, E.C. and Clark, L.F. (1965) Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. Research Monograph No. 35, Cambridge, Mass.: MIT Press.
- Kim, D.O. and Molnar, C.E. (1979) A population study of cochlear nerve fibers: Comparison of spatial distributions of average-rate and phase-locking measures of responses to pure tones. J. Neurophysiol. 42: 16-30.

- Klatt, D.H. (1973) Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. J. Acoust. Soc. Am. 53: 8-16.
- Klein, W., Plomp, R., and Pols, L.C.W. (1970) Vowel spectra, vowel spaces, and vowel identification. J. Acoust. Soc. Am. 48: 999-1009.
- Licklider, J.C.R. (1954) "Periodicity" pitch and "place" pitch. J. Acoust. Soc. Am. 26: 945(A).
- Licklider, J.C.R. (1955) Influence of phase coherence upon the pitch of complex periodic sounds. J. Acoust. Soc. Am. 27: 996(A).
- Licklider, J.C.R. (1959) Three auditory theories. In Psychology: A Study of a Science, Vol. 1, S. Koch (ed.). New York: McGraw-Hill, pp. 41-144.
- Licklider, J.C.R. (1962) Periodicity pitch and related auditory process models. Int. Audiol. 1: 11-35.
- Lindquist, J. and Pauli, S. (1968) The role of relative spectrum levels in vowel perception. Proc. 6th Int. Congr. Acoust. 2: B91-B94.
- Marsh, J.T., Brown, W.S. and Smith, J.C. (1974) Differential brainstem pathways for the conduction of auditory frequency-following responses. Electroenceph. clin. Neurophysiol. 36: 415-424.
- Marsh, J.T., Brown, W.S. and Smith, J.C. (1975) Far-field recorded frequency-following responses: Correlates of low pitch auditory perception in humans. Electroenceph. clin. Neurophysiol. 38: 113-119.
- Marsh, J.T., Smith, J.C. and Worden, F.G. (1972) Receptor and neural responses to auditory masking of low frequency tones. Electroenceph. clin. Neurophysiol. 32: 63-74.
- Marsh, J.T. and Worden, F.G. (1968) Sound evoked frequency-following responses in the central auditory pathway. Laryngoscope 78: 1149-1163.
- Marsh, J.T., Worden, F.G. and Smith, J.C. (1970) Auditory frequency-following response: Neural or artifact? Science 169: 1222-1223.
- Mathes, R. and Miller, R. (1947) Phase effects in monaural perception. J. Acoust. Soc. Am. 19: 780-797.
- Minifie, F.D. (1973) Speech acoustics. In Normal Aspects of Speech, Hearing, and Language, F.D. Minifie, T.J. Hixon, and F. Williams (eds.). Englewood Cliffs, N.J.: Prentice-Hall, pp. 235-284.

- Morest, D.K. (1975) Structural organization of the auditory pathways. In The Nervous System, Volume 3: Human Communication and its Disorders, E.L. Eagle (ed.). New York: Raven Press, pp. 19-30.
- Moore, B.C.J. (1973) Frequency difference limens for short duration tones. J. Acoust. Soc. Am. 54: 610-619.
- Moore, B.C.J. (1977) Effects of relative phase of the components on the pitch of three-component complex tones. In Psychophysics and Physiology of Hearing, E.F. Evans and J.P. Wilson (eds.). London: Academic Press, pp.349-358.
- Moore, T.J. and Cashin, J.L. (1974) Response patterns of cochlear nucleus neurons to excerpts from sustained vowels. J. Acoust. Soc. Am. 56: 1565-1576.
- Moore, T.J. and Cashin, J.L. (1976) Response of cochlear nucleus neurons to synthetic speech. J. Acoust. Soc. Am. 59: 1443-1449.
- Moore, T.J. and Cashin, J.L. (1977) Response patterns of "primary-like" cochlear nucleus neurons to excerpts from sustained vowels. Aerospace Med. Res. Lab. Bull. TR 77-10.
- Moushegian, G. and Rupert, A.L. (1970) Response diversity of neurons in ventral cochlear nucleus of the kangaroo rat to low frequency tones. J. Neurophysiol. 33: 351-364.
- Moushegian, G., Rupert, A.L. and Stillman, R.D. (1973) Scalp-recorded early responses in man to frequencies in the speech range. Electroenceph. clin. Neurophysiol. 35: 665-667.
- Moushegian, G., Rupert, A.L. and Stillman, R. (1978) Evaluation of frequency following potentials in Man: Masking and clinical studies. Electroenceph. clin. Neurophysiol. 45: 711-718.
- Moushegian, G., Rupert, A.L., and Whitcomb, M.A. (1972) Processing of auditory information by medial superior-olivary neurons. In Foundations of Modern Auditory Theory, Vol. 1, J. Tobias (ed.). New York: Academic Press, pp. 263-299.
- Mushnikov, V.N. and Chistovich, L.A. (1969) Formant masking in spectra of steady-state synthetic vowels. ARSO (Kiev-Kanov) 4: 223-228.
- Narins, P.M. and Evans, E.F. (1980) Spectral resolution of comb-filtered noise by cochlear fibers in the cat: Preliminary results of comparison with their rate-level functions. In Psychological, Physiological, and Behavioural Studies in Hearing, G. van den Brink and F.A. Bilsen (eds.). Delft: Delft University Press, pp. 320-325.
- Nierop, D.J.P.J. van, Pols. L.C.W., and Plomp, R. (1973) Frequency analysis of Dutch vowels from 25 female speakers. Acustica 29: 110-118.

