



This is to certify that the

dissertation entitled

PSYCHOACOUSTICAL THEORY AND EXPERIMENTS ON  
HUMAN AUDITORY ORGANIZATION OF COMPLEX SOUNDS  
AND THE CRITICAL BANDWIDTH

presented by

JIAN-YU LIN

has been accepted towards fulfillment  
of the requirements for

PH.D. degree in PHYSICS

WILLIAM M. HARTMANN

Major professor

Date JULY 11, 1996

**LIBRARY**  
**Michigan State**  
**University**

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

**MSU is An Affirmative Action/Equal Opportunity Institution**

c:\olrc\datedue.pm3-p.1

**Psychoacoustical Theory and Experiments on  
Human Auditory Organization of Complex Sounds  
and the Critical Bandwidth**

by

Jian-Yu Lin

**A DISSERTATION**

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

**DOCTOR OF PHILOSOPHY**

Department of Physics and Astronomy

1996

Dr. William M. Hartmann



## **ABSTRACT**

### **Psychoacoustical Theory and Experiments on Human Auditory Organization of Complex Sounds and the Critical Bandwidth**

by

Jian-Yu Lin

Conventional ways of determining the critical bandwidth use masking experiments. Those methods are difficult and inaccurate at low frequencies, since equal loudness contours change dramatically at low frequencies. My experiment uses roughness to determine the critical bandwidth. Data for Maximum Rough Rate (MRR) are collected and used to determine the critical bandwidth at low frequencies. In our model, the MRR is determined by two things: the temporal modulation transfer function (TMTF) and the critical bandwidth. The data could be fitted with a model in which fluctuation is summed over all auditory filters. The data require the critical bandwidth parameter to continue to decrease with decreasing frequency below 500 Hz so that it becomes considerably narrower than critical bands from bark scale, and is even narrower than Moore and Glasberg's formula (1985).

The second chapter discusses a fundamental auditory perception phenomenon - pitch perception. If a single harmonic of a complex tone is mistuned it can be heard as a

separate entity. Pitch matching experiments show that the pitch of a mistuned harmonic is an exaggeration of the frequency mistuning. The differential pitch shift (the amount of exaggeration), for positive vs. negative mistuning, is called the “split.” There are some local-interaction models (Terhardt, 1979; Hartmann and Doty, 1996) which explain the splits. Experiments here measure splits for complex tones having spectral gaps and other anomalies to see whether local-interaction models are correct. The results show that the split may be part of the segregation process instead of a simple local spectral interaction. This means the interaction is at high levels of the auditory system, which in turn tells us that the pitch formation is at high levels of the auditory system.

The last chapter combines the two phenomena in the previous chapters, and applies them to the Duifhuis effect. When a high harmonic of a slow periodic train of narrow pulses is canceled, the absent harmonic is heard. This is the Duifhuis effect. The explanation for the effect by Duifhuis is that when the frequency spacing between harmonics is considerably less than the bandwidth of relevant auditory filters, the auditory system can be said to be broadband. The study first extends the effect to a variety of different conditions. Then, Duifhuis’s explanation for the phenomenon is tested. The test is done by narrowing the bandwidth of the stimulus so that the components whose frequencies are far away from the anomalous harmonic are eliminated. Results show that the Duifhuis effect goes away by this manipulation of the spectrum of the stimuli. which means that Duifhuis’s explanation originating from a single peripheral filter is not right.

## **ACKNOWLEDGMENT**

I would like to thank my thesis advisor Professor William M. Hartmann for his insightful guidance, support, and encouragement to me. I especially thank him for his kindness during the past five years and for his patience during the writing of this thesis .

## TABLE OF CONTENTS

<b>Chapter 1. Roughness and the Critical Bandwidth at Low Frequency .....</b>	<b>1</b>
Introduction .....	2
A. The critical bandwidth concept in psychoacoustics .....	2
B. The shape of the peripheral filter, ERB and the Gamma-tone filter .....	5
C. Low-frequency critical bandwidth .....	9
I. The model of roughness .....	13
A. Roughness: Phenomenon and Model .....	13
B. Temporal Modulation Transfer Function .....	16
C. Quantifying the fluctuation factor in the roughness model .....	17
II. Experiments .....	21
A. Method .....	21
B. Results .....	23
III. Computation and Conclusion .....	24
A. Fitting the model to data .....	19
B. Conclusion and discussion .....	38
Appendix R1 .....	45
References .....	49
 <b>Chapter 2. Pitches of Mistuned Harmonics .....</b>	 <b>50</b>
Introduction .....	51
I. Experiments .....	57
A. Method .....	57
B. Experiment 1 .....	60
C. Experiment 2 .....	64
D. Experiment 3 .....	68
E. Experiment 4 .....	72
F. Experiment 5 .....	76
G. Experiment 6 .....	80
H. Experiment 7 .....	81
I. Experiment 8 .....	84
J. Experiment 9 .....	89
II. Discussion and Conclusion .....	94
References .....	97

<b>Chapter 3. Duifhuis effect - New Measurements Require a Revised Explanation .....</b>	<b>99</b>
Introduction .....	101
I. Experiments exploring the Duifhuis effect .....	103
A. Method .....	103
B. Experiment 1: The range of the Duifhuis effect .....	104
C. Experiment 2: The level effect .....	108
D. Experiment 3: The effect of harmonic phases .....	114
E. Experiment 4: The missing fundamental pitch .....	117
F. Discussion .....	120
G. Traditional Explanation .....	121
H. Challenge to the traditional explanation .....	131
II. Experiments to test the short-term Fourier analysis explanation .....	136
A. Method .....	136
B. Experiment 5: Narrowing the bandwidth of the stimulus .....	139
C. Experiment 6: Narrowing the phase-coherence region .....	144
III. Discussion and Conclusion .....	148
A. Extension of the Duifhuis effect .....	148
B. Implications of the Duifhuis effect .....	154
C. Conjecture on the mechanism of the Duifhuis effect .....	156
Appendix D1. Audiograms of the listeners .....	161
References .....	166

## LIST OF TABLES

<b>Table R1:</b> Experimental results of SB at 60 dB for Roughness experiment .....	26
<b>Table R2:</b> Experimental results of SB at 80 dB for Roughness experiment .....	27
<b>Table R3:</b> Experimental results of SD at 60 dB for Roughness experiment .....	28
<b>Table R4:</b> Experimental results of SD at 80 dB for Roughness experiment .....	29
<b>Table R5:</b> Experimental results of SJ at 60 dB for Roughness experiment .....	30
<b>Table R6:</b> Experimental results of SJ at 80 dB for Roughness experiment .....	31
<b>Table R7:</b> Averaged MRR over listeners and levels, and the predicted MRR from the model .....	37

## LIST OF FIGURES

<b>Figure R1:</b> Plots of the two critical bandwidths, the Bark scale and the ERB scale .....	6
<b>Figure R2:</b> Model for the perception of fluctuations.....	19
<b>Figure R3:</b> Results of the Roughness Experiment for SB .....	32
<b>Figure R4:</b> Results of the Roughness Experiment for SD .....	33
<b>Figure R5:</b> Results of the Roughness Experiment for SJ .....	34
<b>Figure R6:</b> The fittings of model MRR to the measured MRR .....	40
<b>Figure R7:</b> Plot of CB used to make the fittings at low frequency .....	42
<b>Figure R8:</b> Plot of the model predicted MRR .....	43
<b>Figure R9:</b> Plot of the calculation results of Sek and Moore's (1994) flattening effect of CB .....	47
<b>Figure M1:</b> Example of a split .....	54
<b>Figure M2:</b> Spectra for Experiment 1 for mistuned harmonic .....	62
<b>Figure M3:</b> Results of Experiment 1 (Zigzag Effect) for mistuned harmonic .....	63
<b>Figure M4:</b> Spectra for Experiment 2 for mistuned harmonic .....	66
<b>Figure M5:</b> Results of Experiment 2 (Integer Effect) for mistuned harmonic .....	67
<b>Figure M6:</b> Spectra for Experiment 3 for mistuned harmonic .....	70
<b>Figure M7:</b> Results of Experiment 3 (The Proximity Effect) for mistuned harmonic .....	71

## LIST OF FIGURES

<b>Figure R1:</b> Plots of the two critical bandwidths, the Bark scale and the ERB scale .....	6
<b>Figure R2:</b> Model for the perception of fluctuations.....	19
<b>Figure R3:</b> Results of the Roughness Experiment for SB .....	32
<b>Figure R4:</b> Results of the Roughness Experiment for SD .....	33
<b>Figure R5:</b> Results of the Roughness Experiment for SJ .....	34
<b>Figure R6:</b> The fittings of model MRR to the measured MRR .....	40
<b>Figure R7:</b> Plot of CB used to make the fittings at low frequency .....	42
<b>Figure R8:</b> Plot of the model predicted MRR .....	43
<b>Figure R9:</b> Plot of the calculation results of Sek and Moore's (1994) flattening effect of CB .....	47
<b>Figure M1:</b> Example of a split .....	54
<b>Figure M2:</b> Spectra for Experiment 1 for mistuned harmonic .....	62
<b>Figure M3:</b> Results of Experiment 1 (Zigzag Effect) for mistuned harmonic .....	63
<b>Figure M4:</b> Spectra for Experiment 2 for mistuned harmonic .....	66
<b>Figure M5:</b> Results of Experiment 2 (Integer Effect) for mistuned harmonic .....	67
<b>Figure M6:</b> Spectra for Experiment 3 for mistuned harmonic .....	70
<b>Figure M7:</b> Results of Experiment 3 (The Proximity Effect) for mistuned harmonic .....	71



<b>Figure M8: Spectra for Experiment 4 for mistuned harmonic .....</b>	<b>74</b>
<b>Figure M9: Results of Experiment 4 (Local Asymmetry Effect) for mistuned harmonic.....</b>	<b>75</b>
<b>Figure M10: Spectra for Experiment 5 for mistuned harmonic .....</b>	<b>78</b>
<b>Figure M11: Results of Experiment 5 (The Frequency Effect) for mistuned harmonic .....</b>	<b>79</b>
<b>Figure M12: Spectra for Experiment 6 for mistuned harmonic .....</b>	<b>82</b>
<b>Figure M13: Results for Experiment 6 (The Harmonic Number Effect) for mistuned harmonic .....</b>	<b>83</b>
<b>Figure M14: Spectra for Experiment 7 for mistuned harmonic .....</b>	<b>86</b>
<b>Figure M15: Results of Experiment 7 (Controlling the Framework) for mistuned harmonic .....</b>	<b>87</b>
<b>Figure M16: Spectra for Experiment 8 for mistuned harmonic .....</b>	<b>88</b>
<b>Figure M17: Results of Experiment 8 (Instability Check) for mistuned harmonic .....</b>	<b>91</b>
<b>Figure M18: Spectra for Experiment 9 for mistuned harmonic .....</b>	<b>92</b>
<b>Figure M19: Results of Experiment 9 (Top - Bottom effect) for mistuned harmonic .....</b>	<b>93</b>
<b>Figure D1: The Duifhuis effect .....</b>	<b>102</b>
<b>Figure D2: Results for Duifhuis-effect Experiment 1 showing matches .....</b>	<b>107</b>
<b>Figure D3: Waveforms for Duifhuis-effect Experiment 2 .....</b>	<b>110</b>
<b>Figure D4: Results of Duifhuis-effect Experiment 2 .....</b>	<b>113</b>

<b>Figure D5:</b> Illustrative waveforms for Duifhuis-effect Experiment 3 .....	116
<b>Figure D6:</b> Results of Duifhuis-effect Experiment 3 .....	119
<b>Figure D7:</b> Results of Duifhuis-effect Experiment 4 .....	122
<b>Figure D8:</b> Illustrative plots of a filter's operation .....	125
<b>Figure D9:</b> Comparison of a 10-Hz narrow pulse train with a 50-Hz narrow pulse train .....	128
<b>Figure D10:</b> Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 45-harmonic flat-spectrum complex tone with the fundamental frequency at 50 Hz. All the harmonics have cosine phase .....	129
<b>Figure D11:</b> Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 45-harmonic flat-spectrum complex tone with the fundamental frequency at 50 Hz. All the harmonics have cosine phase. The 19th harmonic is omitted .....	132
<b>Figure D12:</b> Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 950-Hz sine tone .....	133
<b>Figure D13:</b> Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 45-harmonic flat-spectrum complex tone with the fundamental frequency at 50 Hz. The harmonics of the complex tone have Schroeder phase .....	134
<b>Figure D14:</b> Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 45-harmonic flat-spectrum complex tone with the fundamental frequency at 50 Hz. The harmonics of the complex tone have Schroeder phase. The 19th harmonic is omitted .....	135

<b>Figure D15:</b> Three-dimensional plot of the output of the peripheral auditory filter bank. The input signal is a 50-Hz complex tone with harmonics 15 to 24. All the harmonics have cosine phase. The 19th harmonic is omitted .....	138
<b>Figure D16:</b> Results for Duifhuis-effect experiment 5, SB .....	141
<b>Figure D17:</b> Results for Duifhuis-effect experiment 5, SD .....	142
<b>Figure D18:</b> Results for Duifhuis-effect experiment 5, SJ .....	143
<b>Figure D19:</b> Plots of the percentages of matches in the central bin (left column) in Figure D16, Figure D17 and Figure D18 as a function of the bandwidth of the stimulus .....	146
<b>Figure D20:</b> Results for Duifhuis-effect experiment 6, SB .....	149
<b>Figure D21:</b> Results for Duifhuis-effect experiment 6, SD .....	150
<b>Figure D22:</b> Results for Duifhuis-effect experiment 6, SJ .....	151
<b>Figure D23:</b> Plots of the percentages of matches in the central bin (left column) in Figure D20, Figure D21 and Figure D22 as a function of the bandwidth of the stimulus .....	152
<b>Figure D24:</b> Illustrative plots for the mechanism of the Duifhuis effect from conjecture .....	159
<b>Figure D25:</b> Audiogram of SB .....	162
<b>Figure D26:</b> Audiogram of SD .....	163
<b>Figure D27:</b> Audiogram of SH .....	164
<b>Figure D28:</b> Audiogram of SJ .....	165

# **CHAPTER 1**

## **Roughness and the Critical Bandwidth at Low Frequency**

## Introduction

### *A. The Critical Bandwidth Concept in Psychoacoustics*

It is known that the periphery of our auditory system has the ability to analyze an incoming acoustic waveform into a spectrum. It functions as a frequency analyzer. We take advantage of this ability when we listen to and separate several simultaneous sounds which are in different frequency ranges. Since it is so important, a lot of research is done in this area. The critical bandwidth concept is a way to describe the resolution of the frequency analyzer of our auditory system. The idea is that stimuli which are separated by more than a critical bandwidth have less interference upon each other than sounds within a critical bandwidth. Zwicker and Fastl (1990) summarized five main methods of determining the critical bandwidth.

Method 1: Measure the threshold and determine the critical bandwidth. A pure tone's, say 920 Hz, threshold level is measured first. Add another tone which is close to the first tone, say, 20 Hz higher. Now the sound contains two components which have the same amplitude. Measure the level of one component when the two-component stimulus is at the hearing threshold. Since the two components fall into the same critical band, the threshold level of the two-component stimulus should be the same as the threshold level of the one-component stimulus. Thus the level of each component goes down by 3dB. More components are added until the level of each component does not change. For example, for 920 Hz, this happens after eight components are added. If more components are added, the threshold level does not change. That means those extra components are outside the critical bandwidth. Therefore, the critical bandwidth at about 1000 Hz is  $8 \times 20 \text{ Hz} = 160 \text{ Hz}$ .

Method 2: Use masking in frequency gaps to determine the critical bandwidth. Two sine tones spaced by  $df$  Hz are used as masker. A narrow band noise is located at the average frequency of the two sine tones. The threshold of the narrow band noise is measured as a function of  $df$ , the frequency difference between the two sine tones. For  $df$  less than some value, the threshold is a constant. When  $df$  is greater than the value, the threshold decreases as  $df$  increases. That means that the two components have less and less effect on the narrow band noise. They are separated by more than a critical band. The  $df$  value at which the threshold begins to fall is taken as the critical bandwidth.

Method 3 is based on the detectability of phase changes, particularly the difference between AM and FM. A waveform of the form  $A(1+m\cos\omega_m t)\cos(\omega_c t)$  is an amplitude modulated signal or AM, where  $m$  is the index of amplitude modulation. There are three spectral components in the AM spectrum, the carrier and the two side-bands. A waveform of the form  $A\cos(\omega_c t + \beta\cos\omega_m t)$  is a frequency modulated signal or FM, where  $\beta$  is the index of frequency modulation. Usually, the spectrum of FM is very complicated and has a wide band. However, if  $\beta$  is very small ( $\ll \pi/2$ ), the FM signal becomes “narrow band” FM, and the power spectrum of the FM has two side-bands just like AM. The only difference of the two spectra is that the two side-bands have different phases for AM and FM. The detection thresholds of the two modulations AM and narrow band FM, which are expressed as  $\beta_c$  and  $m_c$ , are compared at the same modulation frequency  $f_m$ . As the modulation frequency increases, the widths of the two power spectra consisting of the three components increases. For a modulation frequency  $f_m$  less than some value, the two thresholds are different, i.e.  $\beta_c \neq m_c$  or the side-bands of the two signals at their thresholds have different amplitudes. For larger  $f_m$ , the two thresholds merge to one, i.e.  $\beta_c = m_c$  or the side-bands of the two signals

at their thresholds are equal in amplitude. This means that for  $f_m$  less than some value called the critical modulation frequency (CMF), the auditory system is phase sensitive. It is assumed that the components of the sound must be in one critical band in order to get phase sensitivity. Accordingly, twice the value of the modulation frequency at which AM and FM threshold merge is taken as the critical bandwidth because that is the point where the three components could just go into one critical band. Zwicker and Fastl pointed out that this method was effective in determining critical bandwidth at low frequency.

Method 4 is based on the loudness of a constant sound pressure noise. A noise of some bandwidth is presented at a constant sound pressure level. The loudness of the noise is measured as a function of the bandwidth. If the bandwidth of the noise is less than some value, the loudness does not change with the bandwidth. If the bandwidth of the noise exceeds that value, the loudness becomes larger as the bandwidth increases, although the sound pressure level is kept constant. That may be explained as the noise excites more critical bands. Therefore, the bandwidth of the noise at which the loudness begins to increase is taken as the critical bandwidth.

Method 5 stems from binaural hearing. The just-noticeable delay between the envelope of the two tone bursts, each to one ear, is measured. The hearing system is sensitive to the envelope delay as long as the two tone bursts are of high frequency and of nearly the same frequency. The sensitivity decreases if the frequency difference between the two tone bursts becomes larger than the critical bandwidth. This characteristic is used to determine the critical bandwidth.

According to Zwicker and Fastl the critical bandwidth can be approximated as 100 Hz for center frequency below 500 Hz, and 20% of the center frequency for center fre-

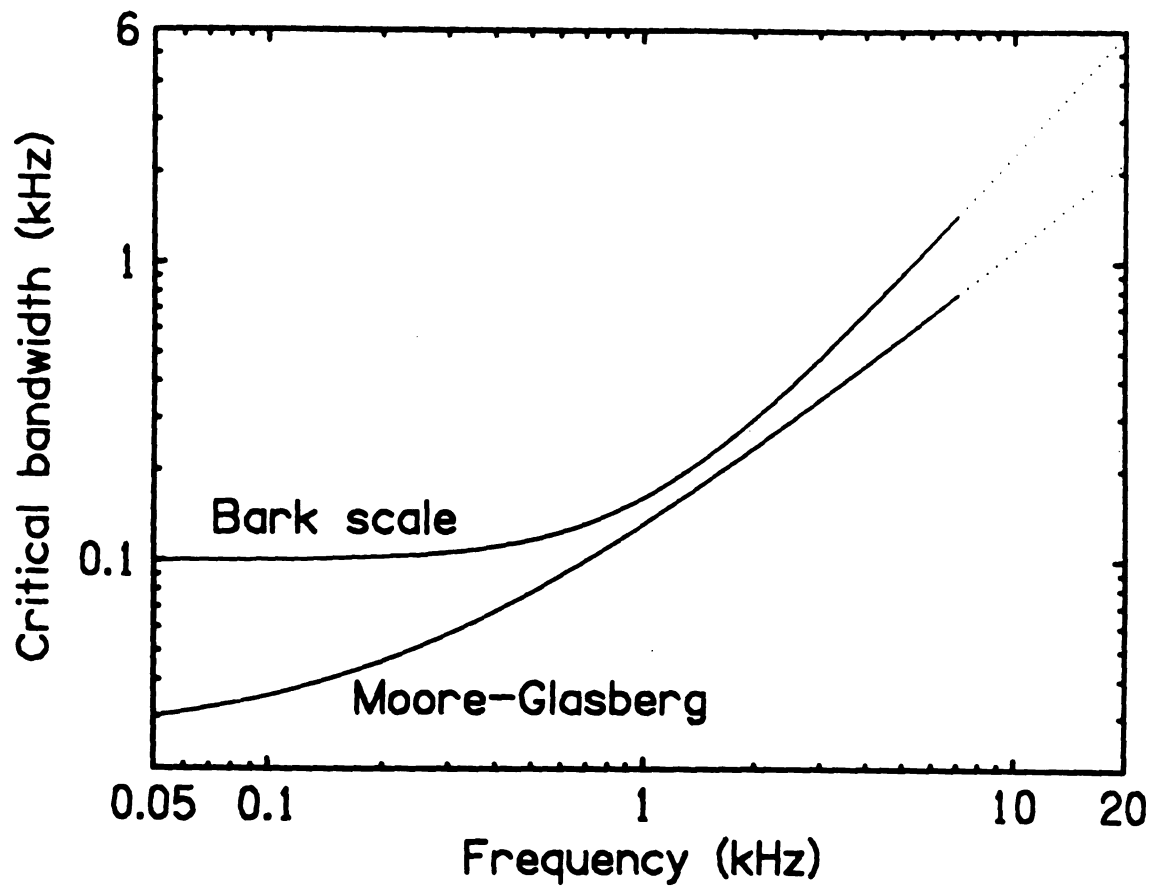
quency higher than 500 Hz. That means for frequencies higher than 500 Hz, auditory filters are of constant Q value. For the frequency range 0-16000 Hz, there are approximately 24 adjacent critical bands, which is also called the Bark scale. Moore and Glasberg (1985) disagree with the approximation. They gave a different formula for the bandwidth of peripheral filters as a function of their center frequencies. The main difference is at low frequency. Figure R1 gives the two different critical bands.

### *B. The Shape of the Peripheral Filters, ERB and the Gammatone Filter*

In the previous discussion, there is an assumption about the shape of the peripheral filter, i.e. the shape of the filter is rectangular and has infinitely sharp edges. This, of course, is not true. We know realistic filters cannot have infinitely sharp edges, they all have some shape, and the shape of filters is a very important factor in describing frequency analyzing characteristic of our auditory system. There were several different experiments done to determine the filter shape. The basic assumption is that at the threshold the power of the signal is proportional to the portion of masker's power which goes into a single filter.

The first method is Houtgast's (1972, 1974) rippled noise method. It makes use of the masking of a sine tone in a rippled noise. When a white noise is comb filtered, its spectrum is rippled. The power density of this rippled noise (rippled spectrum noise) has symmetric peaks and valleys. In fact, the pattern is a DC plus or minus a cosine wave. If a constant phase shift  $\pi/2$  is introduced to every frequency of the noise, the power spectrum becomes a DC plus or minus a sine wave. When the delay of the comb filter is small, the peaks are sparse. When the delay of the comb filter is large, the peaks are dense. If a pure tone is put in the valley, it is masked by the neighboring two





**Figure R1:** Plots of the two critical bandwidths, the Bark scale (Zwicker) and the ERB scale (Moore and Glasberg), as a function of center frequency.

peaks. The denser the peaks, the more masking is obtained, since more noise goes into the filter centered at the signal. By changing the delay of the comb filter, the amount of masking is controlled. The threshold of the signal gives information about the masking and thus about shape of the filter. To calculate the shape of the filter, the transfer function is expanded in trigonometric functions (a Fourier series expansion). Since the power spectrum of the input noise is also a trigonometric function, by the orthogonality property of the trigonometric functions, the coefficients of the Fourier series can be determined by thresholds and the spectra of the noise. Thus the shape of the filter is determined.

The second method is to use notched noise (Patterson 1976). The spectrum of notched noise is constant except at the notch where it is zero. The probe tone is placed in the middle of the notch. The width of the notch is controlled. The threshold of the probe tone is measured as a function of the width of the notch. By the assumption that at the threshold the power of the signal is proportional to the portion of masker's power which goes into a single filter, the amount of masking noise leaked into the filter around the probe is determined. Since the spectrum of the noise is flat, it is not difficult to get the relationship that  $|H(f)|^2$  is proportional to the negative derivative of the threshold with respect to the width of the notch.

To parameterize the shape of the filter, Patterson *et al.* (1982) suggested what they called rounded-exponential or Roex filter. A typical two-parameter *Roex*( $p, r$ ) filter is expressed as:

$$W(g) = (1 - r)(1 + pg)e^{-pg} + r ,$$

where  $g = |f - f_c|/f_c$ . The relationship between the transfer function and  $W(g)$  is  $|H(f)| = W(|f - f_c|/f_c)$ . As the exponential parameter decreases, the function broadens. Parameter  $r$  is intended to approximate the shallow tail section of the filter shape outside the passband.

Now that the shape of the filter is not a rectangle, the critical bandwidth needs to be redefined. One common definition is the width of the filter at which the response has fallen by a factor of two, i.e. by 3dB. An alternative definition is the equivalent rectangular bandwidth, or ERB (Moore, 1982). The ERB of a given filter is equal to the bandwidth of a perfect rectangular filter which has a transmission in its passband equal to the maximum transmission of the specified filter and transmits the same power of white noise as the specified filter.

A realistic filter cannot be perfectly rectangular, because the filter must be causal, i.e. it must have a causal impulse response. Patterson *et al.* (1987) designed a causal impulse response for what they called the gammatone filter to model the peripheral auditory filters. The impulse response is given as:

$$gt(t) \propto t^{n-1} \exp(-2\pi bt) \cos(2\pi f_c t + \phi) \quad (t \geq 0) \quad (1)$$

The corresponding transfer function is:

$$GT(f) \propto [1 + j(f - f_c)/b]^{-n} + [1 + j(f + f_c)/b]^{-n} \quad (-\infty < f < \infty) \quad (2)$$

There are two parameters in the gammatone filter,  $n$  and  $b$ . It is not difficult to see that  $n$  is related to the slope of the filter and  $b$  is proportional to the critical bandwidth of

the filter. The second term in Equation (2) can be ignored when  $f_c/b$  is sufficiently large - which is always the case when modeling the human auditory filter.

According to Patterson *et al.*, the physiological impulse response of a particular auditory nerve fiber is somewhat like the gammatone filter impulse response. If a psychoacoustic tuning curve is obtained at very low level, i.e. just a few fibers are concerned, the two kinds of impulse response may be similar. As Patterson pointed out, the gammatone filter has one notable disadvantage: the amplitude of the transfer function is symmetric. However, the peripheral filters are asymmetric. The comparison of Roex filter and the gammatone filter is also given in Patterson *et al.* (1987). They could make very good fit of the two filters.

Moore *et al.* (1990) used the notched noise method to measure the filter's shape at low frequencies. The improvement of their experiment compared with Patterson's was that they measured the asymmetry of the filter. They did this by putting the probe tone off the center of the notch. They fitted their data by Roex filters. They used different  $p$ 's for the lower and upper side so that the filter is asymmetric. The experiment was done at frequencies: 100 Hz, 200 Hz, 400 Hz, and 800 Hz. The ERBs averaged over all the listeners were listed. The normalized ERB for the center frequency of 100 Hz was 0.36 and for 200 Hz was 0.42, which gives the ERBs at those center frequencies as 36 Hz and 84 Hz respectively. The results from Moore's experiment shows that the critical bandwidths at low frequencies are not constant. This conflicts with Zwicker's theory that the critical bandwidth below 500 Hz is a constant of about 100 Hz, which is the critical bandwidth determined by critical modulation frequency (CMF) method.

### *C. Low-Frequency Critical Bandwidth*

As mentioned before, the CMF can be used to estimate the CB (critical bandwidth). The idea is: when the modulation frequency is low, the thresholds for detecting FM and AM are different. As the modulation frequency increases, the two thresholds merge at some point. That means the auditory system can not distinguish the two modulations any more. Therefore, it is phase insensitive. Schorer (1986) suggested that the CMF corresponded to half the value of the CB. Zwicker and Fastl (1990) mentioned that the CMF was a good way to determine the critical bandwidth at low frequencies. However, as shown by Figure R1, the result from Moore *et al.* (1990) differs a lot from Zwicker's. Later, Sek and Moore (1994) disagreed with the idea of using the CMF to determine the CB at low frequency. Their argument comes from Hartmann and Hnath's theory (1982) that side-band detection task followed a "No Summation Model".

Hartmann and Hnath's work was done with mixed modulation or MM. Mixed modulation is the case in which both AM and FM exist. The advantage of using mixed modulation is that the amplitudes of the lower component and the upper component can be changed independently. This is because in AM the sum of the lower component phasor and the upper component phasor is in phase with the carrier, whereas in FM the sum of the lower component phasor and the upper component phasor is not in phase with the carrier. The variability of the relative level of the lower component and the upper component was used by Hartmann and Hnath (1982) to test the theories of modulation detection.

Hartmann and Hnath suggested and confirmed the No Summation Model for detection of MM. The idea of the model is that modulation is detected by the detection of the most detectable side-band and the other side-band has no effect. For example, suppose at some carrier frequency, the thresholds for detecting lower component and the higher

component are 19 dB and 40 dB separately. According to the model, one would detect the modulation if either the lower side-band is not lower than 19 dB or the higher side-band is not lower than 40 dB. In other words, the threshold for detecting MM is the threshold of detecting the lower side-band if the upper side-band is less than 40 dB; the threshold for detecting MM is the threshold of detecting the upper side-band if the lower side-band is less than 19 dB.

According to the results of Hartmann and Hnath (1982) the auditory system first detects the lower side-band of the modulated signal, which is consistent with masking theory. Now the question arises at low frequency. At low frequencies, the absolute threshold changes very rapidly with frequency. Therefore, the absolute threshold is much higher for the lower component than for the upper component if the carrier frequency is low. Then, its possible that at low frequency the upper side-band is first detected, since its absolute threshold is lower. Sek and Moore (1994) did two experiments to check whether this was true.

The first experiment was to compare the masking thresholds at lower side and upper side of a sine tone masker with the detection thresholds of an AM sine tone. The masker was same as the carrier. The sine tone signal to be detected for the masking thresholds is separated from the masker by the frequency which was the modulation frequency in modulation detection task. What they found was that at low frequencies 125 Hz, 160 Hz, the masking thresholds on the lower side of the masker were much higher (10-40 dB) than the thresholds of modulation detection, whereas the masking thresholds on the upper side of the masker were the same as the thresholds of the modulation detection. That is obviously evidence that at low frequency one first detects the upper side-band of a modulated tone.

The second experiment was a mixed modulation experiment. In mixed modulation, the threshold is determined by three parameters. By manipulating the parameters, they could show that at 125 Hz, the detection threshold was determined by the upper-band side; and at 500 Hz, the detection threshold was determined by the lower-band side.

This causes a problem in using the CMF to measure the CB. Since at high frequency the auditory system uses the lower side-band and at low frequency the auditory system uses the upper side-band, there must be a transition region. Sek and Moore (1994) concluded that the transition region is at 200 - 250Hz. That means that below 200 Hz the CB would be overestimated if the CMF method is used provided that the CB continues to go down at low frequencies. The ERB formula of Moore and Glasberg (1985) continues to go down below 500 Hz. Therefore Sek and Moore (1994) claimed that difference between the ERB and Zwicker's CB at low frequencies came from the overestimation by Zwicker's CMF method.

There may also be some problem with Moore and Glasberg's formula. First, we did a calculation which showed that the overestimation is much smaller than the difference between the Zwicker's CB and the ERB (See Appendix R1). Second, as we mentioned before, Moore and Glasberg used notched noise to determine their ERB. We know that the hearing threshold changes dramatically at low frequency. Therefore, it is difficult to determine the amount of noise leaked into a peripheral filter, which directly affects the results. It seems that both the CMF method and the notched noise method are affected by the hearing threshold at low frequencies. A better method would depend less upon the hearing threshold at low frequency. To use a maximizing task instead of a detecting task would give us more signal, and thus depend less on the hearing threshold. That was our motivation to use maximum roughness (a sensation from rapid beats) to determine the critical bandwidth at low frequencies. In fact, Plomp and Levelt (1965)

suggested that the maximum roughness of two simple tones occurs when the frequency difference is a quarter of critical bandwidth, which is a connection between the critical bandwidth and the maximum roughness. However, it seems difficult to use this relation to cover the whole frequency range, as will be noted in the conclusion. Thus we determined to do further investigations on this issue.

## **I. The Model of Roughness**

### *A. Roughness: Phenomenon and Model*

If the amplitude of a tone, say 1000 Hz, is modulated with a low frequency, say 5 Hz, the amplitude fluctuations are easily recognized as corresponding fluctuations of loudness. This is the first stage of perception of an amplitude modulated tone as the modulation frequency changes from low to high frequency. If the modulation frequency is changed to 10 Hz or higher, another sensation, roughness, starts to increase. At this stage, it is hard to follow the fluctuations of the loudness; rather, some unpleasant beats are perceived. This sensation continues to increase until the modulation frequency reaches 70 Hz. If the modulation frequency increases further, the tone becomes smoother, and the separated components become clearer and clearer. Finally, the roughness is completely gone, and the three spectral components are separately heard.

Zwicker and Fastl (1990) explored the dependence of the roughness on the modulation frequency as a function of the carrier frequency. They used 100% AM tones. Their results showed that, for carriers lower than 1000 Hz, maximum roughness is a function of carrier frequency: A 125 Hz carrier corresponds to about 30 Hz; a 250 Hz carrier corresponds to about 45 Hz; a 500 Hz carrier corresponds to about 55 Hz. For carriers



higher than 1000 Hz, the peaks remain unchanged, i.e. the maximum roughness occurs at about 70 Hz independent of the carrier frequency.

Terhardt (1974) explored the parameters which affected the roughness. The first one was the modulation depth of a AM tone. The method he used in his experiment was as follows: Pairs of AM tones were presented. For each pair of tones, there was a pause in between. The subjects were asked to judge whether the second tone was more or less than half (or twice) rough as the first tone. From the response data, he got the point where more and less had equal probability. That point corresponded to half roughness (or twice roughness). In this way, he got the formula: roughness  $\propto m^2$ , where  $m$  is the modulation index.

The second parameter he studied was sound pressure level. In the experiment, he presented AM tone pairs. The first tone was a standard tone with  $m=1$ . The second tone was 20 or 40 dB higher, with its  $m$  randomly chosen between 0.2 and 1. The subjects were then asked to judge whether the second tone produced more or less roughness. The results showed that high level tones were rougher. However, his results depended upon the order of stimuli presented, and thus could be unreliable. Zwicker and Fastl (1990) mentioned that roughness had only weak dependence on sound level.

Terhardt investigated threshold of roughness as a function of modulation frequency. His experiments showed that high modulation frequency had higher threshold. He compared roughness of FM with that of AM. Results showed that big differences only occurred when the modulation frequency was small compared with the Bark scale, which agreed with Zwicker and Fastl's (1990) CMF method for determining the CB. He also compared the roughness of AM tones with that of beating tones composed of

only two components. He found that roughness was determined by the fundamental component of the envelope, as described Hartmann (1996).

Zwicker and Fastl (1990) gave a model for the roughness phenomenon. They used two factors in their model: the temporal resolution of intensity fluctuation factor and the fluctuation speeding factor, fluctuation factor and speeding factor for short. Roughness was the product of the value of the two factors. The main idea was that at low modulation frequencies, our auditory system could follow the change of intensity very well, i.e. the value of the fluctuation factor was large. However, at low modulation frequencies the change was too slow to be rough, so the value of the speeding factor was small. The value of the product of the two factors was small. For very high modulation frequencies, the value of the speeding factor was large, but the value of the fluctuation factor was small resulting in a small roughness also. A peak of the product of the two factors occurred somewhere between high frequency and low frequency, which corresponded to the modulation frequency for maximum roughness.

It is understandable that the fluctuation factor of an AM sine tone becomes small when the modulation frequency is high because the three spectral components are resolved. They are in different peripheral auditory filters. However, if the auditory filter is the only parameter that can make the AM tone smoother, maximum roughness should occur at higher modulation frequencies for higher carrier frequencies rather than being limited by a 70-Hz modulation frequency no matter how high the carrier frequency is. Therefore, there must be some additional effective parameter. That is the pure temporal constraint, the TMTF parameter. In other words, there are actually three things which affect the maximum roughness, 1) the speeding factor; 2) the CB and 3) the TMTF, which determines the fluctuation factor.

### *B. Temporal Modulation Transfer Function*

The Temporal Modulation Transfer Function was introduced by Viemeister (1979) to describe the ability of the auditory system to resolve pure temporal changes. Viemeister assumed that the Temporal Modulation Transfer Function (TMTF) was an attenuation characteristic of the auditory system. He also assumed linearity of the auditory system in perceiving temporal changes.

The TMTF can be measured using a sine-tone amplitude-modulated sine tone. The threshold of the modulation depth is measured as a function of the modulation frequency. This threshold was supposed to be a measure of the TMTF. However, there is a problem with the stimulus. As the modulation frequency becomes higher, the side-bands of the signal are resolved. This spectral resolution puts a limit to the validity of the stimulus, since the side-bands may be heard before the fluctuation is heard. Viemeister (1979) improved the stimulus by using a wideband noise as a carrier and thus eliminated the possible spectral cues, since the power spectrum of a sinusoidal AM noise is flat.

Viemeister pointed out that TMTFs displayed a lowpass characteristic. That is reasonable, since we expect the auditory system to resolve a temporal change better when the change is slower. He plotted the AM threshold, which is displayed by  $-\log(m)$ , as a function of the modulation frequency  $f_m$ . He expected to see the attenuation characteristic of the TMTFs from the plot. The results showed that the modulation threshold was constant up to about 10 Hz, which is consistent with the fact that the auditory system follows the change of intensity to about 10 Hz as mentioned before. The sensitivity to modulation was reduced by 3 dB at approximately 50 Hz. From 50 Hz to 800 Hz the sensitivity decreased at a rate of 3-4 dB/oct. Viemeister said that at high modulation

frequency ( $f_m > 1000$  Hz), the cues for temporal changes, such as “roughness” or “buzziness” could not be detected.

Viemeister also investigated other properties of the TMTFs. One of the investigations varied the intensity of the stimulus. The results showed that for different intensities the thresholds curves were similar. So he concluded that the characteristics of the temporal processing were intensity independent at least over the range of the levels he investigated, 20 dB SPL to 50 dB SPL.

