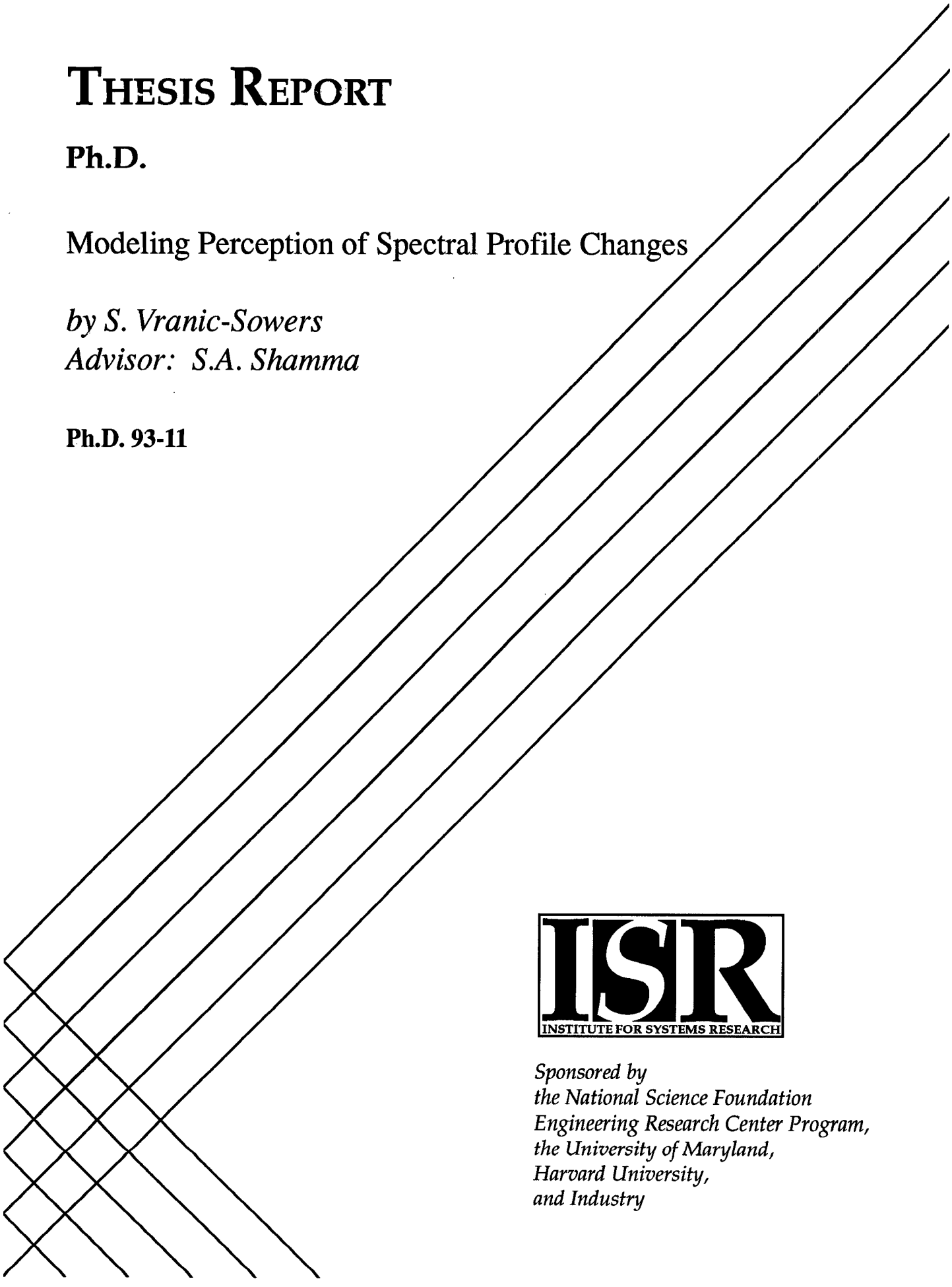
THESIS REPORT

Ph.D.

Modeling Perception of Spectral Profile Changes

by S. Vranic-Sowers
Advisor: S.A. Shamma

Ph.D. 93-11



Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

Abstract

Title of Dissertation: Modeling Perception of Spectral Profile Changes

Svetlana Vranić-Sowers, Doctor of Philosophy, 1993

Dissertation directed by: Associate Professor Shihab Shamma
Department of Electrical Engineering

In this thesis, we explore how the human auditory system represents and detects changes in a spectral profile. First, using profile analysis methods, we measure listeners' sensitivities to changes in spectral peak shapes and ripple phases. More specifically, we measure thresholds to changes in peak symmetry and bandwidth (which respectively are measures of the local evenness or oddness of a peak and of the tuning or sharpness of a peak). The effects of several other manipulations are also studied. It is found that the thresholds are constant for almost all initial peak shapes. Second, these changes in symmetry and bandwidth are interpreted as changes in the phase and magnitude of the profile's Fourier transform. In this light, the last set of experiments measured the sensitivity to (ripple) phase changes in spectral sinusoids. We find that the thresholds obtained are similar to the above-mentioned symmetry thresholds.

A fundamental conclusion arising from this analysis is that spectral peaks are represented along two largely independent axes: the magnitude and phase of their Fourier transforms. More specifically, it is argued that, along these two dimensions, the auditory system analyzes an arbitrary spectral pattern in a localized Fourier transform domain. This is closely analogous to spatial frequency transformations in the visual system. Within this general framework, we propose a model of profile analysis in which a spectral profile is represented by a weighted sum of sinusoidally modulated spectra (ripples). The first part of the analysis is performed by a bank of bandpass filters, each tuned to a particular ripple frequency and ripple phase. The parameters of the model are estimated using data from several ripple discrimination experiments. The second part of the model is a detection stage which operates on the the magnitude and phase of the computed transform, and varies with the type of perceptual task. The results of the detection operations are compared to experimental data from various profile analysis tasks. The model accounts well for the perceptual results in these tests. We propose two types of psychoacoustical experiments involving any arbitrary spectral pattern, which should further verify the predictions of the model.

Modeling Perception of Spectral Profile Changes

by

Svetlana Vranić-Sowers

Dissertation submitted to the Faculty of the Graduate School
of The University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1993

Advisory Committee:

Associate Professor Shihab Shamma, Chairman/Advisor
Professor Nariman Farvardin
Professor P. S. Krishnaprasad
Associate Professor Adrian Papamarcou
Professor John J. Benedetto
Professor Robert J. Dooling

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and motivation

The understanding of speech is the most important auditory function for humans. Speech is remarkably robust to interference from other sounds, and its intelligibility is impervious to a variety of physical transformations. While some of the robustness is due to factors specific to speech (e.g., context), other reasons are those shared with any arbitrary sound. Perception of a sound (natural, or synthesized) is intrinsically linked to its internal representation at the auditory cortex level. Perceptually relevant spectral features are encoded along the various stages of the auditory pathway, and carried up to the auditory cortex level. The primary auditory cortex (AI) is essential for the localization and processing of complex sounds. Nevertheless, except for the recent set of psychoacoustical experiments [2], studies of auditory perception have been largely unrelated to the possibly relevant physiological results. The approach taken here was to relate studying and modeling perception of various spectral shapes to new developments of relevant theories of spectral feature representations in AI.