- Ohm, G.S. (1843) Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. Ann D. Phys. 59: 497-565.
- Papcun, G. (1980) How Do Different Speakers Say the Same Vowels?: Discrimination Analyses of Four Imitation Dialects. UCLA Working Papers in Phonetics 48.
- Patterson, R.D. (1973) The effects of relative phase and the number of components on residue pitch. J. Acoust. Soc. Am. 53: 1565-1572.
- Patterson, R.D. and Wightman, F. (1976) Residue pitch as a function of component spacing. J. Acoust. Soc. Am. 59: 1450-1459.
- Peterson, G.E. and Barney, H.L. (1952) Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24: 175-184.
- Pfeiffer, R.R. and Kim, D.O. (1975) Cochlear nerve fiber responses: Distribution along the cochlear partition. J. Acoust. Soc. Am. 58: 867-869.
- Pick, G.F. (1977) Comment on "Critical bandwidth at high intensities" by B. Scharf and C.H. Meiselman, In Psychophysics and Physiology of Hearing, E.F. Evans and J.P. Wilson (eds.). London: Academic Press, pp. 233-234.
- Pickett, J.M. (1980) The Sounds of Speech Communication. Baltimore: University Park Press.
- Plomp, R. (1964) The ear as a frequency analyzer. I. J. Acoust. Soc. Am. 36: 1628-1638.
- Plomp, R. (1965) Detectibility threshold for combination tones. J. Acoust. Soc. Am. 37: 1110-1123.
- Plomp, R. (1967) Pitch of complex tones. J. Acoust. Soc. Am. 41: 1526-1533.
- Plomp, R. (1970) Timbre as a multidimensional attribute of complex tones. In Frequency Analysis and Periodicity Detection in Hearing, R. Plomp and G. Smoorenburg (eds.). Leiden: Sijthoff, pp. 397-411.
- Plomp, R. (1975) Auditory psychophysics. Ann. Rev. Psych. 26: 207-232.
- Plomp, R. (1976) Aspects of Tone Sensation. London: Academic Press.
- Plomp, R. and Mimpen, A.M. (1968) The ear as a frequency analyzer. II. J. Acoust. Soc. Am. 43: 764-767.
- Plomp, R., Pols, L.C.W., and Geer J.P. van de (1967) Dimensional analysis of vowel spectra. J. Acoust. Soc. Am. 41: 707-712.