The TMTF can be modeled as a lowpass filter. A simple lowpass filter can be expressed by its amplitude response:

$$|H(\omega_m)| = \frac{1}{\sqrt{1 + \omega_m^2 \tau^2}}, \quad (3)$$

where  $\omega_m$  is the modulation frequency, and  $\tau$  is the time constant of the lowpass filter. According to Viemeister, the time constant is in the range of 2 to 3 ms.

### *C. Quantifying the Fluctuation Factor in the Roughness Model*

The model given by Zwicker and Fastl must be quantified to compare with the data. There are two factors, the fluctuation factor and the speeding factor. The speeding factor is a sensation that does not have much to do with CB or TMTF, because there is no direct relation. Furthermore, it is difficult to use an experiment to measure the speeding factor. Therefore, the speeding factor is completely unknown. We want to avoid this factor in our calculation, because the CB is an unknown that we are interested in, and we want to have as few unknowns as possible.

Since we want to avoid the speeding factor, we define a new term - the maximum rough rate (MRR) as the rate of the most rapid beats without a dramatic decrease in fluctuations of the beats. The “most rapid” and the “without a dramatic decrease” make the two factors, fluctuation and speeding, large. On the other hand, from this definition we can drop the speeding factor in our calculation. If we plot the fluctuation as a function of modulation rate (beat rate), a dramatic decrease would correspond to the maximum slope (absolute value) of the curve, since the curve is monotonically decreasing. Therefore, we choose the modulation rate of maximum slope as MRR. Thus, the speeding factor is avoided, only one factor, the fluctuation factor, is needed in our calculation.

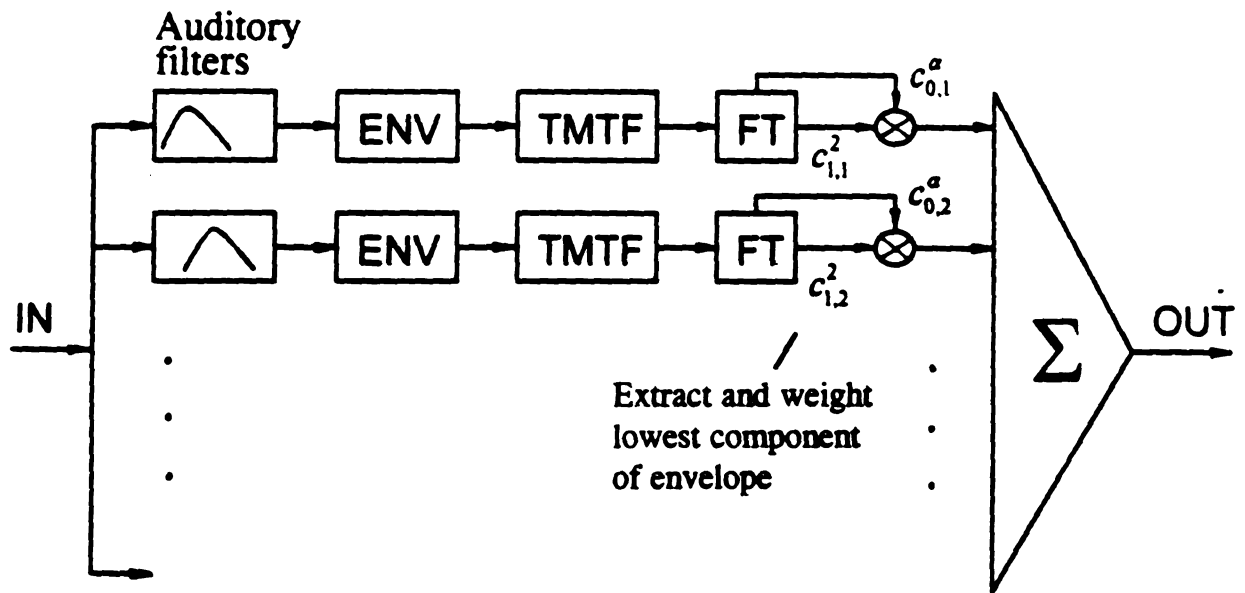
The fluctuation factor describes the ability to follow the change of the intensity, or it describes the perceived sound intensity fluctuations. The common ways of producing beats are by AM and Two-Tone Beats (Beats for short). Our model for calculating the perceived fluctuation for AM and Beats is given in Figure R2.

The incoming sound waveform first goes through the filters of the peripheral filter bank. The output of a filter, which forms a channel of signal, has three spectral components for AM and two spectral components for Beats with a certain ratio of amplitudes. The waveforms of the output have certain envelopes. The envelopes are:

$$E_a(t) = \sqrt{[a_c + (a_l + a_h) \cos \Delta\omega t]^2 + [(a_l - a_h) \sin \Delta\omega t]^2} \quad (4)$$

for AM, and

$$E_b(t) = \sqrt{a_1^2 + a_2^2 + 2a_1a_2 \cos(\Delta\omega t)} \quad (5)$$



**Figure R2:** Model for the perception of fluctuation. The critical bandwidth is a parameter in the auditory filters.

for Beats, where  $a_1$ ,  $a_c$  and  $a_h$  are amplitudes of the three spectral components from AM output, and  $a_1$  and  $a_2$  are amplitudes of the two spectral components from Beats output,  $\Delta\omega$  is the frequency spacing between the components. The amplitudes of the spectral components at the output of a peripheral auditory filter depends on the CB of the filter. The envelopes, which are determined by the amplitudes of spectral components, describe the fluctuation of the output and are supposed to be extracted by the auditory system. Then, the TMTF puts a high frequency limit to the fluctuations. It puts the extracted envelopes (4) or (5) as signals through the lowpass filter (3). We know that the auditory system cannot follow the fine structure of the envelope change. From Aures's data (Aures, 1985; Hartmann, 1996), the roughness is determined by the fundamental component of the envelope. This assumption was confirmed by comparing the roughness of AM and the roughness of the Beats. So in our model, only the DC value and the fundamental frequency component of the filtered envelope are used to determine the fluctuation of this channel. The amplitude of the fundamental frequency component is a measure of the absolute fluctuation in this channel. The DC level plays two roles: 1) it acts as a smoothing source; 2) it is an approximate measure of the average power in the channel. Therefore, the fluctuation in the channel is assumed to be the ratio of the power of the fundamental component to the power of the DC component  $F_i = (c_{1,i}^2 / c_{0,i}^2)$ , where  $F_i$  is the fluctuation of the channel,  $c_{1,i}$  is the amplitude of the fundamental of the envelope, and  $c_{0,i}$  is the amplitude of the DC. Finally, the fluctuations from all channels are weighed by the average power of their channel and summed up as the fluctuation perceived:

$$F = \sum_i F_i \left( \frac{c_{0,i}^2}{P} \right)^\beta = \left( \frac{1}{P} \right)^\beta \sum_i c_{1,i}^2 c_{0,i}^{2\beta} \quad (6),$$

where  $P$  is the total power of the signal,  $\beta$  is the energy weighting index, and  $\alpha=2\beta-2$ . Power  $P$  is used as a normalizing factor so as to make  $F$  a unitless variable. However, in our calculation, we do not have a way to determine the number of filters we should use, i.e. we can not determine how wide the frequency spacing between the adjacent filters. But when the number is larger than a certain number (20/oct), the curve of  $F$  converges to a certain shape. Therefore, we just choose a large number. So, the values of the calculated  $F$  are relative to each other, and it is not necessary to put the  $P$  in the calculation.

## II. Experiment

### A. Method

#### 1. Procedure

At the beginning of a run, the listener was seated in a sound-treated room and given a response box with which to control the experiment. The listener pushed a green button to start a trial. After the green button was pushed, a beating stimulus either an AM or a Beats was presented. The fluctuation rate could be adjusted by a ten-turn potentiometer. Turning the potentiometer changed the modulation frequency in the AM stimulus, or changed the frequency of one sine tone in the Beats stimulus so that the beating rate was changed. The listener was asked to adjust the potentiometer to the point where the most rapid beats are heard without decreasing the fluctuations substantially. For the AM stimulus, the listener just had one frequency to tune. For the Beats stimulus, there were two, low and high. Since one component had fixed frequency, beats could be produced if the adjustable component was either higher or lower in frequency than the fixed component. When the listener made his decision, he pressed the green button



again to finish the trial. The frequency chosen by the listener was recorded, and then the next trial with a different AM carrier or a different fixed frequency for Beats began. The whole process was controlled by a 486 computer through the TDT II system. There was no feedback to the listener.

Trials were blocked into runs. Each experimental run included 15 trials for the AM, or 30 trials for the Beats (high and low for each fixed frequency). It took about 7 minutes to do an AM run, and 15 minutes to do a Beats run. After a run was completed, the listener could come out to rest. In the results reported below, the data from the final eight runs were used for each data point. The experiment was done at two levels, 60 dB and 80 dB.

## 2. Stimuli

There were two kinds of stimuli in the experiment, AM and Beats. For the AM stimulus, the sine-tone carrier was generated by the WG1 waveform generator of the TDT II system. The modulation tone was generated by a Wavetek VCG116 function generator. The frequency of the modulation tone was controlled by a control voltage. The two tones were multiplied by a multiplier to generate the AM signal. The modulation percentage was 100%.

The Beats stimulus was generated more easily. The fixed sine tone was generated by the WG1 at the same frequencies as the carrier in AM. The other frequency-changeable sine tone was generated by the Wavetek function generator like the modulation tone in the AM. The two equal-amplitude sine tones went through a mixer to form the Beats.

When the listener made his decision in each trial, the chosen frequency for the AM modulation tone or for the frequency-changeable sine tone in Beats were read by the computer through a Metrabyte CTM5 card. There were altogether 15 frequencies for the AM carrier and the Beats fixed frequency. They were: 70, 85, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500 and 2000 Hz. Since we were interested in low frequency, eight frequencies were less than or equal to 500 Hz, which were considered to be low frequencies.

### 3. Listeners

Three male listeners SB, SD, SJ participated in the experiment. Their ages were 56, 20 and 34 respectively. All of them had negative otological histories and had some training as performers of musical instruments.

### *B. Results*

The data from the experiment are listed in Tables R1 to R6. To make them easier to see, plots of the data are shown in Figures R3, R4 and R5, each for one subject. There are two plots in each figure, one for the 60 dB experiment and one for 80 dB. The horizontal axis, "center frequency," represents the carrier frequency in the AM experiment, and the fixed frequency in the Beats experiment. The vertical axis, "modulation rate," is the beat rate of the stimuli, i.e. it is the modulation frequency for the AM stimulus, and is the frequency spacing of the two sine tones for the Beats stimulus. The points plotted in a graph are the MRR's the listener chose at different frequencies. There were two data points for each center frequency in the Beats experiment, low and high. As shown by Tables R1 - R6, they are usually different. We average the two data points to get one data point, which is the plotted data point.

There were similarities among the three listeners. The overall patterns are the same although there are individual differences. First, the MRR increases with the center frequency. The value increases faster at low center frequencies than at high center frequencies. At center frequencies higher than 1000 Hz, it increases very slowly around 70 Hz as if it entered a saturated region. This is consistent with Zwicker and Fastl (1990). Second, it seems that there is no obvious level effect. Third, at low frequencies, it seems that the MRR for AM is larger than that for Beats.

### III. Computation and Conclusion

#### *A. Fitting the Model to Data*

To fit the data, there are three parameters in our model. First, we can change the critical bandwidth of the peripheral filter. The CB is the parameter we are interested in. In our calculation we use the gammatone filter (with  $n=4$ ) mentioned in the introduction as our model peripheral filter bank. Second, the time constant  $\tau$  for the TMTF in equation (3) is another parameter in our calculation. Third, we can change the energy weighting index  $\beta$  in equation (6).

Next, how many constraints are there on our calculation? First, we need to fit the MRR for AM and Beats, which are our data. Second, according to Aures (1985) and Zwicker and Fastl (1990), there is a quantitative relationship between the roughness (fluctuation in our calculation) of AM and Beats at a particular modulation rate. Aures (1985) did an experiment showing that at 1000 Hz, the roughness of a equal-amplitude Beats stimulus was equal to the roughness of an partially modulated AM stimulus with

modulation percentage  $m=2/3$ . Zwicker and Fastl (1990) showed that at 1000 Hz an AM with  $m=2/3$  is about  $1/2$  as rough as the completely modulated AM ( $m=1$ ). In our experiment, we used 100% modulation for AM, and equal-amplitude components for Beats. Therefore, our AM stimulus should be twice as rough as the Beats stimulus at 1000 Hz at the same modulation rate. If we extend the Aures and Zwicker-Fastl results to other frequencies, AM should be twice as rough as Beats at the same modulation rate. That is our second constraint. The third constraint, which is also from the data, is that the roughness is independent of level. This constraint is automatically satisfied by our model.

	MRR of Beats (Hz)				MRR of AM (Hz)	
$f_c$ (Hz)	Lower		Upper			
	Avg.	Sd.	Avg.	Sd.	Avg.	Sd.
2000	92.1	10.2	94.8	18.4	56.6	9.0
1500	82.4	12.5	78.2	9.1	65.1	10.5
1000	69.4	7.9	74.2	11.5	52.8	3.0
900	76.2	10.7	73.6	9.1	61.5	12.0
800	66.9	5.0	79.8	4.7	57.4	4.2
700	61.6	13.7	70.5	4.9	59.8	10.4
600	58.4	8.5	70.2	5.1	50.9	4.9
500	50.8	4.0	59.6	6.1	47.6	7.4
400	41.1	4.2	45.2	3.8	40.1	4.9
300	30.5	3.0	39.8	4.5	30.9	2.4
200	24.0	2.0	26.1	2.7	24.9	2.0
150	20.1	1.5	21.5	2.1	23.2	2.8
100	14.8	1.9	18.0	1.9	18.0	2.4
85	12.6	0.7	17.1	2.8	17.0	1.3
70	12.5	1.3	15.0	2.6	16.0	2.5

**Table R1:** Experimental results of SB at 60 dB.

	MRR of Beats (Hz)				MRR of AM (Hz)	
$f_c$ (Hz)	Lower		Upper			
	Avg.	Sd.	Avg.	Sd.	Avg.	Sd.
2000	60.4	18.2	62.6	14.7	65.2	8.1
1500	62.4	11.2	57.4	11.9	60.8	11.0
1000	58.5	11.6	59.2	17.8	53.8	10.6
900	60.2	7.4	56.8	10.4	55.8	11.1
800	55.9	17.0	59.8	13.3	55.9	11.6
700	50.5	13.0	56.6	18.5	55.2	8.8
600	52.4	9.8	48.6	11.7	50.6	7.3
500	44.9	10.9	46.8	11.4	49.8	5.3
400	34.2	8.6	37.2	7.9	38.2	3.9
300	28.4	4.1	31.6	8.3	31.9	4.4
200	23.1	3.1	23.4	2.4	24.9	2.8
150	20.0	3.1	21.4	3.3	24.0	2.0
100	15.1	1.6	16.9	1.9	19.9	1.9
85	13.8	2.1	16.0	2.4	20.2	1.0
70	12.2	1.3	13.4	1.8	18.1	2.6

**Table R2:** Experimental results of SB at 80 dB.

	MRR of Beats (Hz)				MRR of AM (Hz)	
$f_c$ (Hz)	Lower		Upper			
	Avg.	Sd.	Avg.	Sd.	Avg.	Sd.
2000	60.6	11.4	69.4	12.1	63.9	11.5
1500	61.6	16.9	64.4	15.5	65.5	6.2
1000	60.5	8.6	60.9	8.2	63.6	6.0
900	59.0	10.5	58.6	10.7	61.1	5.8
800	68.1	10.9	58.3	11.9	60.8	5.8
700	53.1	11.1	61.3	9.6	58.2	5.5
600	55.0	14.3	64.0	16.3	57.7	5.5
500	46.2	8.9	53.8	9.2	55.1	5.2
400	39.9	4.6	47.7	3.8	49.1	4.7
300	42.6	3.4	37.0	4.2	40.0	3.8
200	36.0	3.5	31.4	3.0	31.4	3.0
300	31.7	3.0	26.3	3.0	25.1	2.4
100	18.5	1.7	19.5	1.9	21.9	2.1
80	16.6	1.6	17.0	1.6	17.7	1.7
70	13.8	1.5	15.0	1.5	17.4	1.7

**Table R3:** Experimental results of SD at 60 dB.

	MRR of Beats (Hz)				MRR of AM (Hz)	
$f_c$ (Hz)	Lower		Upper			
	Avg.	Sd.	Avg.	Sd.	Avg.	Sd.
2000	60.5	12.7	63.5	12.1	59.7	8.5
1500	57.5	11.7	64.5	18.7	56.5	5.8
1000	60.0	7.3	61.3	9.2	46.3	3.3
900	57.5	10.9	62.0	11.3	47.7	6.2
800	66.8	12.4	59.5	14.7	47.2	9.0
700	53.8	11.0	52.5	5.7	46.7	7.0
600	53.5	7.9	55.5	6.4	43.0	3.1
500	50.0	5.2	50.0	6.1	46.7	11.1
400	40.5	6.5	47.2	3.8	38.0	4.6
300	38.0	6.6	41.5	8.1	35.7	3.1
200	33.7	6.5	33.7	6.9	32.7	3.9
150	25.8	5.8	32.2	4.7	30.0	1.8
100	16.3	1.8	21.7	4.9	26.2	2.8
80	10.2	2.3	24.0	4.9	17.0	3.2
70	10.8	2.5	17.0	3.6	20.3	3.4

**Table R4:** Experimental results of SD at 80 dB.

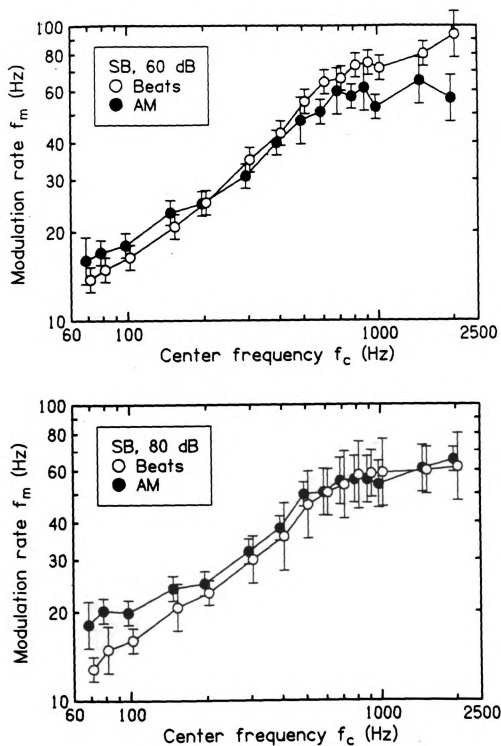


	MRR of Beats (Hz)				MRR of AM (Hz)	
$f_c$ (Hz)	Lower		Upper			
	Avg.	Sd.	Avg.	Sd.	Avg.	Sd.
2000	72.8	9.6	71.2	7.2	66.1	11.3
1500	65.1	19.7	63.8	11.5	54.6	11.5
1000	52.4	11.1	60.2	14.4	49.0	8.9
900	48.1	10.3	54.2	12.2	48.8	8.1
800	42.4	9.8	47.9	10.0	46.2	3.4
700	40.4	10.8	44.5	8.6	38.9	8.9
600	36.4	7.9	38.8	12.3	38.5	4.8
500	37.4	5.7	38.9	7.3	36.6	4.0
400	35.1	2.6	36.9	2.7	34.6	2.7
300	32.4	4.2	33.2	3.4	31.1	3.6
200	29.8	3.8	28.4	3.2	26.4	3.1
150	25.8	1.8	30.8	3.7	27.4	2.8
100	17.0	1.6	22.4	1.7	20.0	3.2
85	13.8	1.8	19.0	1.9	14.9	3.0
70	9.2	1.6	17.5	1.8	14.9	2.5

**Table R5:** Experimental results of SJ at 60 dB.

	MRR of Beats (Hz)				MRR of AM (Hz)	
$f_c$ (Hz)	Lower		Upper			
	Avg.	Sd.	Avg.	Sd.	Avg.	Sd.
2000	72.0	18.0	72.5	13.3	51.6	9.2
1500	59.8	8.3	62.1	10.3	44.6	12.6
1000	48.8	9.3	55.8	14.9	47.8	6.7
900	52.1	12.0	56.6	7.7	47.4	9.5
800	52.9	20.5	53.1	14.4	42.6	5.1
700	50.1	13.7	54.0	21.6	39.4	5.3
600	43.6	9.7	49.2	13.2	38.0	5.4
500	36.6	6.0	38.8	5.8	35.1	3.3
400	35.6	6.1	41.1	8.0	33.8	3.5
300	29.9	4.5	32.9	4.1	31.9	4.2
200	28.0	6.1	29.4	7.5	26.1	4.1
150	21.9	3.1	24.8	4.5	25.2	2.9
100	14.8	0.7	18.0	2.1	19.2	3.9
85	13.0	1.8	16.9	2.0	18.5	3.0
70	11.9	1.7	14.4	2.3	16.6	2.5

**Table R6:** Experimental results of SJ at 80 dB.



**Figure R3:** Results of the experiment for SB. The upper panel is the results at 60 dB, and the lower panel is the results at 80 dB. The open circle is for the Beats. Each point represents an MRR at one center frequency. The upper point and the lower point at each center frequency from Tables 1 and 2 are averaged to make this plot. Therefore, there is only one point at each frequency for the Beats. The filled circle is for AM.

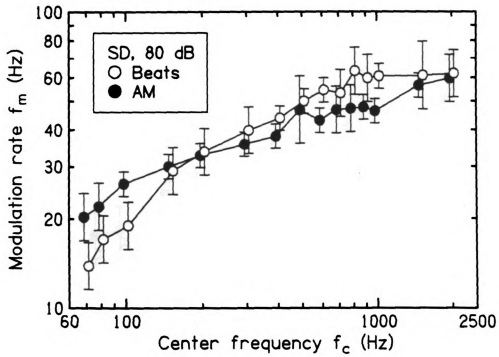
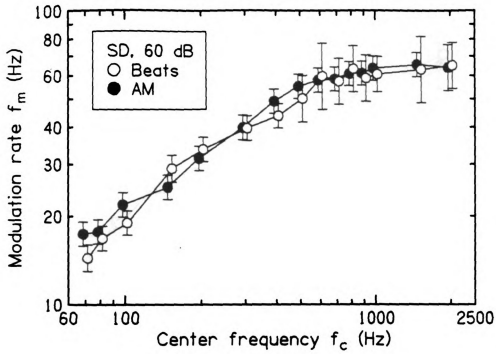


Figure R4: Results of the experiment for SD.

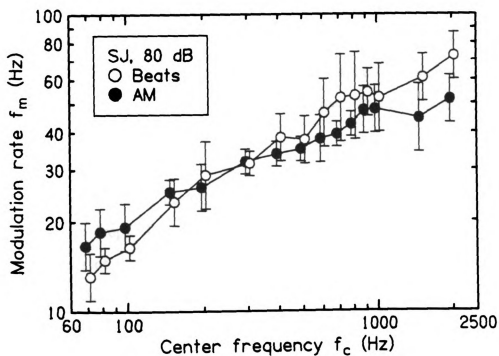
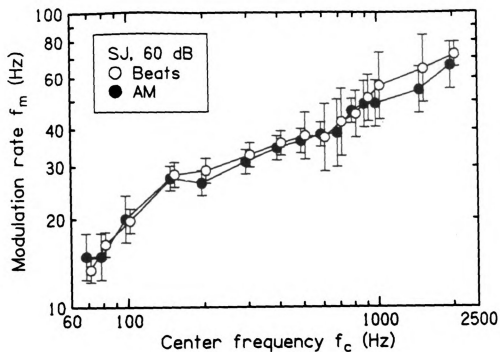


Figure R5: Results of the experiment for SJ.

The calculation was done using a C-program. The purpose was to fit the model MRR to the measured MRR averaged over subjects and levels. The averaged measured MRR and the fitted model MRR are listed in Table R7. The fitted model MRR is also plotted in Figure R8. Figure R6 shows plots of fitted fluctuation or roughness curves at each center frequency.

The fitting procedure is determined by the functions of the parameters. The energy weighting parameter  $\beta$  changes the alignment of the MRR's for AM and Beats, it also changes the relative sizes of the two fluctuation curves. For smaller  $\beta$ , the alignment of MRR's for AM and Beats at low center frequencies is closer to the data, i.e. the MRR for AM is larger than that for the Beats. However, for smaller  $\beta$  the roughness (fluctuation in our calculation) at a particular modulation rate for AM and Beats does not follow the constraint that AM is about twice as rough as Beats (Beats is too small). For this constraint, larger  $\beta$  makes a better fit. It seems that  $\beta=0.5$  is a good compromise between the two constraints. This result weights the fluctuation with the square root of the power in a channel, which happens to agree with a maximum-likelihood estimator.

The other two parameters, the CB and the  $\tau$  of the TMTF, determine how fast the fluctuation curve drops. Therefore, they determine the modulation rate at which the maximum slopes of the fluctuation curves occur. Those are the calculated MRR's. Narrowing the TMTF filter or narrowing the CB results in a faster drop of the fluctuation curves and thus smaller MRR's.

According to Viemeister (1979), the time constant  $\tau$  is about 2 to 3 ms, which means the cut off frequency of the TMTF filter is about 100 Hz or so. Therefore the TMTF does not have much influence on the calculated fluctuation at low modulation rate (say

below 40 Hz). Since the MRR for low center frequencies always occurs at low modulation rates, the TMTF does not have much influence on the fitting results at low frequencies, i.e. errors in  $\tau$  do not affect much the fitting results at low center frequencies. Thus the CB is the dominant parameter at low frequencies and can be well determined by the MRR from this model. This is an excellent result, since we are primarily interested in the critical band at low frequencies.

At high frequencies, the MRR occurs at high modulation rate. In the frequency range 600 - 2000 Hz, both the TMTF and the CB influence the calculated MRR in a similar

	Averaged MRR (Hz)				Model MRR (Hz)	
$f_c$ (Hz)	Beats		AM		Beats	AM
	Avg.	Sd.	Avg.	Sd.		
2000	71.0	11.9	60.5	5.7	68.65	71.98
1500	64.9	7.7	57.8	7.8	63.93	67.04
1000	60.1	6.6	52.2	6.3	55.45	58.15
900	59.6	8.2	53.7	6.6	52.88	56.78
800	59.3	9.7	51.7	7.3	50.43	52.88
700	54.1	7.7	49.7	9.3	48.10	50.43
600	52.1	9.5	46.4	7.9	44.79	46.97
500	46.1	7.0	45.2	7.8	41.72	42.72
400	40.2	3.9	39.0	5.5	37.83	38.73
300	34.8	4.2	33.6	3.6	33.60	36.07
200	28.9	4.3	27.7	3.4	27.08	27.73
150	25.2	4.0	25.8	2.5	23.26	23.26
100	17.8	1.7	20.1	2.9	18.79	18.35
85	15.8	1.0	18.0	2.5	16.76	16.37
70	13.6	0.6	16.5	1.9	14.89	14.54

**Table R7:** Averaged MRR over listeners and levels, and the predicted MRR from the model.





way. Thus we cannot determine the two unknowns,  $\tau$  of the TMTF and the CB, at the same time by fitting the MRR. However, our goal is to check the CB at low frequencies since the main disagreement on CB is at low frequencies. So we assume that the CB is the ERB for frequencies higher than 500 Hz. Thus, we first fit our data at high frequencies to determine the time constant  $\tau$  of the TMTF, and then use the obtained  $\tau$  to do fittings at low frequencies to determine the CB at low frequencies.

With this procedure,  $\tau$  was determined to be 2 ms. Figure R6 shows plots of the fittings, one plot for each center frequency. The curve with a star is the fluctuation curve for AM. The star marks the maximum slope of the curve, and hence the MRR. The curve with a circle is for Beats. The circle labels the MRR. Figure R7 is a plot of the CB's used to make these fittings at low frequency. The dotted line is the Bark scale. The dashed line is the ERB from Moore and Glasberg (1985). The solid line is our fitting. Figure R8 is the plot of the fitted MRR's, i.e. the fifteen star-circle pairs from the fifteen plots in Figure R6.

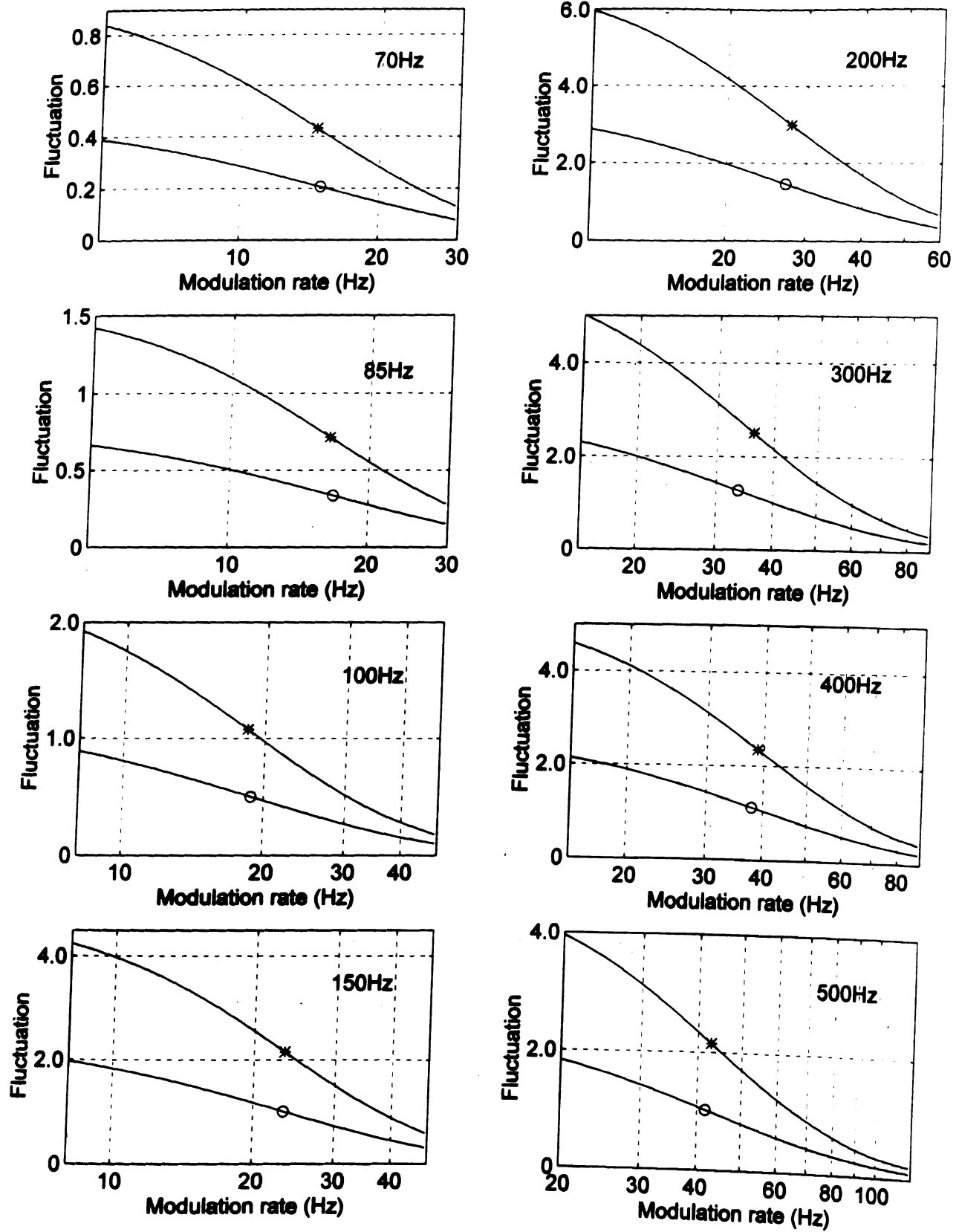
### *B. Conclusion and Discussion*

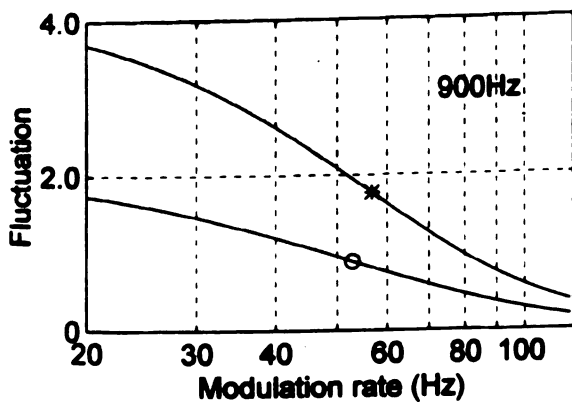
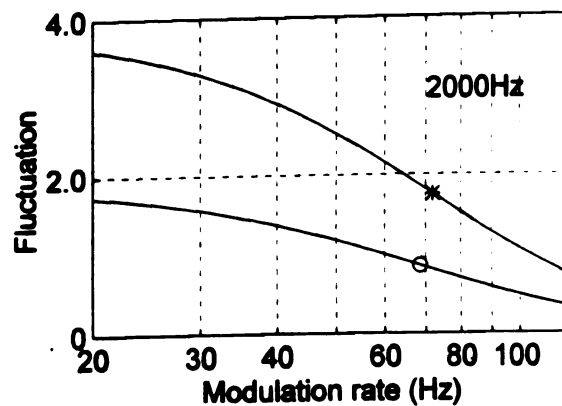
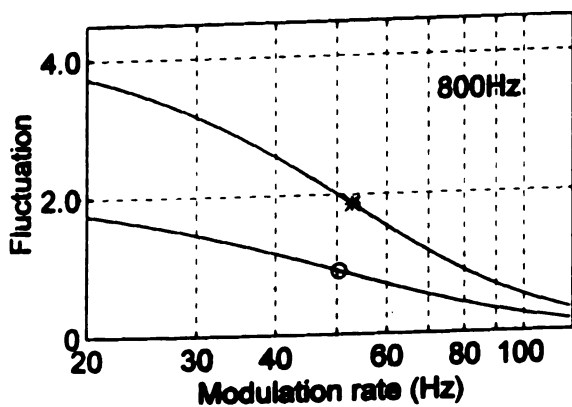
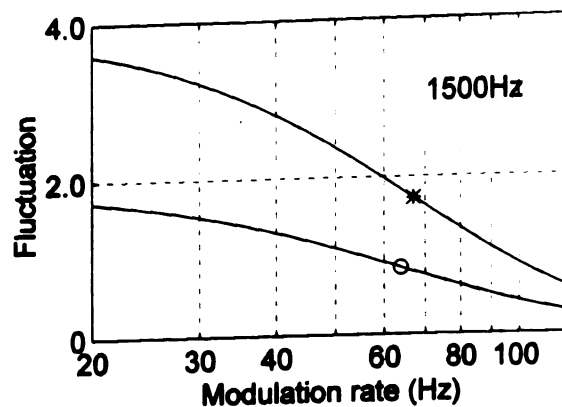
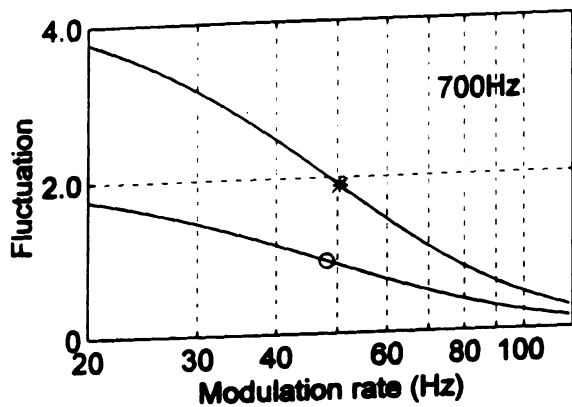
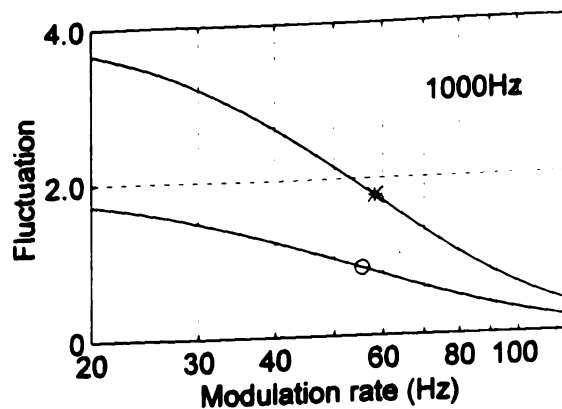
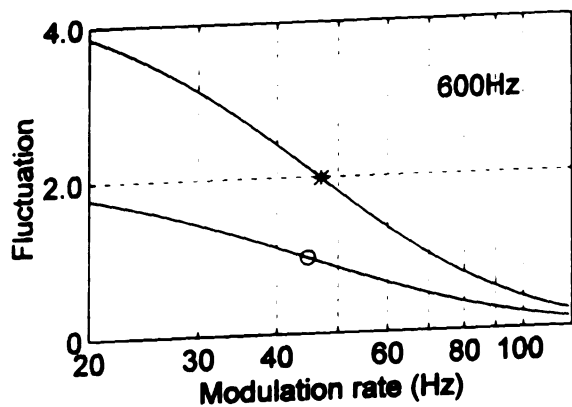
Figure R7 is the plot of the results of using the MRR to determine the CB (mainly at low frequencies). As we can see, for center frequency of 300, 400 and 500 Hz our fitted critical bandwidths agree with the ERB of Moore and Glasberg (1985). For center frequency below 300 Hz (70-300 Hz) our fitted critical bandwidths are smaller than Moore and Glasberg's ERB.