Until recently, only two general organizational principles of AI have been identified, both of which are known to originate at much lower levels of the auditory pathway. They are: the spatially ordered tonotopic axis, which originates at the cochlea [3, 4, 5, 6]; and the alternating bands of the binaural response properties that run parallel to the tonotopic axis [7, 8], which originates at the inferior colliculus level. Recent series of physiological experiments have explored the detailed organization of cortical cell responses in order to discover what aspects of the stimuli are mapped along the isofrequency plane. Cells along the isofrequency axis are tuned to the same best frequency (BF). Best frequencies are organized in a logarithmic fashion along the tonotopic axis (Fig. 1.1 (left)).

Experimental results reveal three basic types of responses along the isofrequency line, depending on the distribution of the inhibitory responses around the BF (see Fig. 1.1(a) (right)). Such distribution of impulse responses suggests several new organizational principles (besides the tonotopic map and binaural columns). The one of immediate relevance to stationary (non-varying) stimuli is the mapping of the locally averaged gradient of the acoustic spectrum along the isofrequency planes (gradient map). More precisely, the symmetry of the AI unit responses correlates well with the responses to spectrally shaped noise bursts [9] (Fig. 1.1)(b). Since the response area symmetry is ordered along the AI, then so is the local symmetry of the spectral envelope of the most effective stimulus. Other maps are the selectivity to the direction of frequency modulated sweeps [9] (Fig. 1.1)(b); and response area bandwidth and tuning maps [10].

The organizational similarity of visual and auditory cortices [11] suggests the existence of similar mappings in both. Thus, the analogs in the visual cortex to the tonotopic and binaural column maps are the retinotopic maps and the

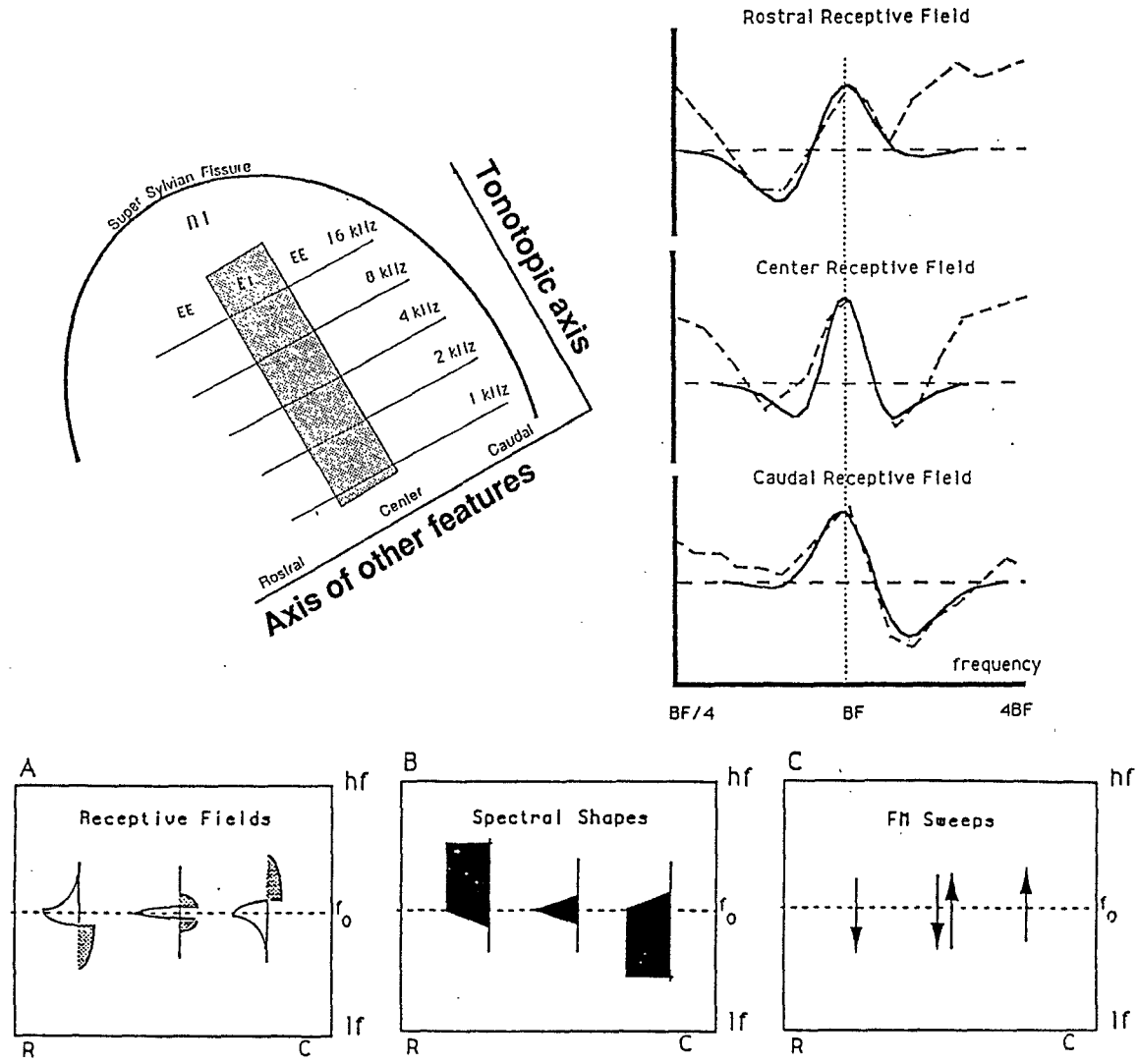


Figure 1.1: (a) A schematic organization of the primary auditory cortex (left). Three types of response areas are shown on the right. At the center of AI, units have symmetric response areas with a narrow excitation band, flanked by the inhibitory side-bands. Away from the center, the response areas become asymmetric, with the inhibitory side-bands becoming relatively stronger above or bellow the cell's BF, in two opposite directions (caudal and rostral). (b) right: Organization of the response areas along the isofrequency line. center: Neurons with inhibition to frequencies below (above) their best frequency (BF), respond optimally to noise bursts with least spectral energy below (above) BF. left: The AI unit response properties correlate well with the preferred direction of frequency modulated sweeps.

ocular dominance columns of the primary visual cortex. Similarly, the gradient map can be viewed as a one-dimensional analog of the orientation columns of the visual cortex, since the orientation of a two-dimensional edge amounts to specifying its gradient in two directions.

The experimental results described above suggest that specific features of the shape of the acoustic spectrum are being extracted and mapped in the cortex. The existence of such ordered maps has certain perceptual implications. The gradient map enhances and explicitly represents such perceptually important features as the shape of spectral peaks, edges and spectral envelope. This, in turn, suggests that in characterizing the perceptual quality of an arbitrary spectral pattern (e.g., that of a speech vowel), one has to take into account not only its peaks' frequencies and levels, but also the local gradients around, and tuning of, the peaks [12].

The first set of psychoacoustical experiments reported here was carried out in order to explore further the implications of these findings [13]. The stimuli used were spectral peaks, for which the local gradient is directly related to the symmetry of the peak. Since the shape of a peak can be effectively described by its symmetry and bandwidth, our goal was to examine the perceptual sensitivity of, and interdependence between, these features. The range of symmetries and bandwidths tested was comparable to those of vowel formant peaks. These experiments differ from previously published profile analysis experiments in the choice of a non-flat standard (a spectral peak), and in the nature of the manipulations applied to it, i.e., changes in bandwidth and symmetry, rather than amplitude. Neurophysiological experiments with the same stimuli reveal that the responses to different peak symmetries correlate well with the cell's specific

inhibition asymmetries [14].