- Pollack, I. and Pickett, J.M. (1958) Masking of speech by noise at high sound levels. J. Acoust. Soc. Am. 30: 127-130.
- Pols, L.C.W. (1970) Perceptual space of vowel-like sounds and its correlation with frequency spectrum. In Frequency Analysis and Periodicity Detection in Hearing, R. Plomp and G. Smoorenburg (eds.). Leiden: Sijthoff, pp. 463-470.
- Pols, L.C.W. (1971) Real-time recognition of spoken words. IEEE Trans. Comput. C-20: 972-978.
- Pols, L.C.W., Kamp, L.J. Th. van der, and Plomp, R. (1969) Perceptual and physical space of vowel sounds. J. Acoust. Soc. Am. 46: 458-467.
- Pols, L.C.W., Tromp, H.R.C., and Plomp, R. (1973) Frequency analysis of Dutch vowels from 50 male speakers. J. Acoust. Soc. Am. 53: 1093-1101.
- Ribaupierre, F. de, Goldstein, M.H. Jr., and Yeni-Komshian, G. (1972) Cortical coding of repetitive acoustic pulses. Brain Research 48: 205-225.
- Ritsma, R.J. (1962) Existence region of the residue. I. J. Acoust. Soc. 34: 1224-1229.
- Ritsma, R.J. (1963) Existence region of the residue. II. J. Acoust. Soc. Am. 35: 1241-1245.
- Ritsma, R.J. (1967) Frequencies dominant in the perception of pitch of complex sounds. J. Acoust. Soc. Am. 42: 191-198.
- Ritsma, R.J. and Bilsen, F.A. (1970) Spectral regions dominant in the perception of repetition pitch. Acustica 23: 334-339.
- Ritsma, R.J. and Engel, F. (1964) Pitch of frequency-modulated signals. J. Acoust. Soc. Am. 36: 1637-1644.
- Ritsma, R. J. and Hoekstra, A. (1974) Frequency selectivity and the tonal residue. In Facts and Models in Hearing, E. Zwicker and E. Terhardt (eds). Berlin: Springer Verlag, pp. 156-163.
- Rose, J.E. (1970) Discharges of single fibers in the mammalian auditory nerve. In Frequency Analysis and Periodicity Detection in Hearing, R. Plomp and G. Smoorenburg, G. (eds.). Leiden: Sijthoff, pp. 176-188.
- Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E. (1967) Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol. 30: 769-793.

- Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E. (1969) Some possible neural correlates of combination tones. J. Neurophysiol. 32: 402-423.
- Rose, J.E., Gross, N.B., and Geisler, C.D. (1966) Some neural mechanisms in the inferior colliculus of the cat which may be relevant to the localization of a sound source. J. Neurophysiol. 29: 288-314.
- Rose, J.E., Hind, J.E., Anderson, D.J., and Brugge, J.F. (1971) Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. J. Neurophysiol. 34: 685-699.
- Rose, J.E., Kitzes, L., Gibson, M. and Hind, J.E. (1974) Observations on phase-sensitive neurons of anteroventral cochlear nucleus of the cat: Nonlinearity of cochlear output. J. Neurophysiol. 37: 218-253.
- Rosenberg, A.E. (1965) Effect of masking on the pitch of periodic pulses. J. Acoust. Soc. Am. 38: 747-758.
- Rosenberg, A.E. (1966) Pitch discrimination of jittered pulse trains. J. Acoust. Soc. Am. 39: 920-928.
- Rouiller, E., Ribaupierre, Y. de, and Ribaupierre, F. (1979) Phase-locked responses to low-frequency tones in the medial geniculate body. Hearing Research 1: 213-226.
- Rupert, A.L., Caspary, D.M., and Moushegian, G. (1977) Response characteristics of cochlear nucleus neurons to vowel sounds. Ann. Otol. 86: 37-48.
- Rutherford, W. (1886) A new theory of hearing. J. Anat. Physiol. 21: 166-168.
- Sachs, M.B. and Abbas, P.J. (1976) Phenomenological model for two-tone suppression. J. Acoust. Soc. Am. 60: 1157-1163.
- Sachs, M.B. and Kiang, N.Y-S. (1968) Two-tone inhibition in auditory nerve fibers. J. Acoust. Soc. Am. 43: 1120-1128.
- Sachs, M.B. and Young, E.D. (1979) Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. J. Acoust. Soc. Am. 66: 470-479.
- Sachs, M.B. and Young, E.D. (1980) Effects of nonlinearities on speech encoding in the auditory nerve. J. Acoust. Soc. Am. 68: 858-875.
- Sachs, R.M. and Zurek, P.M. (1979) Contralateral-probe measurements of auditory vowel spectra. J. Acoust. Soc. Am. 65: S55(A).