It seems that our data do not agree with Plomp and Levelt's (1965) 1/4 critical bandwidth rule. For example, at 1000 Hz, the MRR's is about 60 Hz, which gives CB=240 Hz, larger than the Bark scale, Moore-Glasberg's ERB and the prediction of our

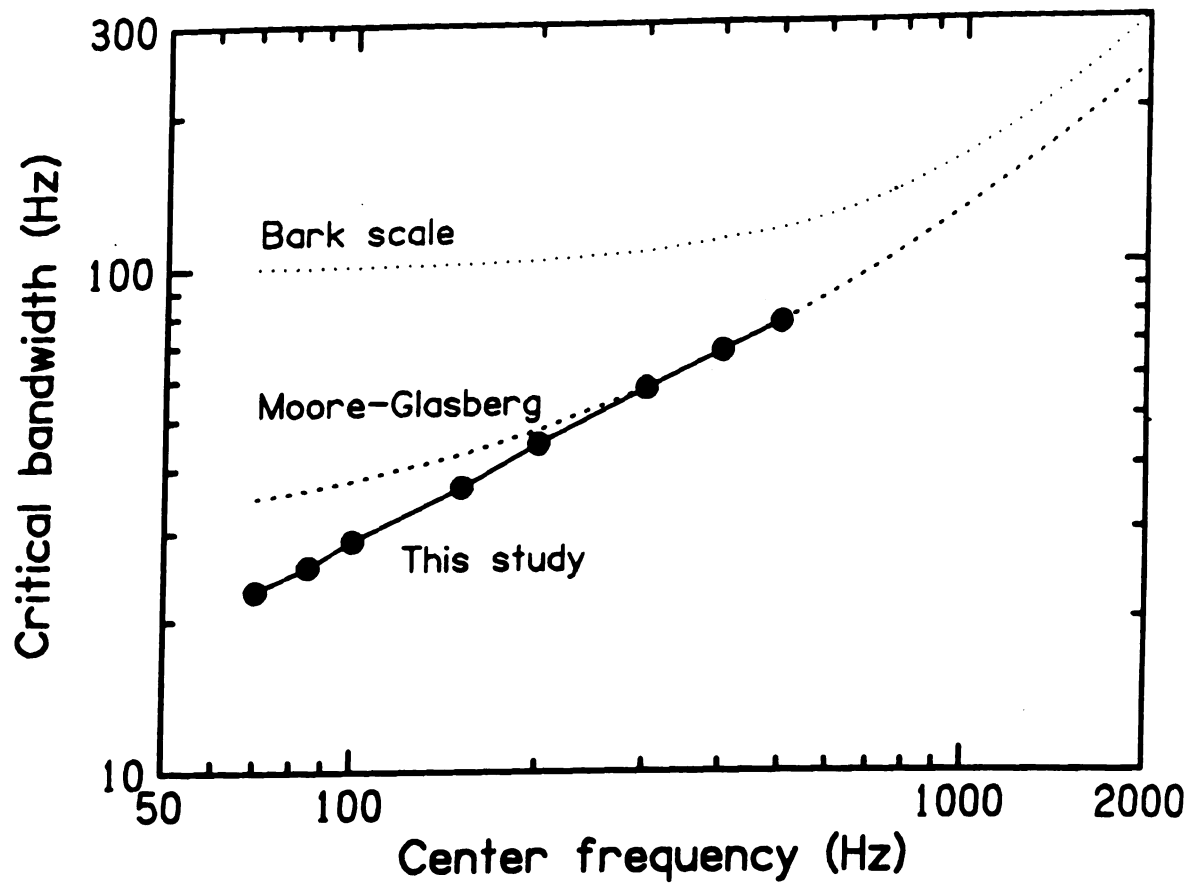
model. At 100 Hz, the MRR is about 20 Hz, which gives  $CB=80$  Hz. This is close to the Bark scale, but differs from Moore-Glasberg's ERB and the prediction from our model.

There are some problems with our fitting. First, we cannot align well the MRR of the two stimuli and at the same time adjust the relative roughness (fluctuation) of the two stimuli to reasonable values. The compromise is to choose  $\beta=0.5$ . Figure R8 is a plot of fitted MRR's at different frequencies. As shown in the Figure and Table 7, at low frequencies the alignment does not agree with the data in that the data show that the

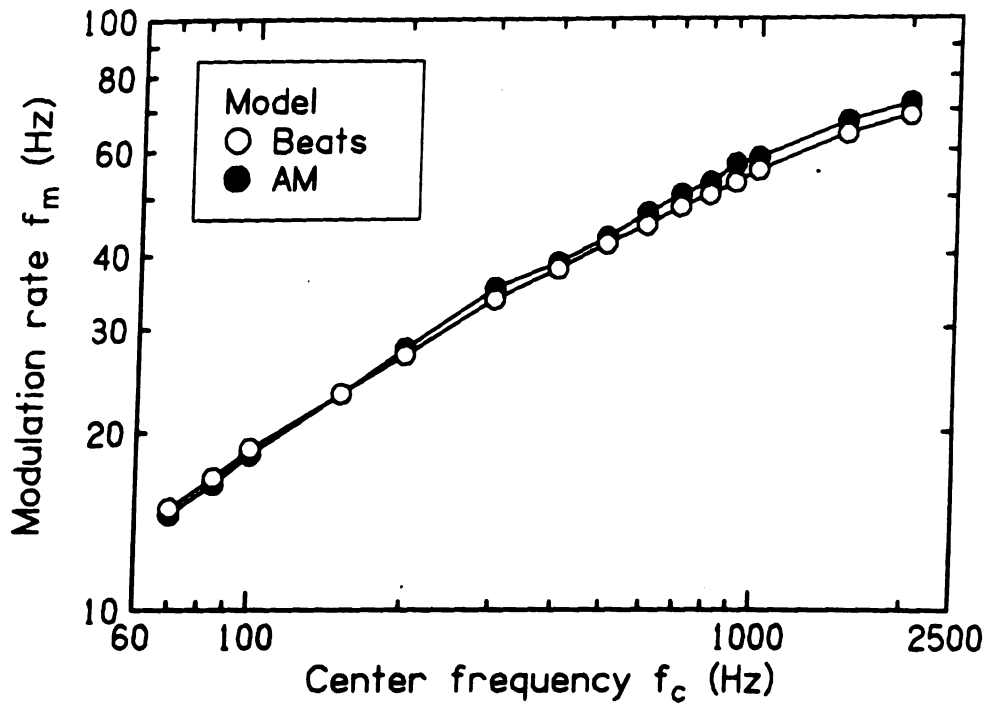




**Figure R6:** The fittings of model MRR to the measured MRR. There are fifteen panels, each for a calculation result at one center frequency.



**Figure R7:** Plot of CB used to make the fittings at low frequency. The dotted line is the bark scale. The dashed line is the ERB from Moore and Glasberg (1985). The solid line is our fitting.



**Figure R8:** Plot of the model predicted MRR.

MRR for AM is larger than that for the Beats whereas the model predicts that the two MRR's are about the same. This is probably something that cannot be solved within this model.

The second problem is that the fitted fluctuation curve decreases too slowly. It seems that the listeners could perceive a sharper turning point than that shown in Figure R6. Consider the 70-Hz plot in the figure. The MRR for Beats is about 15 Hz. From the same plot, the fluctuation of Beats at MRR is the same as the fluctuation of the AM at a modulation rate of about 23 Hz. This seems to be unlikely, because at modulation rate=23 Hz almost no fluctuation could be perceived for the 70-Hz AM stimulus. This shows that the curve for AM does not drop fast enough. Since the curve for Beats is not steeper than that for AM, it also drops too slowly. With those existing problems, we cannot say that our results from this model are very reliable. It is the best we can do at this stage. There may be some other mechanism which we do not understand very well at this time.



## Appendix R1

According to Sek and Moore (1994), the CMF is determined by the lower side-band at high frequency and by the upper side-band at low frequency, thus confounding the CMF as a measure of the critical bandwidth. The transition region is around 200-250 Hz. They claimed that this was the reason that CB obtained from Zwicker's CMF method flattens out at low frequency while the ERB still goes down to small values at low frequencies.

Based upon this idea, we calculated the influence of the transition of the side-band on the CMF. We chose carrier frequencies from 50 Hz to 500 Hz. We calculated the CMF in two conditions: 1) the CMF determined by the lower side-band; 2) The CMF determined by the upper side-band. The equations used to determine the two CMF's are as follows:

$$\begin{cases} f_u - f_c = \frac{1}{2} ERB(f_u) \\ CMF_u = \frac{1}{2} ERB(f_u) \end{cases} \quad \text{for the upper side-band,}$$

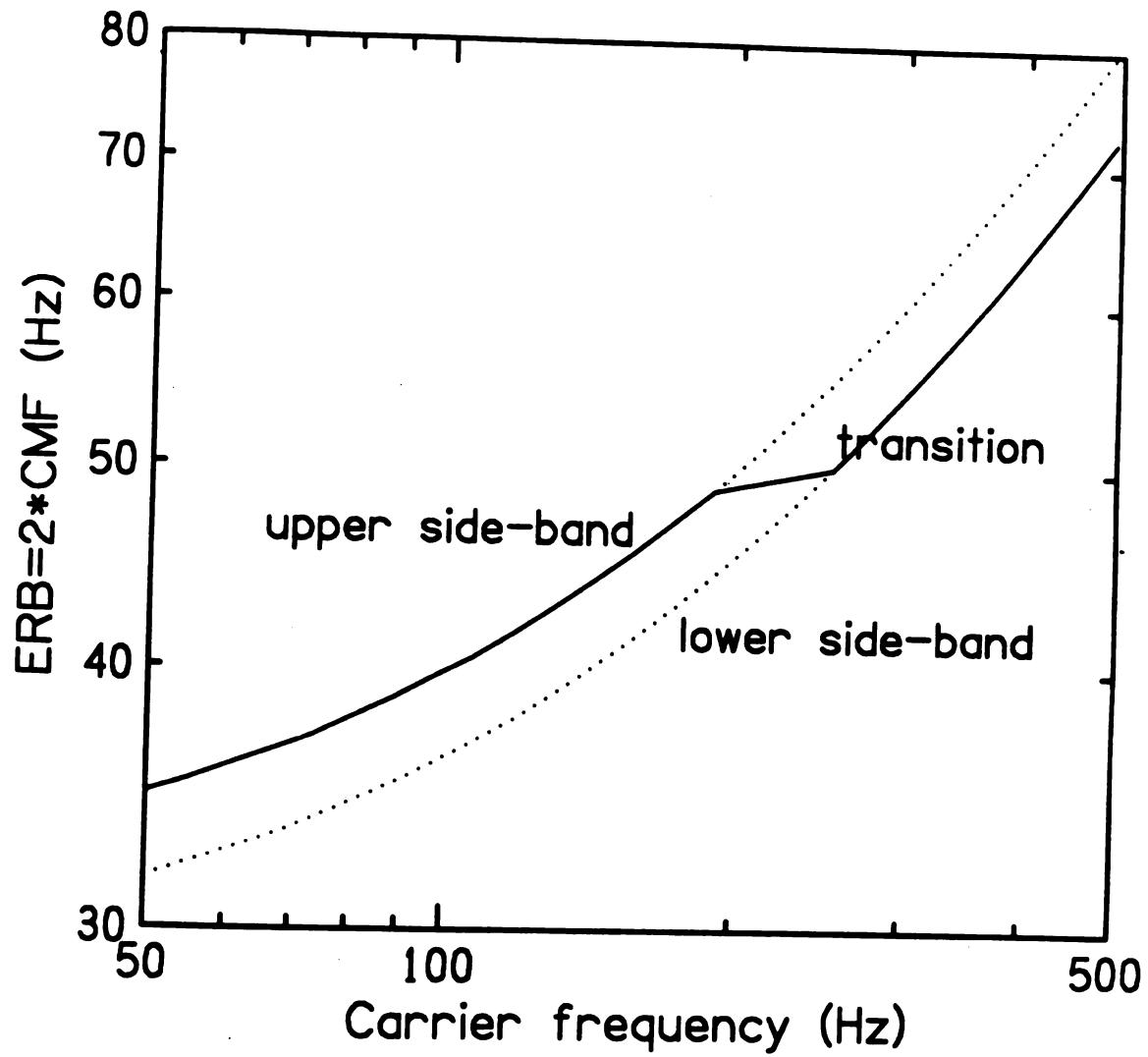
$$\begin{cases} f_c - f_l = \frac{1}{2} ERB(f_l) \\ CMF_l = \frac{1}{2} ERB(f_l) \end{cases} \quad \text{for the lower side-band,}$$

where  $ERB(f) = 6.23f^2 + 93.39f + 28.52$  (Moore and Glasberg, 1985),  $f_u$ ,  $f_l$  and  $f_c$  are in kHz;  $CMF_u$ ,  $CMF_l$  and  $ERB$  are in Hz.

As shown in Figure R9, the two CMF's are close together. That means that the flat region due to the CMF transition is very small and the effect suggested by Sek and

Moore would seem to be very small. In fact, from the figure, it is much easier to see the ERB ( $=2 \times \text{CMF}$ ) go down either above the transition region or below the transition region.

It is also notable that Zwicker's critical band results, which are flat at low frequency are self-consistent. They are undisturbed by the transition between side-bands because



**Figure R9:** Plot of the calculation results of Sek and Moore's (1994) flattening effect of CB.

the bandwidth evaluated at low side-band frequency is the same as the bandwidth evaluated at the upper side-band frequency.

## REFERENCES

- Aures, V.W. (1985). "Ein Berechnungsverfahren der Rauigkeit," *Acustica*, **58**, 268-272.
- Hartmann, W.M. and Hnath, G.M. (1982). "Detection of Mixed Modulation," *Acustica*, **50**, 297-312.
- Hartmann, W.M. (1996) "*Signals, Sound and Sensation*," AIP Press, New York.
- Moore, B.C.J. (1982). "*An Introduction to the Psychology of Hearing*," Academic Press, pp. 82 (London).
- Moore, B.C.J. and Glasberg, B.R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.*, **74**, 750-753.
- Moore, B.C.J., Peters, R.W. and Glasberg, B.R. (1990). "Auditory filter shapes at low frequencies," *J. Acoust. Soc. Am.*, **88**, 132-140.
- Patterson, R.D. (1976). "Auditory filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.* **59**, 640-654.
- Patterson, R.D., Nimmo-Smith, I., Weber, D.L. and Miltroy, R., (1982). "Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.* **72**, 1788-1802.

Patterson, R.D., Nimmo-Smith, I., Holdsworth, J. and Rice, P. (1987). "An efficient auditory filterbank based on the gammatone function," Annex B of the SVOS Final Report.

Plomp, R. and Levelt, W.J.M. (1965). "Tonal Consonance and Critical Bandwidth," J. Acoust. Soc. Am., **38**, 548-560.

Sek, A. and Moore, B.C.J. (1994). "The critical modulation frequency and its relationship to auditory filtering at low frequencies," J. Acoust. Soc. Am., **95**, 2606-2615.

Shorer, E. (1986). "Critical modulation frequency based on detection of AM versus FM tones," J. Acoust. Soc. Am. **79**, 1054-1057.

Terhardt, E. (1974). "On the perception of Periodic Sound Fluctuations (Roughness)," Acoustica, **30**, 201-213.

Viemeister N.F. (1979). "Temporal modulation transfer function based upon modulation threshold," J. Acoust. Soc. Am. **66**, 1364-1380.

Zwicker, E. and Fastl, H., (1990). *"Psychoacoustics Facts and Models,"* Springer-Verlag (Berlin).

## **CHAPTER 2**

### **Pitches of Mistuned Harmonics**

## Introduction

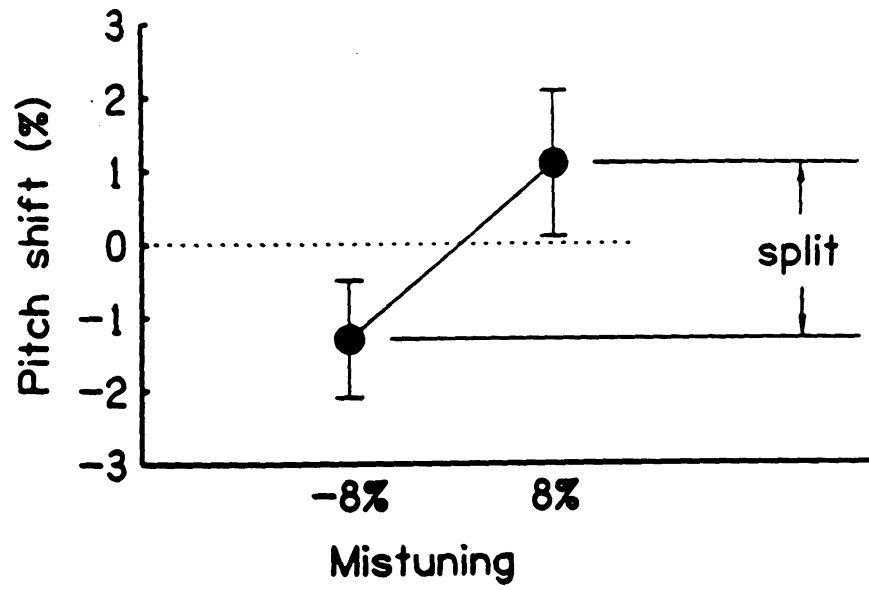
Our auditory system has the ability of integrating the components of a complex tone into a single entity. It is a good thing that we have this ability, because all the components of a complex tone come from one sound source. If we did not have this ability, we would have trouble in our daily life, since almost all the sound sources in nature are complex. We would perceive individual harmonics, which very likely would cause information overload.

The pitch associated with a complex sound is the product of the process in which our auditory system integrates the harmonic components coming from the single source. Therefore, a good way to study pitch perception of a complex tone might be to study how our auditory system integrates the harmonic components. In other words, it is important to study the relationship, which is built in our auditory system, between the complex tone as an entity and its harmonic components. The mistuned harmonic experiment is a very good experiment in this respect.

When a harmonic of a complex tone is mistuned, it is heard separately from the complex tone as if there are two sound sources: the complex tone and a sine tone (the mistuned harmonic). Moore *et al.* (1986) measured the thresholds of mistuning for mistuned harmonics. In the Mistuned Harmonic Matching Experiment (Hartmann *et al.*, 1990), subjects were asked to match the pitch of the mistuned harmonic with a pure tone. The data showed significant pitch shifts. For example, if the third harmonic of a 200 Hz fundamental was mistuned +8% so that its frequency was 648 Hz (mistuned +48 Hz away from its original frequency 600 Hz), a typical matching would be around 655 Hz instead of around 648 Hz, resulting in a pitch shift of +1.1%; if the harmonic was mistuned -8% so that its frequency was 552 Hz, a typical matching would be



around 545 Hz, resulting in a pitch shift of  $-1.3\%$ . The sign of the pitch shift was generally correlated with the sign of the mistuning of the mistuned harmonics, i.e. positive mistuning made positive pitch shifts and negative mistuning made negative pitch shifts. In other words, pitch shifts exaggerate the mistuning. The differential pitch shift, for positive vs. negative mistuning, will be called the “split.” Splits describe the exaggeration. In the example above, the split for the third harmonic of the 200 Hz fundamental mistuned at  $\pm 8\%$  is  $[+1.1 - (-1.3)] = 2.4\%$ . Figure M1 shows the example.



**Figure M1:** Example of a split.

In 1979 Terhardt presented a semi-empirical algorithm for calculating the pitch shifts of the harmonics of a complex tone (see also Terhardt *et al.*, 1982b). It is mainly based on the concept of the excitation pattern. When a sine tone is partially masked by a sound having lower frequency, its pitch will be shifted higher. Similarly, maskers having higher frequency than the sine tone will tend to shift the sine tone pitch lower. Thus, the excitation of a harmonic would mask the neighboring harmonics, and thus “push away” the adjacent harmonics in pitch. Therefore, the algorithm predicts a negative pitch shift for the fundamental of a complex tone and positive pitch shifts for the rest of the harmonics regardless of their mistunings, since for the fundamental masking can only come from higher harmonics (mainly the 2nd) and for the other harmonics more masking comes from the adjacent lower harmonic than from the higher harmonic because of the asymmetry of the excitation pattern (Terhardt *et al.*, 1982b).

Terhardt’s theory can also be used to predict the pitch shifts for inharmonic components, since it is mainly a place theory. For example, Terhardt *et al.*, (1982a) used the algorithm to predict the pitch of church-bell tones. However, in applying Terhardt’s algorithm to the plus-minus mistuning pair mentioned above, it always predicted positive pitch shifts no matter whether the mistuning was positive or negative. It even gives splits with wrong sign in most conditions, i.e. instead of exaggerating the mistuning it compensates the mistuning (see Hartmann and Doty, 1996). Obviously, this algorithm does not agree with data and cannot explain the splits phenomenon. Therefore, Hartmann and Doty (1996) developed an alternative theory based upon timing.

The model began with the idea that pitch should be determined by the peaks of the interspike interval (ISI) histogram, as suggested by Goldstein and Sruлович (1977). Pitches were shifted when excitation corresponding to neighboring harmonics appeared in the same auditory filter, leading to a complicated temporal pattern. In general terms,

the effect of such interaction on the ISI histogram was to cause a harmonic to be attracted in pitch to its nearest neighbors, one higher and one lower than the harmonic. The higher neighbor tends to make a positive pitch shift, and the lower neighbor tends to make a negative pitch shift, an “attraction” effect which is opposite to the Terhardt’s “repulsion” effect. Decreasing the spacing in frequency between the harmonic and a neighbor will result in an increasing of attraction between them. Now consider a  $\pm 8\%$  mistuned 3rd pair. The  $+8\%$  mistuned 3rd harmonic is closer in frequency to the higher neighbor - the 4th harmonic than the  $-8\%$  mistuned 3rd. Therefore, the  $+8\%$  mistuned 3rd gets more attraction from the 4th than the  $-8\%$  mistuned 3rd, a tendency of making more positive pitch shift to the  $+8\%$  mistuned 3rd. Similarly, the  $-8\%$  mistuned 3rd is more attracted by the 2nd than the  $+8\%$  mistuned 3rd, a tendency of making more negative pitch shift to the  $-8\%$  mistuned 3rd. The differential pitch shift for  $\pm 8\%$  mistuned pair - the split, is thus produced.

The model predictions agreed well with data for all harmonics, but there was one glaring exception, the mistuned fundamental. The model said that the pitch shifts should be unusually small and opposite in sign to the mistuning (positive mistuning for the fundamental leads to negative pitch shift). Experiment showed, without question, that the pitch shifts are just like any other harmonic, namely substantial and in the same direction.

This crisis leads us to rethink the theory anew. Both the Terhardt Model and the timing model of Hartmann and Doty are local-interaction models in that the split depends almost entirely on neighboring harmonics. The two closest neighbors are especially important. A problem with a local-interaction model is that it gives no recognition to the role of the entire spectrum in integrating a correctly tuned harmonic into a complex tone. In fact, we know a harmonic stands out if it is mistuned. It is reasonable to say

that this prominence is caused by the relationship between the mistuned harmonic and the whole complex tone, which is composed of the other correctly tuned harmonics as an entity. It is reasonable to conjecture that the split is related to this prominence, in other words, the split might be related to the whole complex tone instead of to only neighboring harmonics. The complex tone may produce some periodic “framework” or pattern to make the frequencies of the harmonics special in the brain. In other words, whether a harmonic exists or not in those frequencies, the frequencies are special in the brain once the “framework” is formed by a background complex tone. In fact, there are several pitch perception models that suggest a template. For example, the DWS pitch meter uses a harmonic pattern recognizer to extract a periodicity pitch (Duifhuis *et al.*, 1982; Scheffers, 1983). Terhardt (1974) suggested that a tone would tend to cue all its sub-harmonics, i.e. making possible places for a pattern. If we think of a complex tone with a mistuned harmonic as a perfectly-tuned complex tone which omits that mistuned harmonic and a sine tone which has the mistuned harmonic frequency, then the picture in frequency domain is more like the sine tone dropped into the “framework” of the complex tone. An exaggeration of the mistuning in pitch results in a contrast between the two sounds: the complex tone and the sine tone (the mistuned harmonic), which may make it easier for us to segregate one sound source from the other. Having those thoughts in mind, we did a line of experiments to see whether a pervasive frame-like periodicity can be supported.

## **I. Experiments**

### **A. METHOD**

#### **1. Procedure**

The listener was seated in a sound-treated enclosure, holding a response box that controlled the events of an experimental trial. When the listener pressed a yellow button there was a pause of 300 ms, and then there was an 800 ms complex tone, with one of its harmonics mistuned. When the listener pressed an orange button there was a pause of 300 ms, and then there was an 800 ms sine tone with a frequency that could be adjusted by means of a ten-turn potentiometer on the box. The potentiometer allowed the listener to make the pitch match. The listener could call up the complex tone or the matching tone in any sequence, as often as he liked. When the listener was satisfied with his match he pressed the green button to finish the trial. The stimulus and matching frequencies were then recorded, and then the next trial with a different complex tone (either with different mistuning of the mistuned harmonic, or different fundamental of the complex tone, or a different number of the mistuned harmonic) began. There was no feedback to the listener.

There were many different experiments with different stimuli and different numbers of stimuli. In any experiment, there were as few as three or as many as six complex spectra. Each experimental run included one or two presentations of each with mistuned harmonic mistuned at +8% and -8%. Therefore a run included 6 to 12 trials. After a run was completed, the listener could come out to rest. It took about 10 to 20 minutes for a listener to finish a run. The data from the last 12 matches for each data point were used.

## 2. Stimuli

There were two kinds of stimuli in the experiment, complex tones and sine tones. The complex tones contained 8 to 16 partials of equal amplitude. The fundamental frequency varied from 150 Hz to 800 Hz for different complex tones.

Complex tones were generated by sound files which were specific as to the fundamental frequency, the mistuned harmonic, the harmonic content and the percentage of mistuning - either plus or minus 8%. For a given trial, the appropriate sound file was loaded into a digital buffer 16k (16384) samples long and was converted by a 16-bit DAC at a normal sample rate of 16k/s. Therefore, the period of the signal was 1s. To prevent the listener from using his memory for pitch to do the task, the sample rate was actually different on every trial. It was randomized over a range of +10% to -10%, with a rectangular distribution.

The analog signal was low-pass filtered at 7KHz, -115dB/octave. The complex tone, as presented to the listener, was shaped by computer-controlled amplifier to give it an envelope with a 10 ms raised-cosine onset and offset and a full-on duration of 1 sec. The electrical signal was presented to the listener via SENNHEISER HD480 headphones at a level of 58dB per component. For instance, a 16-component complex tone would be at 70dB SPL.

The matching tone was generated through repeatedly cycling a buffer using fractional-addressing technology. One cycle of a sine wave was loaded into a 16k buffer. It was sampled at the rate of 32k/s. The fractional address increment was determined by the potentiometer on the control box, as read by a 12-bit ADC. An exponential frequency control law was applied in software, leading to a frequency resolution of 0.06%. The matching tone was filtered and enveloped like the complex tone. The matching tone was at 55dB, i.e., 3dB lower than the mistuned component. As a result, the matching

tone was approximately equally as loud as an individual component. It was expected that equal loudness would make the pitch matching task easier.

### 3. Listeners

There were four listeners, three male, B, J, T, and one female, C. The ages ranged from 19 to 54. All of them could perform accurate pitch-matching in the range 150 to 1000 Hz. All the listeners had negative otological histories and had some training as performers of musical instruments.

## *B. EXPERIMENT 1*

### 1. Spectra of the Complex Tones

The spectra of the two complex tones in this experiments are shown in Figure M2. The idea is to create a complex tone which is composed of a perfectly-tuned part with three of its successive components omitted, the so called "gap," and a mistuned harmonic in the gap. In the upper panel, harmonics 1, 2, 3 are omitted. Then we do experiments with mistuned harmonic 1 (fundamental), 2, and 3 separately. We use the notation "-8M2" to denote a mistuned second harmonic, mistuned by -8%. There is only one mistuned harmonic in any stimulus. We call this experiment "mistuned 1/2/3 in a gap." Therefore, there are six matches to the six different stimuli for the "mistuned 1/2/3 in a gap." The lower panel of Fig. 2 is the same as the upper one except that the gap is created by omitting harmonics 2, 3, and 4. The corresponding mistuned components are harmonics 2, 3, and 4. We call this experiment "mistuned 2/3/4 in a gap."

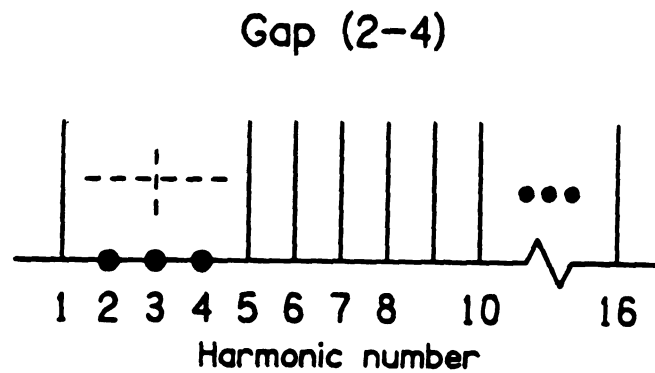
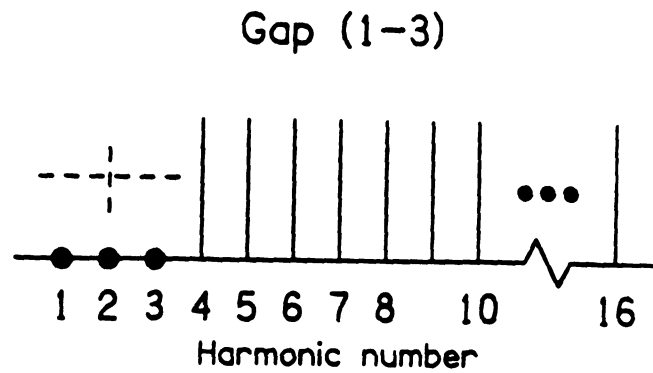


## 2. Results - The Zigzag Effect

Figure M3 shows the results for mistuned 1/2/3 and mistuned 2/3/4. Each panel shows results from one subject. There are several observations to be made. Firstly, there is a split for each plus-minus-mistuning pair of mistuned harmonics. There are overall shifts for some of the mistuning pairs. The overall shift differs from subject to subject. It might be caused by individual differences and the excitations of other harmonics.

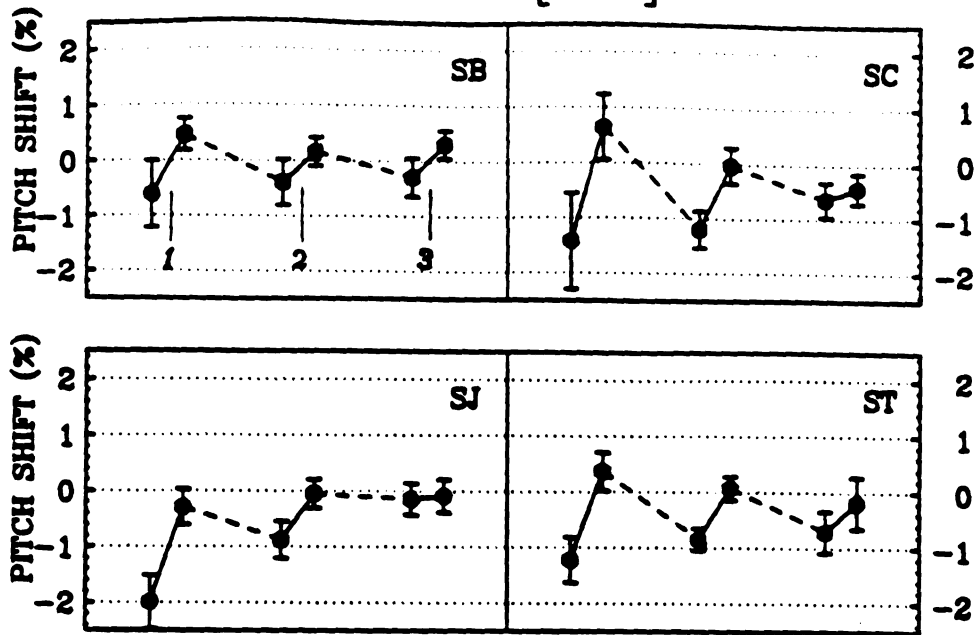
Secondly, the sizes of the splits differ from subject to subject, but there is a general tendency for low mistuned harmonics to give bigger splits. This is due to some other effects: the frequency effect and the harmonic number effect (see *F.* and *G.*).

Finally, the pitch shifts are not monotonic as a function of frequency. Since the mistuned harmonic is in a gap, it can be viewed in another way. We can organize the six stimuli in the following way (for mistuned 1/2/3 in a gap):  $-8M1$ ,  $+8M1$ ,  $-8M2$ ,  $+8M2$ ,  $-8M3$ ,  $+8M3$ . The frequency domain picture now is a component moving in several steps from low frequency ( $-8M1$ ) to high frequency ( $+8M3$ ) in the gap. If a



**Figure M2:** Spectra for Experiment 1. Spectral lines are represented by vertical solid lines. A vertical dashed line represents a mistuned harmonic. A dot is put on the horizontal axis if a harmonic is omitted.

## ZIGZAG [1-3]



## ZIGZAG [2-4]

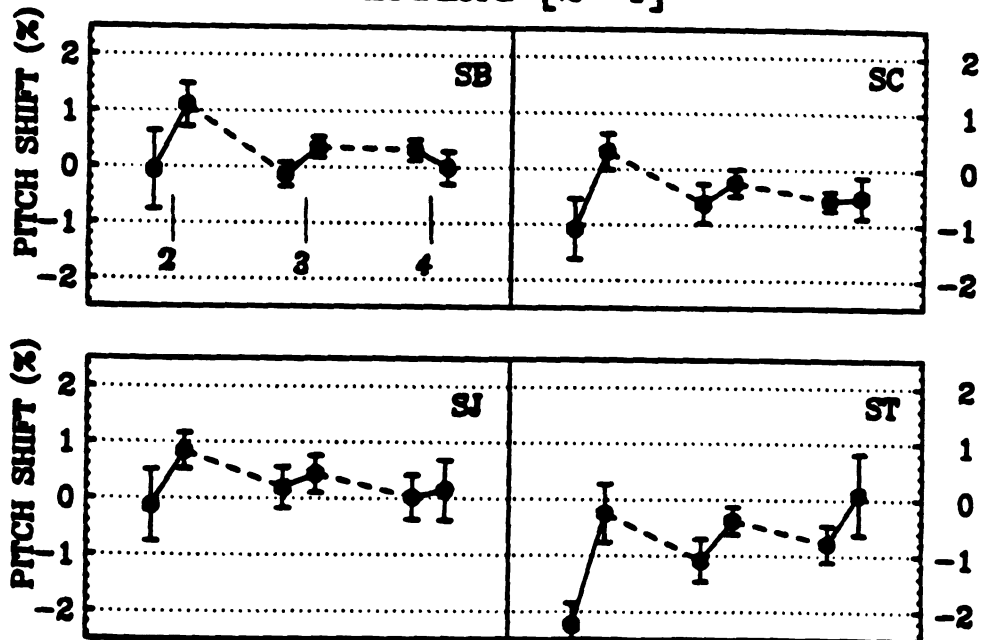


Figure M3: Results of Experiment 1 (Zigzag Effect). There are four panels for experiment "mistuned 1/2/3 in a gap" and four panels for "mistuned 2/3/4 in a gap" separately. Each panel shows results from one subject. There are three pairs of data in each panel representing the three  $\pm 8\%$  mistuned pairs in each sub-experiment. A pair is connected with a solid line so that one can compare the split of the pair with the split of another pair by comparing the slopes of the two solid lines. The end of a harmonic pair is connected to the beginning of the following harmonic pair by a dashed line so that one can see the zigzag pattern, which is the important result of the experiment.

local-interaction model is valid, the shifts of the six matches should be a monotonic function of frequency. However, a zigzag pattern is formed. The zigzag pattern shows that a local-interaction model is not appropriate. It suggests that a split can only occur when the pair straddles the special position where a harmonic is supposed to be. In the gap there are three of these special positions, therefore, three splits occur. This is evidence that the complex tone forms a template in the frequency domain.

### *C. EXPERIMENT 2*

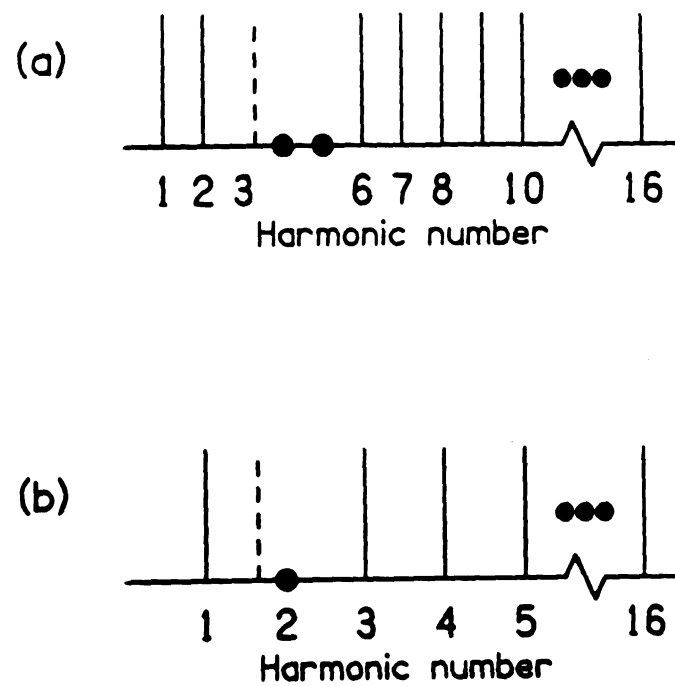
#### 1. Spectra of the Complex Tones

The spectra of the complex tones in this experiment are shown in Figure M4. We want to compare the two conditions (a) and (b). The idea here is to create the same local structures but different frameworks in the frequency domain. For condition (a), the components of the complex tone are perfectly tuned with three successive harmonics 3, 4, 5 omitted creating a gap as in experiment 1. The fundamental frequency is 200 Hz, and the highest harmonic is the 16th. Then we do an experiment with mistuned 3rd harmonic in this gap, i.e. two matches for  $\pm 8\%$  mistunings separately, whose frequencies are 552 and 648 Hz. For condition (b), the complex tone is an octave higher with fundamental frequency at 400 Hz. The second harmonic is omitted. A component whose frequency is the same as the mistuned 3rd harmonic in condition (a) i.e. 552 Hz or 648 Hz is added to the complex tone spectrum. We may call the component a mistuned "1.5," since the fundamental frequency is twice that in condition (a). Now the local frequency structure for the mistuned 3rd harmonic in case (a) is the same as that for mistuned harmonic "1.5" in case (b). Notice that the 2nd harmonic and the 6th

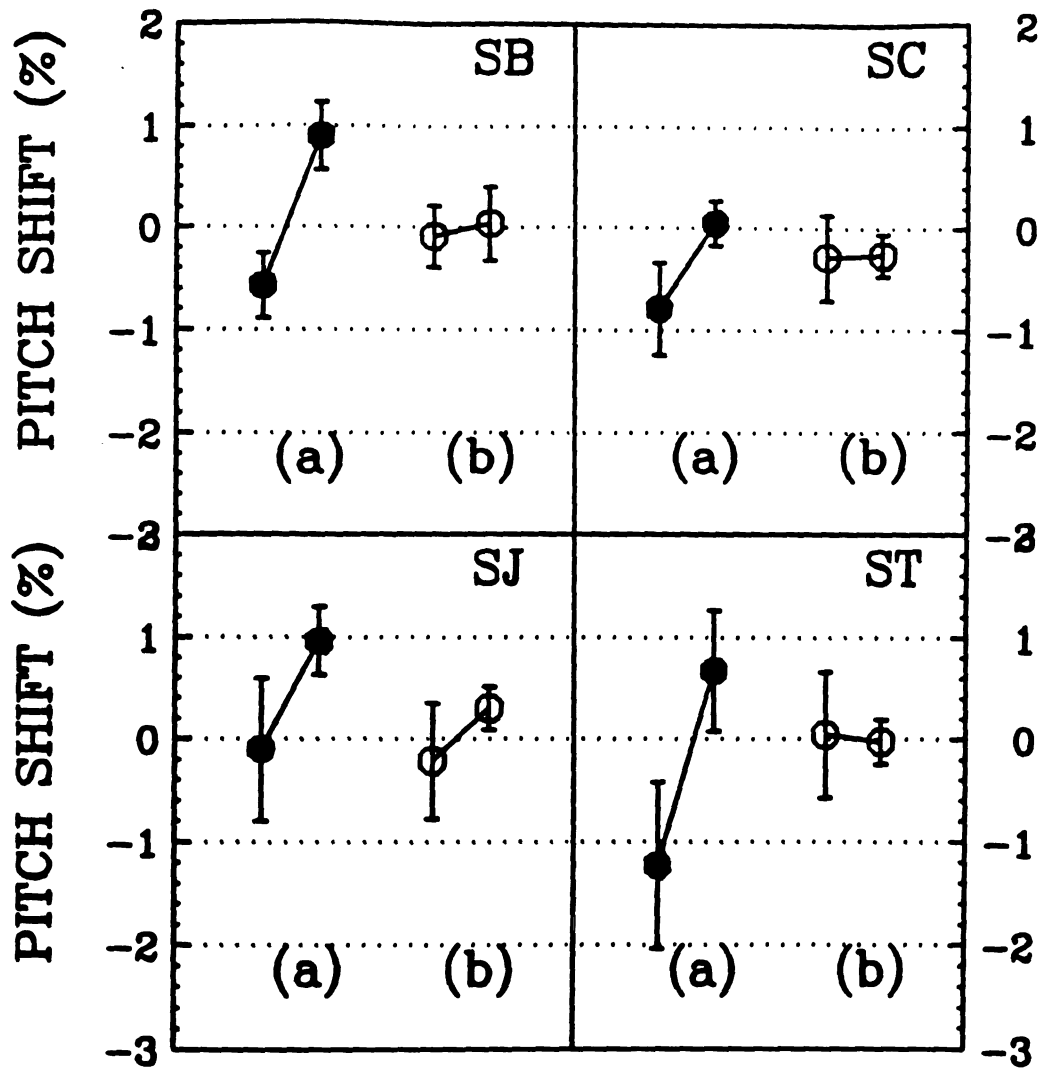
harmonic in condition (a) correspond to the fundamental and the 3rd harmonic in condition (b).