The next set of experiments reported here tested the human subject's sensitivity to changes in the phase of ripple spectra (spectral sinusoids). These experiments were motivated by: the results of symmetry changes in spectral peaks which indicate that the detection is constant for almost all the conditions; the mathematical relationship that exists between the symmetry and the phase of the Fourier transform of the peak; and, again, by the newest findings of neurophysiological recordings. These recordings, in both, visual and auditory systems, suggest that spectral profiles are represented in the auditory system by their Fourier transformation. The idea that the brain analyzes and perceives its sensory patterns in this manner is relatively common in the vision literature, where it has been variously called multi-resolution or multi-scale representation, and spatial frequency analysis [15, 16]. Neurophysiological recordings in the auditory cortex have confirmed that cortical cells are indeed tuned to specific ripple frequencies and phases [17, 18]. Auditory cortical cells are able to perform the necessary ripple analysis.

We next show that the two manipulations of peak profile shapes can be extended to any arbitrary spectrum. They can be precisely defined in the Fourier transform domain of a profile: changing the symmetry is approximately analogous to adding a constant phase angle to all components of the profile transform; and changing the bandwidth is equivalent to dilating a profile, or, in the Fourier transform domain, to shifting the profile's transform along the logarithmic ripple frequency axis.

Within this generalized framework, a computational model of profile analysis is proposed in which a spectral profile is assumed to be represented by a weighted

sum of sinusoidally modulated spectra. The first part of the analysis is performed by a bank of bandpass filters, each tuned to a particular ripple frequency and ripple phase. The parameters of the model are estimated using data from several ripple discrimination experiments. The second part of the model is a detection stage which operates on the magnitude and phase of the computed transform, and varies with the type of perceptual task. These detection operations are applied to the filter outputs and the results compared to experimental data from various profile analysis tasks. The model accounts well for the perceptual results in these tests. Finally, two types of psychoacoustical experiments are proposed, which involve any arbitrary spectral pattern, and which should further verify the predictions of the model.

1.2 Psychoacoustics and profile analysis

Psychophysics is the study of relationship between the perceptual and physical aspect of the stimulus. It deals with procedures employed in asking questions about the presented stimulus, in order to estimate subjects' sensitivity (i.e. threshold) to changes in the stimulus. Several paradigms and procedures are used for collecting data, and estimating thresholds. In this dissertation, a common "two-interval forced-choice" (2IFC) procedure (or paradigm), and a "two-down, one-up" adaptive method was used to perform a variety of discrimination tests. The discrimination task involves presenting two (or more) stimuli and estimating the smallest difference in a particular parameter of a sound, which the subjects can distinguish with a certain probability of correct responses. We briefly review the paradigm and the method used in this work.

"Two-interval forced-choice" paradigm

A paradigm (also called procedure, or technique) is a sequence of events which occur in presenting stimuli during a single trial. The main procedure in use today is the "two-interval forced choice" (2IFC), although the "yes-no" paradigm is also used especially for estimating absolute thresholds [19]. The 2IFC paradigm has two intervals, one containing the same stimulus (called standard) over a block of trials, and the other, containing the varying stimulus (signal) which is to be distinguished. The signal and standard occur with the same *a priori* probability in one of the two intervals.

Methods

Psychoacoustical methods are concerned with estimating a physical value of the stimulus at threshold. There are several threshold tracking procedures (or methods) for collecting data. The three classical methods are: the method of limits, method of constant stimuli, and the adjustment method [20]. Common to these methods is that they are prone to subject's biases and require either blank trials or feedbacks in order minimize the bias effect.

Newer methods are based on the signal detection theory which provides a way for separating the relevant changes in subjects' performance from the bias effects. This involves computing the Receiver Operating Characteristic curves (ROC), and, from them, some measure of listener's performance. For example, the area under a ROC depends solely on listener's sensitivity, i.e., is a bias-free measure of listener's performance. Another measure often used is the probability of correct responses $p(C)$. The area or $p(C)$ plotted against the corresponding signal presentation level, defines a psychometric function. Threshold is then estimated from this function. A psychometric function in general, relates some measure of

listener's performance (e.g., $p(C)$) to the physical value of the stimulus.

Although reliable, this method requires a large number of trials and is rarely used for estimating thresholds. Prevalent methods in use today are the adaptive procedures which combine the reliability of the methods based on signal detection theory and the speed of classical methods. In particular, the adaptive "up-down" methods, which use the percentage of correct responses as a measure of listener's performance, are highly efficient, small-sample reliable, and simple.

A "two-down, one-up" adaptive method

In adaptive methods, the stimulus level is adjusted based on listener's response. After an incorrect response, the detection is made easier by increasing the difference between the two stimuli. Similarly, the difference is decreased after a fixed number of correct responses in a raw. For instance, in a "two-down, one-up" procedure, stimuli are made less distinguishable after two correct answers in a row. After several reversals (i.e., presentation levels at which the stimulus is changed in the opposite direction than in the previous case), the signal stabilizes at the value for which the probability of decreasing and increasing the difference between the stimuli becomes equal [21]. For the "two-down, one-up" case, this estimates the threshold at the level which produces 70.7% correct answers [21]. An important practical problem is deciding on the size of the first step. Although decreasing step sizes after each reversal maximizes the convergence rate to the target thresholds value [22], this procedure is for practical reasons never used. Instead, a good approximation is obtained by halving the step size once (after the third run). Simple mid-run estimates are used for estimating the threshold, i.e. the threshold is computed as an average of last even number of reversals, excluding the first three. An even number is used in order to reduce estimation

bias.

d' measure

In this section we briefly review the d' measure since it is used later in Sec. 2.10.2, and throughout Chapter 3. In modeling a sensory system it is often assumed that the internal representation of the stimulus is corrupted by some internal noise, which is, for both theoretical and practical reasons, assumed to have a Gaussian distribution. The separation between the means of the distributions for two different signals normalized by their standard deviation defines a discriminability index or sensitivity measure, d'. This measure parametrizes the ROC curves, since each ROC corresponds to a given separation of the two distributions [19]. For an ideal observer it is possible to relate the probability of correct responses for a "two-interval forced-choice" procedure to the d' measure (as $\Phi(d'/\sqrt{2}) = p(C)$, where $\Phi(\cdot)$ is the area under the normalized Gaussian distribution) (values are given in Table II of Appendix 1 in [20]). The constancy of d' for a given $p(C)$ (0.71 here) is evaluated in Sec. 2.10.2.

Profile analysis

Profile analysis is a type of test in which a subject detects a change in spectral shape rather than a change in just the overall presentation level [23]. The main mechanism involved in performing such a task is the simultaneous comparison of the energy levels of different frequency regions at the same time. Listeners can also perform successive comparisons of the same frequency regions at different times. In order to preclude the subject from making such comparisons in a profile analysis task, the overall presentation level is randomized within and across the trials (Appendix A in [23]).

1.3 Experimental setup

Two sets of equipments were used to synthesized the sound. First, the sounds were generated at 25 kHz sampling rate, via a Data Acquisition/Control Unit – HP3852A, and two 16 bit 2-Channel Arbitrary Waveform DAC – HP44726A. They were low-pass filtered at 10 kHz and passed through an equalizer (IEQ One/Third Octave Intelligent Programmable) for level adjustment.

Experiments were also conducted using a Measurement Hardware HP3565S. The executable was downloaded from an application software through an Interface/Signal Processor HP3565B1 module, and the sounds were generated via a programmable 16 bit DAC HP35656A module. The stimuli were low-pass filtered at 10 kHz using KH3905A multichannel filter.