- Scharf, B., Florentine, M. and Meiselman, C.H. (1976) Critical band in auditory lateralization. Sens. Process. 1: 109-126.
- Scherg, M. and Brinkman, R.D. (1979) Least-square-fit technique applied to the frequency following potential: A method to determine components, latencies, and amplitudes. In Models of the Auditory System and Related Signal Processing Techniques, M. Hoke and E. de Boer (eds.). Scand. Audiol. Suppl. 9: 197-203.
- Schouten, J.F. (1938) The perception of subjective tones. Proc. K. Ned. Akad. Wet. 41: 1086-1093.
- Schouten, J.F. (1940a) The residue, a new component in subjective sound analysis. Proc. K. Ned. Akad. Wet. 43: 356-365.
- Schouten, J.F. (1940b) The residue and the mechanism of hearing. Proc. K. Ned. Akad. Wet. 43: 991-999.
- Schouten, J.F. (1940c) The perception of pitch. Phillips Tech. Rev. 5: 286-294.
- Schouten, J.F., Ritsma, R.F. and Cardozo, B.L. (1962) Pitch of the residue. J. Acoust. Soc. Am. 34: 1418-1422.
- Searle, C. Jacobson, C., and Rayment, S. (1979) Stop consonant discrimination based on human audition. J. Acoust. Soc. Am. 65: 799-809.
- Seebeck, A. (1841) Beobachtung uber einige Bedingungen der Entstehung von Tonen. Ann. Phys. Chem. 53: 417-436.
- Seebeck, A. (1844) Uber die Definition des Tones. Ann. Phys. Chem. 63: 353-368.
- Siebert, W.M. (1970) Frequency discrimination in the auditory system: Place or periodicity mechanisms? Proc. IEEE 58:723-730.
- Small, A.M. Jr. and Campbell, R.A. (1961) Masking of pulsed tones by bands of noise. J. Acoust. Soc. Am. 33: 1570-1576.
- Smith, J.C. (1976) Neural Periodicity and the Perception of Low Pitch: Physiological Correlates of Complex Sound Perception. Doctoral Dissertation, University of California, Los Angeles.
- Smith, J.C., Marsh, J.T. and Brown, W.S. (1975) Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. Electroenceph. clin. Neurophysiol. 39: 465-472.
- Smith, J.C., Marsh, J.T., Greenberg, S. and Brown, W.S. (1978) Human auditory frequency-following responses to a missing fundamental. Science 201: 639-641.

- Smooenburg, G.F. (1970) Pitch perception of two-frequency stimuli. J. Acoust. Soc. Am. 48: 924-942.
- Smooenburg, G.F., Gibson, M., Kitzes, L., Rose, J. and Hind, J.E. (1976) Correlates of combination tones observed in the response of neurons in the anteroventral cochlear nucleus of the cat. J. Acoust. Soc. Am. 59: 945-962.
- Sohmer, H. and Pratt, H. (1976) Recording of the cochlear microphonic potential with surface electrodes. Electroenceph. clin. Neurophysiol. 40: 253-260.
- Sohmer, H. and Pratt, H. (1977) Identification and separation of acoustic frequency following responses (FFRs) in man. Electroenceph. clin. Neurophysiol. 42: 493-500.
- Sohmer, H., Pratt, H. and Kinarti, R. (1977) Sources of frequency following responses (FFR) in Man. Electroenceph. clin. Neurophysiol. 42: 656-664.
- Starr, A. and Hellerstein, D. (1971) Distribution of frequency following responses in cat cochlear nucleus to sinusoidal acoustic signals. Brain Research 33: 367-377.
- Stillman, R.D., Crow, G., and Moushegian, G. (1978) Components of the frequency-following potential in man. Electroenceph. clin. Neurophysiol. 44: 438-446.
- Stotler, W.A. (1953) An experimental study of the cells and connections of the superior olivary complex of the cat. J. Comp. Neurol. 98: 401-431.
- Tasaki, I. (1954) Nerve impulses in individual auditory nerve fibers of the guinea pig. J. Neurophysiol. 17: 97-122.
- Teas, D.C., Eldredge, D.H., and Davis, H. (1962) Cochlear responses to acoustic transients: An interpretation of whole-nerve actions potentials. J. Acoust. Soc. Am. 34: 1438-1459.
- Terhardt, E. (1970) Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch. In Frequency Analysis and Periodicity Detection in Hearing, R. Plomp and G. Smooenburg (eds.). Leiden: Sijthoff, pp.278-287.
- Terhardt, E. (1974) Pitch, consonance and harmony. J. Acoust. Soc. Am. 55: 1061-1069.
- Terhardt, E. (1975) Influence of intensity of the pitch of complex tones. Acustica 34: 344-348.
- Thurlow, W.R. and Small, A.M. Jr. (1955) Pitch perception for certain periodic auditory stimuli. J. Acoust. Soc. Am. 27: 132-137.