## 2. Results - The Integer Effect

The results are shown in Figure M5. Each panel is for one subject. As we can see, there are substantial splits for condition (a), and negligible splits for condition (b) compared to (a). And that is true for all subjects. In condition (a) the pair straddles harmonic 3 whereas in condition (b) the pair is on the low side of harmonic 2. In other words, condition (b) is a mistuned "1.5," but there is no such thing as a 1.5 harmonic.



**Figure M4:** Spectra for Experiment 2.



**Figure M5:** Results of Experiment 2 (Integer Effect). As before, each panel is for one subject. There are two pairs in each panel, one for one condition.

Therefore the effect is called the integer effect (a split occurs only if the mistuning pair straddles a harmonic). This clearly shows that there is a template formed by the complex tone.

For local interaction models, the pitch shifts of a mistuned harmonic are mainly determined by the closest components. For example, for the masking model of Terhardt, the pitch shift of the mistuned harmonic is mainly from the masking of the neighboring two components. Therefore, the model would predict approximately the same pitch shifts for both conditions (a) and (b), since the local spectra for both cases are the same. The same thing can be expected to happen for any other local-interaction models (including the timing model), because nearest neighbors are likely to be dominant. Thus, the integer effect cannot be explained by local-interaction models.

#### *D. EXPERIMENT 3*

##### 1. Spectra of the Complex Tones

The main purpose of this experiment is to see whether spectrally distant harmonics make contributions to splits of the mistuned harmonics. The spectra of the five complex tones are shown in Figure M6. They are: (a) 16-component complex tone with its fundamental mistuned. The fundamental frequency is 200 Hz; (b) same as (a) except that the 2nd harmonic is omitted; (c) same as (a) except that harmonics 2, 3 are omitted; (d) same as (a) except that harmonics 2, 3, 4 are omitted; (e) 3-component complex tone with its fundamental mistuned.

##### 2. Results - The Proximity Effect



The results are shown in Figure M7. All the subjects show the same effect. As the neighboring harmonics 2, 3, 4 are omitted one by one, the split becomes smaller and smaller. But some split is still there. This means that the far-away components do create a splitting effect. For conditions (b), (c), (d), the mistuned fundamental and the complex tone are very far apart. They are definitely in different peripheral auditory filters, since the critical bandwidth at the fundamental is much less than the spacing. Excitation interactions between them should be negligible compared to condition (a), which therefore, predicts much smaller pitch shifts compared to (a). But that is not the

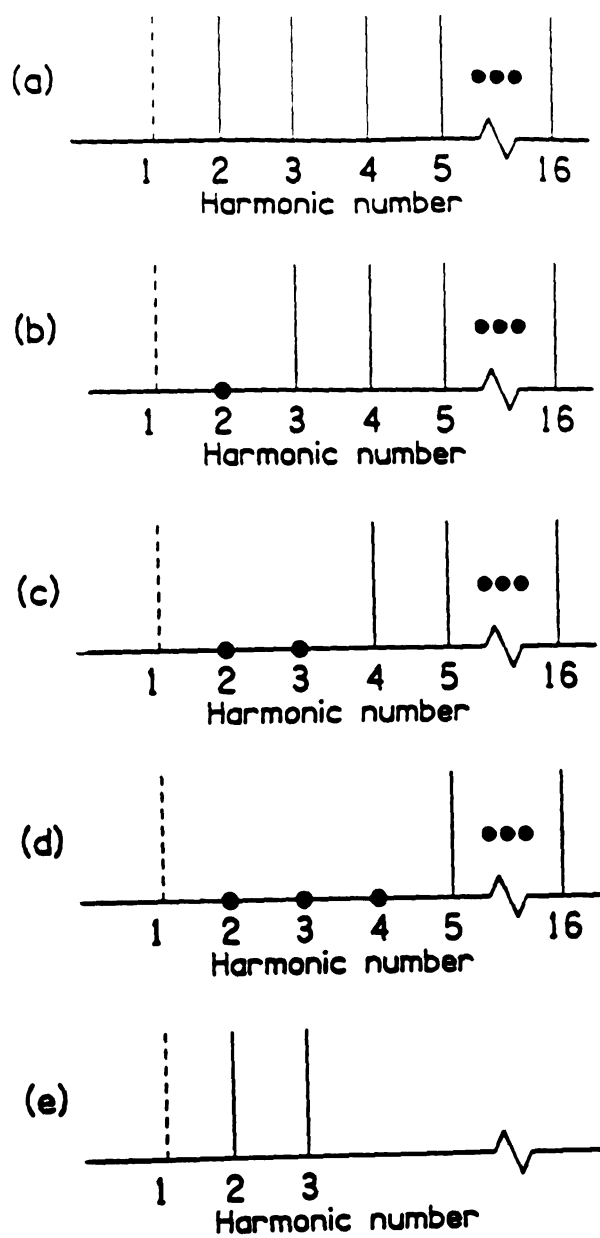
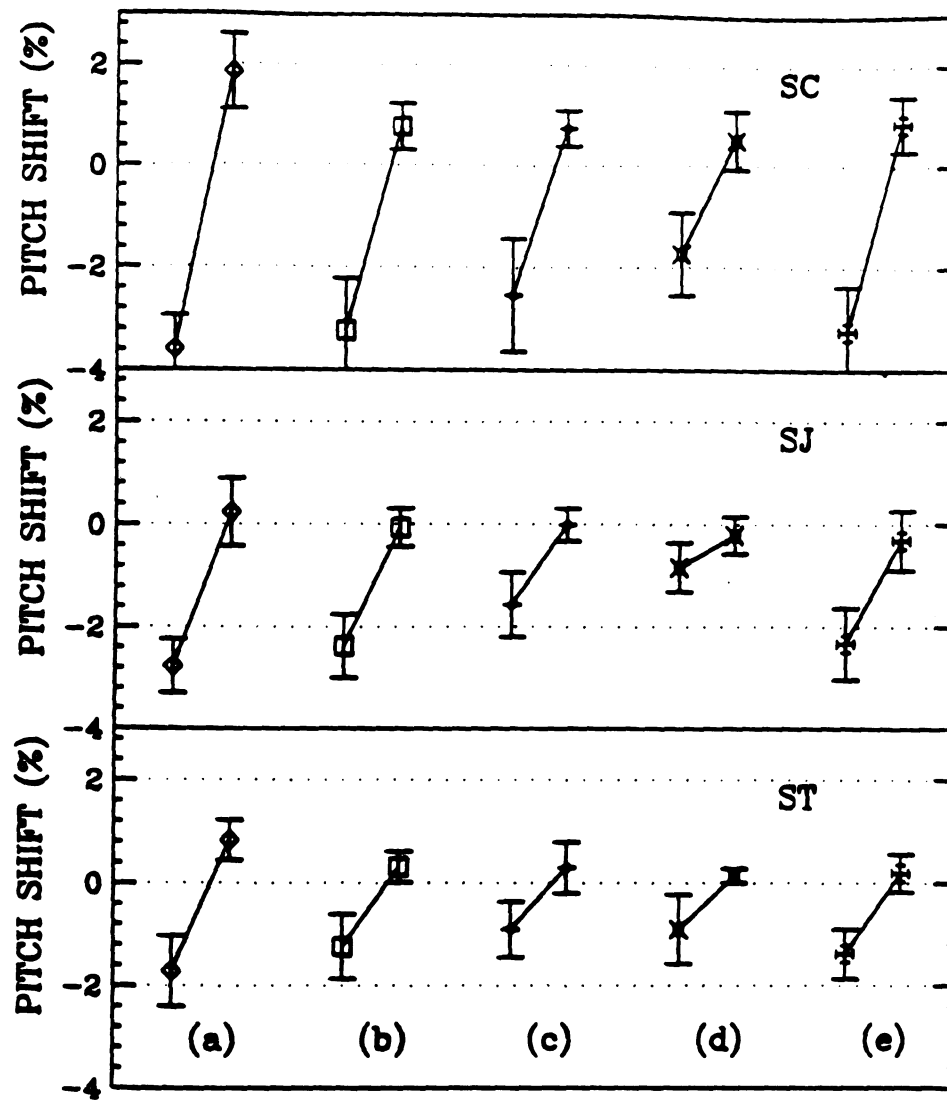


Figure M6: Spectra for Experiment 3.



**Figure M7: Results of Experiment 3(The Proximity Effect).**

case. The existence of splits for conditions (b), (c), (d) implies that the interaction is at a high level of the auditory system, which indicates that the formation of pitch is at high level of our auditory system. Somehow, the outputs of the peripheral filters are combined at high levels of the auditory system to form the pitch.

On the other hand, if we compare the split in (c) and (d) with the split in (e), we find the neighboring harmonic(s) are much more important in making the split. In other words, the closer the harmonic, the more contribution it makes to the framework.

#### *E. EXPERIMENT 4*

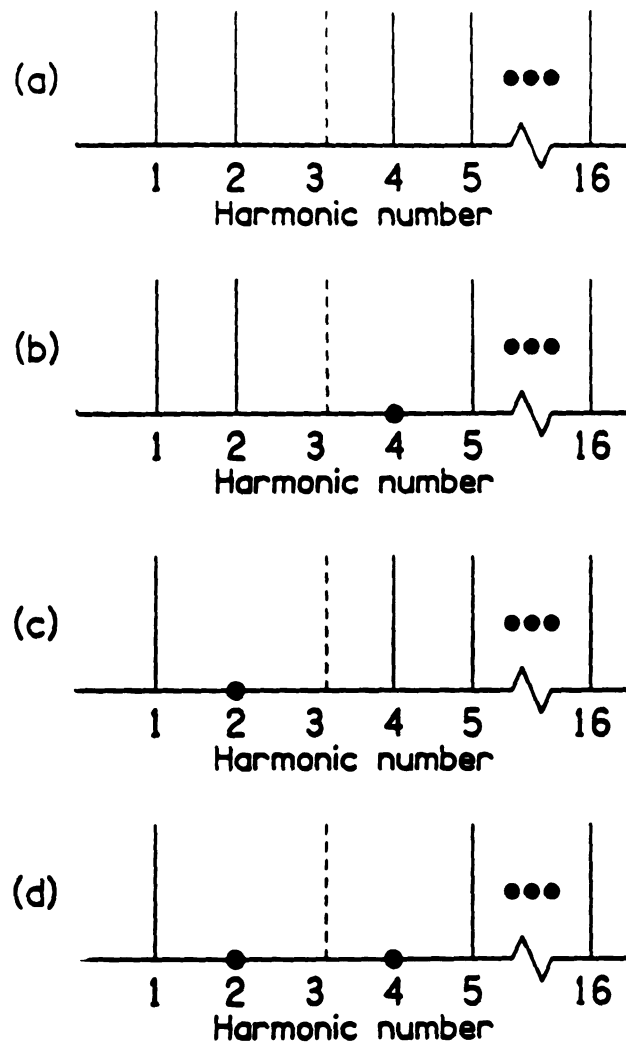
##### 1. Spectra of the Complex Tones

In this experiment spectral asymmetries are created for the mistuned harmonic. There are two purposes of the experiment. One is to see the importance of the neighboring harmonics in producing the splits. The other is to see whether the local-interaction models can explain the data for the asymmetric conditions. The spectra used here are shown in Figure M8. There are four conditions: (a) 16-component complex tone with its 3rd harmonic +8% or -8% mistuned; (b) same as condition (a) except that the 4th harmonic, which is a neighboring harmonic of the mistuned component, is omitted; (c) same as condition (a) except that the 2nd harmonic is omitted; (d) same as condition (a) except that both the 2nd and the 4th harmonics are omitted. There are eight matches in the experiment, a pair ( $\pm 8\%$  mistuning) for each condition.

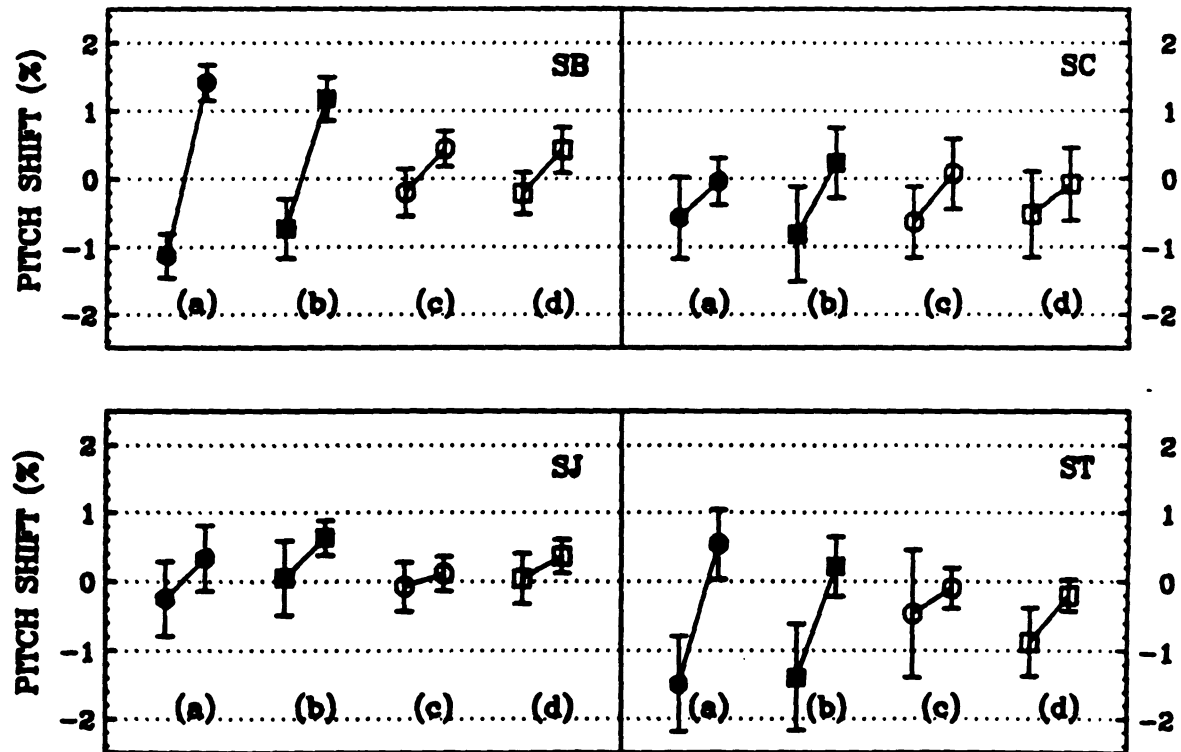
##### 2. Results - The Local Asymmetry Effect

The data are shown in Figure M9. As before, each panel is for one subject. For all the four conditions, splits are observed. For two subjects, SB and ST, removing the nearest harmonics dramatically reduces the size of the split, a proximity effect. All the subjects show that the lower neighboring harmonic (in this case the 2nd harmonic) leads to a larger split than the higher harmonic (in this case the 4th harmonic).

One observation is to be made. In conditions (b) and (c), the neighboring harmonics of the mistuned harmonic are asymmetric. For a symmetric condition, say (a), the mistuned component is at the center of the gap between the two neighbors - 2nd and 4th.



**Figure M8:** Spectra for Experiment 4.



**Figure M9:** Results of Experiment 4 (Local Asymmetry Effect). Four conditions are labeled for each panel.

Once a neighbor is omitted, asymmetry is created, and the mistuned component is located off the center of the gap between the two new neighbors. In a local-interaction model, the pitch shift is almost entirely determined by the two neighbors. If a local-interaction model is true, the asymmetry caused by omitting one neighbor (only lower neighbor for the excitation model, since the excitation comes mostly from the lower side) would dramatically change the balance between the attractions (or repulsion) from high side and low side, because the pair is far from the center of the gap due to the asymmetry. For example, in condition (c), the neighboring harmonics are 1st (the fundamental) and 4th instead of 2nd and 4th, since the 2nd is omitted. This would cause a large uni-directional pitch shift (compared to the split) for the mistuned pair. For the excitation model the shift should be negative, since there is now less “repulsion” from below (the 2nd is replaced by the 1st). For the timing model the shift should be positive, since there is now less “attraction” from below. However, that is not observed. Except for their sizes, the splits are not changed by local asymmetry. That is because the pair are still straddling the special position - the position where the 3rd harmonic is supposed to be. There are no substantial uni-directional shifts compared to the sizes of splits for the asymmetric pairs - condition (b) and (c). That is true for all four listeners. Therefore, local-interaction models do not apply to this effect.

## *F. EXPERIMENT 5*

### *1. Spectra of the Complex Tones*

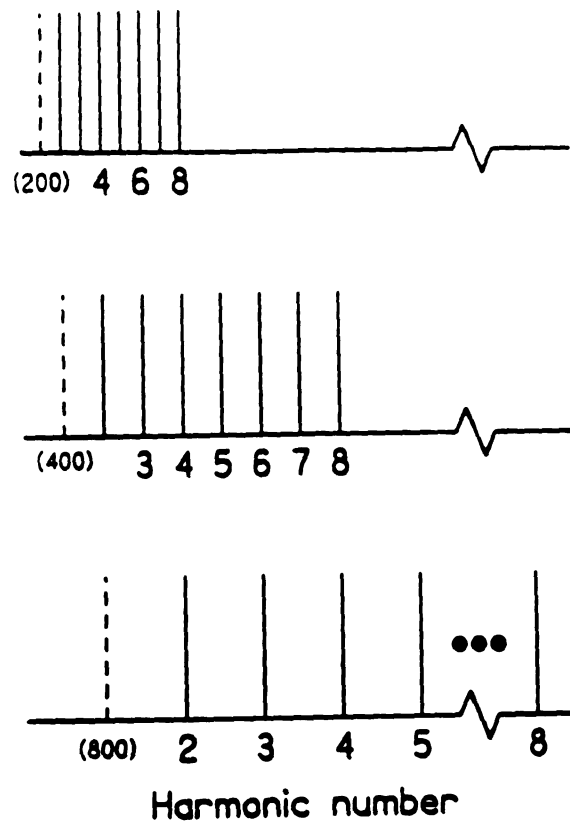
This experiment was done to see the how splits for the mistuned fundamental depended upon frequency. The spectra of the three complex tones used in this experiment are shown in Figure M10. All the complex tones here are 8-component complex tones with



their fundamentals mistuned. The only difference among them is the fundamental frequency: (a) 200 Hz; (b) 400 Hz; (c) 800 Hz.

## 2. Results - The Frequency Effect

Results are shown in Figure M11. The results for all the subjects are very similar and obvious: The split decreases with increasing fundamental frequency of the complex tone. The splits at 800 Hz are almost zero. Modern pitch perception models suggest that the periodicity pitch is extracted by a harmonic pattern recognizer (e.g. the DWS



**Figure M10:** Spectra for Experiment 5.

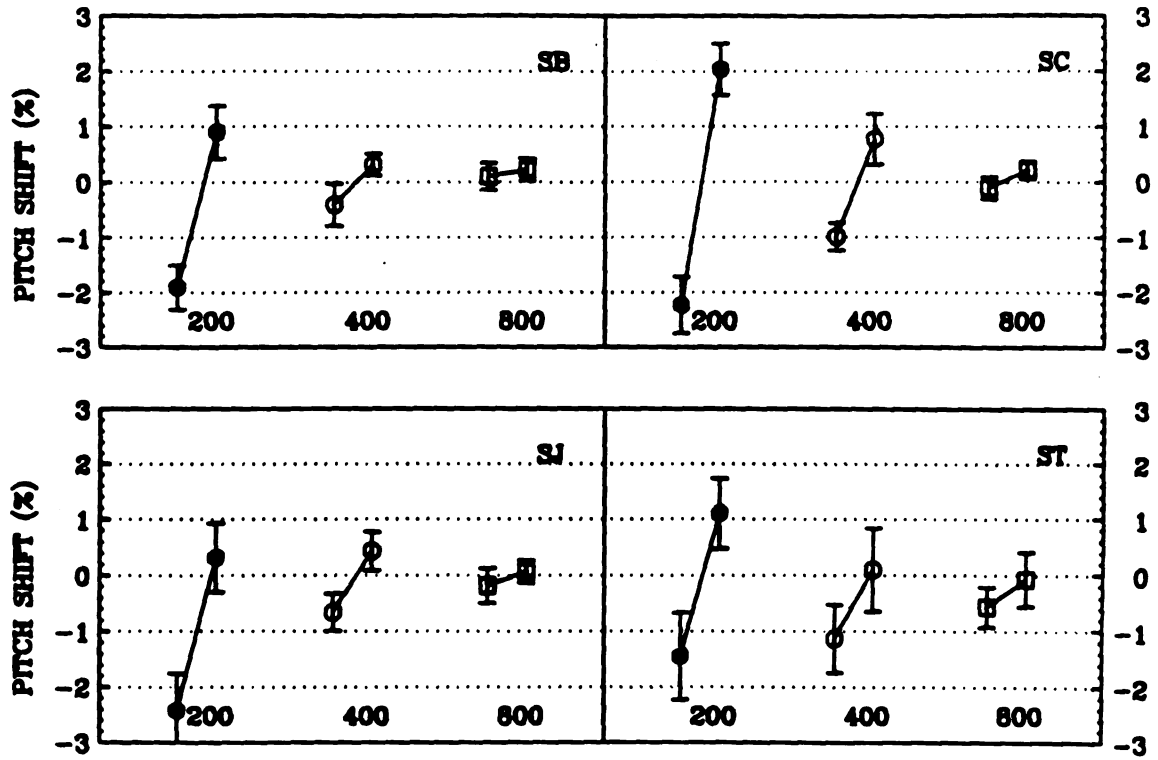


Figure M11: Results of Experiment 5 (The Frequency Effect).

pitch meter, Scheffers 1983). The results of our experiment, that the framework is almost absent for a fundamental above 800 Hz, would imply that 800 Hz is the limitation for the fundamental of the periodicity pitch. This is consistent with Ritsma's results for the existence region of the sensation of periodicity pitch (Ritsma, 1962). Therefore, splits appear to be related to pitch formation of a complex tone.

## *G. EXPERIMENT 6*

### **1. Spectra of the Complex Tones**

The goal of this experiment is to explore the dependence of splits on harmonic number. The idea is to create the same local spectral environment for two harmonics having different harmonic number. The spectra are shown in Figure M12. Condition (a): the second harmonic of the complex tone is mistuned and its neighboring harmonics, the fundamental and the third harmonic, are omitted. The fundamental frequency is 200 Hz, and the highest harmonic is the 12th. Condition (b): The stimulus in this condition is a 8-component complex tone with its fundamental mistuned. The fundamental frequency of the complex tone is 400 Hz. The local structures in the frequency domain for the mistuned harmonic in both cases are the same, but the harmonic number in (a) is one (the fundamental) and the harmonic number in (b) is two (the 2nd harmonic). The power of the two stimuli is the same.

### **2. Results - The Harmonic Number Effect**

Results are shown in Figure M13. All the subjects show that the lower harmonic number (in this case the fundamental) suffers the larger split. Visually, it seems that the

effect is not as strong as other effects. Therefore, we did a one-tailed t-test on the differences in each split. The test showed that at the  $\alpha=0.01$  level the lower harmonic number has the bigger split.

## ***H. EXPERIMENT 7***

### **1. Spectra of the Complex Tones**

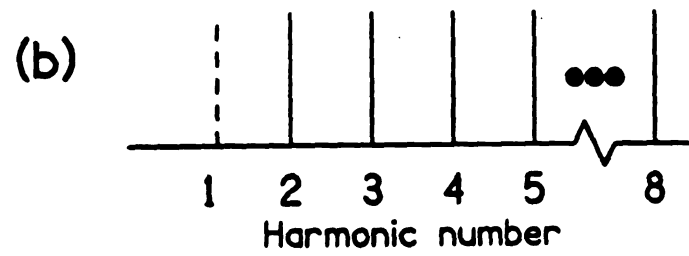
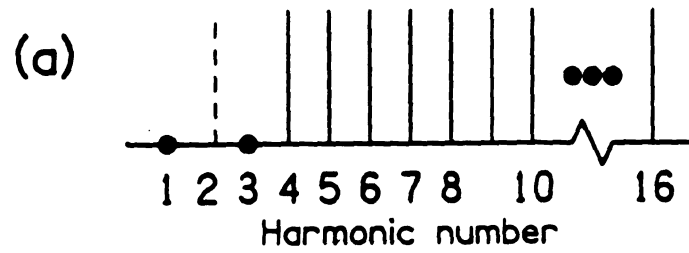


Figure M12: Spectra for Experiment 6.

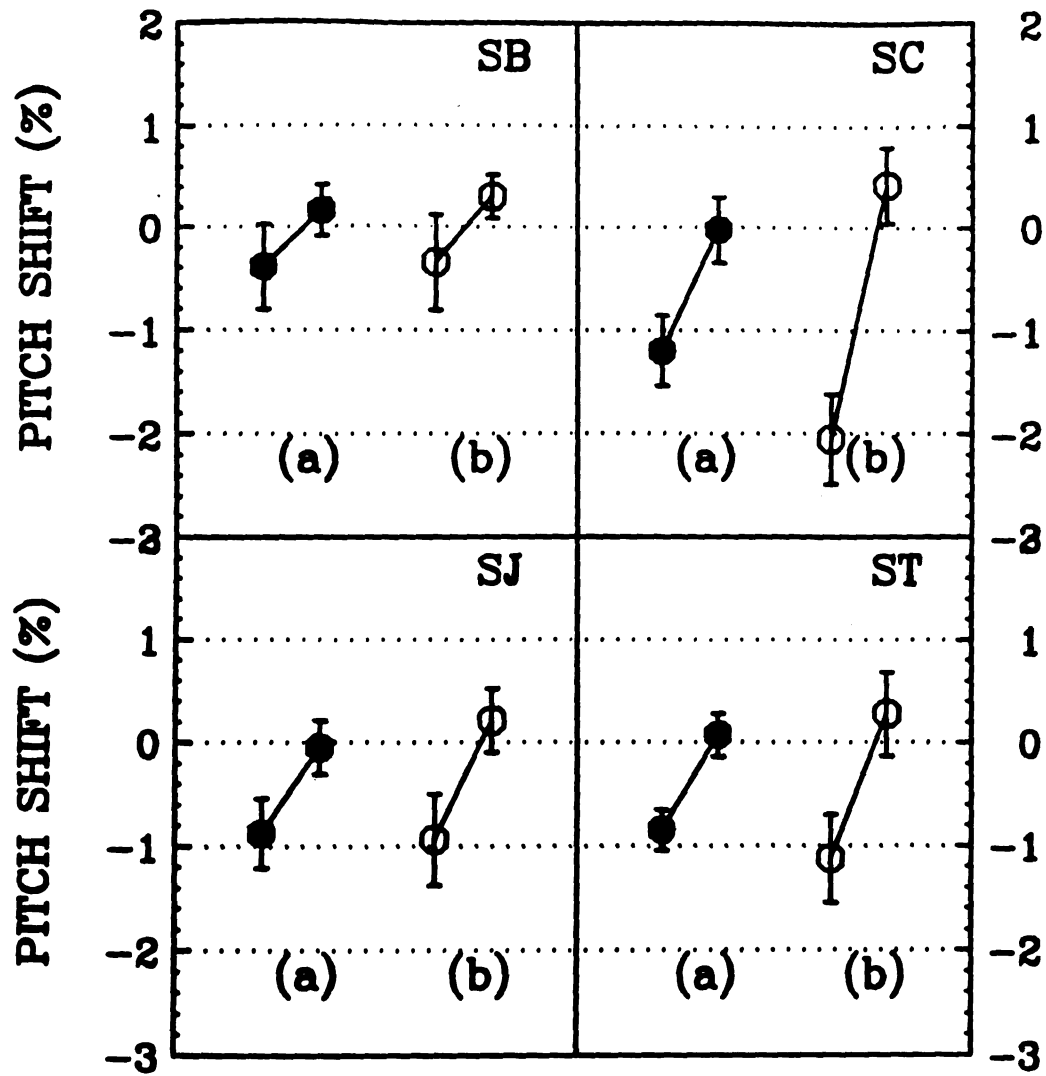


Figure M13: Results for Experiment 6 (The Harmonic Number Effect).

Previous experiments have indicated that the framework is important in the pitch of mistuned harmonics. In this experiment, we try to control the framework based upon the conjecture that if the 3rd is mistuned then the multiples of the 3rd will contribute more to the framework than other harmonics. This conjecture comes from the idea that the bases of a template may lie in the timing pattern of neural firing. Those harmonics that are integral multiples of three will synchronize with the 3rd harmonic and therefore may form a more effective template. Note in this experiment only harmonics higher than 3rd are considered since we know that lower harmonics are much more important than the higher harmonics (see local asymmetry effect and the top-bottom effect). The spectra of the stimuli in this experiment are shown in Figure M14. There are four conditions. All of them are mistuned 3rd with harmonics higher than 3rd up to 21st. Conditions (a) and (b) periodically omitted harmonics which are not multiples of 3rd. Conditions (c) and (d) omitted harmonics which are the multiples of the 3rd, i.e. 6th, 9th, 12th, etc.

## 2. Results - Controlling the framework

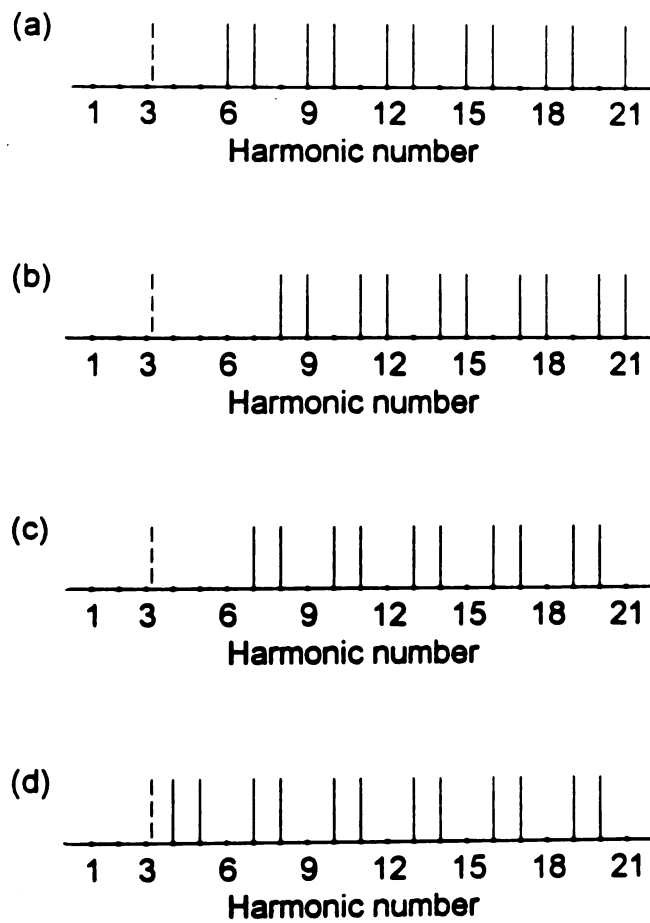
The data are shown in Figure M15. All the subjects except for SB show the effect that (a) and (b) give a larger split than (c) and (d). This confirms our conjecture that multiples of the harmonics leads to a stronger framework.

### *1. EXPERIMENT 8*

#### 1. Spectra of the stimuli



The pitch shift made by the brain is an exaggeration of the mistuning. When there is no mistuning, the exaggeration of 0% mistuning is expected to be zero. This may be a “neutral” or “equilibrium” position. However, if this equilibrium point exists, it should be an unstable equilibrium, because any deviation (perturbation) from the point would be “amplified” by the brain, since the exaggeration of mistuning is a positive feedback. This would result in a instability at 0% mistuning. In this experiment, we want to check whether this kind of “instability” exists. Four conditions of stimuli (Figure M16) are made to check the instability. All of them are mistuned fundamental. The mistunings are: -8%, +8%, 0%. Condition (a) is mistuned fundamental with



**Figure M14:** Spectra for Experiment 7.

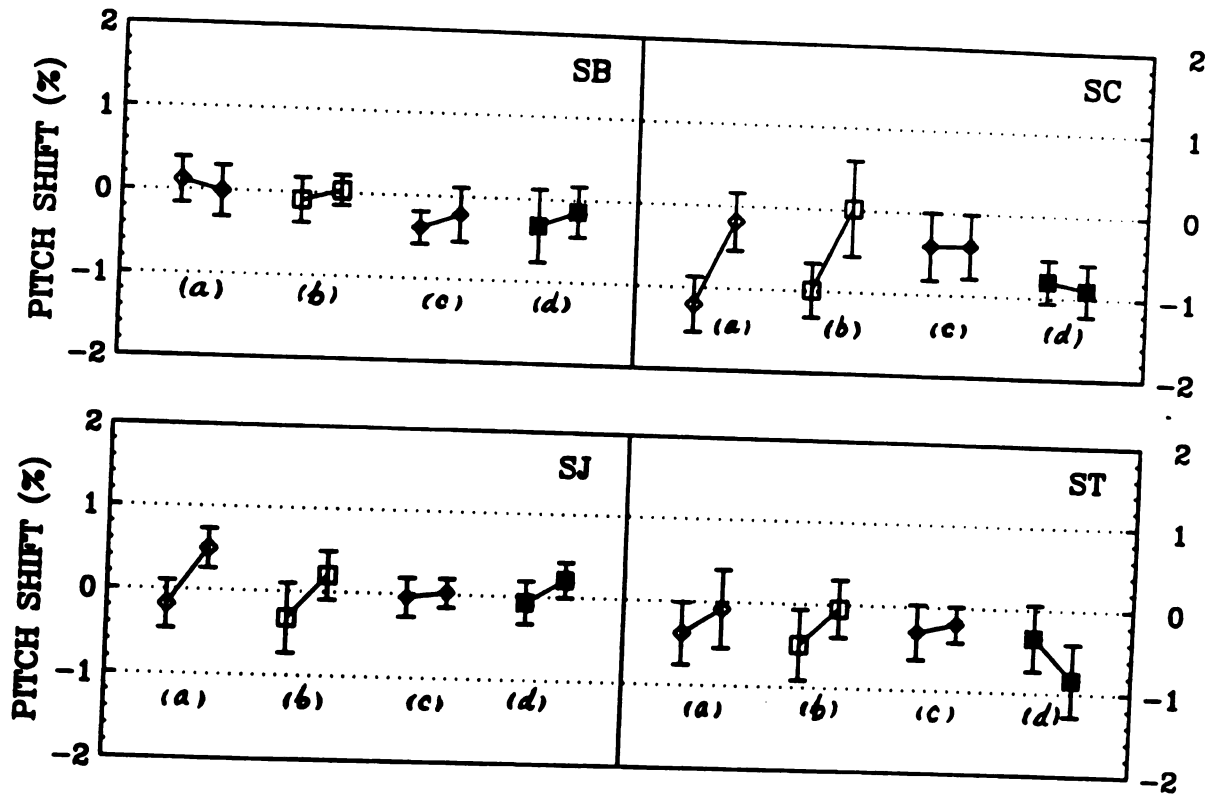
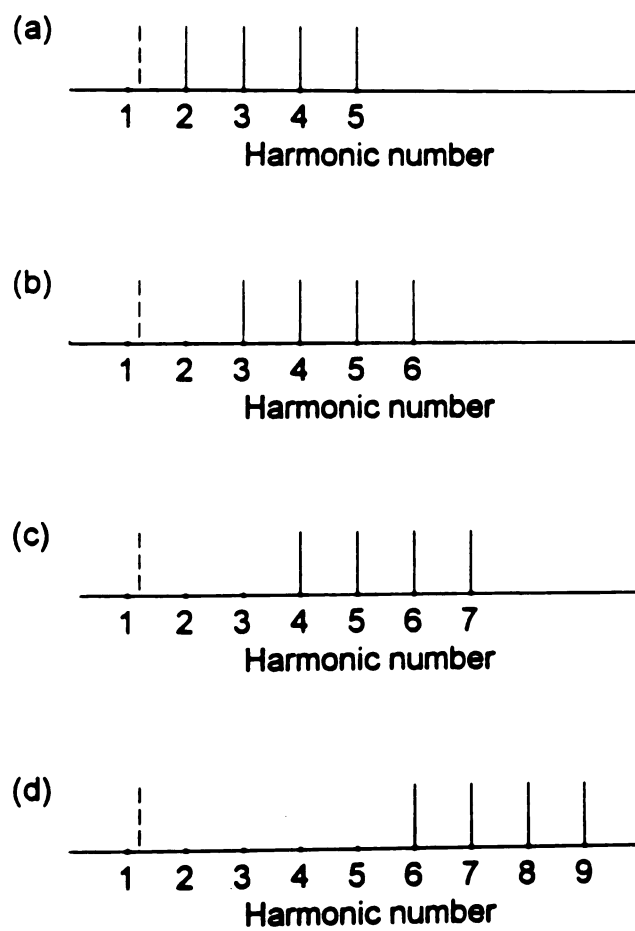


Figure M15: Results of Experiment 7 (Controlling the Framework).



**Figure M16:** Spectra for Experiment 8.

harmonics 2 to 5; (b) is mistuned fundamental with harmonics 3 to 6; (c) is mistuned fundamental with harmonics 4 to 7; (d) is mistuned fundamental with harmonics 6 to 9.

## 2. Results - The Instability Check

A instability might be recognized by two effects. One would be a large error bar for 0% mistuning. The other effect would be that the pitch shift for 0% mistuning deviates a lot from the center of pitch shifts of the corresponding  $\pm 8\%$  mistuning pair.