In both cases, before presentation to listeners, sounds were gated for a 110 ms duration, including 10 ms rise and decay ramps. Sounds were delivered inside an acoustic chamber through a speaker (ADS L470), i.e., without headphones. After hearing the stimuli, the listeners responded by pressing a key button (GraphOn GO-140 terminal was used), after which a short visual feedback was provided and a paradigm was repeated.

1.4 Organization

In Chapter 2, we describe all the psychoacoustical experiments carried out for this dissertation. We start with the description of the stimuli and the general procedure followed in the tests. Next we present and discuss the results of subjects' sensitivities to changes in the shape of spectral peaks and in the phase

of spectral sinusoids. Results are briefly discussed within a general theoretical framework. The threshold measurements for above-mentioned manipulations are interpreted in the context of two existing profile analysis models [24, 25]. Based on these results, and on further physiological and psychoacoustical evidence, a new prediction model for profile analysis tasks, called "the ripple analysis model", is proposed and defined in Chapter 3. The model is used in Chapter 4 to predict the results of a wide range of profile analysis experiments which have been performed here and elsewhere. Two predictions of the model applicable to a wide range of stimuli are stated. In the last chapter we summarize the main results of this work, and discuss further research direction.

Chapter 2

Detection of spectral peak shapes and ripple phases

2.1 Introduction

The shape of the acoustic spectrum is a fundamental cue in the perception and recognition of complex sounds. It is largely uncertain, however, how this spectrum is represented in the auditory system, and what specific features are extracted and emphasized by such a representation. This issue was explored in a recent series of physiological mappings in the primary auditory cortex, AI [9]. The findings from these experiments revealed that the responses along the isofrequency planes of AI potentially encode an explicit measure of the locally averaged gradient of the acoustic spectrum.

The existence of such ordered maps has certain perceptual implications. For instance, it is likely that the perception of a spectral peak (such as a vowel formant) would be significantly affected by its symmetry and bandwidth. This, in turn, suggests that in characterizing the perceptual quality of an arbitrary

spectral pattern, one has to take into account not only its peaks' frequencies and levels, but also the local gradients around, and tuning of, the peaks. In order to explore further this and other possibilities, psychoacoustical experiments were carried out to test directly the sensitivity of human subjects to changes in spectral peak shapes. Specifically, our aim was to measure the sensitivity to symmetry and bandwidth changes in single spectral peaks under a variety of conditions, such as different spectral compositions, peak levels, peak locations, and peak frequency randomization.

The experiments reported here are similar in methodology to previously reported profile analysis experiments [24, 26, 27, 28]. They also share the same overall goals of the phonetic distance measure experiments described in [12] and [29]. Our experiments, however, differ from previously published profile analysis experiments in the choice of a non-flat standard (a spectral peak). They also differ in the nature of the manipulations applied to it, i.e., changes in bandwidth and symmetry, rather than amplitude.

These two deformations of the peak profile are somewhat more general than would appear at first glance. Specifically, if one imagines the peak profile drawn on a flat stretchable square sheet, then changing the bandwidth is equivalent to dilating the profile or pulling apart the opposite sides of the sheet. Changing the symmetry is approximately analogous to pulling apart opposite corners of the sheet, thus causing the profile to appear skewed or tilted. Clearly, such deformations of the spectral peak can be applied to, and thresholds measured and compared for any arbitrary profile drawn on the sheet. Moreover, as we shall elaborate in Sec. 2.7, these manipulations of the profile can be precisely defined in another domain – the Fourier transform domain of the profile. This

view, combined with the physiological evidence and the psychoacoustical data presented here regarding subjects' sensitivities to these manipulations, suggests that it is the transform, and not the profile itself, that is represented in the central auditory system.

In the following section, the acoustic stimuli and general experimental procedures are described in detail. Then, we present the results of subjects' sensitivities to changes in the symmetry (Sec. 2.3) and bandwidth (Sec. 2.4) of peak profiles, for different peak shapes, levels, spectral densities, and peak locations. Two control experiments are described in Sec. 2.6 in which the relevance of pitch cues and peak energy changes in the above discrimination tasks are evaluated. In Sec. 2.7, the results are briefly discussed within a general theoretical framework and further experiments with rippled spectra are performed (Sec. 2.8). We end with a general discussion of the results in relation to other profile analysis models and experiments.

2.2 General procedure

2.2.1 Method

A two-alternative, two-interval forced choice adaptive procedure was used to estimate the thresholds. Each trial consisted of two 110 ms long observation intervals separated by 500 ms pause. After listener's response, a short visual feedback was provided and a new trial started until all 50 trials that comprise one block were presented.

The discrimination task for spectral peak stimuli, was to distinguish between

the standard, which did not change over a block of trials, and the signal, which resembled the standard except for an adaptive change in spectral peak shape in each trial. The step size was defined in terms of changes in the right slope of the peak in decibels, and it differed across the testing conditions. For spectral sinusoidal stimuli, the discrimination task and stimulus parameters are described in Sec. 2.8.

On the first trial the signal was three step sizes away from the standard. On each subsequent trial the signal was changed according to the "two-down, one-up" procedure in order to estimate the level that produces 70.7% correct answers [21]. The step size was halved after 3 reversals and the threshold was estimated as the average of the signal across the last even number of reversals excluding the first three. Signal and standard occurred with equal a priori probability in one of the two intervals.

The overall presentation level was randomized across trials and within a trial over a 20 dB range in 1 dB resolution, in order to ensure that listeners base their judgement on a change in spectral shape rather than on absolute level change in a particular frequency band [23].

The results reported are based on data from two to five normal hearing subjects, depending on the particular test. Subjects were trained for about a week (four days a week, 60 – 90 minutes per day), before the actual recording took place. In order to ensure a large enough sample collection, we collected approximately sixteen threshold estimates per subject and per testing condition [30].