- Tsuchitani, C. and Boudreau, J. (1964) Wave activity in the superior olivary complex of the cat. J. Neurophysiol. 27: 814-827.
- Voight, H.F., Sachs, M.B., and Young, E.D. (1980) Effects of masking noise on auditory-nerve coding of vowels. J. Acoust. Soc. Am. 67: S76(A).
- Watanabe, T. and Sakai, H. (1978) Responses of the cat's collicular auditory neuron to human speech. J. Acoust. Soc. Am. 64: 333-337(L).
- Wegel, R.L. and Lane, C.E. (1924) The auditory masking of one pure tone by another and its possible relation to the dynamics of the inner ear. Phys. Rev. 23: 266-285.
- Weinstein, C., McCandless, S., Mondschein, L., and Zue, V. (1975) A system for acoustic-phonemic analysis of continuous speech. IEEE Trans. ASSP 23: 54-67.
- Weiss, T.F., Mulroy, M.J., Turner, R.G., and Pike, C.L. (1976) Tuning of single fibers in the cochlear nerve of the alligator lizard: Relation to receptor morphology. Brain Research 115: 71-90.
- Wernick, J. and Starr, A. (1968) Binaural interaction in the superior olivary complex of the cat: An analysis of field potentials evoked by binaural beat stimuli. J. Neurophysiol. 31: 428-441.
- Wever, E.G. (1949) Theory of Hearing. New York: Wiley.
- Wever, E.G. and Bray, C.W. (1930a) Action currents in the auditory nerve in response to acoustical stimulation. Proc. Nat. Acad. Sci. 16: 344-350.
- Wever, E.G. and Bray, C.W. (1930b) The nature of the acoustic response: The relation between sound frequency and frequency of impulses in the auditory nerve. J. Exp. Psych. 13: 373-387.
- Whitfield, I. (1970a) Central nervous processing in relation to spatio-temporal discrimination of auditory patterns. In Frequency Analysis and Periodicity Detection in Hearing, R. Plomp and G. Smoorenburg, (eds.). Leiden: Sijthoff, pp. 136-147.
- Whitfield, I. (1970b) Neural integration and pitch perception. In Excitatory Synaptic Mechanisms, P. Andersen and P. Janssen (eds.). Oslo: Universitat Sovlaget, pp. 277-285.
- Wier, C.C., Jesteadt, W., and Green, D.M. (1977) Frequency discrimination as a function of frequency and sensation level. J. Acoust. Soc. Am. 61: 178-184.

- Wightman, F. (1973a) Pitch and stimulus fine structure. J. Acoust. Soc. Am. 54: 397-406.
- Wightman, F. (1973b) The pattern transformation model of pitch. J. Acoust. Soc. Am. 54: 407-416.
- Wilson, J.P. (1974) Psychoacoustical and neurophysiological aspects of auditory pattern recognition. In The Neurosciences: Third Study Program, F.O. Schmitt and F.G. Worden (eds.). Cambridge, Mass.: MIT Press, pp. 147-153.
- Worden, F. and Marsh, J.T. (1968) Frequency-following (microphonic-like) neural responses evoked by sound. Electroenceph. clin. Neurophysiol. 25: 45-52.
- Wundt, W. (1880) Grundzuge der physiologischen Psychologie, Vol. 1. Leipzig: Verlag W. Engelmann.
- Yamada, O., Kodera, K., Hink, R.F. and Suzuki, J. (1979) Cochlear distribution of frequency following response initiation: A high-pass masking noise study. Audiology 18: 381-388.
- Yamada, O., Kodera, K., Hink, R.F., and Yamane, H. (1978) Cochlear initiation site of the frequency-following response: a study of patients with sensorineural hearing loss. Audiology 17: 489-499.
- Yost, W.A. and Hill, R. (1978) Strength of the pitches associated with ripple noise. J. Acoust. Soc. Am. 64: 485-492.
- Yost, W.A. and Hill, R. (1979) Model of the pitch and pitch strength of ripple noise. J. Acoust. Soc. Am. 66: 400-410.
- Young, E.D. and Sachs, M.B. (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. J. Acoust. Soc. Am. 66: 1381-1403.
- Zwicker, E. (1975) Scaling. In Handbook of Sensory Physiology V/2, W.D. Keidel and W.D. Neff (eds.). Berlin: Springer Verlag, pp. 401-448.
- Zwicker, E., Terhardt, E. and Paulus, E. (1979) Automatic speech recognition using psychoacoustic models. J. Acoust. Soc. Am. 65: 487-498.