Figure M17 is the plot of the data. No larger error bars are observed for 0% mistuning points. There do exist some 0% mistuning points whose pitch shifts are greatly off the "center point." However, there are also some points whose shifts are at the "center point." Therefore, it is hard to say whether an instability is observed or not. The data from Hartmann and Doty (1996) showed no instability.

Another result from this experiment is that it gives additional evidence for the proximity effect. As we can see, the split decreases with increasing distance between the mistuned harmonic and the harmonics producing the framework.

## *J. EXPERIMENT 9*

### 1. Spectra of the stimuli

The spectra are plotted in Figure M18. There are four conditions: (a) mistuned 3rd harmonic as the top frequency, i.e. the complex tone has only 3 components harmonic 1 to harmonic 3; (b) mistuned 3rd harmonic as the bottom frequency, i.e. the complex

tone has components 3 to 16; (c) mistuned 5th as the top frequency; (d) mistuned 5th as the bottom frequency.

## 2. Results - The Top-Bottom Effect

Figure M19 shows the data for this experiment. There are three results: (1) Lower harmonics makes larger splits than higher harmonics. It is easy to see that top conditions, (a) and (c), give larger split than bottom conditions. (2) There are some masking effects which give some overall pitch shifts to each pair. As expected, this effect usually gives positive shifts to the top conditions and negative shifts to the bottom

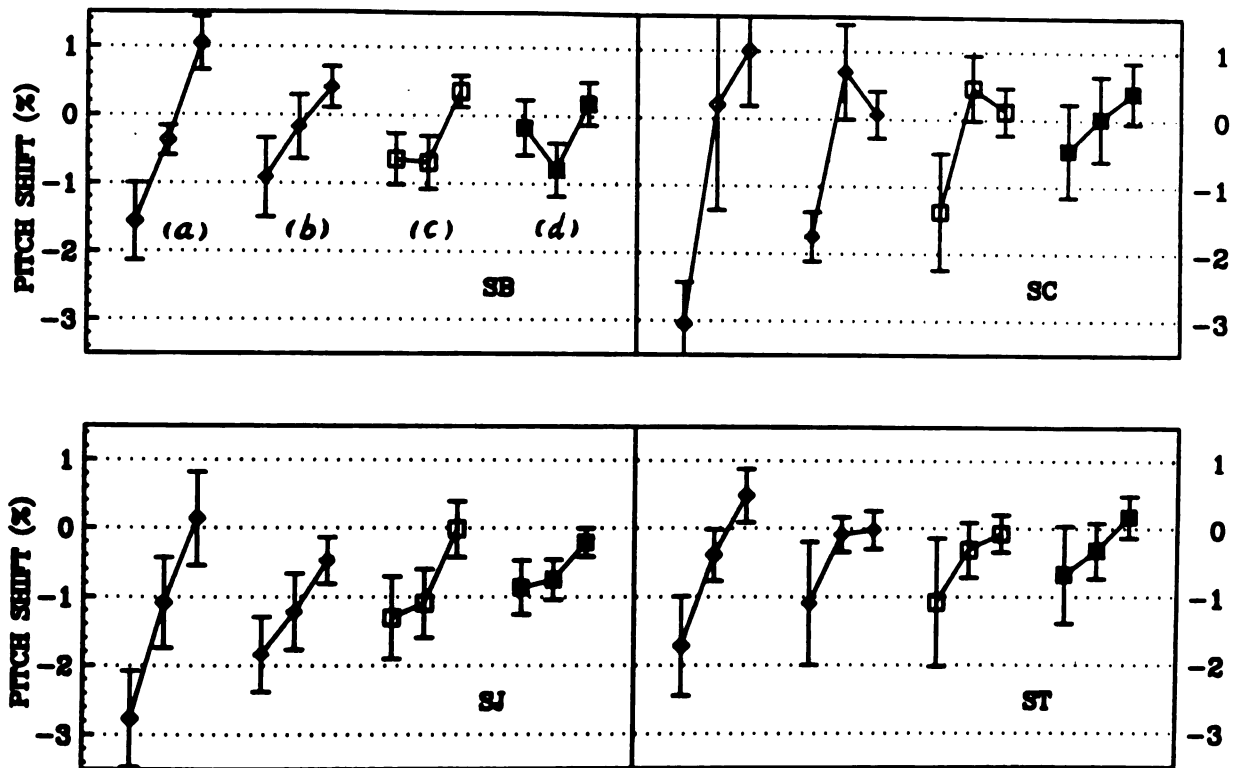
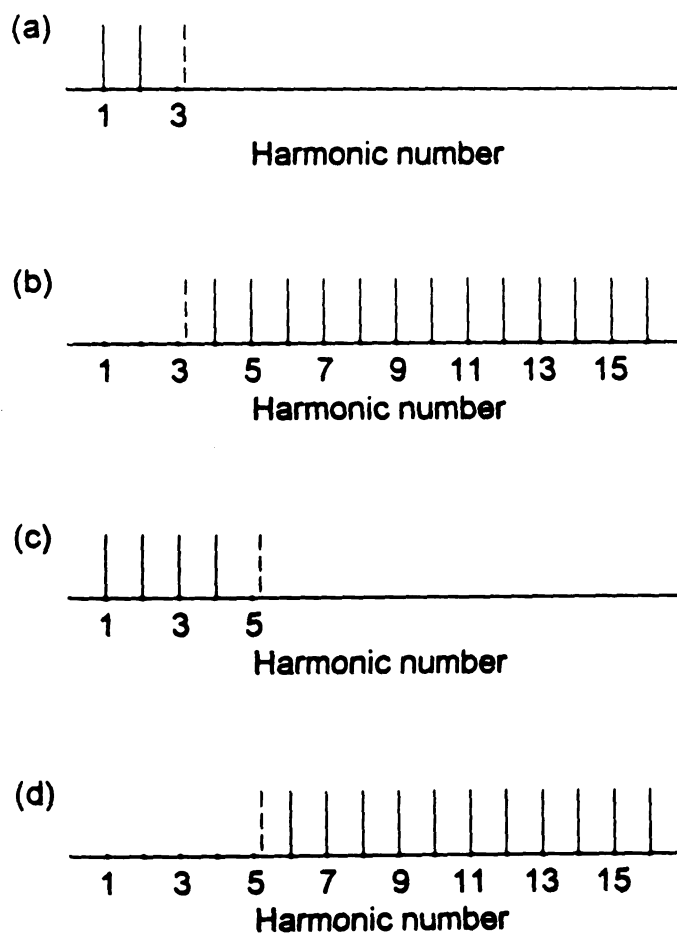


Figure M17: Results of Experiment 8 (Instability Check).



**Figure M18:** Spectra for Experiment 9.



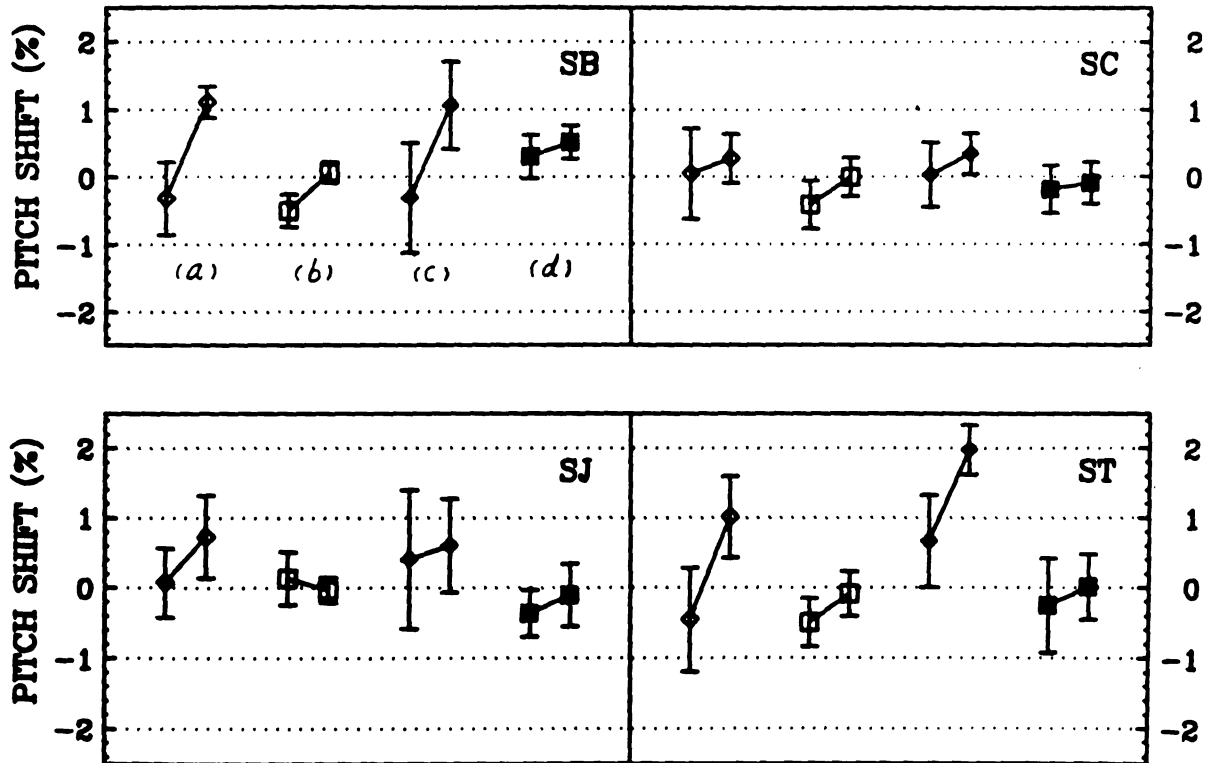


Figure M19: Results of Experiment 9 (Top - Bottom effect).

conditions - a “repulsion” effect. (3) The experiment is an additional case of the local asymmetry effect. The data show again that local asymmetry does not destroy the split, which conflicts with local-interaction models.

## II. Discussion and Conclusion

This report gives details about the pitch shifts of mistuned harmonics for low harmonics (harmonic number less than or equal to five) under various conditions. We choose low harmonics, because we believe for harmonic number less than six the splits are more stable and reproducible than for higher harmonics. The error bars are much smaller for lower harmonics (Hartmann *et al.*, 1990).

It is important to point out that in some of the experiments the results are a combination of many effects. For example, the effect of experiment 1 (the zigzag effect) is, in fact, a combination of three effects: (1) the harmonic number effect; (2) the frequency effect; (3) the (Terhardt’s) masking effect. The first two effects cause the splits to become smaller for higher harmonics. The third effect causes an overall shift for each split pair (see Figure M3). This can be seen by comparing the same mistuned harmonic in different gaps: 1-2-3 gap vs. 2-3-4 gap. Also, this masking effect makes the last split, i.e., the one on the high frequency side of the gap, less obvious. Since the highest mistuned components in the series are close to the high frequency edge of the gap, they tend to be shifted lower by masking from higher components.

From experiment 2, we find that splits occur only when the two compared mistuned components straddle the point where a harmonic is supposed to be (The integer effect). When the mistuned component is in a gap (the case where three successive harmonics

are omitted), the pitch shift of the mistuned component is not a monotonic function of frequency. In other words, as the mistuned component starts from the low edge of the gap, and continuously goes upwards toward the other end, its pitch shift changes direction several times (The zigzag effect). Consistent with the previous argument, we see splits where harmonics are supposed to be. Further, this zigzag feature shows that local-interaction is probably not the main mechanism for the split. The split cannot be mainly caused by masking. Both the integer effect and the zigzag effect strongly suggest that a template is formed by a complex tone.

Right now, we think that the split is a perceptual contrast enhancement in the process of segregation. When the component is perfectly tuned, it is a harmonic of the complex tone. Therefore, it is integrated into the complex tone. When the component is mistuned, it does not belong to the complex tone. It should come from a different sound source. The exaggeration of the mistuning made by the brain enhances the contrast between the complex tone and the sine tone. It makes it easier for us to segregate the two sounds.

It may be argued that although a mistuned harmonic can be heard out, it still makes a contribution to the complex tone, since the mistuning of the component affects the pitch of the complex tone (Moore and Glasberg 1985). We don't think that is a problem. First, we believe that there is no definite border to judge whether a component is a perfectly tuned harmonic or not. If cued, a perfectly tuned harmonic can be heard out. Second, while the effects of the mistuned component on the complex tone can be regarded as the contribution of the component to the complex tone, it can also be regarded as the interaction between the complex tone as an entity and the pure tone (the mistuned harmonic) as another entity.

The integration of a perfectly-tuned harmonic and the segregation of a mistuned harmonic are directly related to the mechanism of pitch perception of complex tones. The reason that the excitation masking model and the timing model cannot predict splits perfectly is that they ignore the role of the entire spectrum in integrating a correctly-tuned harmonic into the complex tone. Results of the experiments here show that interactions between components and the complex tone are not local. At high levels of the auditory system, a template, which is supposed to be related to the mechanism of pitch perception, is formed by the complex tone. In order to explain the splits more appropriately, the mechanism of template formation at the high level must be understood. In other words, in order to know how the contrast enhancement is made when a harmonic is mistuned we must understand how our auditory system integrates the (correctly tuned) harmonics to form a complex tone as an entity. At this time, there is no satisfactory theory of pitch perception.

There is another phenomenon which may reflect the same mechanism: poststimulatory pitch shifts for pure tones (Rakowski & Hirsh, 1980). When a long (500ms) pure tone is followed by a short (25ms) pure tone whose frequency is close to the long pure tone frequency, the pitch of the short tone is "pushed" away from the long tone's pitch, making a split around the long tone's pitch. The pattern of pitch shifts (Rakowski & Hirsh, 1980) is very similar to the pattern for the mistuned harmonics. The template in a complex tone may play the same role as the pure tone leading-stimulus. The contrast enhancement for distinguishing the two sources may make the split in both cases.

## REFERENCE

Duifhuis, H., Willems L.F. and Sluyter R.J. (1982) "Measurement of pitch in speech: An implimentation of Goldstein's theory of pitch perception," *J. Acoust. Soc. Am.* **71**, 1568-1580.

Goldstein, J. L. (1977) "Auditory-nerve spike intervals as an adequate basis for aural spectrum analysis," in *Psychophysics and Physiology of hearing*, ed. E.F. Evans and J.P. Wilson, (Academic, New York) pp. 337-345.

Hartmann, W.M. (1988) "Pitch perception and the segregation and integration of auditory entities," in *Auditory Function*, ed. G.M. Edelman, W.E. Gall and W.M. Cowan (Wiley, New York) pp.623-645.

Hartmann, W.M., McAdams, S., and Smith, B.K. (1990) "Matching the pitch of mistuned harmonic in an otherwise periodic complex tone," *J. Acoust. Soc. Am.* **88**, 1712-1724.

Hartmann, W.M. and Doty, S.L. (1996) "On the pitches of the components of a complex tone" *J. Acoust. Soc. Am.* in press

Moore, B.C.J. and Glasberg, B. R. (1985) "Relative dominance of individual partials in determining the pitch of complex tones," *J. Acoust. Soc. Am.* **77**, 1853-1860.

Moore, B.C.J., Glasberg, B. R. and Robert W.P. (1986) "Thresholds for hearing mistuned partials as separate tones in harmonic complexes" *J. Acoust. Soc. Am.* **80**, 479-483.

Rakowski, A. and Hirsh, I.J. (1980) "Post-stimulatory pitch shifts for pure tones," J. Acoust. Soc. Am. **43**, 764-767.

Ritsma, R.J. (1962) "Existence region of tonal residue I," J. Acoust. Soc. Am. **34**, 1224-1229.

Scheffers, M.T.M. "Simulation of auditory analysis of pitch: An elaboration on the DWS pitch meter," J. Acoust. Soc. Am. **74**, 1716-1725.

Terhardt, E. (1974) "Pitch, consonance and harmony," J. Acoust. Soc. Am. **55**, 1061-1069.

Terhardt, E. (1979) "Calculating virtual pitch," Hearing Research, **1**, 155-182.

Terhardt, E., Stoll, G. and Seewann, M. (1982a) "Pitch of complex signals according to virtual-pitch theory: Test, examples, and predictions," J. Acoust. Soc. Am. **71**, 671-678.

Terhardt, E., Stoll, G. and Seewann, M. (1982b) "Algorithm for extraction of pitch and pitch salience from complex tonal signals," J. Acoust. Soc. Am. **71**, 679-688.

## **CHAPTER 3**

**The Duifhuis effect -  
New Measurements Require a Revised Explanation.**

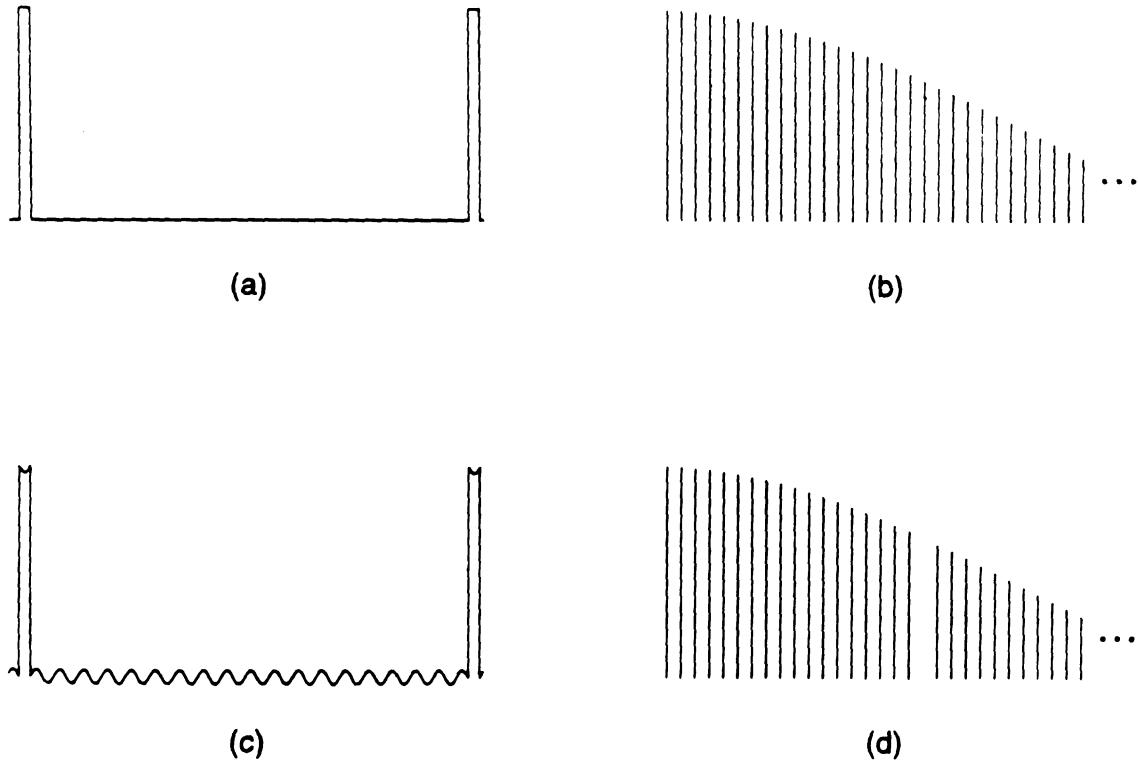
## Introduction

Duifhuis (1970, 1971) described an effect that might be called “the pitch of the tone that is not there.” When we listen to a slow, periodic train of narrow pulses with its period at, say 20ms, we hear a 50Hz complex tone with a buzzy timbre. It is interesting that when one of the high harmonics, for example the 19th, of the complex tone is omitted, the absent harmonic is heard out. What is heard by listeners is exactly what is not present in the signal. However, an oscillographic tracing of the waveform shows a clear small sine-tone oscillations in the time gap (see Figure D1). This is understandable, because the signal now can be thought of as the sum of two signals: the complex tone (periodic narrow pulse train) including the canceled harmonic and a cancellation tone (180° out of phase with the canceled harmonic). Below, when we talk about a cancellation tone, we have this picture in mind, no matter how our signal is generated. In our experiments, it is generated digitally by omitting one harmonic, and not by adding a cancellation tone. But a cancellation tone can always be imagined to be there, and it is useful to think of it that way.

Traditionally, the phenomenon is ascribed to the peripheral frequency analysis of our auditory system (Duifhuis, 1970, 1971; Alcantara and Moore, 1995). The basic idea is that the filters in the peripheral auditory filter bank are wide compared to the spacing of the harmonics. Therefore, the small oscillations of the cancellation tone in time gaps of the waveform result in some sine-tone-like output at the frequency of the cancellation tone within some duration in each period. An equivalent way of saying this is that the windowing time (the impulse response time of the filters) of the peripheral auditory system is short compared to the period of the complex tone so that during a part of each period the output of the peripheral auditory system comes mainly from the small



oscillations of the cancellation tone, which of course results in a peak in the output spectrum at the frequency of the cancellation tone. This is the concept of short-term Fourier analysis (for short-term Fourier analysis, see Deller, 1993). These traditional ideas will be thoroughly illustrated and discussed later. From informal experiments, we found that the traditional way of explaining the phenomenon might not be correct. Therefore, we explored the phenomenon and tested the traditional peripheral explanation by experiments.



**Figure D1:** Duifhuis effect: (a) A periodic narrow pulse train with a period of 20 ms; (b) Spectrum of the waveform in (a); (c) The narrow pulse train with the 19th harmonic omitted; (d) Spectrum of the waveform in (c).

## **I. Experiments Exploring the Duifhuis Effect**

### *A. Method*

#### **1. Procedure**

The listener was seated in a sound-treated room, holding a response box that controlled the events of an experimental trial. When the listener pressed a yellow button there was a pause of 200 ms, and then a complex tone, with one or more of its harmonics either omitted or phase inverted. When the listener pressed an orange button there was a pause of 200 ms, and then a sine tone with a frequency that could be adjusted by means of a ten-turn potentiometer on the box. The potentiometer allowed the listener to make the pitch match. The listener could call up the complex tone or the matching tone as often as he liked. When the listener was satisfied with his match he pressed the green button to finish the trial. The stimulus and the matching frequencies were recorded, and then the next trial with different omitted or phase-inverted harmonic(s) began. There was no feedback to the listener.

When level matching was not required in a experiment, the level of the matching tone could be easily controlled through a knob. If level matching was required, as it was in some experiments, the listener was asked to adjust the level through another box on which there two buttons - one to increase the level, the other to decrease the level. Every time one of the buttons was pushed the level of the sine tone was increased or decreased by 1 dB.

Trials were blocked into runs. Each experimental run included four trials. After a run was completed, the listener could come out to rest. It took about 1 to 5 minutes for a

listener to finish a run. In the results to be reported below, the data from the final five matches were used for each data point.

## 2. Stimuli

There were two kinds of stimuli in the experiments, complex tones and sine tones. The complex tones were generated digitally by a TDT System II. Sinusoidal waveforms of the harmonics were summed up to produce the waveform of the complex tone. One of the harmonics was omitted or phase inverted in order to produce the Duifhuis effect. The high spectral ends of the stimuli are tapered to reduce edge pitches. The fundamental frequencies of all the complex tones in this paper were 50 Hz. The level of the stimuli was at 45 dB/component for a flat-spectrum complex tone.

The matching tone was generated by a Wavetek function generator. The frequency was controlled by a voltage from the listener's control box and was read by the computer through a Metrabyte CTM5 card. The matching tone went through the PA4 attenuator of the TDT system so that its level could also be controlled and read by the computer.

## 3. Listeners

Three male listeners SB, SD, SJ participated in the experiments. Their ages were 56, 20 and 34 respectively. All of them could perform a pitch-matching task accurately. All the listeners had negative otological histories and had some training as performers of musical instruments. Audiograms in a crucial frequency range appear in Appendix A.

### *B. Experiment 1: The range of the Duifhuis effect*

## 1. Spectrum

Experiment 1 is our baseline experiment. Instead of using periodic narrow pulse train, we used a flat-spectrum complex tone as our baseline spectrum. The complex tone had 170 harmonics. Every harmonic had a cosine phase. To create the Duifhuis effect, one harmonic, which could be as low as the 8th harmonic and as high as 145th harmonic, was omitted from the complex tone spectrum. The purpose of this experiment was to see how low and how high the Duifhuis effect exists for a 50-Hz complex tone.

There are some advantages to the flat spectrum. Firstly, the amplitude of each harmonic is the same. Therefore, when different harmonics are canceled, the required cancellation tones have the same amplitude. So for different harmonics the oscillations of the cancellation tone in the time gaps of the waveform have the same amplitude. This makes it easier to do a level experiment later. Secondly, with a flat spectrum it is easier to make a more complicated phase manipulation, such as the Schroeder phase.

## 2. Results

The results are shown in Figure D2. There are three panels in the figure, one for each subject. The y-axis of a panel is called "Ratio." The ratio is defined as the listener's matching frequency divided by the frequency of the omitted harmonic (or the frequency of the cancellation tone). Therefore, when the ratio is unity the listener has made a perfect match. The symbol "h.." tells which harmonic is omitted in the complex tone.

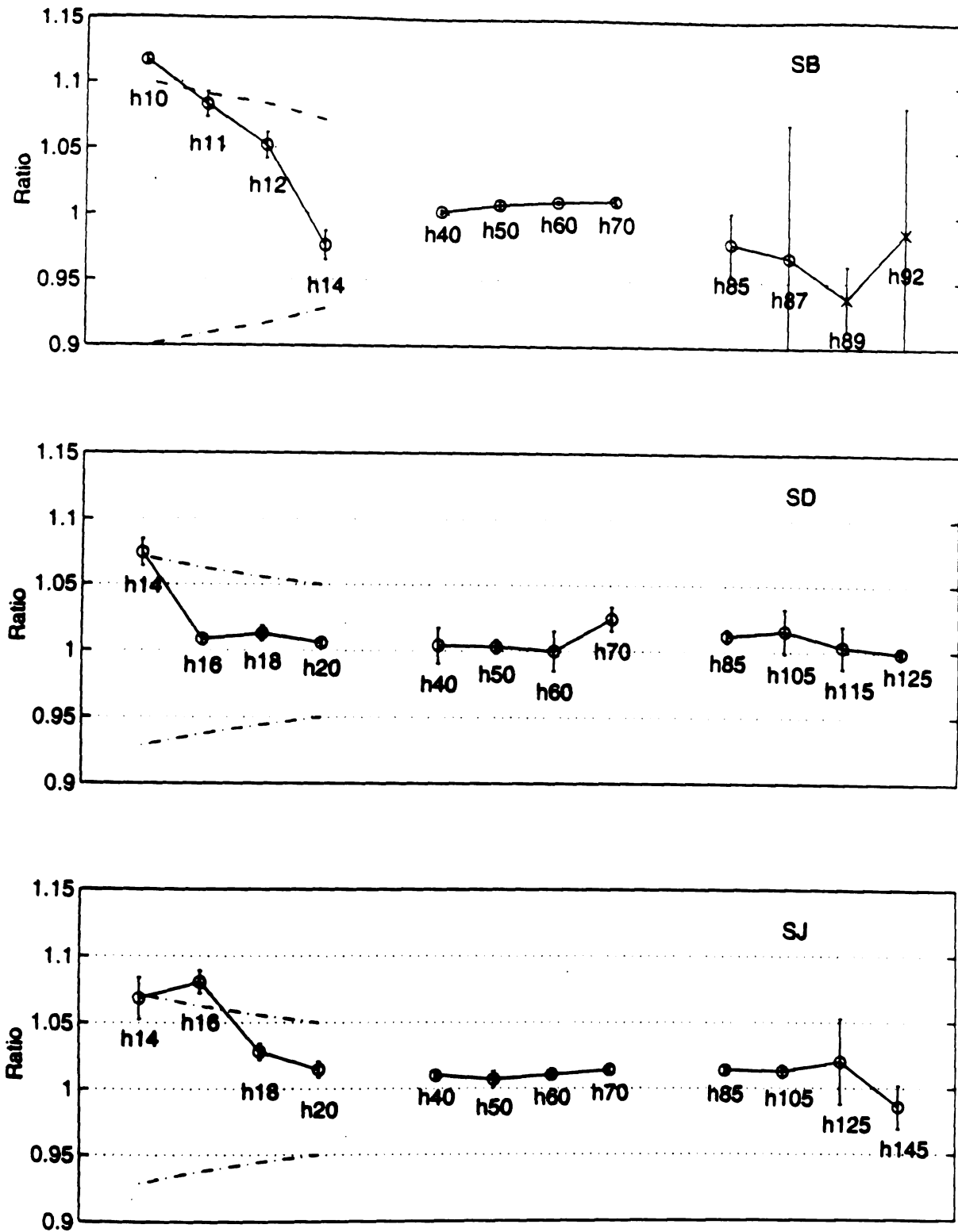
The left four data points were tests of the listener's low ends of Duifhuis effect. The dot-dashed lines are the edges of the spectral gap created by omitting a harmonic. For

example, when harmonic 10 is omitted, the edges of the spectral gap are harmonic 9 (corresponding to a ratio 0.9) and harmonic 11 (corresponding to a ratio 1.1). At very low harmonics, listeners hear the edges and match them. Our criterion for the low end of the effect is that the listener's match is closer to the omitted harmonic than to one of the edges. Thus we can see that three different listeners have different low ends. Subject B seems to begin at harmonic 13. Subject D starts at about harmonic 15. Subject J starts at about harmonic 19.

The middle four points test the performance of the listeners at the middle range. It seems that all the listeners could do the task well except that there is a little systematic positive pitch shift. From the descriptions by the listeners, this was the best range for the phenomenon. The pitch was clear and loud, and there was no ambiguity.

The last four points are a test for the high end of the effect. There are big differences among the three listeners. For SB, reliable matches occur only for harmonic numbers lower than 87. For "h87" the error bar is huge. The "x" symbol for "h89" and "h92" means that out of five trials there were only two or three trials in which the listener could hear the pitch. In the other trials, the listener could not hear the pitch at all. We can probably say that the high end of the effect for SB is at about harmonic 86, which corresponds to frequency 4300 Hz.

For the other two listeners, it seems difficult to find their high ends using the matching technique. Unlike SB, they could hear the pitch very clearly up to much higher harmonics than shown in the figure. However, it is difficult to match a pitch higher than 5000 Hz. For SD, it makes the listener too slow for matching pitch higher than harmonic 125 (6250 Hz) so that we could not use the task as our criterion. The same thing happened for SJ at harmonics higher than 145. It is obvious that the criterion does not



**Figure D2:** Results for Experiment 1 showing matches. Each panel is for one subject. The "Ratio" is defined as the matching frequency divided by the frequency of the omitted harmonic. The harmonic number of the omitted harmonic is labeled as "h..." The left four points are for the low-frequency end of the effect. The middle four points are in best range of the effect. The right four points are points to explore the high-frequency end of the effect. The dashed lines for the left four points represent the edges of the spectral gap.

work at these high frequencies. It is limited by the pitch matching task. Pitch matching becomes unreliable above 5000 Hz, probably because there is no neural timing at high frequencies (Pickles, 1982; Moore, 1973). A sine tone match to a sine tone would result in the same difficulty. So we did not find the high end for SD and SJ. However, from the data we get, we can at least conclude that SD and SJ have much higher high ends than SB.

SB has normal hearing at low frequencies. However, there is tremendous difference between SB and the other two listeners at high frequencies. In order to check whether SB had some hearing loss, an audiogram was done by every listener (see Appendix 1). Compared to the other listeners as standard normal hearing, SB does have some hearing loss at high frequencies. Compared with SJ, SB has approximately 16 to 17 dB losses in both ears at 4500 Hz, which is about his high-end frequency for the Duifhuis effect. It is possible that this hearing loss caused the high-end limit for SB.

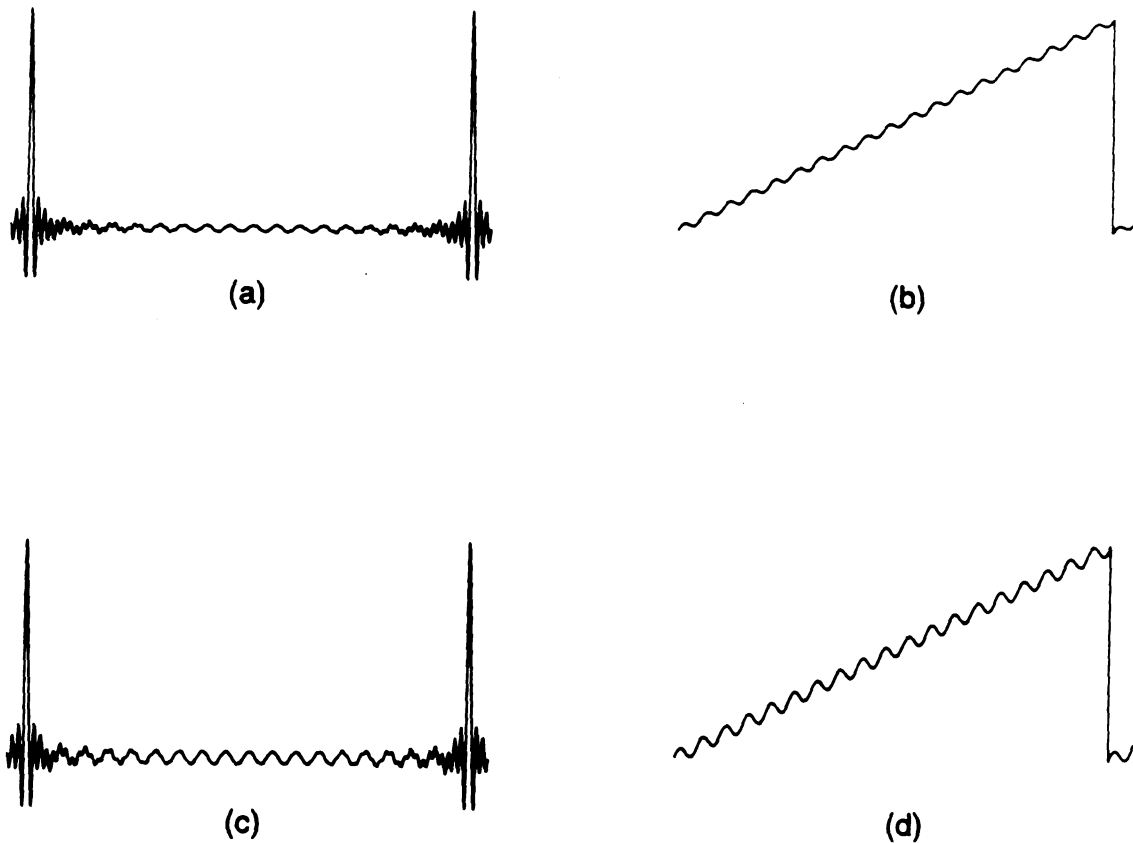
We want to point out that we did not use a two-interval forced-choice task to test the high end, because we found in an informal experiment that listeners tend to use cues from comparison of the two stimuli. For example, when a harmonic of a complex tone with random phases is omitted, nothing is heard. Five trials for each listener were done. It was impossible to match the omitted harmonics. There is no Duifhuis effect. However, using two-interval forced-choice, one can easily hear the tone by comparison of the two stimuli even if the interval between the two signals is as long as 700 ms.

### *C. Experiment 2: The level effect*

#### **1. Spectrum of the stimuli**



In this experiment, we want to see whether the strength of the Duifhuis pitch is determined by the level of the cancellation tone. There are four conditions for comparison. The illustrative waveforms are plotted in Figure D3. *(a)* A 70-component complex tone with one of its harmonics omitted. The power spectrum is flat and is tapered at the end of the spectrum to reduce the edge pitch (Klein and Hartmann, 1981). The phases of all harmonics are cosine. *(b)* Same as condition (a) except that the target harmonic is phase inverted instead of omitted. Therefore, there is no spectral gap in the power spectrum. Note that in this case, the cancellation tone's level is 6 dB higher than condition (a). This can be easily seen in the waveform. The oscillations between the pulses in (b)



**Figure D3: Waveforms for Experiment 2. (a)** Waveform of a flat-spectrum complex tone with the 19th harmonic omitted; **(b)** Sawtooth waveform with 19th harmonic omitted; **(c)** Waveform of flat-spectrum complex tone with 19th harmonic phase inverted; **(d)** Sawtooth waveform with 19th harmonic phase inverted. The damped high frequency oscillations for flat spectrum cases (a) and (c) are from the high-frequency end of the spectra (the edge effect). In fact, the high-frequency ends of the spectra were tapered to reduce the effect.

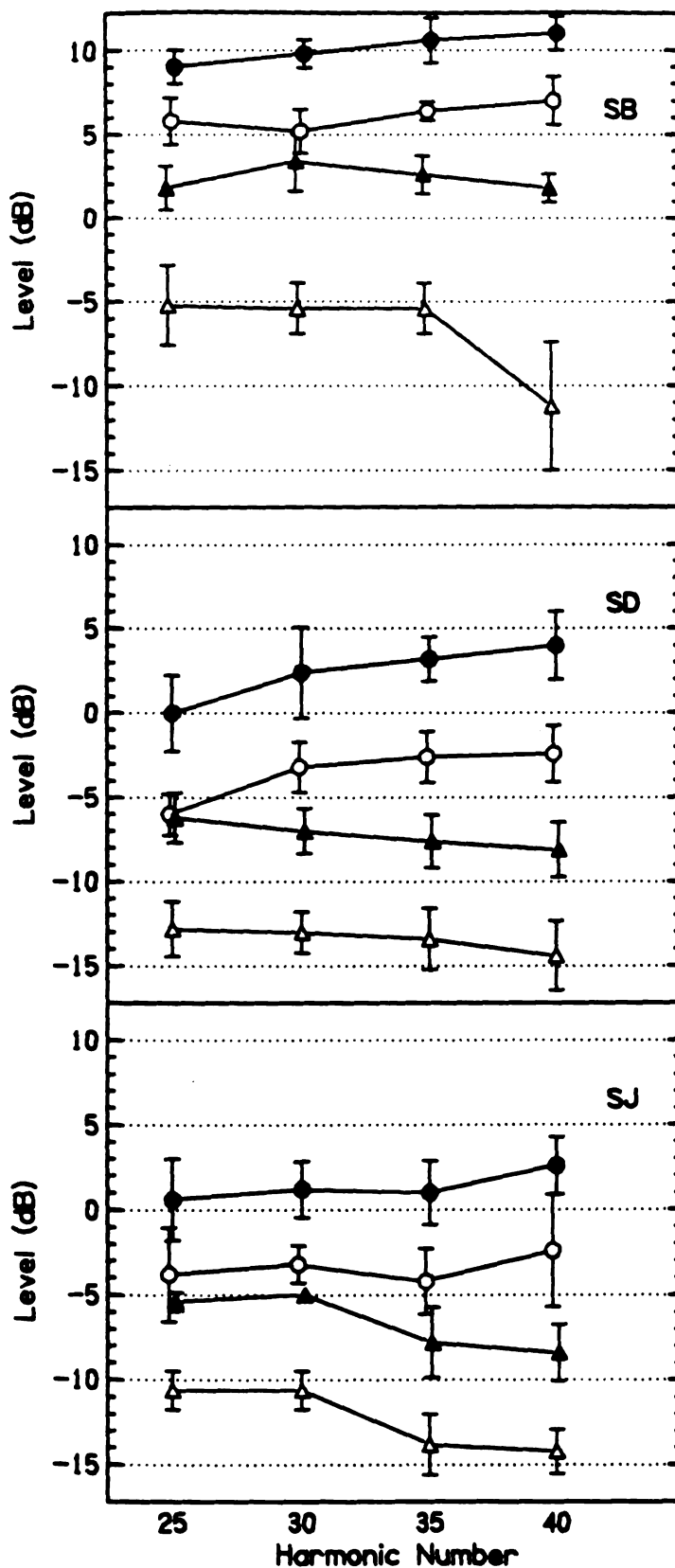
have twice the amplitude as in (a). (c) This is a sawtooth wave with one of its harmonics omitted. The highest harmonic is number 70. We know that the amplitude of the harmonics of a sawtooth wave is inversely proportional to the harmonic number. For example, the 10th harmonic has twice the amplitude as the 20th harmonic. So we hope to see some perceived level effect with harmonic number, since the cancellation tone for different harmonic numbers has a different amplitude. (d). Same as condition (c) except that phase inversion replaces omission. We used a 70-component complex tone in this experiment instead of 170, because we want to make the Duifhuis pitch as clear as possible so that its level could be matched. The harmonics chosen were harmonic 25, 30, 35 and 40.