2.2.2 Spectral peak stimulus parameters

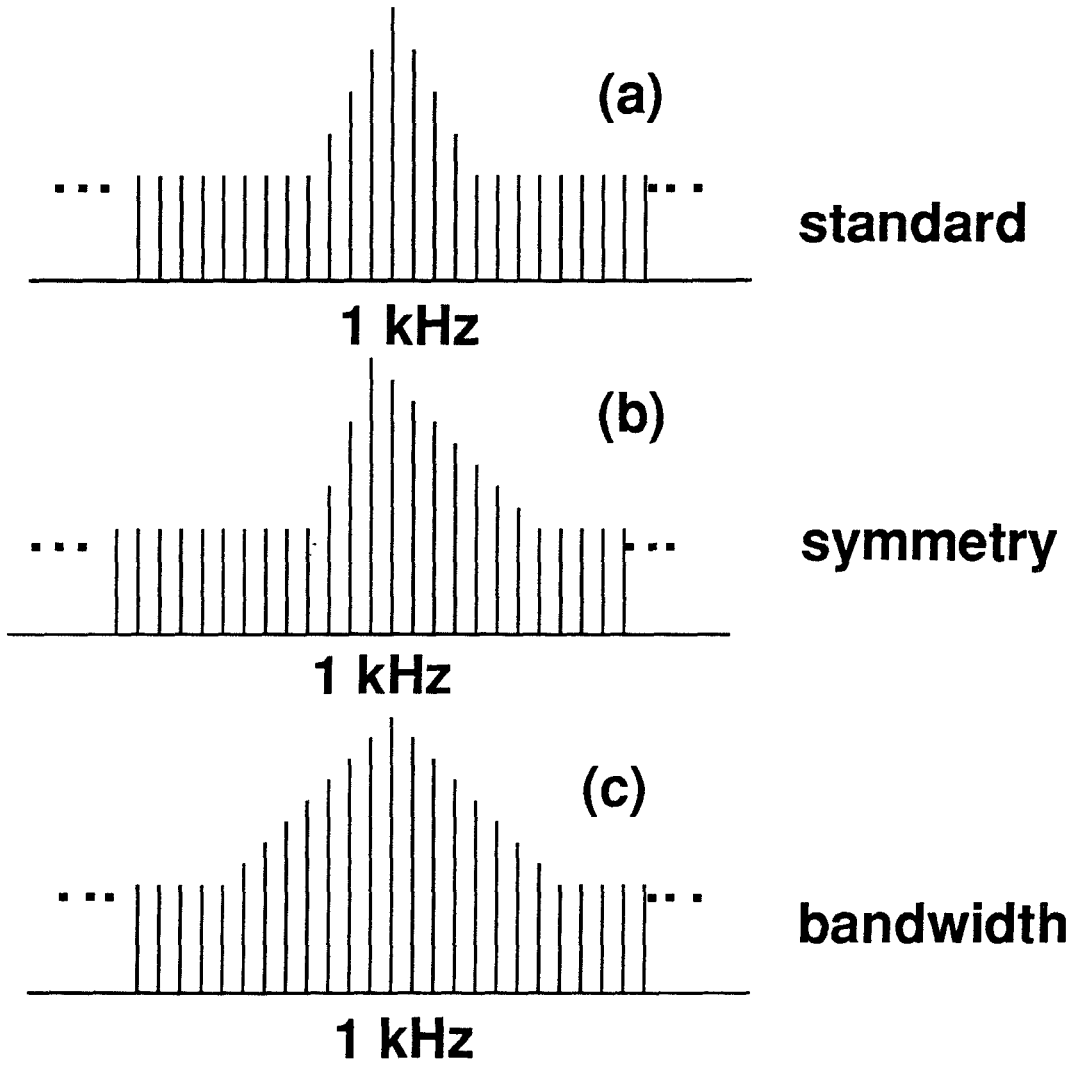


Figure 2.1: (a) Complex waveform consists of a flat base and a peak added to it. Peak takes different symmetries (b) and bandwidths (c).

Both of the multicomponent standard and signal peak profiles consist of two portions, the base and the peak. The base components were all equal in amplitude and they were added in phase to peak components of different symmetries and bandwidths to form peak profiles as shown in Fig. 2.1. The peak profile was defined against a logarithmic frequency axis (ω) in octaves, $\omega = \log_2 (f/f_o)$, where f is the frequency in (kHz), and f_o is the frequency of the largest peak component. The peak profile is defined in terms of the following parameters (Fig. 2.2(a)):

- ω_o is the location of the peak's maximum. Since the peak is always located at 1 kHz, $\omega_o = 0$.
- S is the slope of the profile near the peak's maximum (in dB/octave). For $\omega \leq \omega_o$, $S = L$ (the left slope), and for $\omega > \omega_o$, $S = R$ (the right slope).
- $b(\omega) = b$ is the flat base of the peak profile.
- $a(\omega) = a_{max} \cdot 10^{\frac{S}{20}(\omega - \omega_o)}$, is the amplitude of the peak portion of the profile. a_{max} is the maximum amplitude of the peak profile (at $\omega = \omega_o$). It is also defined in dB as $A_{max} = 20 \log_{10}(\frac{a_{max}}{b})$.

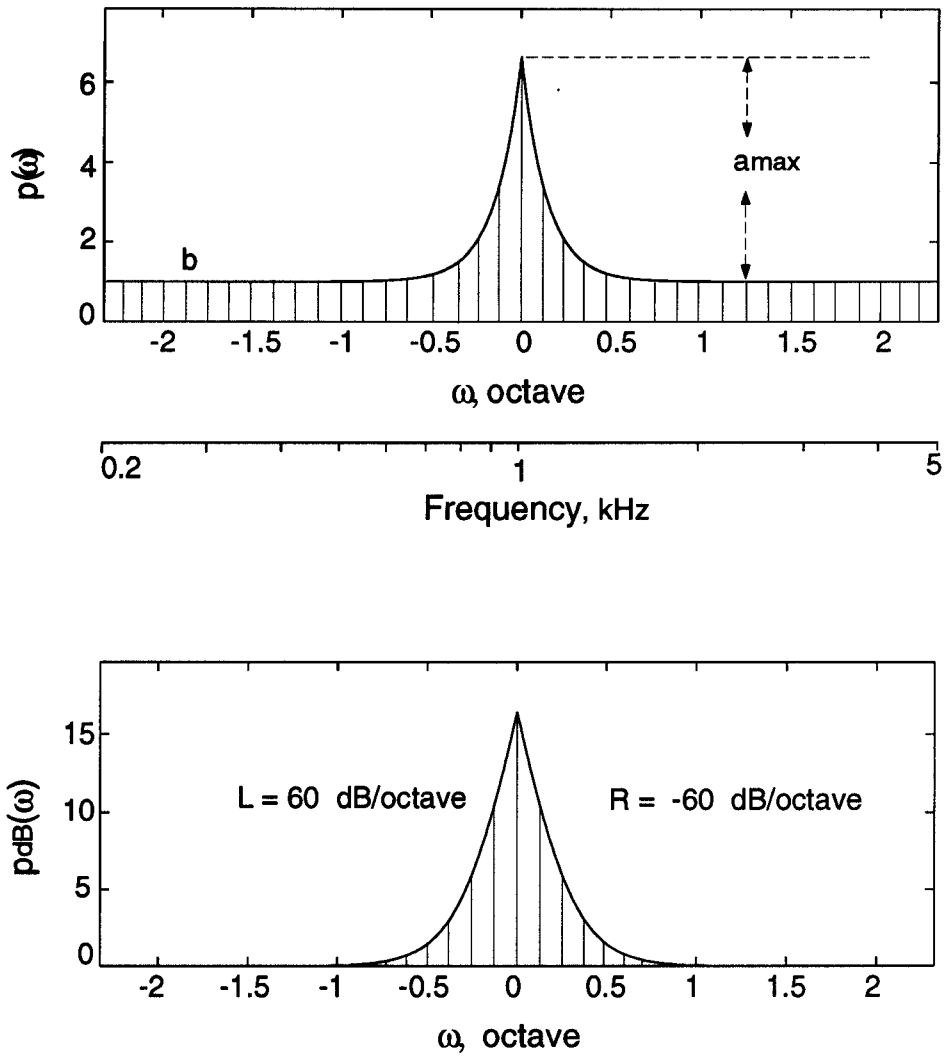


Figure 2.2: (a) Peak profile plotted on a linear (top) and logarithmic (bottom) amplitude scale. Peak level (A_{max}) is 15 dB, and BWF = 0.1 and SF = 0.

Therefore, the overall peak profile (on a linear scale) is given by:

$$p(\omega) = b(\omega) + a(\omega) = b + a_{max}10^{\frac{S}{20}(\omega-\omega_o)}, = b\ (1 + 10^{\frac{A_{max}}{20}+\frac{S}{20}(\omega-\omega_o)}), \qquad (2.1)$$

and on the dB scale:

$$p_{dB}(\omega) = 20\ \log_{10}(b + a_{max}10^{\frac{S}{20}(\omega-\omega_o)}) = 20\ \log_{10}(b\ (1 + 10^{\frac{A_{max}}{20}+\frac{S}{20}(\omega-\omega_o)})).$$

For example, the peak in Fig. 2.2(a) (plotted on linear and dB scales) is 15 dB in level (A_{max}) with slopes $L = 60$ dB/octave and $R = -60$ dB/octave around the peak. Note that around ω_o, the peak profile can be approximated by:

$$p_{dB}(\omega) \approx 20\ \log_{10}(b \cdot 10^{\frac{A_{max}}{20}+\frac{S}{20}(\omega-\omega_o)}) = 20\ \log_{10} b\ +\ A_{max}\ +\ S(\omega - \omega_o),$$

i.e., the peak has approximately a triangular profile as shown in Fig. 2.2(a).