## 2. Results

The results are shown in Figure D4. There are three panels in the figure, one for each subject. The listeners' pitch matches were very consistent with the Duifhuis pitch. The standard deviations were less than 1% of the matching frequency, and the pitch shifts were less than 2%. However, only the results for level matches are plotted, because we are interested only in the level effect. The level used in Figure D4 is a relative level, and the reference is the level of one component in a flat-spectrum complex tone i.e. 45 dB.

Let's first look at the results from SJ. First, compare the harmonic-inverted case with the harmonic-omitted case. For the flat-spectrum, the inverted case is about 6 dB higher than the omitted case. The same thing happens for the sawtooth waveform condition. Note 180° phase inversion of a harmonic requires 6 dB higher cancellation tone than omitting the harmonic. Therefore, it seems the strength of Duifhuis pitch is consistent with the level of cancellation tone. Next, let's look at the results for different

harmonic in the sawtooth waveform condition. The observed level for harmonic 40 is about 3.0 dB lower than the level for harmonic 25 in both the omission case and the phase inversion case. Theoretically, the level of the cancellation tone for harmonic 40 should be  $20\log(40/25)=4.1$  dB lower than the level of the cancellation tone for harmonic 25 in both harmonic-omitting and phase inversion cases. However, by examining the level changes in the two flat-spectrum cases, we find that harmonic 40 sounds about 1 dB louder than harmonic 25 for flat spectrum cases. Therefore, making the amplitude of the harmonics inversely proportional to harmonic number (sawtooth



**Figure D4:** Results of Experiment 2. Each panel is for one subject. Matching levels of the Duifhuis pitch are plotted as a function of the harmonic number of the anomalous harmonic. The open circle symbol is for an omitted harmonic in a flat-spectrum complex tone. The filled circle is for a phase inverted harmonic in a flat-spectrum complex tone. The open triangle is for an omitted harmonic in a sawtooth wave. The filled triangle is for a phase inverted harmonic in a sawtooth wave.

waveform), does make a corresponding perceived level change consistent with the level of cancellation tone.

SD seems to have the same level effect as SJ except that there are differences in the pattern. SB is different from the other two listeners in absolute levels perceived for the Duifhuis pitch. He matched much higher levels. However, there are some similar level effects in SB, but the level effects are not so strong as for the other two listeners. Generally speaking, level seems to be determined by the level of the cancellation tone.

Informal experiments show that the level of the Duifhuis pitch can be matched by the best beats method, in which the listener adjusts the level of a added sine tone to maximize beats. However, the level range for best beats is wide, especially in the omitted-harmonic condition. So the result has a large variability ( $\pm 1$  to 4 dB). It is significant though that a real sine tone can form beats with the Duifhuis pitch, which implies that the Duifhuis pitch is not something absent, it is a segregated sine tone, i.e. the cancellation tone. The amplitudes of harmonics in a triangle wave are inversely proportional to the square of the harmonic number. Informal experiment shows no Duifhuis pitch can be heard.

#### *D. Experiment 3: The effect of harmonic phases*

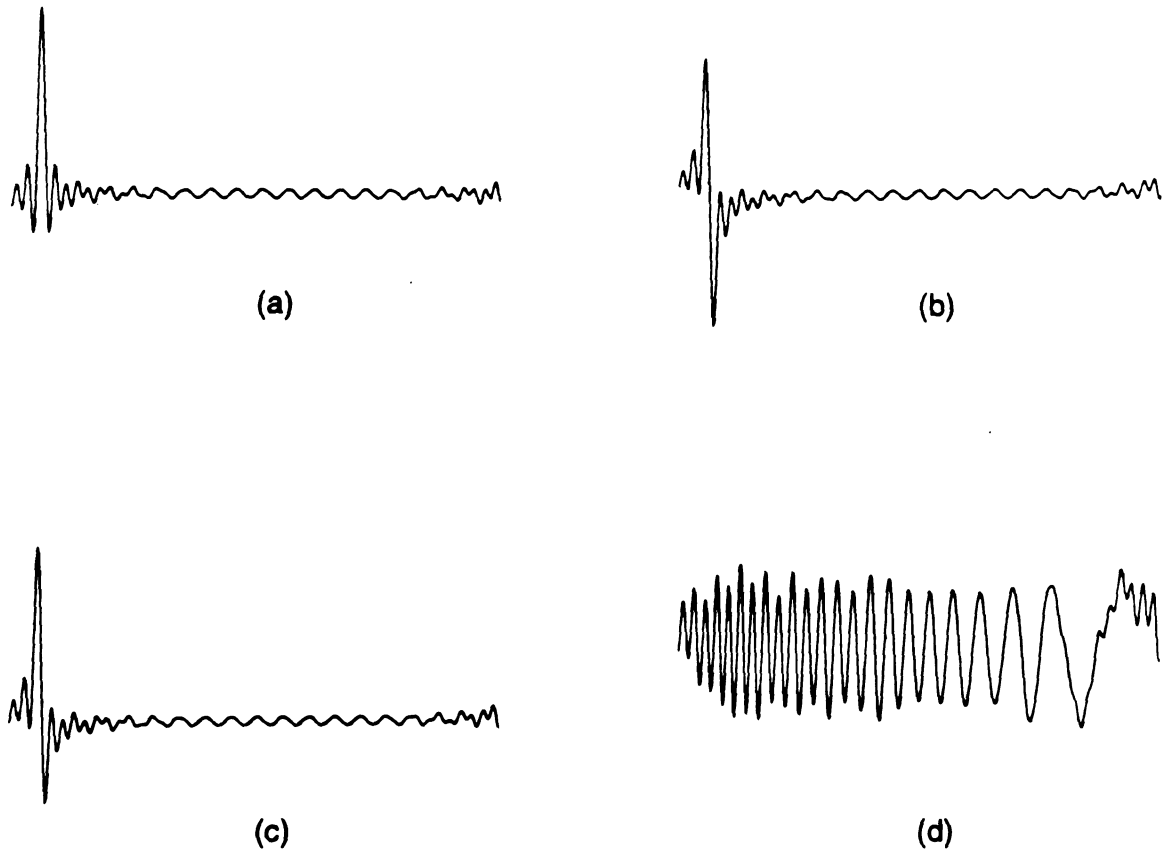
##### **1. Spectra of the stimuli**

In this experiment, we use four different phases for the complex tone to explore the possible existence of the Duifhuis effect with different waveforms. We have a 100-component complex tone with four different phase conditions: *(a)* All the harmonics are cosine phase. *(b)* All the harmonics are sine phase. *(c)* All the harmonics have a  $45^\circ$

initial phase. **(d)** Schroeder phase (Schroeder, 1970). For an equal-amplitude complex tone Schroeder phases of the harmonics are given by:  $\phi_n = -\pi n^2 / N$ , where  $N$  is the total number of harmonics in the complex tone, and  $n$  is the harmonic number.

We chose harmonics 40, 50, 60 and 70, i.e. in the best range of the effect, and the Duifhuis effect was created by omitting one of the harmonics. Figure D5 shows the waveforms for the four conditions.

## 2. Results



**Figure D5:** Illustrative waveforms for Experiment 3. (a) Flat-spectrum complex tone with all the harmonics in cosine phase; (b) Flat-spectrum complex tone with all the harmonics in sine phase; (c) Flat-spectrum complex tone with all the harmonics at  $45^\circ$  initial phase; (d) Flat-spectrum complex tone with Schroeder phase.



The consistency in matching the pitch is used to determine whether the Duifhuis pitch is heard or not. Figure D6 shows the results for the three listeners. The consistency of matches to the Duifhuis pitch shows that all the three listeners can hear the tone equally well in all four conditions, because none of them is special. It is interesting that, unlike the other three waveforms, there is no time gap in the waveform of the complex tone with Schroeder phase. So the cancellation tone cannot be seen in the waveform. Yet, the effect does exist. (See part G. for a possible traditional explanation of the Duifhuis effect with a Schroeder-phase complex tone.)

When all the harmonics had random phases, the three listeners agreed that no Duifhuis pitch could be heard. In order to show this, SJ and SB tried to match a Duifhuis pitch in a random phase complex tone. The results of their matches had wide distribution. The standard deviation was about 0.15 of the stimulus spectral width, which is half of the fraction 0.29 expected for random guessing.

#### *E. Experiment 4: The missing fundamental pitch*

##### 1. Spectra of the stimuli

When harmonics 18, 24 and 30 are deleted, the spectrum of the cancellation tones resembles the 3rd, 4th and the 5th harmonics of 300 Hz. Informal experiments showed that this stimulus produced a pitch at 300 Hz. This experiment tests this phenomenon. The complex tone used in the experiment was a 100-component equal-amplitude cosine-phase complex tone. There were four conditions: *(a)* Phases of harmonic 18, 24 and 30 are inverted, which resembles the 3rd, 4th and the 5th harmonics of 300 Hz. *(b)* Phases of harmonic 27, 36 and 45 are inverted, which resembles the 3rd, 4th and the 5th har-

monics of 450 Hz. *(c)* Phases of harmonic 36, 48 and 60 are inverted, which resembles the 3rd, 4th and the 5th harmonics of 600 Hz. *(d)* Phases of harmonic 42, 56 and 70 are inverted, which resembles the 3rd, 4th and the 5th harmonics of 700 Hz.

We used phase inversion in this experiment because we wanted to have a flat stimulus spectrum for comparison with predictions of some contemporary pitch perception models. In fact, the effect can also be produced by omitting the harmonics.

## 2. Results

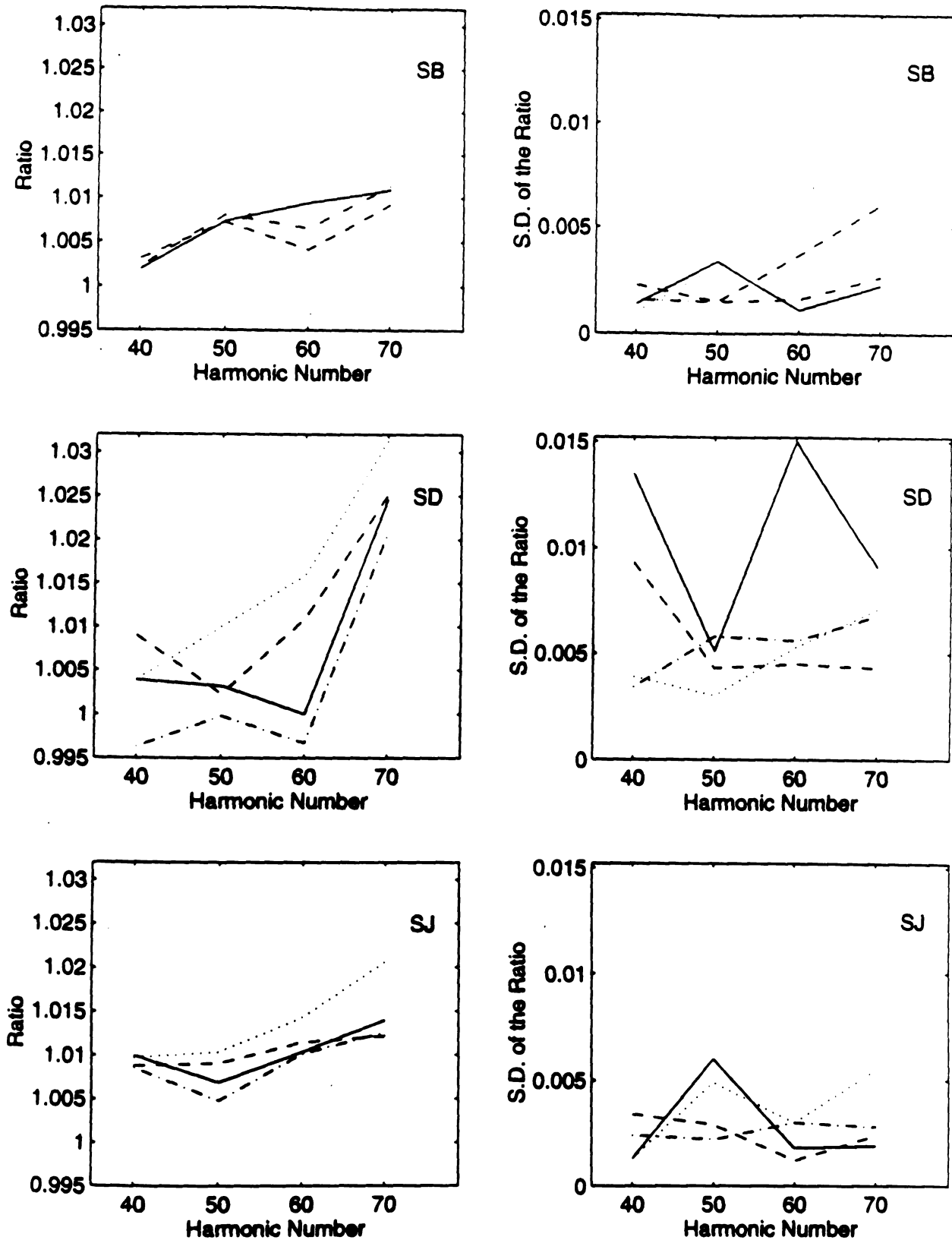


Figure D6: Results of Experiment 3. The left three panels are plots of the matches, each for one subject. The right three panels are plots of the standard deviation of the matches. The solid line represents cosine phase. The dashed line represents sine phase. The dot-dashed line represents 45° initial phase. The dotted line represents Schroeder phase.

Results from the three listeners are plotted in Figure D7, as before, one panel for each subject. The “Ratio” for the y-axis is defined as the ratio of the matching frequency to the frequency of the missing fundamental. The results show that all the listeners could match the missing fundamental frequency. The matches of the listeners are flat, and SB is especially flat. Compared with the result of Terhardt (1971) that the pitch of the missing fundamental of a complex tone is about 2% flatter than the pitch of a sine tone with the frequency of the fundamental, SJ and SD are less flat, SB is more flat.

One observation is to be made. In contemporary pitch perception theories, pitch is extracted only from the magnitude of spectral components. Among these theories are the DWS pitch meter (Duifhuis *et al.* 1982) and Terhardt’s virtual pitch extraction algorithm (Terhardt *et al.* 1982). In this experiment, every component has the same amplitude, and the pitch is made entirely with phase inversion. Therefore, these models do not apply directly.

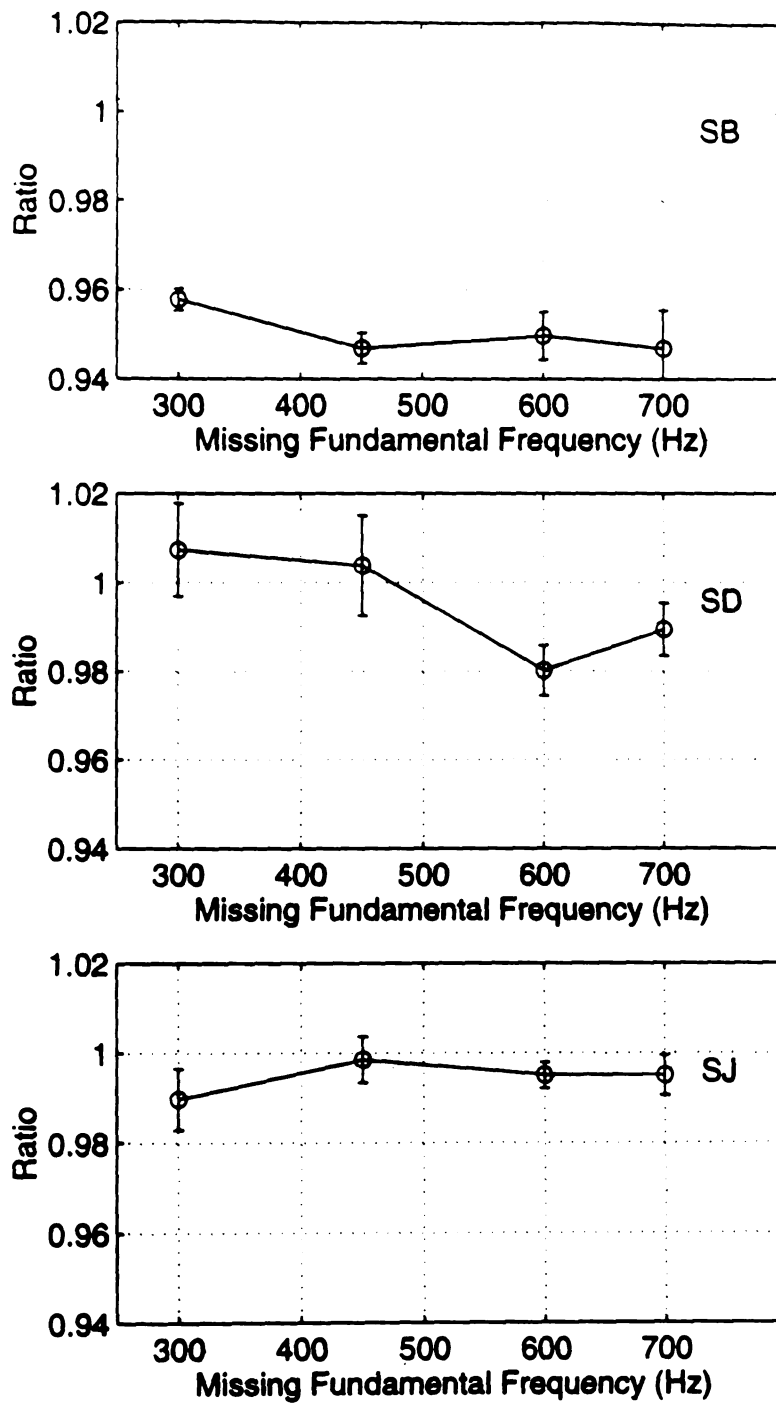
#### *F. Discussion*

In the above experiments, the Duifhuis effect was extended to a variety of new conditions. We want to emphasize that the effect we are interested in is a “steady state” effect. By steady state, we mean that the spectrum of the stimulus does not change during its presentation. It is well known that changes in the spectrum or the waveform (turning a harmonic on and off or suddenly shifting the phase of a harmonic) can cause a harmonic to be heard out from a complex tone. (Pierce, 1960; Kubovy and Jordan, 1979; McAdams, 1984; Hartmann, 1988; Moore and Glasberg, 1989; Alcantara and Moore 1995). Such effects occur both for harmonics that are spectrally resolved and for harmonics that aren’t. They are “attention” effects caused by the contrast between

“before” and “after” conditions, and are not the effect we are interested in. In our experiment such temporal effects are not involved. Listeners heard only one single periodic complex tone.

### *G. Traditional Explanation*

The explanation for the effect by Duifhuis (1970, 1971) is that when the frequency spacing between harmonics is considerably less than the bandwidth of relevant peripheral auditory filters, the peripheral filters can be said to be broad band. Generally



**Figure D7:** Results of Experiment 4. Matches to the missing fundamental tone are shown in three panels, one for each subject. The matches are plotted in ratio of matching frequency to missing fundamental frequency.

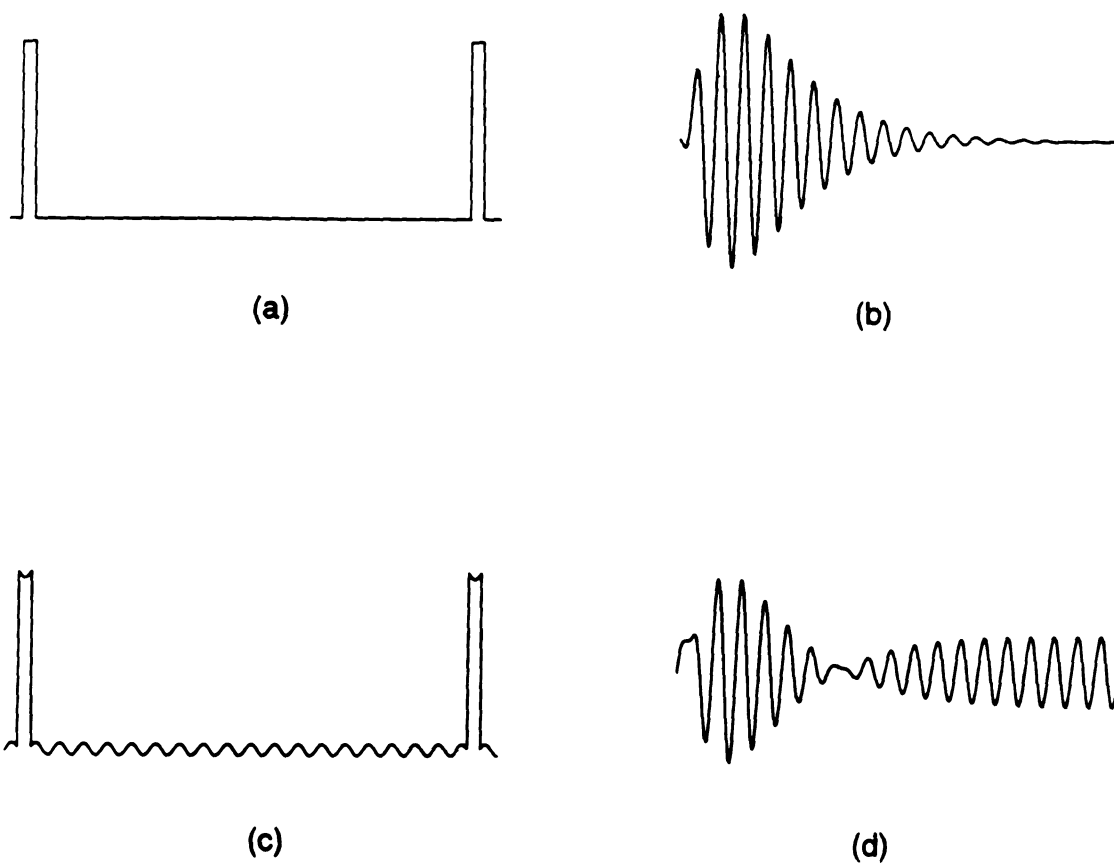
speaking, a broad-band frequency analyzing system has a short impulse response time, i.e. the windowing time for analyzing an incoming signal is short. For the broad-band peripheral filter mentioned above, the window time is less than the period of the input signal. Therefore, there is a portion of time in each period during which the peripheral filter windows mainly the small oscillations of the cancellation tone, and hence the output is the cancellation tone. That portion of output in each period is detected as a sine tone. This is an effect from the short-term Fourier analysis of the peripheral auditory filters due to their relatively broad band. Thus we call it the short-term Fourier analysis explanation.

The idea is more understandable if we take an extreme as an example: If the 60th harmonic of a 10-Hz periodic narrow pulse train is omitted, you will hear two sounds. One is the pulses repeated at 10 Hz. The other is a sine tone at 600 Hz, which is the cancellation tone. Because the bandwidths of the whole peripheral filter bank are wide compared to the frequency spacing of the components of the signal (10 Hz), the window time for analyzing the spectrum is much less than the signal's period (100 ms). In fact, the pulses you perceive are resolved at this low frequency (10 Hz), which simply means that you can even "feel" that the window time of your auditory system is less than the period of the stimulus. Therefore, in each period (100 ms), most of the time the window is on the cancellation tone, which results in an output of the cancellation tone. In other words, the power spectrum has a peak at the frequency of the cancellation tone for most of the time, or the corresponding region on the tonotopic axis is excited. Thus for most of the time the brain receives a real sine tone output and hence you hear the sine tone. The essence of short-term Fourier analysis is that although there is a gap in the spectrum of the stimulus implying that there is no input signal at that frequency, the output after a short-term Fourier analysis has a peak in the spectrogram,

i.e. a real signal output at the frequency of the input spectral gap. Therefore, you are not detecting something absent, you are detecting a real output.

Returning to our 50-Hz stimulus, there are similarities to the above extreme condition. Look at a filter centered at the omitted harmonic, say the 19th harmonic. When all the harmonics are present, the response of the filter is just a ringing followed by a quiet duration for each period (see Figure D8). The quiet duration exists because the filter is wide compared to the frequency spacing of the components, or the impulse response time is less than the period. When the harmonic is omitted, the quiet duration is replaced by the response to the small oscillations of the cancellation tone (Figure D8), a





**Figure D8:** Illustrative plots of a filter's operation. (a) Input waveform of a 50-Hz narrow pulse train. (b) Output of the waveform in (a) at a filter whose center frequency is 950 Hz. (c) The periodic narrow pulse train with the 19th harmonic omitted. (d) Output of the waveform in (c) at the same filter.

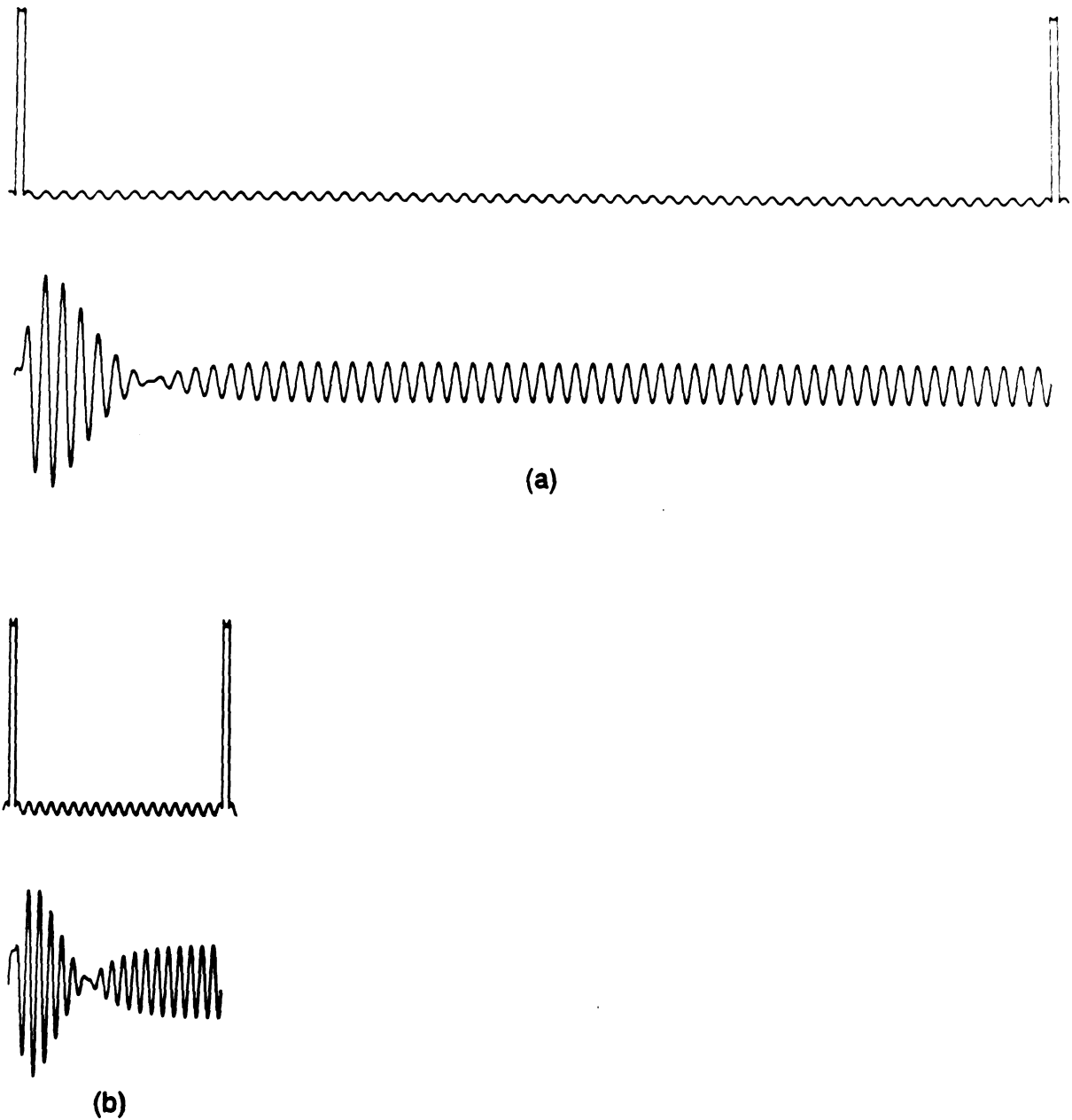
real output of a sine tone (cancellation tone) which is same as the 10-Hz extreme condition mentioned above. This similarity can be seen by a comparison of the two conditions: 60th harmonic of the 10 Hz complex tone versus the 19th harmonic of the 50 Hz complex tone. The signals and the outputs of the peripheral filters centered at the omitted harmonics are given in Figure D9. The two outputs are very similar in that for each period both have a response to the cancellation tone after the ringing part. The only difference is that for the 50-Hz complex tone condition, the fraction, in each period, of response time to the cancellation tone is much less. However, the short-term Fourier analysis explanation suggests that the two conditions are basically equivalent.

The ringing part has a bigger amplitude than the response to the cancellation tone part. Why does the part with small amplitude produce the dominant percept? It is important to point out that the short-term Fourier analysis explanation is not based upon one single filter. The time gap exists in the output of the entire filter bank. And the response of the entire filter bank to the cancellation tone in the gap is the same as a response to a sine tone, i.e. filters whose frequencies are close to the omitted harmonic are all responding to the cancellation tone.

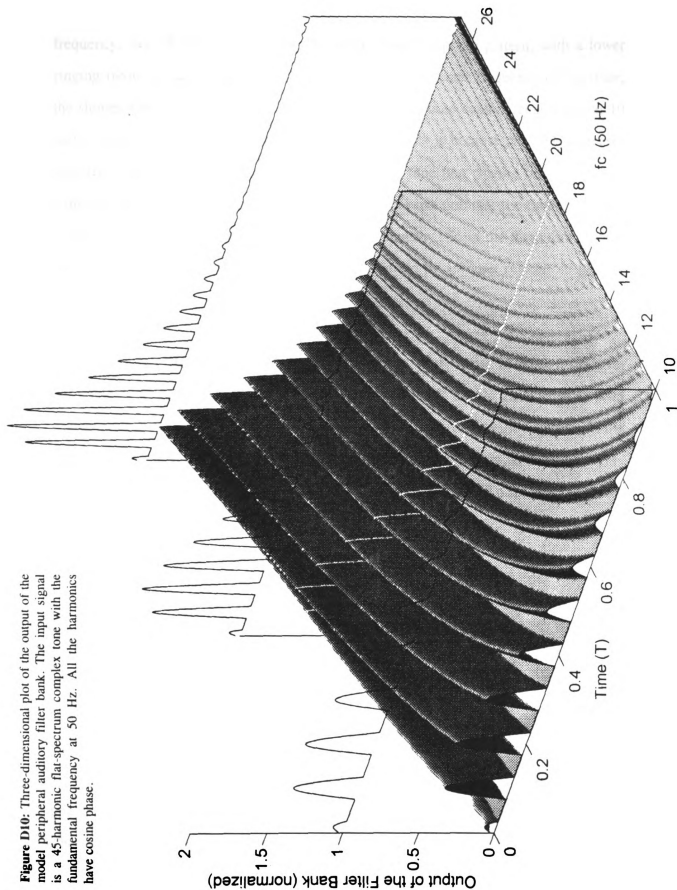
To get a clearer picture of the idea of the short-term Fourier analysis, we implement the idea with a realistic auditory filter bank - the gammatone filter bank (Patterson, 1987). Figure D10 is a three-dimensional plot of the output of a gammatone filter bank. The bandwidth of the filter bank is determined by the ERB from Moore and Glasberg (1987). The input to the filter bank is our 50-Hz complex tone. The complex tone has harmonics from 1 to 45. The high-frequency end of the spectrum is tapered. The x-axis is a time axis. Its unit is the period of the input signal, or 20 ms. Therefore the plot is just one period of the output. The y-axis is the center frequency of a filter in the filter bank. The unit is chosen to be 50 Hz so that at an integer number on the y-axis a har-

monic of the input signal occurs and its harmonic number is equal to the integer. The z-axis shows the amplitude of the output. It is normalized so that the maximum amplitude has an output of one. The output of the filter bank is half-wave rectified to improve the three-dimensional impression in the figure.

Let's analyze what we get from Figure D10. First, look at an output of a filter at a particular center frequency, say 27 ( $50 \times 27 = 1350$  Hz). What we get is an impulse response to the pulse (i.e. a ringing of the filter) followed by a quiet duration or a time gap, which we are already familiar with. Next look at a filter with a lower center



**Figure D9:** Comparison of a 10-Hz narrow pulse train with a 50-Hz narrow pulse train. (a) The upper plot in (a) is a plot of the waveform of a 10-Hz narrow pulse train with the 60th harmonic omitted. The lower plot in (a) is the output of a filter whose center frequency is at the omitted harmonic ( $10 \times 60 = 600$  Hz); (b) The upper plot in (b) is a plot of the waveform of a 50-Hz narrow pulse train with the 19th harmonic omitted. The lower plot in (b) is the output of a filter whose center frequency is at the omitted harmonic ( $50 \times 19 = 950$  Hz). All plots have the same time axis.



**Figure D10:** Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 45-harmonic flat-spectrum complex tone with the fundamental frequency at 50 Hz. All the harmonics have cosine phase.



frequency, say 19 ( $50 \times 19 = 950$  Hz). We get a similar response pattern, with a lower ringing frequency and a shorter time gap. The lower the center frequency of the filter, the shorter the time gap. As we can see, for the filter whose center frequency is at 10 units, there is almost no time gap after the ringing. This is because the ERB increases with the center frequency, i.e. the impulse response time (the ringing time) decreases with increasing center frequency. Finally, we want to mention that the time gap exists in the outputs of all the filters with high center frequency. So in the three-dimensional plot of the output, we see a “ringing region” followed by a “quiet region” in the x-y plane.

Figure D11 is a three-dimensional plot of the output of the filter bank with the 19th harmonic (950 Hz) of the input complex tone omitted. It should be compared with Figure D10 in which all harmonics are present. As we can see, a valley appears in the ringing region of the output and a “ridge” appears in the previously quiet region. This ridge is the same as the response of the filter bank to a sine tone. To see this the three-dimensional plot of the response to a sine tone, whose frequency is 950 Hz, is shown in Figure D12.

Now it is very clear that the explanation of short-term Fourier analysis is that the auditory system detects the response to the cancellation tone in the quiet region of the response of the filter bank. In other words, the auditory system detects the sine tone like “ridge” response in Figure D11.

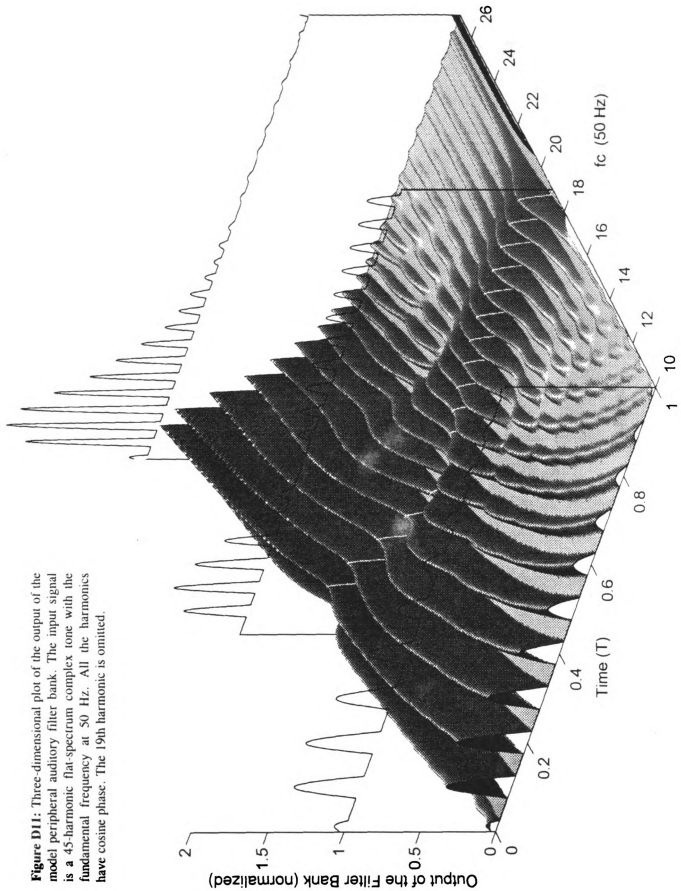
The traditional explanation may also applied to the Schroeder-phase condition Duifhuis pitch. The Schroeder phase leads to a frequency sweep with the duration of a period. The sweep begins at the highest frequency in the band and moves continuously to the lowest. Therefore, there exists some duration within which the instantaneous frequency

(Schroeder, 1970) of the stimulus differs a lot from the center frequency of a filter in the filter bank. Thus a “tilted” quiet region is produced in the output of the filter bank by the Schroeder-phase complex tone (see Figure D13). When a harmonic is omitted, a ridge appears across the quiet region (Figure D14).

#### *H. Challenge to the Traditional Explanation*

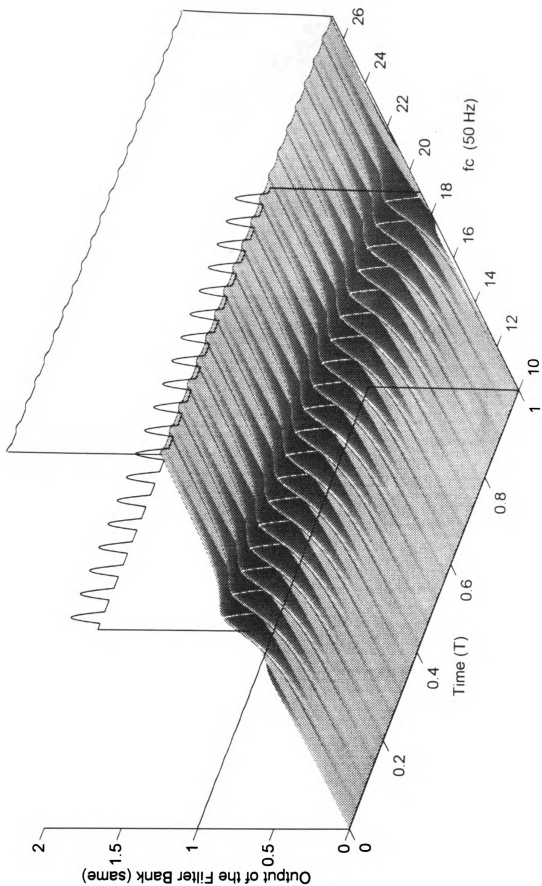
In the traditional explanation, a sine-tone-like ridge response is required in the quiet region of the peripheral output in order to form the Duifhuis pitch. If the ridge is



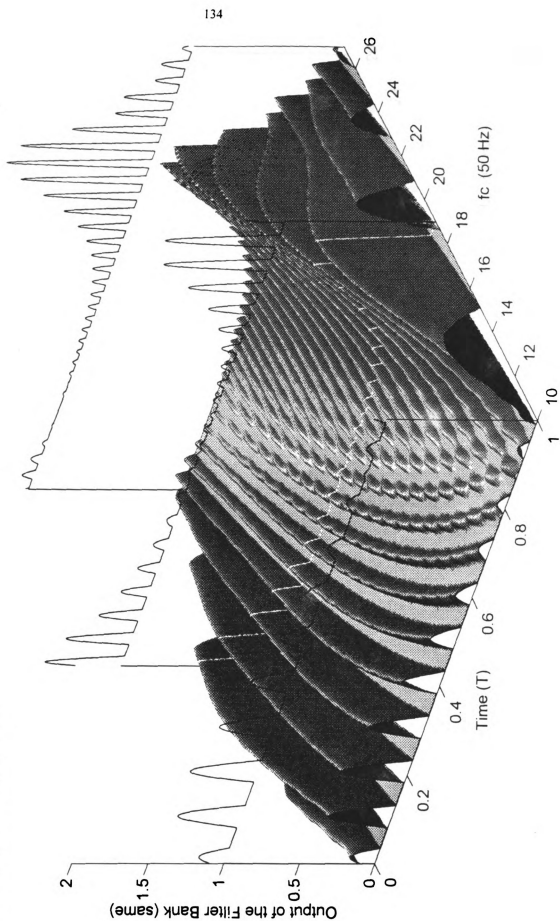


**Figure D11:** Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 45-harmonic flat-spectrum complex tone with the fundamental frequency at 50 Hz. All the harmonics have cosine phase. The 19th harmonic is omitted.

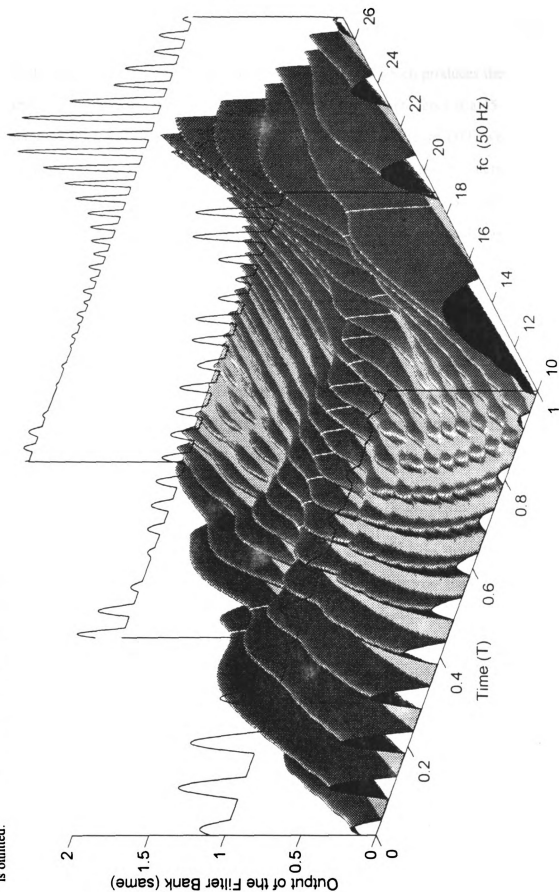
**Figure D12:** Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 950-Hz sine tone.



**Figure D13:** Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 45-harmonic flat-spectrum complex tone with the fundamental frequency at 50 Hz. The harmonics of the complex tone have Schroeder phase.



**Figure D14:** Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 45-harmonic flat-spectrum complex tone with the fundamental frequency at 50 Hz. The harmonics of the complex tone have Schroeder phase. The 19th harmonic is omitted.



critical in detecting the sine tone, can we produce some other signal which produces the same ridge and thus produce the same effect? For Figure D11, the input signal is a 45-component complex tone with the 19th harmonic omitted. Examining Figure D11, we can easily find that the ridge is mainly within a limited frequency range, approximately harmonics 17 to 22. Therefore, harmonics whose frequency are far from this range contribute little to the response of the filter bank in this range. In other words, deleting those “distant” harmonics will not change the response of the filter bank in this region.

Figure D15 is a three-dimensional plot of the output of the filter bank with an input complex tone which consists of harmonics from the 15th to the 24th with equal amplitude. The 19th harmonic is omitted. All the present harmonics have a cosine phase. Therefore, the input complex tone is a band-pass filtered version of the input signal in Figure D11. The high and low edges of the spectrum are tapered. As we can see, the sine-tone-like ridge does not change much from Figure D11. The major difference from Figure D11 is that the ringing region is narrow band now. According to the short-term Fourier analysis explanation, this narrow band complex tone will give the same Duifhuis effect as the wide band complex tone, because the effect comes from the sine-tone-like ridge. If no Duifhuis pitch can be detected in the narrow-band complex tone, the traditional explanation cannot be correct. Therefore, we used this kind of stimulus to test the explanation.

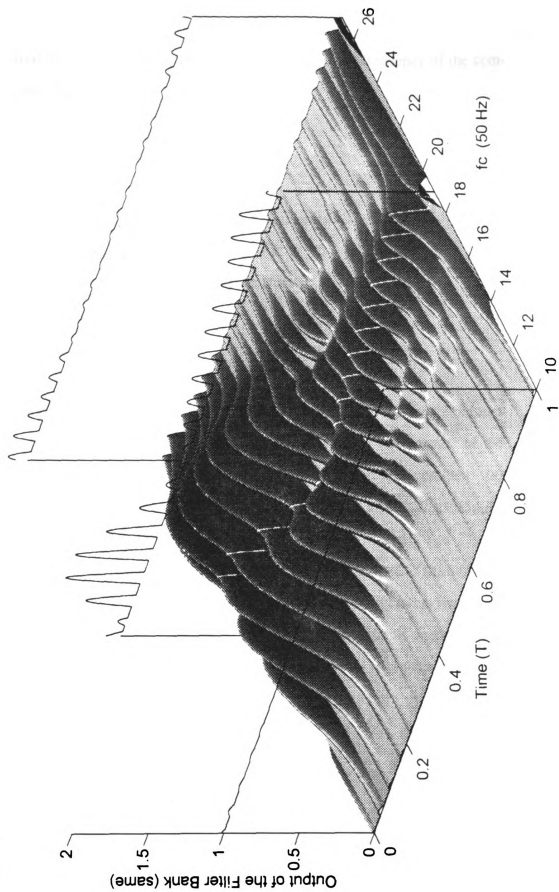
## **II. Experiments to Test the Short-term Fourier Analysis Explanation**

### ***A. Method***

#### **1. Procedure**

The procedure was the same as for previous experiments except for one difference. That was that the listener could listen to the stimuli as many times as he wanted. However, once he decided to do matching and listened to the sine tone to adjust its pitch, he was not allowed to listen to the stimulus again. Therefore, he must first listen to the stimulus and remember the pitch, and then do the pitch matching task from memory. We did this to avoid a cueing effect, because the matching sine tone could cue a sine tone out of a stimulus. In this experiment, we want to see whether the effect exists or not, therefore we do not want any cue. The sampling rate of the DA converter

**Figure D15:** Three-dimensional plot of the output of the model peripheral auditory filter bank. The input signal is a 50-Hz complex tone with harmonics 15 to 24. All the harmonics have cosine phase. The 19th harmonic is omitted.



was randomized by  $\pm 5\%$  so that the listener could not expect the frequency of the component from previous matching.

## 2. Stimuli and listeners

The stimuli and listeners were same as previous experiments.

### *B. Experiment 5: Narrowing the bandwidth of the stimulus*

#### 1. Spectra of the stimuli

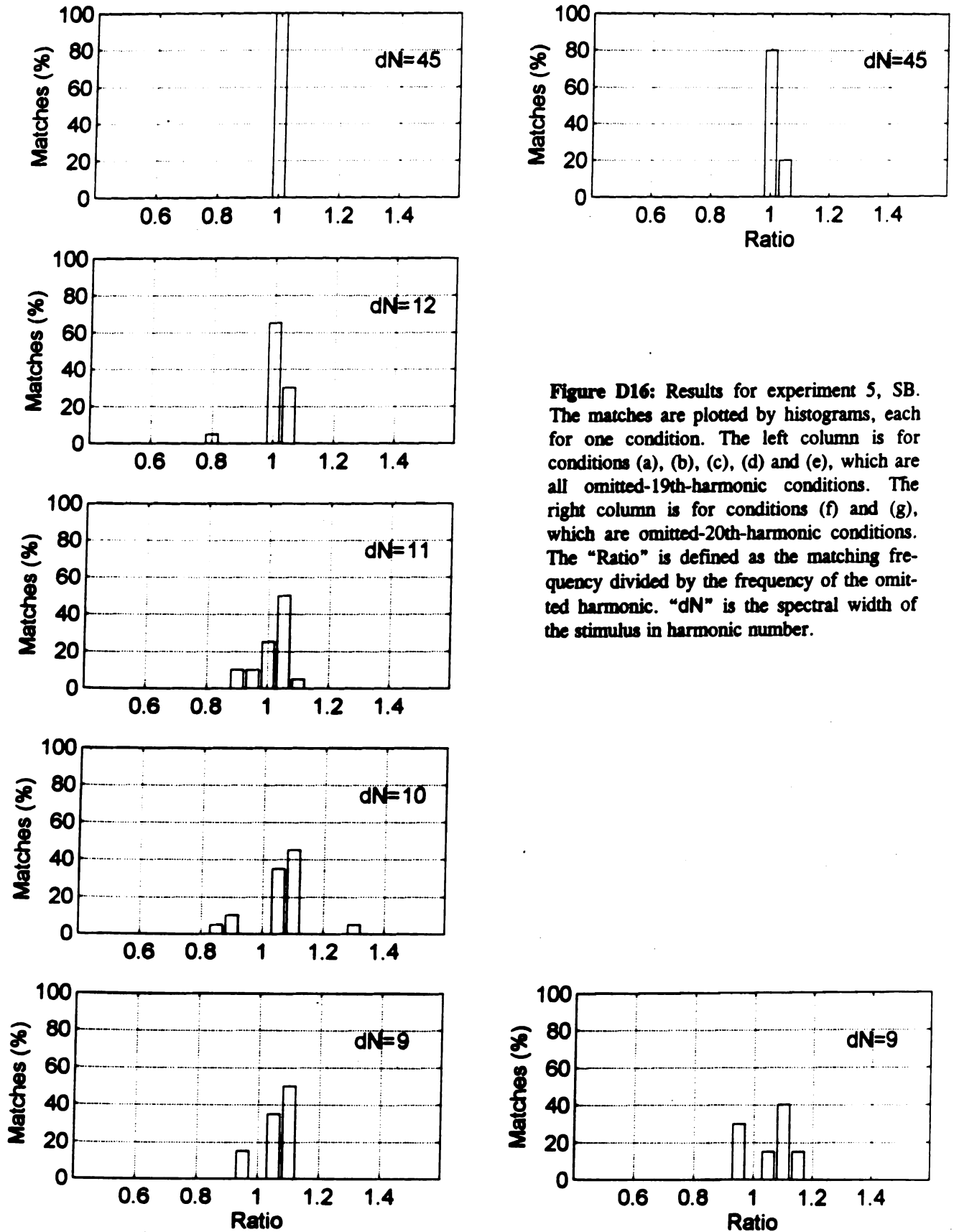
There were seven stimuli. The high and low ends of the spectra of the complex tones were tapered to reduce edge pitches, unless the low end was the fundamental of the complex tone. Five stimuli were complex tone with 19th harmonic omitted: *(a)* Stimulus one was a complex tone with harmonics from 1 to 18 and 20 to 45; *(b)* Stimulus two was a complex tone with harmonics from 14 to 18 and 20 to 25; *(c)* Stimulus three was a complex tone with harmonics 15 to 18 and 20 to 25; *(d)* Stimulus four was a complex tone with harmonics 15 to 18 and 20 to 24; *(e)* Stimulus five was a complex tone with harmonics 16 to 18 and 20 to 24. Two of the stimuli were complex tones with 20th harmonic omitted: *(f)* Stimulus six was a complex tone with harmonics from 1 to 19 and 21 to 45; *(g)* Stimulus seven was a complex tone with harmonics 17 to 19 and 21 to 25.

## 2. Results

The results are plotted in Figures D16, D17 and D18. Each figure is for one subject. There were altogether 20 matches. The consistency of the matches is plotted by histo-



grams, one panel for each stimulus. The left column is for conditions (a), (b), (c), (d) and (e), which are all omitted 19th harmonic conditions. The right column is for conditions (f) and (g), which are omitted 20th harmonic conditions. The variable “Ratio” is defined as the matching frequency divided by the frequency of the omitted harmonic. The bin width is 0.05 which is approximately the frequency spacing between harmonics of the complex tone near harmonics 19 and 20. We use this bin size because we want the central bin to allow a maximum pitch shift of half a harmonic spacing.



**Figure D16:** Results for experiment 5, SB. The matches are plotted by histograms, each for one condition. The left column is for conditions (a), (b), (c), (d) and (e), which are all omitted-19th-harmonic conditions. The right column is for conditions (f) and (g), which are omitted-20th-harmonic conditions. The "Ratio" is defined as the matching frequency divided by the frequency of the omitted harmonic. " $dN$ " is the spectral width of the stimulus in harmonic number.

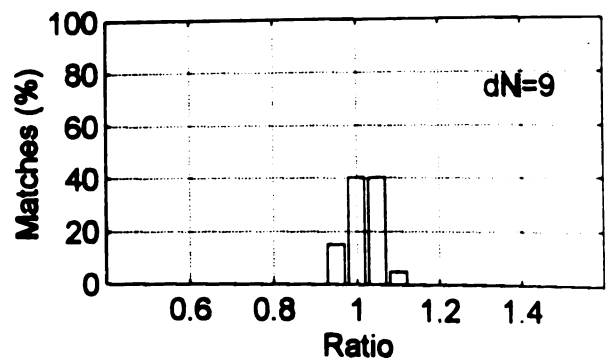
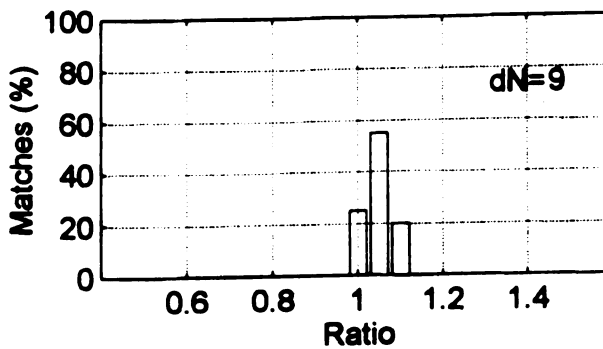
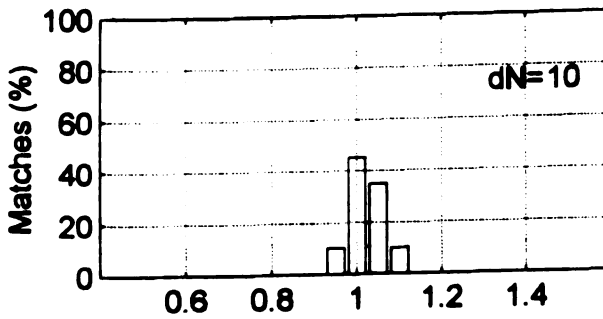
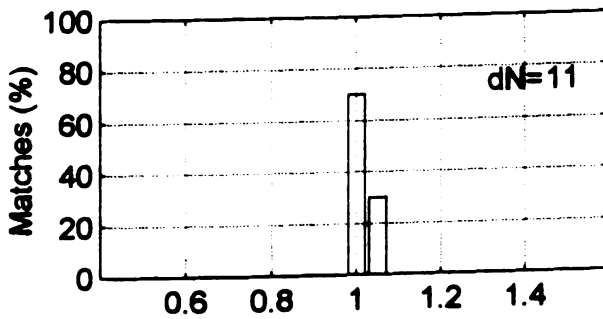
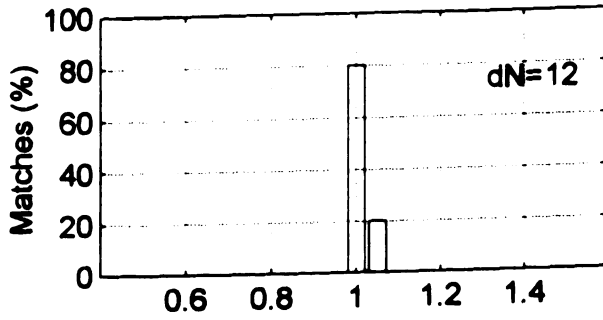
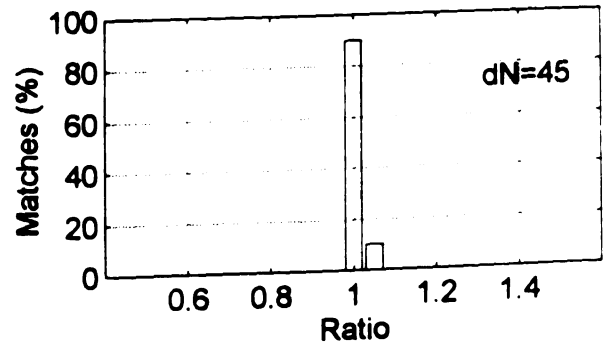
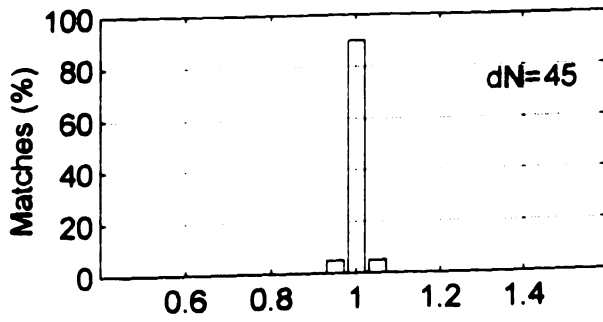


Figure D17: Results for experiment 5, SD.

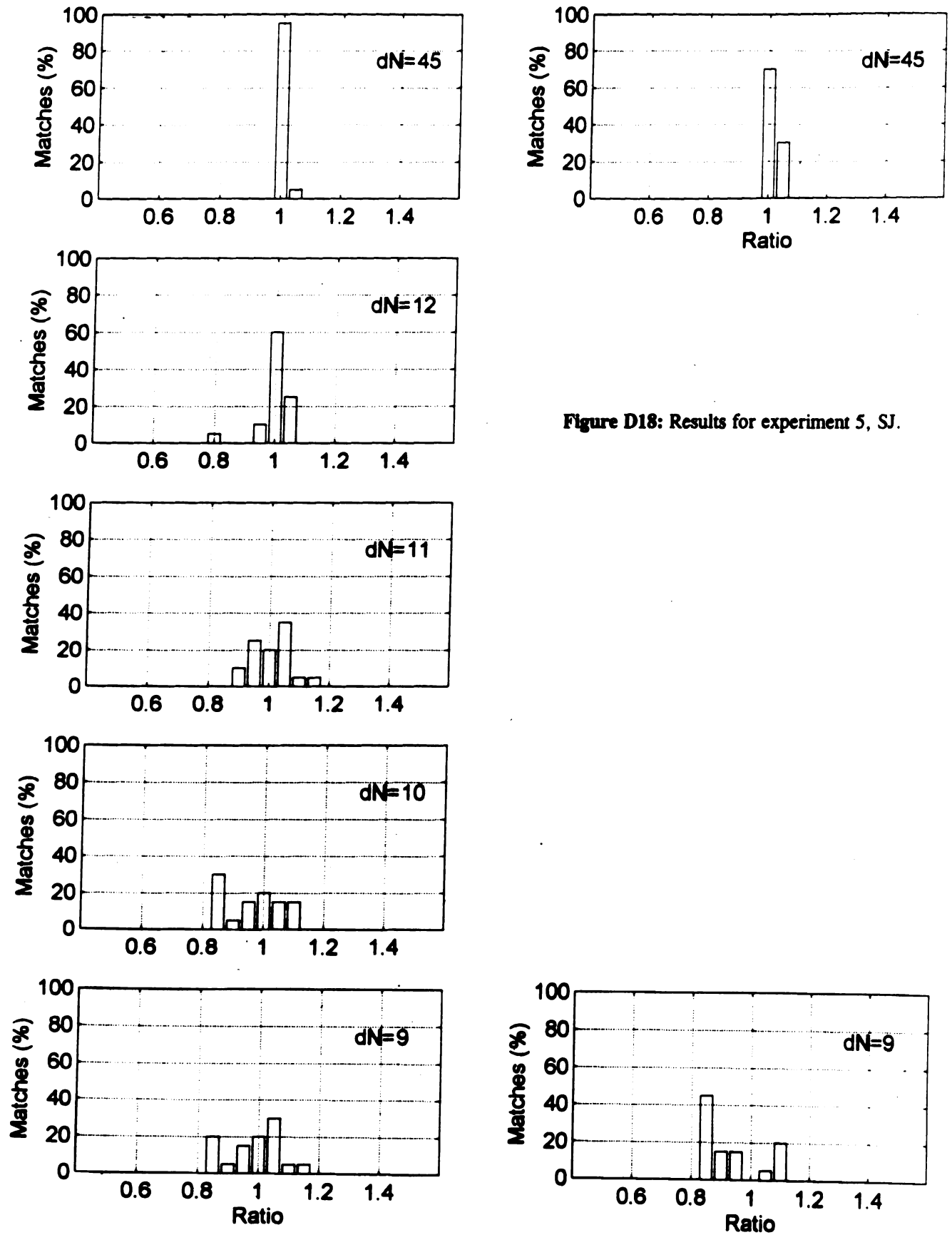


Figure D18: Results for experiment 5, SJ.

From the first glance at the figures, we can easily see that all the listeners made perfect matches with wide band stimulus (1-45 harmonics), and the matches became worse and worse as bandwidth of the stimulus decreased. To determine whether a pitch is heard or not, the percent of matches in the central bin, at ratio=1, for the omitted 19th is plotted in Figure D19 as a function of stimulus bandwidth (marked by the number of harmonics of the complex tone). As we can see, the percentage decreases as the bandwidth decreases. For all the listeners the steepest slope occurs when the graph crosses the 50% point. It is reasonable to use the steepest slope as our criterion for judging whether the pitch is heard or not. Therefore, we use 50% as a boundary, i.e. if the number of matches in the bin at ratio=1 is less than 50% of the total matches, the pitch is said to be “not heard.” Applying this criterion to our results, we find none of the listeners could hear the pitch resulting from Duifhuis effect for conditions (d), (e) and (g). This is consistent with the descriptions of the listeners that no sine tones were heard for conditions (d), (e) and (g) and a very clear sine tone from the Duifhuis effect was heard for conditions (a) and (f). This clearly shows that the short-term Fourier analysis explanation is not correct, since the output of the auditory filter bank does not differ much for the response to the cancellation tone for all the stimuli in our experiment as we discussed earlier.

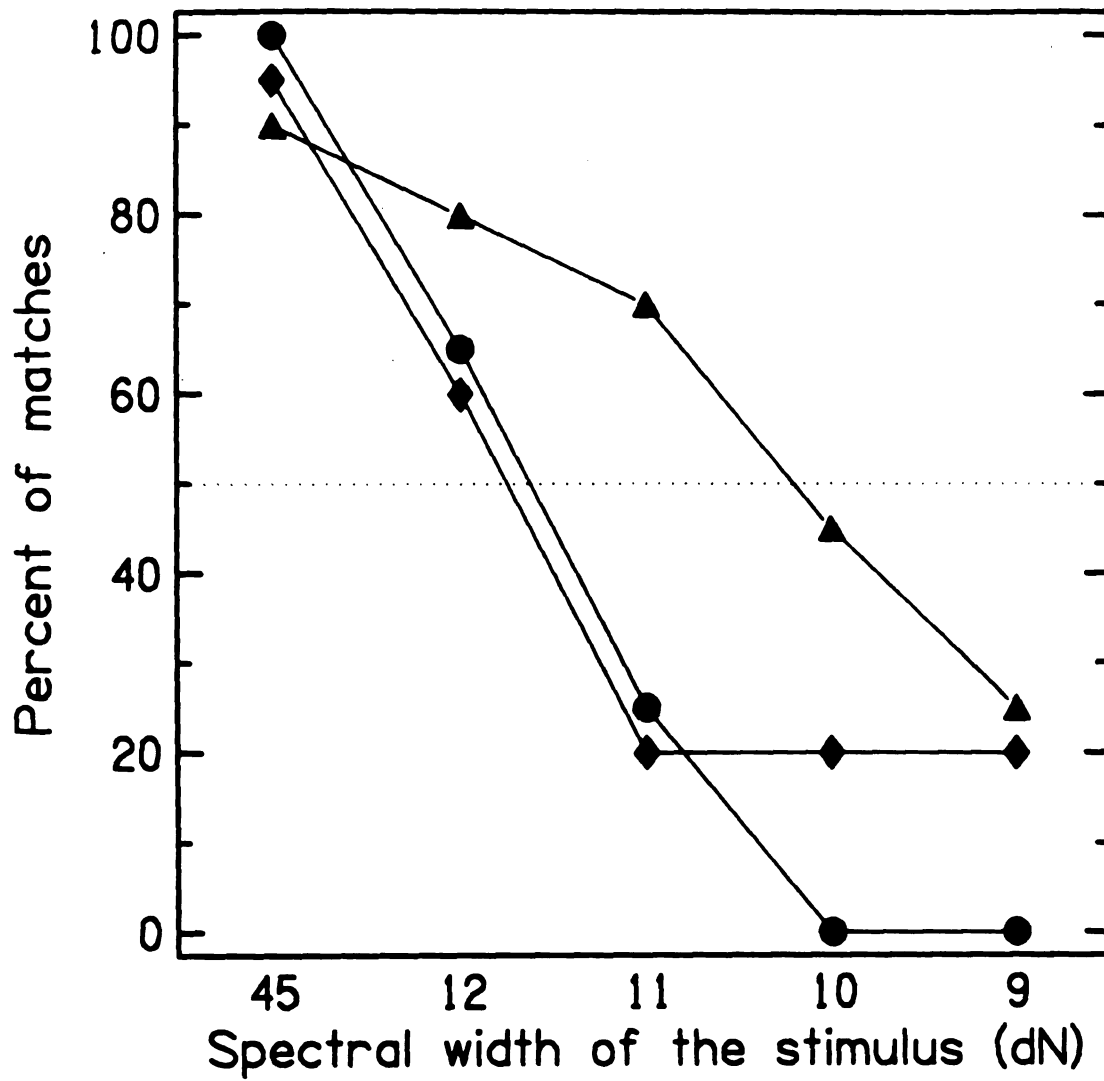
### *C. Experiment 6: Narrowing the phase-coherence region*

#### **1. Spectra of the stimuli**

In Experiment 5 we showed that components that were spectrally distant from the omitted harmonic were important for the existence of Duifhuis effect. In Experiment 6 we wanted to show that not only the amplitudes of those components were important for the existence of the effect but also the phases of those components. Therefore the

stimuli in this experiment had cosine phase for components spectrally close to the omitted harmonic. The size of the cosine phase region was changed to see how big this region must be to produce the Duifhuis effect.

There were seven stimuli. All of them were flat-spectrum 45-harmonic complex tones with high spectral end tapered. Five stimuli were complex tones with the 19th harmonic omitted: **(a)** Same as (a) in Experiment 5, i.e. all the remaining (the 19th was omitted) 44 harmonics had cosine phase; **(b)** Harmonics 14 to 18 and 20 to 25 had cosine phase, the others had random phase; **(c)** Harmonics 15 to 18 and 20 to 25 had



**Figure D19:** Plots of the percentages of matches in the central bin (left column) in Figure D16, Figure D17 and Figure D18 as a function of the bandwidth of the stimulus. Filled circle is for SB. Filled triangle is for SD. Filled diamond is for SJ.

cosine phase, the others had random phase; **(d)** Harmonics 15 to 18 and 20 to 24 had cosine phase, the others had random phase; **(e)** Harmonics 16 to 18 and 20 to 24 had cosine phase, the others had random phase. Two stimuli were complex tones with 20th harmonic omitted: **(f)** All the harmonics had cosine phases; **(g)** Harmonics 17 to 19 and 21 to 25 had cosine phase, the others had random phase.

## 2. Results

The results are plotted in Figures D20, D21 and D22. Each figure is for one subject. There were altogether 20 matches for each stimulus. The consistency of the matches is plotted by histograms, one panel for each stimulus. As in Experiment 5, we choose 0.05 as our bin size. The left column is for conditions (a), (b), (c), (d) and (e), which are all omitted-19th-harmonic conditions. The right column is for conditions (f) and (g), which are omitted-20th-harmonic conditions.

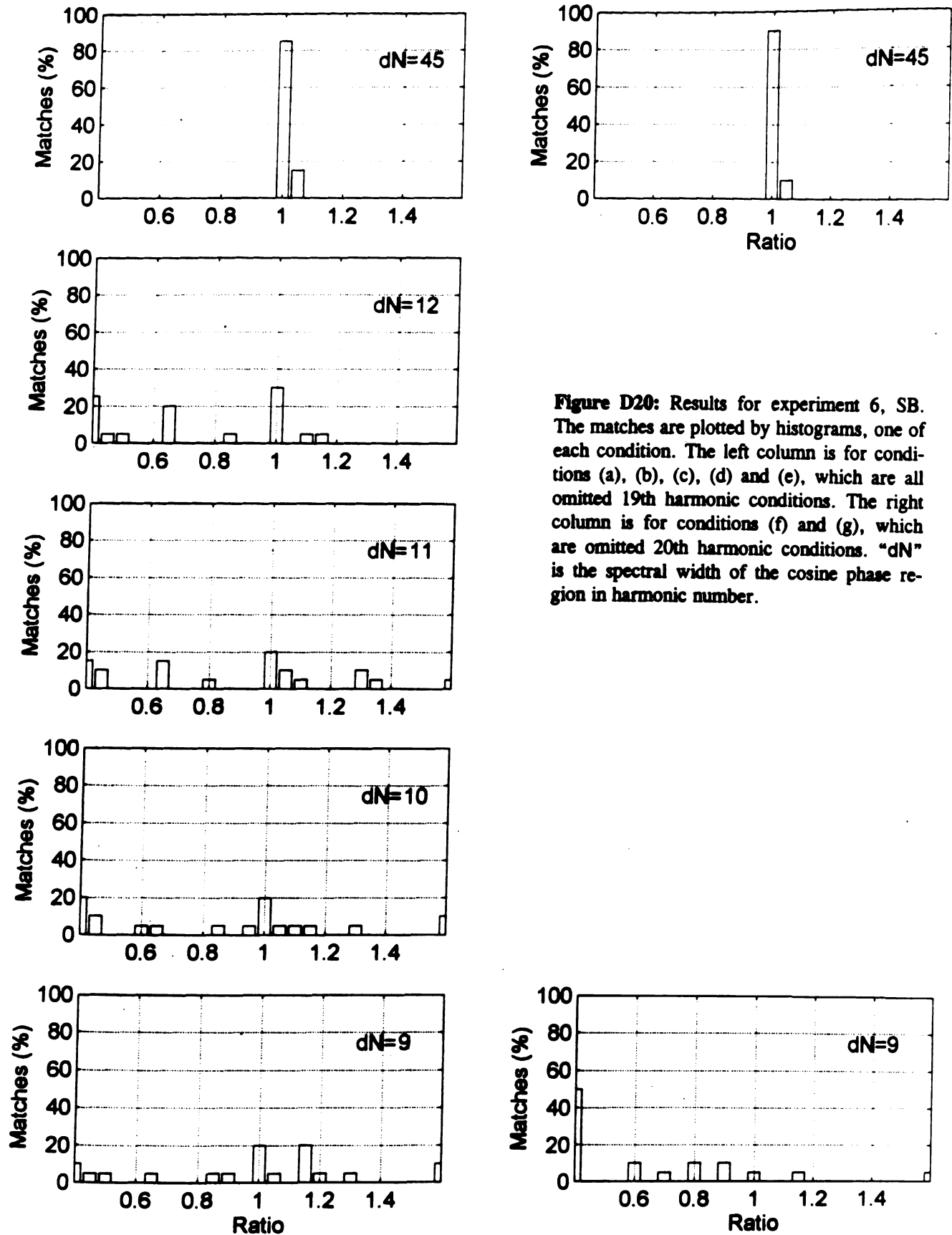
The percent of matches in the central bin, at ratio=1, for the omitted 19th is plotted in Figure D23 as a function of the bandwidth of the cosine region (marked by the number of harmonics of the complex tone). The number in the bin decreases as the bandwidth decreases. The steepest slope for all the listeners crosses the 50%. So we use 50% as a boundary, i.e. if the number of matches in the bin at ratio=1 is less than 50% of the total matches, the pitch is said to be “not heard.” Applying this criterion to our results, we find none of the listeners could hear the pitch resulting from Duifhuis effect for conditions (c), (d), (e) and (g). This shows that the phase information of the components that are spectrally far from the omitted harmonic is needed for the Duifhuis effect.



### **III. Discussion and Conclusion**

#### *A. Extensions of the Duifhuis Effect*

The Duifhuis pitch is a pitch created by a missing harmonic which is a simple spectral gap. In section I of this chapter, it is extended to several new conditions including a lot of interesting cases. Some of the extended cases might be explained by the traditional explanation although the traditional explanation is strongly contradicted by the results of our band-narrowing experiments (Experiments 5 and 6).



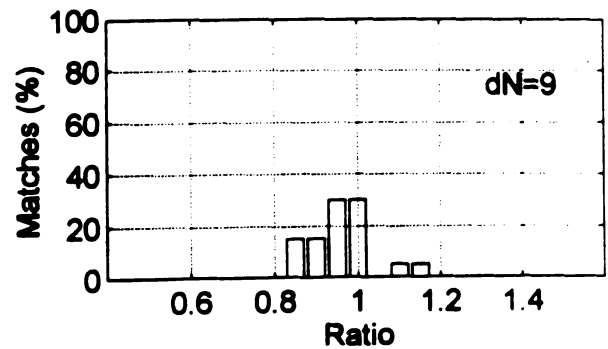
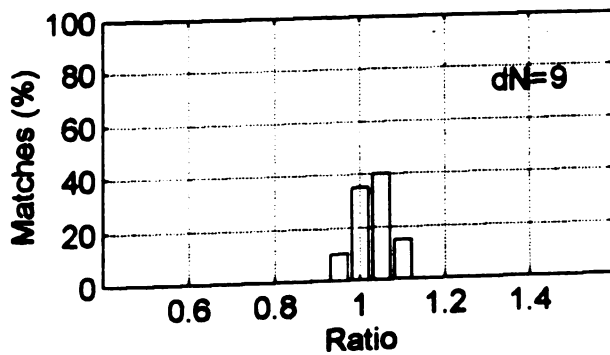
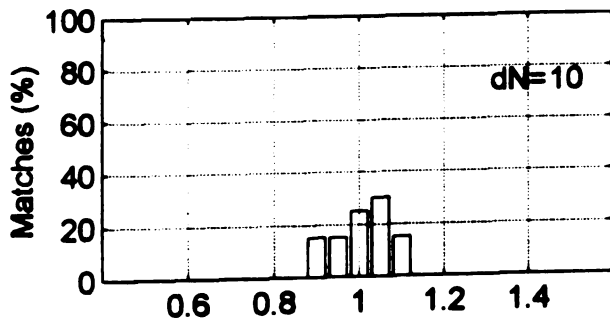
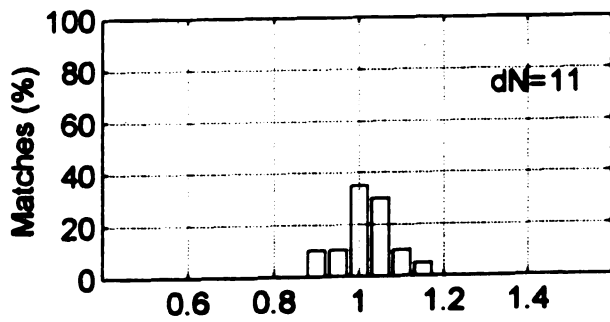
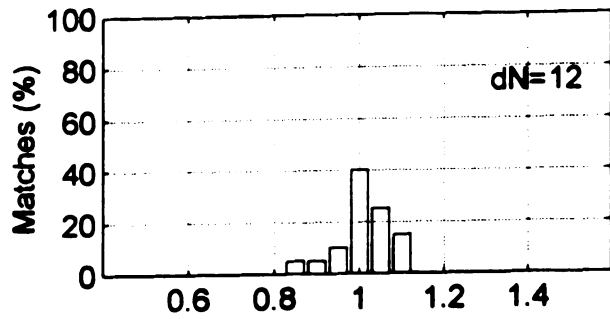
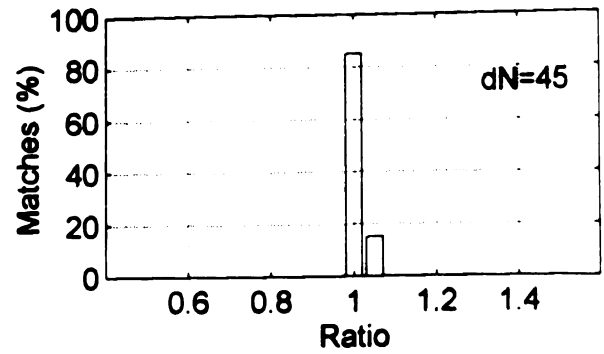
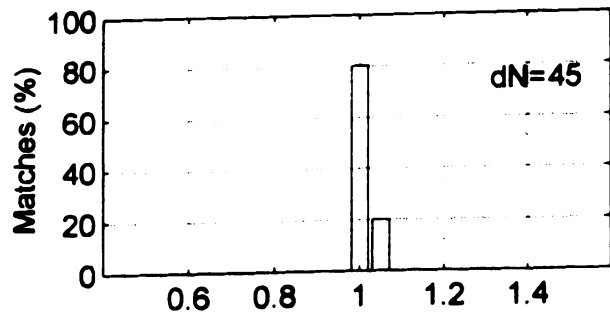


Figure D21: Results for experiment 6, SD.