From the above definitions, the amplitude of each component p_i in the stimulus can be computed from:

$$p_i = b + a_{max}10^{l(i-i_o)}),\ \text{for}\ i \leq i_o,$$

and

$$p_i = b + a_{max}10^{r(i-i_o)},\ \text{for}\ i > i_o,$$

where i is the component index, i_o is the index of the highest component located at the peak's maximum, $l = (L/20) \cdot (M/N)$, $r = (R/20) \cdot (M/N)$, M is the frequency range of the spectrum in octaves, and N is the (odd) number of components. For our centered peaks $i_o = (N + 1)/2$.

In order to vary the shape of the peaks, the peak profile was parametrized uniquely in terms of a symmetry factor (SF) and a bandwidth factor (BWF).

These parameters reflect the difference and the average, respectively, of the slopes around the peak. They are defined as: (1) SF = (L + R) / (L – R); (2) BWF = 3 (1/L – 1/R) octave. Thus, the peak in Fig. 2.2(a) has SF = 0 and BWF = 0.1 octave. Peaks with various other SF's and BWF's are shown in Fig. 2.2(b) covering the full range of profiles used in our experiments. Conversely, given any SF and BWF, the slopes around the peak can be computed as: R = –6/(BWF (1 + SF)) dB/octave, and L = 6/(BWF (1 – SF)) dB/octave. Note that BWF is not strictly the bandwidth of the peak, but rather is analogous to the inverse of the Q-factor of the peak. A third parameter – the peak level (A_{max}) is also required to define the peak completely with respect to the baseline.

To make the spectral peaks asymmetric, they were always tilted towards higher frequencies (or to the right). This, together with choosing the peak frequency at 1 kHz (except for the tests in Secs. 2.3.1 and 2.4.1) and limiting the range of BWF values under 0.4, ensured that the spectral peaks were located above 500 Hz where the cochlear frequency axis is assumed largely logarithmic. This is an important consideration since the peak shapes used were explicitly defined in terms of spectral slopes along such an axis. The range of SF and BWF values tested also correspond to those that might be computed from the spectral envelope of speech sounds, as shown in Fig. 2.2(c).

In all experimental conditions, standard and signal consisted of N = 11, 21, or 41 zero phase spectral components equally spaced on a logarithmic scale between 0.2–5 kHz, (ω in the range ±2.32 octaves), i.e., M = 4.64 octave. The peak was always centered at 1 kHz ($\omega = 0$ octaves) (with exception of the tests in Secs: 2.3.1 and 2.4.1). The waveform was turned on 10 ms following the onset in order to suppress the large amplitudes due to zero phases. No other

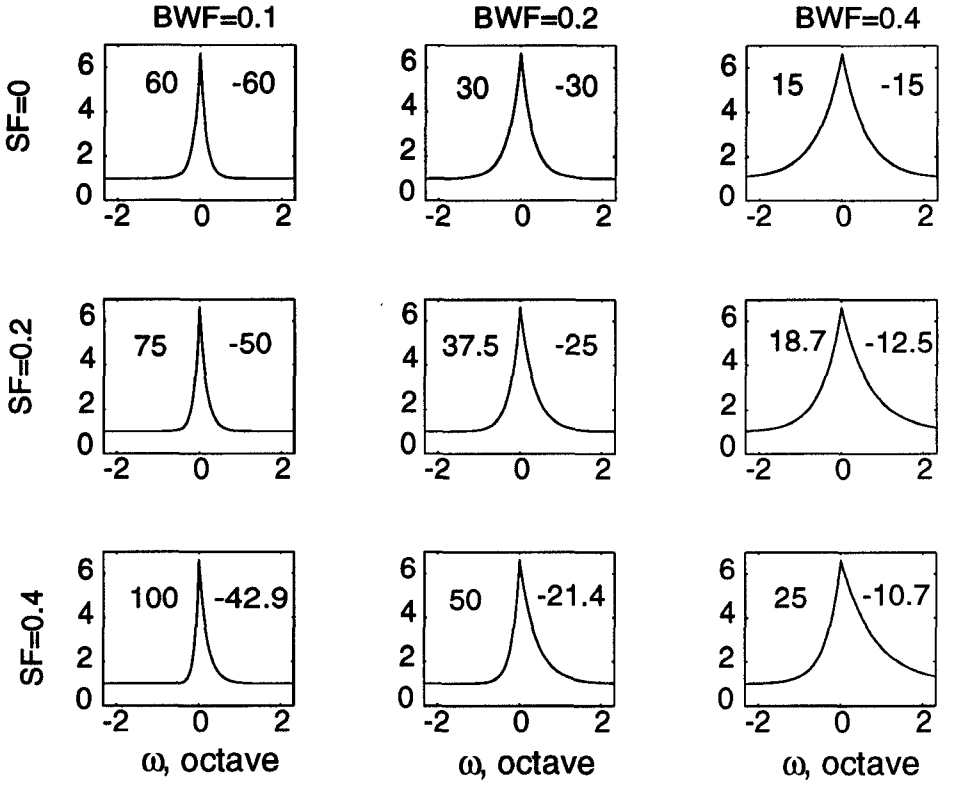


Figure 2.2: (b) Envelopes of various peak profiles plotted on a linear amplitude axis. Columns share the same BWF's, and rows share the same SF's. Corresponding left and right slope values (in dB/octave) are shown for each case.

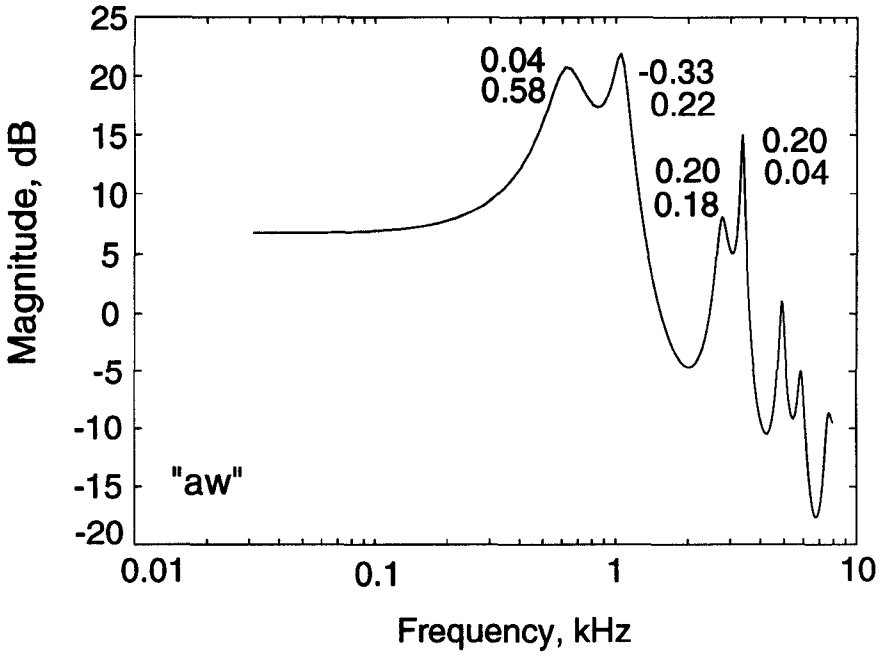


Figure 2.2: (c) SF's and BWF's for the spectral peaks of a naturally spoken vowel "aw".

phase conditions were tested since numerous previous experiments have shown that phase effects on signal detection are minimal [31, 27].