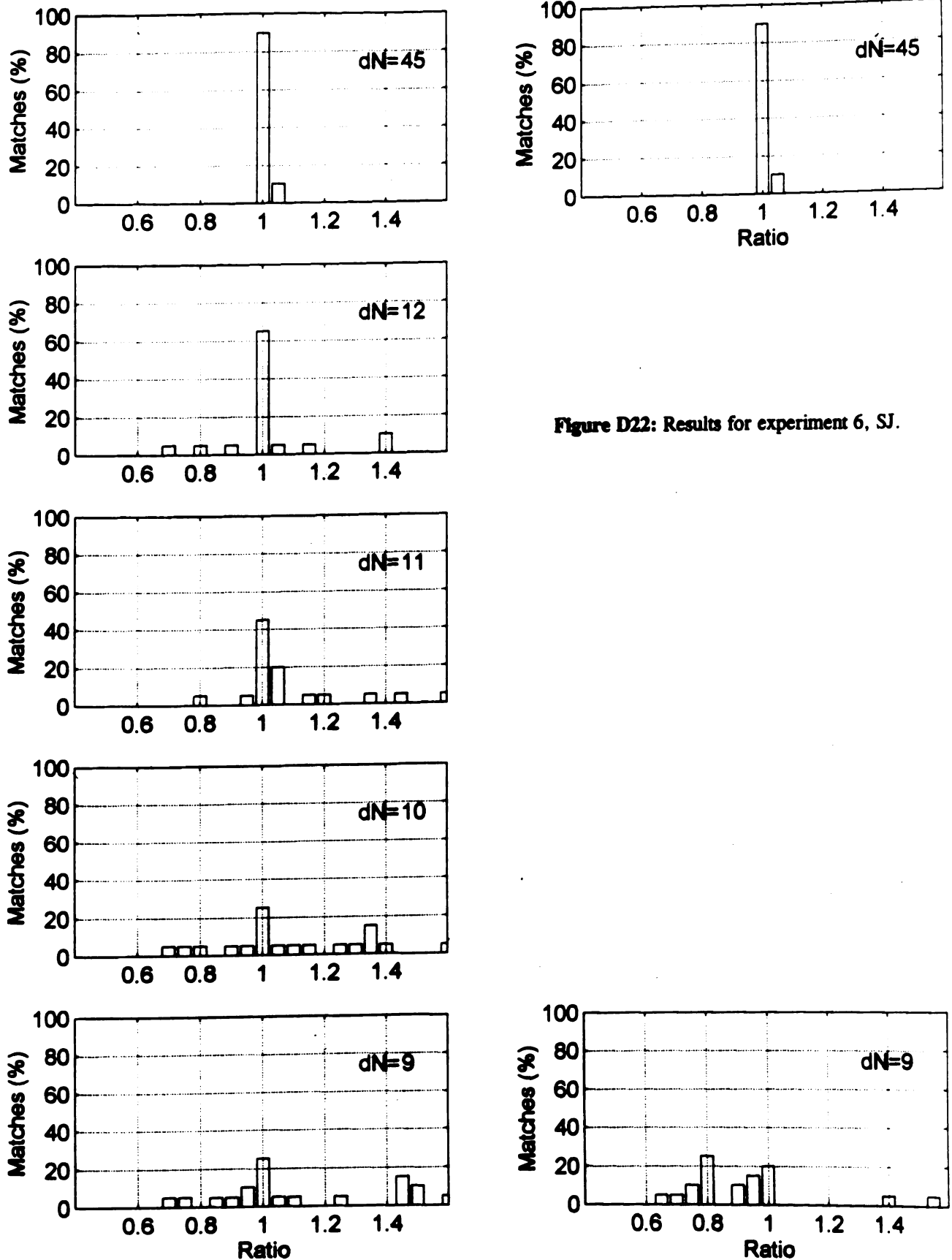
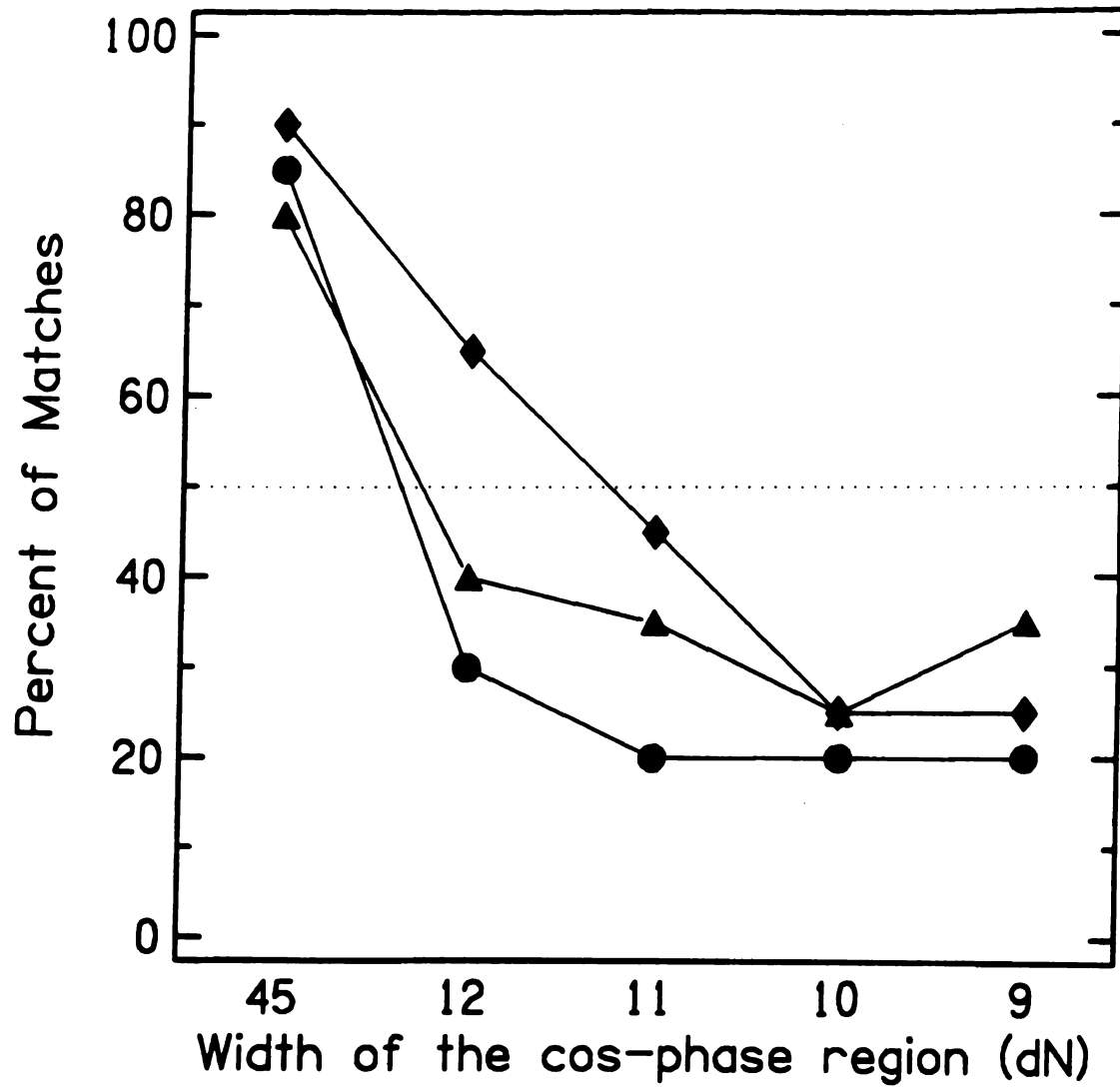


Figure D22: Results for experiment 6, SJ.



**Figure D23:** Plots of the percentages of matches in the central bin (left column) in Figure D20, Figure D21 and Figure D22 as a function of the bandwidth of the stimulus. Filled circle is for SB. Filled triangle is for SD. Filled diamond is for SJ.

The traditional explanation of the Duifhuis effect needs a relatively wide critical bandwidth of the peripheral filter compared to the spacing of the spectral components of the stimulus. It predicts a low frequency end for the effect, since at low frequencies the critical bandwidth decreases. We did find a low frequency end for the effect.

The phase-inversion Duifhuis pitch in Experiment 2 makes no change to the power spectrum and hence no change to the autocorrelation function. Thus the Duifhuis effect cannot be explained by the autocorrelation function. Using the traditional explanation, the level of the inverted Duifhuis pitch should be 6 dB higher than the omitted case, since the height of the ridge in Figure 11 for the inverted case should be increased by a factor of two. That prediction is consistent with data.

The Duifhuis effect using a Schroeder-phase complex tone makes it difficult to use the time gaps from other channels (i.e. the outputs of other auditory filters) as a time-gap cue to detect the cancellation tone, because the time gap of the output for different peripheral filters is at different times (refer to Figures D13 and D14 and see the “tilted” quiet region). However, as mentioned before, it is possible to explain the effect using the traditional explanation, because a sine-tone-like response is produced in the tilted quiet region.

The virtual Duifhuis pitch phenomenon could also be explained by combining the traditional explanation for simple Duifhuis effect (one component case) and the existing virtual pitch theory. The problem is that two separate processes, the peripheral frequency analysis process (for segregating components) and the formation of the virtual pitch process (synthesizing the segregated components), are involved in explaining this

Duifhuis virtual pitch. Yet it is possible that the segregation and the pitch formation process is just one single process.

### *B. Implications of the Duifhuis Effect*

#### 1. Duifhuis Effect and the Spectrum

The Duifhuis pitch might be called an “anomalous” pitch in that it cannot be explained based upon the power spectrum - a usual way of describing pitch perception. The effect seems anomalous because we are hearing something absent in the power spectrum. People are very familiar with and attached to the commonly accepted mode of pitch perception, i.e. analyzing the power spectrum, extracting spectral components and combining them as harmonics to synthesize a pitch (for a sine tone the same procedure just gets simpler). One of the criteria for extracting a spectral component for a pitch, either the pitch of the component (sometimes called spectral pitch) or the pitch of the complex tone, is that the larger the magnitude of the component, the bigger the contribution of the component to the pitch. Now spectral gaps which are zero-magnitude components cannot make any contribution to a pitch of any kind. This makes the Duifhuis pitch abnormal to the traditional ways of thinking.

The traditional explanation tried to make the “anomalous” effect appear normal to our commonly accepted mode of pitch perception. Thus the short-term Fourier analysis process of the peripheral filters was invoked so that at the output of the filters a real component (spectral peak) is made out within some duration by the process. However, our band-narrowing experiments show that the explanation is not correct. This implies that the Duifhuis effect is not some effect from the peripheral analysis (short-term Fourier analysis) of our auditory system. This implication is a main conclusion of this

chapter. Now both the long-term spectrum (from long-term Fourier analysis) and the short-term spectrum (from short-term Fourier analysis) cannot be used to explain the effect. Therefore, I conjecture that the segregation of the Duifhuis pitch does not take place in a spectral-like domain.

## 2. Duifhuis Effect and Segregation

In modern pitch perception theories, segregation of pitches is done by using only the power spectrum of the stimulus. For example, the DWS pitch meter (Duifhuis *et al.*, 1982) is basically a sieve on the power spectrum. In Terhardt's algorithm (Terhardt, 1982), calculations are also done by using only the power spectrum. No phase information is concerned in their segregation process. The fatal problem is that zero amplitude components cannot be segregated out by any of the above pitch theories unless some other segregation process, such as the short-term Fourier analysis, is invoked. As mentioned before, this means that the explanation of the Duifhuis effect consists of two processes. Furthermore, from the main conclusion mentioned before, the short-term Fourier analysis is not an appropriate explanation of the Duifhuis effect.

For the Duifhuis effect, listeners perceive two sound images that have two different pitches and two different timbres. One is the low buzzy complex tone, the other is a sine tone - the cancellation tone. Therefore, the Duifhuis effect involves some kind of segregation process to separate the two sounds. The segregation process and the pitch formation process might just come from one single mechanism, since the two segregated sounds form two pitches with two different timbres and in parallel the two different pitches and the two different timbres might cause the segregation. It is not parsimonious to say that a Duifhuis pitch, which is from the cancellation tone, is formed under a different pitch mechanism from that for a sine tone, because the segregation



process is so common in everyday life and the cancellation tone is just a segregated sine tone. This challenges the value of spectral analysis as a conceptual point of departure for the pitch of a normal sine tone.

It is shown in Experiment 4 that a virtual pitch can be formed by the Duifhuis effect. Provided that the mechanism for a simple Duifhuis effect (sine tone case) is not a spectral component extracting process, the perception of the Duifhuis virtual pitch should not involve any spectral component extracting process either (i.e. no sieve process or sub-harmonic cueing process). Since segregation happens all the time in our daily life, the virtual pitch of a single complex tone should be mediated by the same pitch mechanism as the Duifhuis virtual pitch. This leads us to abandon the concept of spectral component extracting and then pitch synthesizing - a traditional way of thinking about pitch perception of normal complex tones.

### *C. Conjecture on the Mechanism of the Duifhuis Effect*

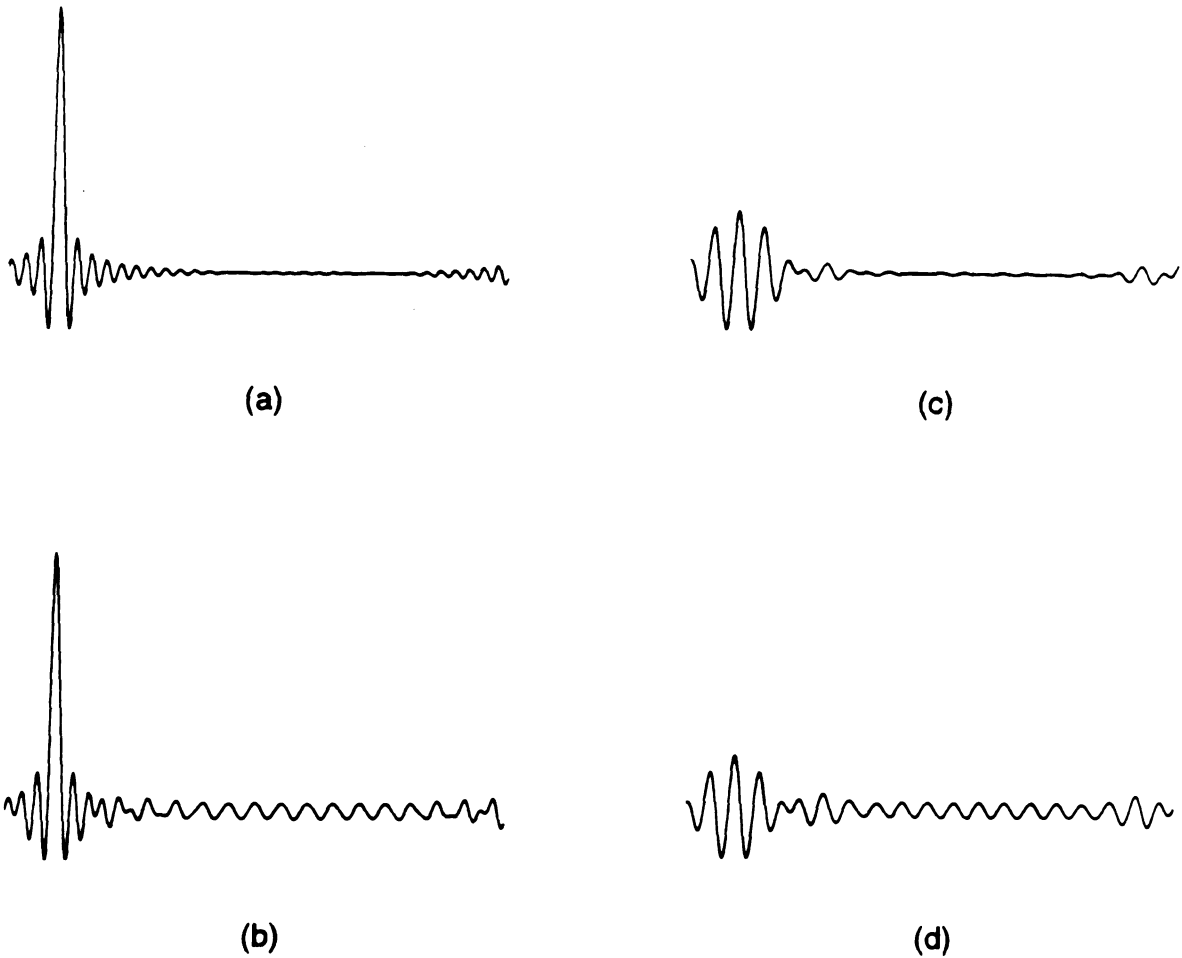
How is the Duifhuis pitch formed? The band-narrowing experiment and the cos-phase region narrowing experiment show that in order to segregate the cancellation tone, both phase and amplitude information of distant spectral components is required. So the information is equivalent to the information of the original waveform. Since we know that there is a peripheral analysis, the result tells us two things. 1) Phase information outside a critical band is not lost in this effect, which contradicts contemporary theories (Zwicker and Fastl, 1990). 2) In order to get the equivalent information as the original waveform, all the outputs of the peripheral filters must be recombined. We may call this a “waveform-restoring-like” process. Therefore, the segregation in the Duifhuis effect must be a two-waveform segregation rather than spectral components segregation.

How is the “restored waveform” (or the information equivalent to the original waveform) analyzed and segregated into two sound sources? At this time, we have no rigorous explanation for the effect. However, our conjecture is that the segregation is done by capturing different features of different sound sources. Take the cos-phase Duifhuis pitch as an example. By “scrutinizing” the “restored waveform,” the brain decides that the small oscillations and the sharp peaks come from different sound sources, since the contrast between the two parts of the waveform is strong. See figure D24 (a) and (b) for the waveforms in this case. For the narrow band condition (see Figure D24 (c) and (d) ), like we had in Experiment 5, there are no sharp peaks any more. In this case, although the small oscillations in the time gap are not changed much (in fact, it is difficult to say whether the small oscillations are weaker than those in the wide band stimulus), the contrast of the two parts becomes much lower. Thus, the waveform cannot be segregated into two sound sources. For other extended Duifhuis pitches, say the Schroeder phase in Experiment 3, there may exist some features of the two segregated waveforms which are not so obvious for us to see at this stage. Anyway, our naive explanation of the effect is: waveform restoring plus feature capturing.

Two remarks are made here. 1) The abandonment of the spectral components concept for the Duifhuis pitch, and by extension, for the pitch of a normal tone, is, in fact, concomitant with the waveform restoring, analyzing and segregating. In other words, the two arguments are actually of the same essence. The reason that we don't have a completely satisfactory pitch perception theory may lie in the fact that we need some deeper understanding of the pitch mechanism, and perhaps need to abandon some old ideas - such as the concept of assigning spectral components as harmonics. However, at this stage, more evidence is required to make this assertion, although this study could be the starting point. 2) On the other hand, the old ideas can be efficient in describing

the pitch phenomenon in certain circumstances. For example, looking at the power spectrum of a sine tone to determine its pitch is certainly very efficient. Similarly, the present pitch perception theories, such as the sieve theory, are very efficient in describing and calculating pitch of complex tones under some circumstances. So those theories are still very good theories.

Now this study shows that the original waveform or something equivalent can be “restored” at some high level of the auditory system. What do our brains extract from the waveform to produce a pitch? They must extract a common property whether the



**Figure D24:** (a) Waveform of a flat-spectrum complex tone with harmonics 1 to 45. The high spectral end is tapered. (b) The 19th harmonic of the stimulus in (a) is omitted. (c) Waveform of a flat-spectrum complex tone with harmonics 15 to 24. Both high and low spectral ends are tapered. (d) The 19th harmonic of the stimulus in (c) is omitted.

pitch is from a single frequency stimulus (pure tone pitch), a missing fundamental complex tone stimulus (virtual pitch), a comb-filtered noise (repetition pitch) or missing high harmonic (the Duifhuis pitch). To search for other kinds of “anomalous” pitches will certainly be helpful in finding the common property and in building up new theories in the pitch perception mechanism and signal segregation mechanism of our auditory system.

## **Appendix D1. Audiograms of the listeners**

Since there were differences in the frequency range of the effect among the three listeners, audiograms were done to show their thresholds of hearing. The audiograms were done by using a Bekesy tracking method.

Listeners were seated in a sound-treated room holding a box with a push button. The listener pushed the button to start. Then 300-ms tones with starting frequency (3000 Hz) and starting level (30 dB) were presented repeatedly with a 250-ms pause between them. If the listener pushed the button again, the level of the tones started to decrease, 1 dB for each tone, and the frequency of the tones began to sweep. The level continued to decrease until the listener released the button. Then the level began to increase, 1 dB for each tone. The listener was instructed to push the button as long as he heard the tones and to release the button as long as he could not hear the tone. The duration of the whole sweep from 3000 to 6000 Hz took five minutes.

The whole procedure described above was controlled by a computer through a TDT II system. The tones were generated by a WG1 waveform generator of the TDT II system. The level of the tones were changed by a PA4 attenuator of the TDT II system. The tones were presented through Sennheiser 480 II headphones, the same headphones used to run the main experiments of this chapter.

Figures D25 - D28 are results of four listeners (one more beside the three listeners). The signal level is shown by a dashed line. The middle points of each line segment are connected by solid lines, and the solid lines are assumed to be the thresholds.

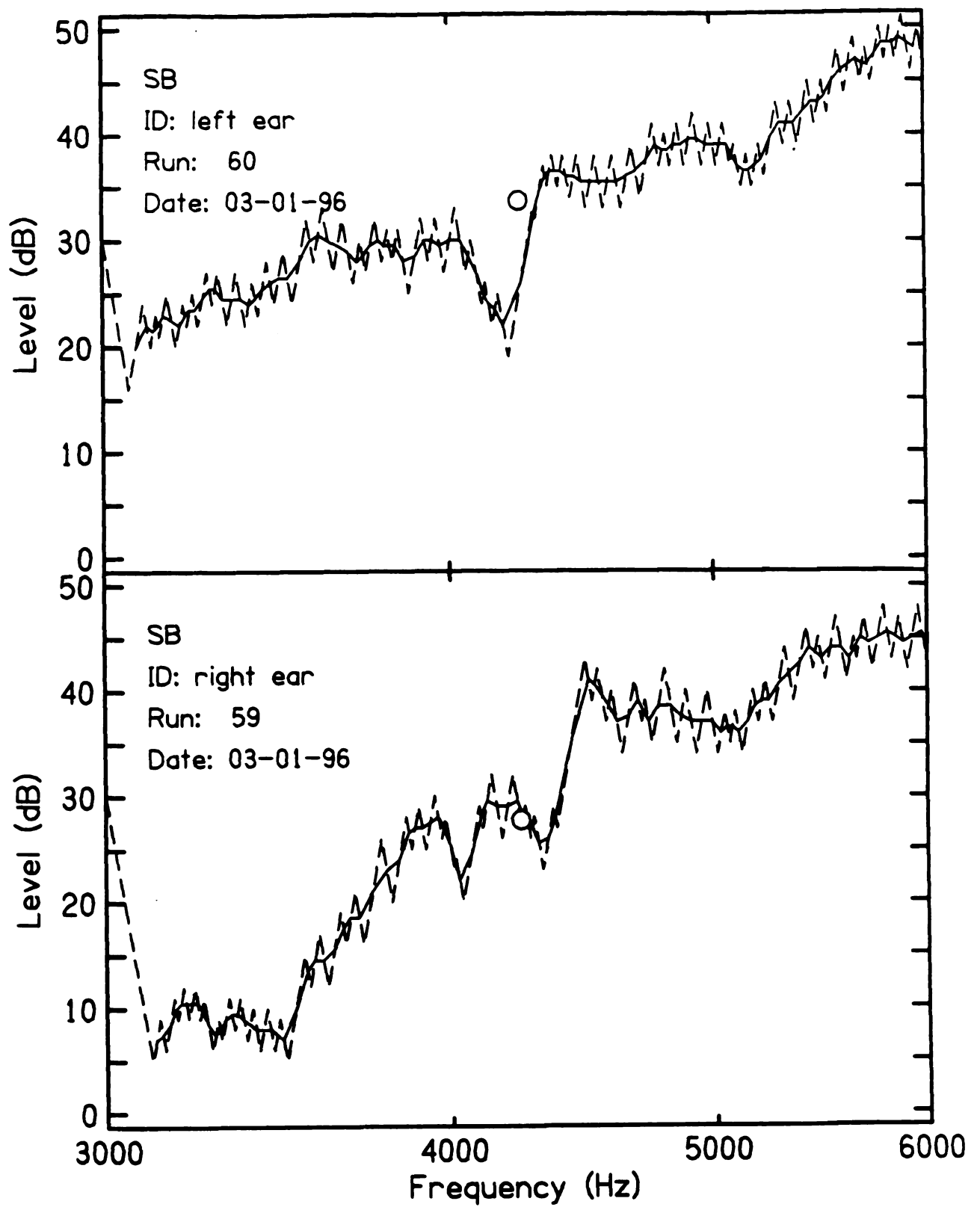


Figure D25: Audiogram of SB.

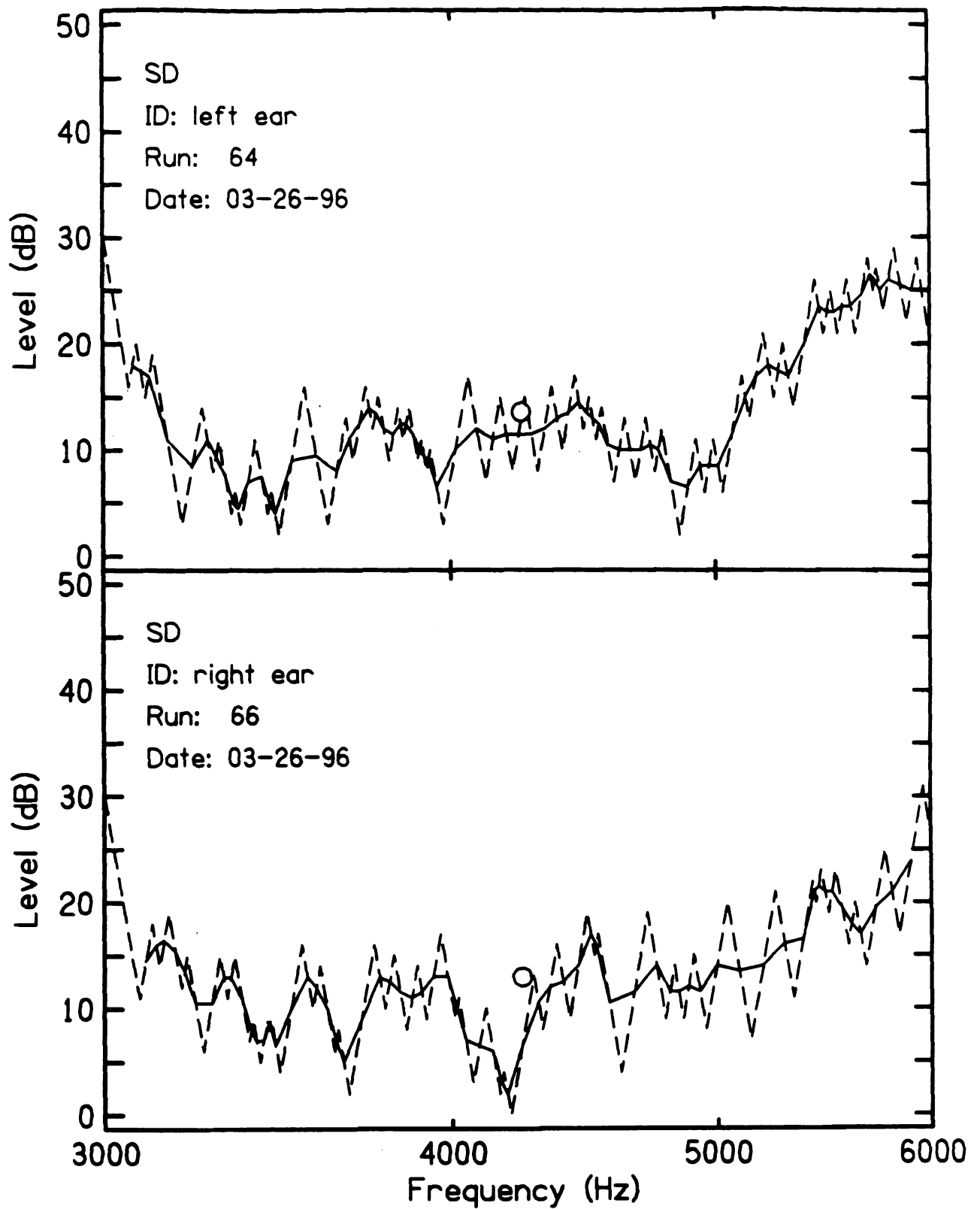


Figure D26: Audiogram of SD.



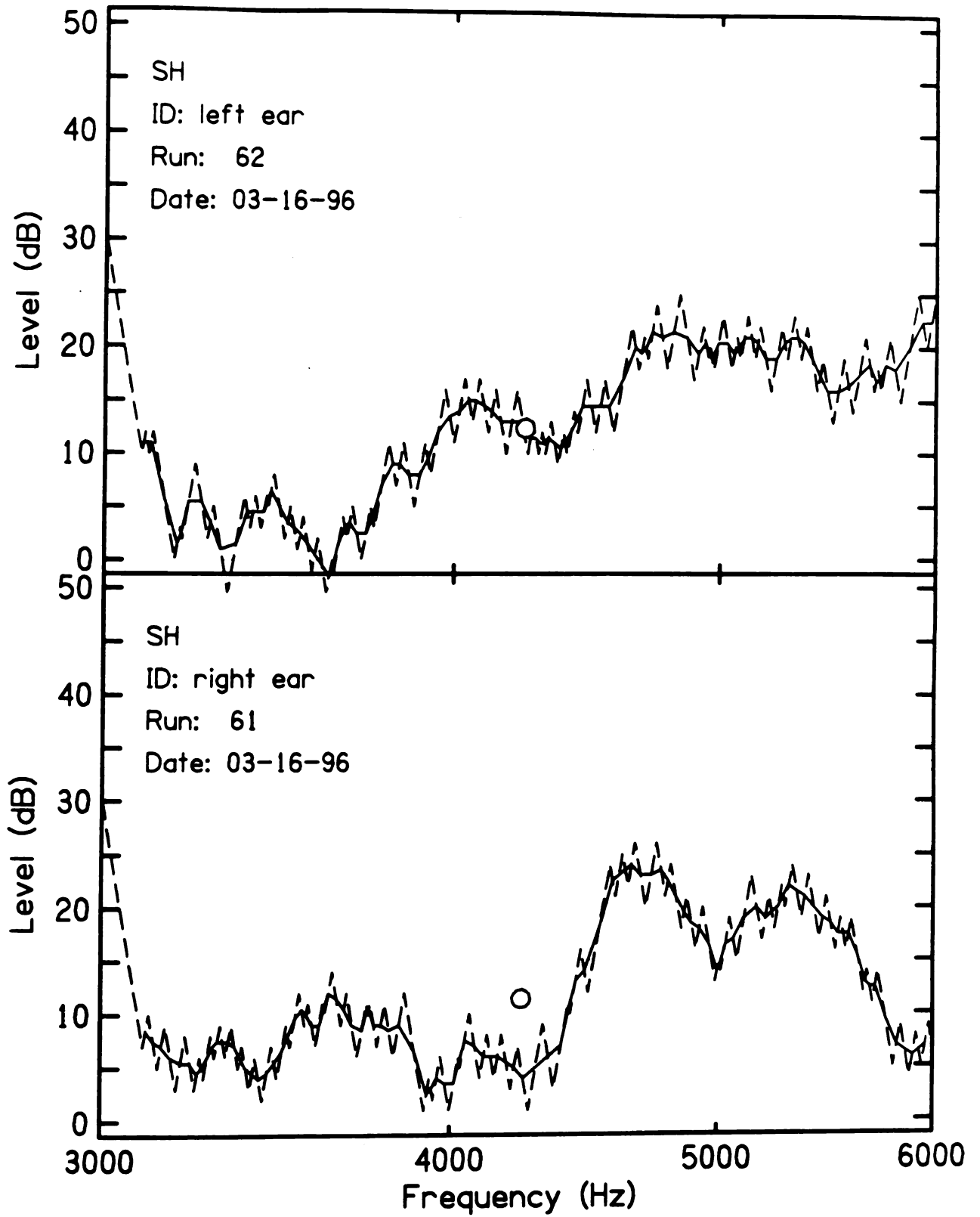
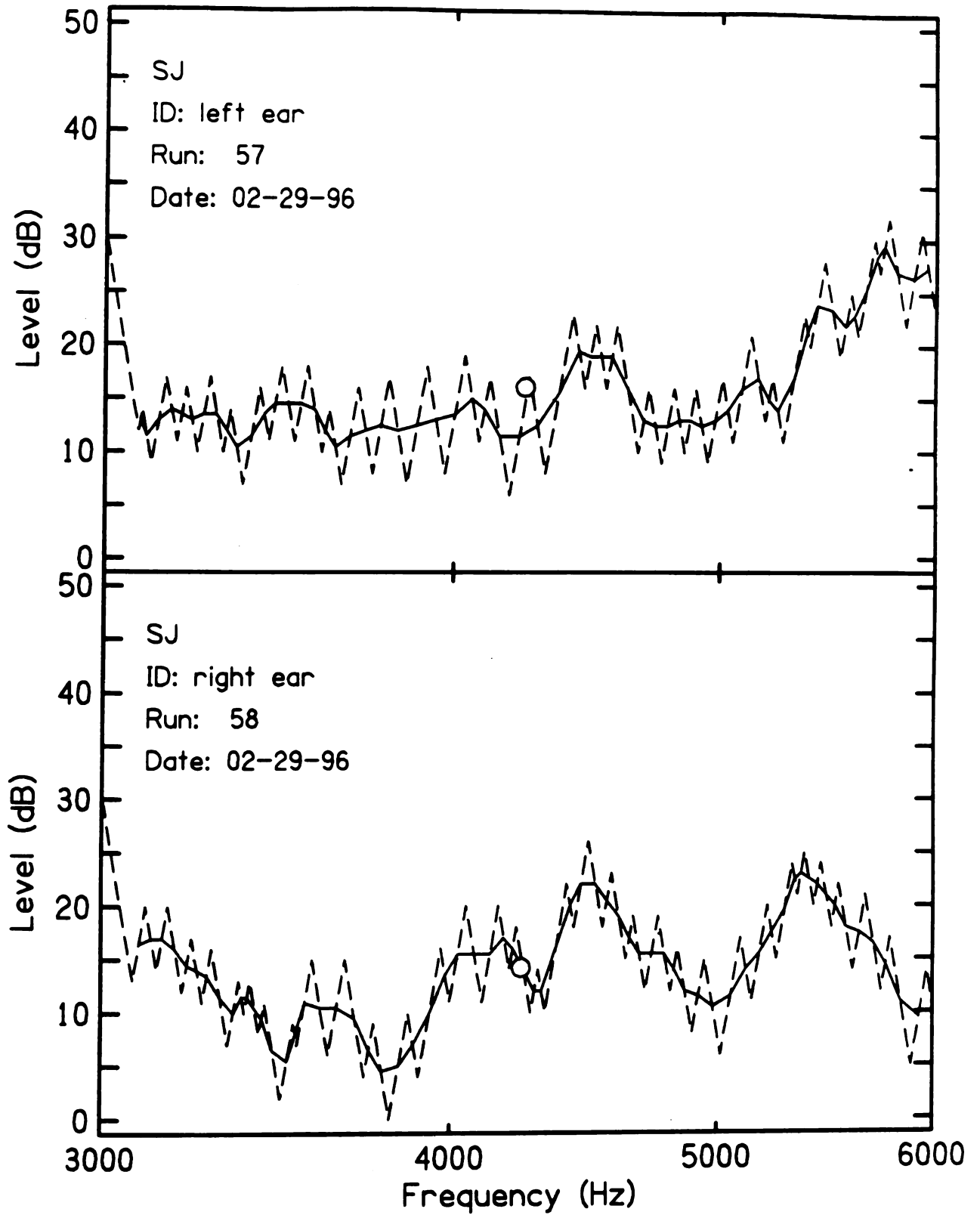


Figure D27: Audiogram of SH.



**Figure D28:** Audiogram of SJ.

## References

- Alcantara, J. I. and Moore, B. C.J. (1995) "The identification of vowel-like harmonic complexes: Effects of component phase, level, and fundamental frequency," *J. Acoust. Soc. Am.* **97**, 3813-3824.
- Deller, J.R. (1993) "Discrete-Time Processing of Speech Signals," Macmillan Publishing Company, 251-256.
- Duifhuis, H. (1970) "Audibility of high harmonics in a periodic pulse," *J. Acoust. Soc. Am.* **48**, 888-893.
- Duifhuis, H., (1971) "Audibility of high harmonics in a periodic pulse II, Time effect," *J. Acoust. Soc. Am.* **49**, 1155-1162.
- Duifhuis, H., Willems L.F. and Sluyter R.J. (1982) "Measurement of pitch in speech: An implimentation of Goldstein's theory of pitch perception," *J. Acoust. Soc. Am.* **71**, 1568-1580.
- Hartmann, W.M. (1988) "Pitch perception and the segregation and integration of auditory entities," in *Auditory Function*, ed. G.M. Edelman, W.E. Gall and W.M. Cowan (Wiley, New York) pp.623-645.
- Klein, M.A. and Hartmann, W.M. (1981) "Binaural edge pitch," *J. Acoust. Soc. Am.* **66**, 51-61.

- Kubovy, M. and Jordan, R. (1979) "Tone segregation by phase: On the phase sensitivity of the single ear," *J. Acoust. Soc. Am.* **66**, 100-106.
- McAdams, S. (1984) "The auditory image: A metaphor for musical and psychological research on auditory organization," in *Cognitive Processes in the Perception of Art*, North Holland, Amsterdam.
- Moore, B.C.J. (1973) "Frequency difference limens for short-duration tones," *J. Acoust. Soc. Am.* **54**, 610-619.
- Moore, B.C.J. and Glasberg, B.R., (1983) "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750-753.
- Moore, B.C.J. and Glasberg, B.R., (1989) "Difference limens for phase in normal and hearing impaired listeners," *J. Acoust. Soc. Am.* **86**, 1351-1365.
- Patterson, R., Nimmo-Smith, L., Holdworth, J. and Rice P. (1987) "An effective auditory filterbank based on the gammatone function," paper presented at a Speech-Group meeting for the Institute of Acoustics on Auditory Modelling, which was held at RSRE, Malvern, 14-15 December 1987.
- Pickles, J.O. (1982) "An Introduction to the Physiology of Hearing," Academic Press, 82-83.
- Pierce, J.R. (1960) "Some work on hearing," *Am. Scientist* **48**, 40-45.

Schroeder, M.R. (1970) "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. on Information Theory, IT-16, 85-89.

Terhardt, E. (1971) "Die Tonhöhe Harmonischer Klänge und das Oktaveintervall," Acustica, **24**, 126-136.

Terhardt, E., Stoll, G. and Seewann, M. (1982) "Algorithm for extraction of pitch and pitch salience from complex tonal signals," J. Acoust. Soc. Am. **71**, 679-688.

MICHIGAN STATE UNIV. LIBRARIES



31293014137669