2.2.3 Spectral peak threshold measures

Threshold measures reported here were derived from threshold estimates of the signal given in terms of the right slope. This, together with the paradigm conditions (a constant SF or BWF) defines uniquely the corresponding left slope, and therefore the SF and BWF of the peak at threshold.

Two types of measures were defined and computed: (1) The first is in terms of the amount of change in SF or BWF needed for detection, i.e., δSF or δBWF. In the case of BWF change tests, thresholds are normalized by the peak's BWF

(i.e., δBWF/BWF). (2) The second measure is the root-mean-square of the change in peak energy needed for detection (see Appendix A). It is referred to as the rms–threshold.

The two types of threshold measures described above imply different detection models. We shall emphasize in this dissertation the presentation and interpretations of the first type of threshold. The rms–thresholds for all tests are compiled in Appendix A, mostly to facilitate comparisons with results from other profile analysis experiments previously reported.

2.3 Detection of changes in spectral peak symmetry

For all testing conditions in this section, peak bandwidth factor (BWF) was kept constant over a set of trials, so that both standard and signal were of the same BWF. This forced listeners to base the signal detection on a change in peak symmetry factor (SF).

2.3.1 Results and discussion

Dependence on symmetry and bandwidth factors of the standard

A 41 component complex was used in this set of experiments. The peak amplitude was fixed at a level which allowed it to be heard clearly (15 dB above the baseline). The detection threshold was measured for standard peaks of four

different bandwidth factors (BWF = 0.1, 0.13, 0.2, 0.4), and five different symmetries (SF = 0, 0.1, 0.15, 0.2, 0.4), i.e., a total of 20 tests were run. The averaged results for five subjects are presented in Figs. 2.3. In Fig. 2.3(a) the data are averaged over the four BWF's and plotted against SF. In Fig. 2.3(b), they are averaged over the five SF's and plotted as a function of BWF.

The fundamental result that emerges from these data is that, in the range of SF's and BWF's tested, the detection of a change in peak symmetry (δSF) is largely independent of the peak shape of the standard. Thus, δSF does not vary as a function of SF (Fig. 2.3(a)). However, there is a slight consistent decrease in threshold as a function of BWF (Fig. 2.3(b)). This is mostly evident for the narrowest peaks as δSF drops by 0.04 for the first 0.38 octave change in BWF (from BWF = 0.1 to 0.13), and by 0.03 for the next 1.62 octaves (from BWF = 0.13 to 0.4). For all other conditions, the δSF at threshold is near 0.11.

Plots of the rms–thresholds of these tests are shown in Appendix A. They are independent of SF and BWF, with average detection threshold at $\approx$ –8.5 dB.

The subjects trained relatively quickly to distinguish signal from standard for all test conditions above. To make the distinction, they reported that they were listening for the "higher" sounding complex tone (signal). This pitch-like change is intrinsic to the symmetry detection task as defined here, because the signal was tilted to the right from the standard, i.e., towards the higher frequencies. This "pitch" effect is further explored in Sec. 2.6.1.

Dependence on peak amplitudes

In order to determine how the detection threshold depended on peak levels, the tests described in Sec. 2.3.1 were repeated at two other peak levels: 10 dB

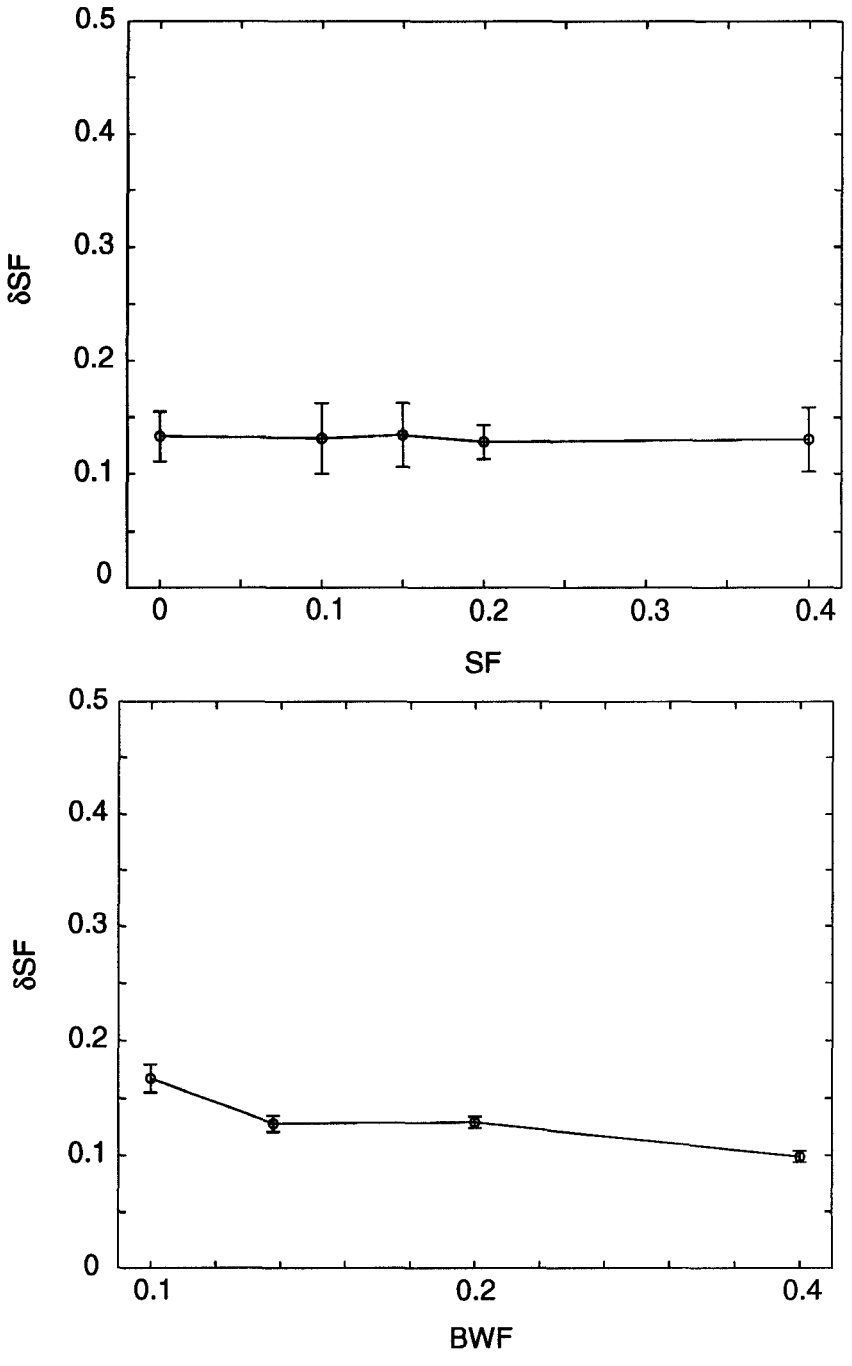


Figure 2.3: Symmetry change detection δSF thresholds for 41 component complex and 15 dB peak amplitude, averaged over five subjects and: four BWF's in (a), and five SF's in (b). The δSF threshold measure is defined as the change in SF between the signal at threshold and the standard. In (b), the δSF increases for the narrowest BWF. The error bars are the standard deviations of the means.

and 20 dB above the baseline. To account for the fact that two new subjects participated in this series of tests, experiments at 15 dB level were repeated as well. A total of 9 different conditions were tested at each peak level: three SF's (0, 0.2, 0.4) and three BWF's (0.1, 0.2, 0.4). The data obtained are presented in Figs. 2.4. As in Figs. 2.3, data are averaged over the BWF's in Fig. 2.4(a) and over the SF's in Fig. 2.4(b), for each of the levels.

Two conclusions can be derived from these data:

(1) The same trends described earlier hold regardless of peak levels. Thus, except for the narrowest peak, all δSF thresholds are the same regardless of peak shapes studied. The rms–thresholds are independent of peak's SF and BWF (Appendix A). Note that on average, the two subjects here exhibited uniformly higher thresholds than the earlier five in Sec. 2.3.1.

(2) δSF thresholds as a function of BWF (Fig. 2.4(b)) deteriorate faster at the narrowest peaks with decreasing peak level. This rise is largely responsible for the upward shift in the mean of δSF's in Fig. 2.4(a) with decreasing peak levels. The overall slight rise in thresholds may reflect the masking of the peak by the base, which presumably increases for lower peak levels.

Spectral density dependence

These experiments explored threshold dependence on the spectral density of the complex while keeping total base bandwidth constant (0.2–5 kHz). The following signal parameters were tested with four subjects: 41, 21, and 11 spectral components, two SF's (0, 0.4), and three BWF's (0.1, 0.2, 0.4). For two of the subjects, additional SF's were tested: SF = 0.1, 0.15, and 0.2 in the 41 component tests, and SF = 0.2 in the 21 component case. Peak level was always set at

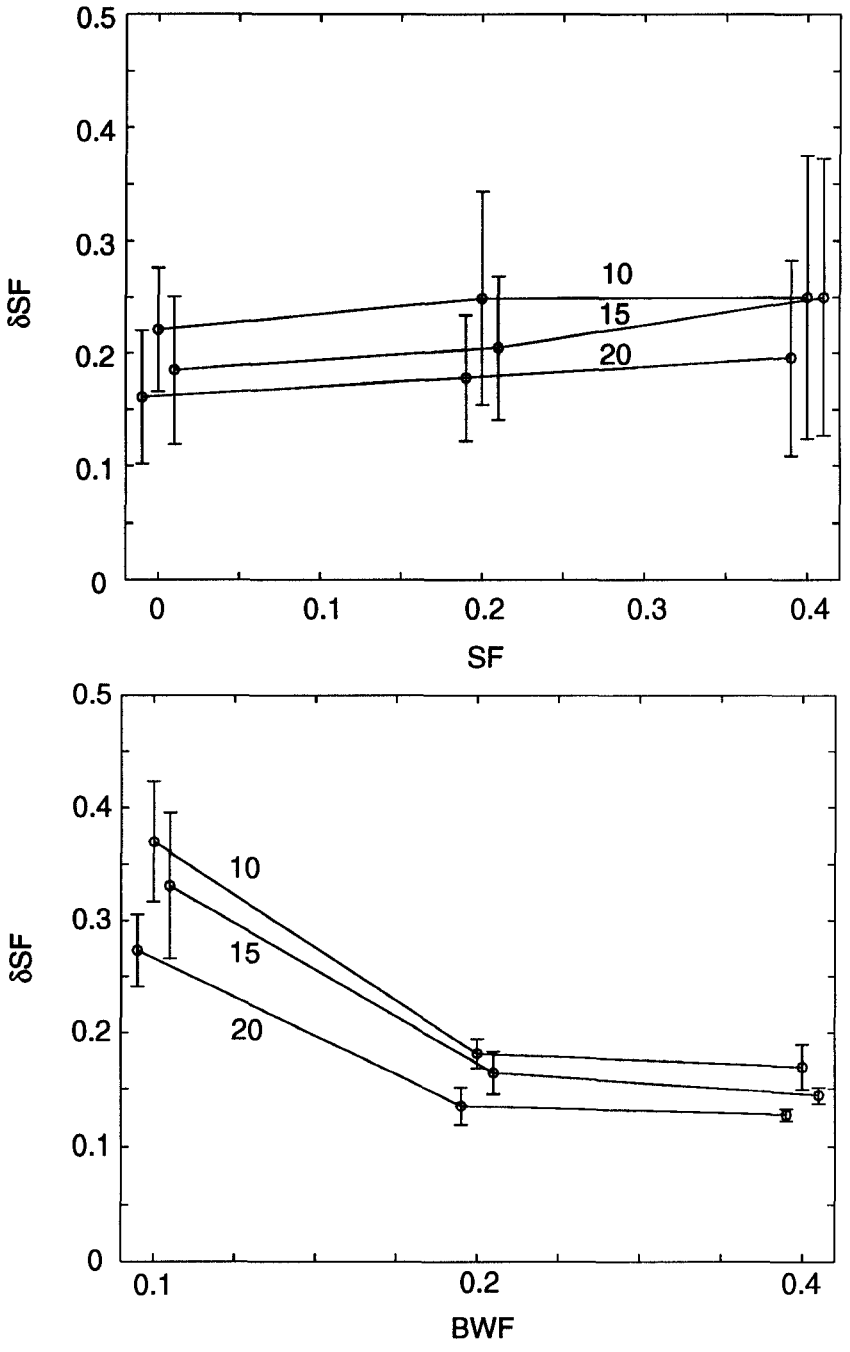


Figure 2.4: Symmetry change detection δSF thresholds for 41 component complex and 3 peak amplitudes: 10 dB, 15 dB, and 20 dB, relative to baseline. The data are averages of three subjects and: three BWF's in (a), and three SF's in (b). The values along the ordinates are defined as in Fig. 2.3. The large error bars in (a) are due to the δSF threshold increase at the narrowest BWF seen in (b). Points are slightly offset along the abscissa for clarity.

15 dB above the baseline.

Once again, all δSF values and trends described earlier largely hold regardless of spectral densities (Figs. 2.5). The most prominent change in δSF thresholds occurs as a function of spectral density at the narrowest peak (Fig. 2.5(b)). The threshold here deteriorates rapidly as the spectral density decreases and, as in Figs. 2.4, it is largely this accelerated rise that is responsible for the upward shifts in the mean δSF in Fig. 2.5(a).

Note that the rms–threshold plots in Appendix A do not immediately present a comparable picture since the rms–threshold directly reflects also the change in overall peak energy as the density is varied.

Dependence on location of the largest peak component

This set of experiments explored the threshold dependence on the location of the maximal peak component. A 21 component complex was used, with peaks located at $7^{th}, 9^{th}, 10^th$ (centered peak), 11^{th}, and 13^{th} component. Four subjects were tested for 6 different conditions (SF's = 0, 0.4; and BWF's=0.1, 0.2, 0.4), at each peak location. The peak amplitude was 15 dB above the baseline. The averaged results for four subjects are presented in Figs. 2.6. In Fig. 2.6(a) the data are averaged over the three BWF's and presented for SF=0 (left) and SF=0.4 (right). In Fig. 2.6(b) they are averaged over the two SF's and shown for the three BWF's.

Two conclusions can be derived from these data:

(1) Same trends described earlier for centered peaks (Sec. 2.3.1) hold regardless of the peak location. Thus, the δSF thresholds are independent of the SF's over the range of the SF's tested (Fig. 2.6(a)). They are also independent

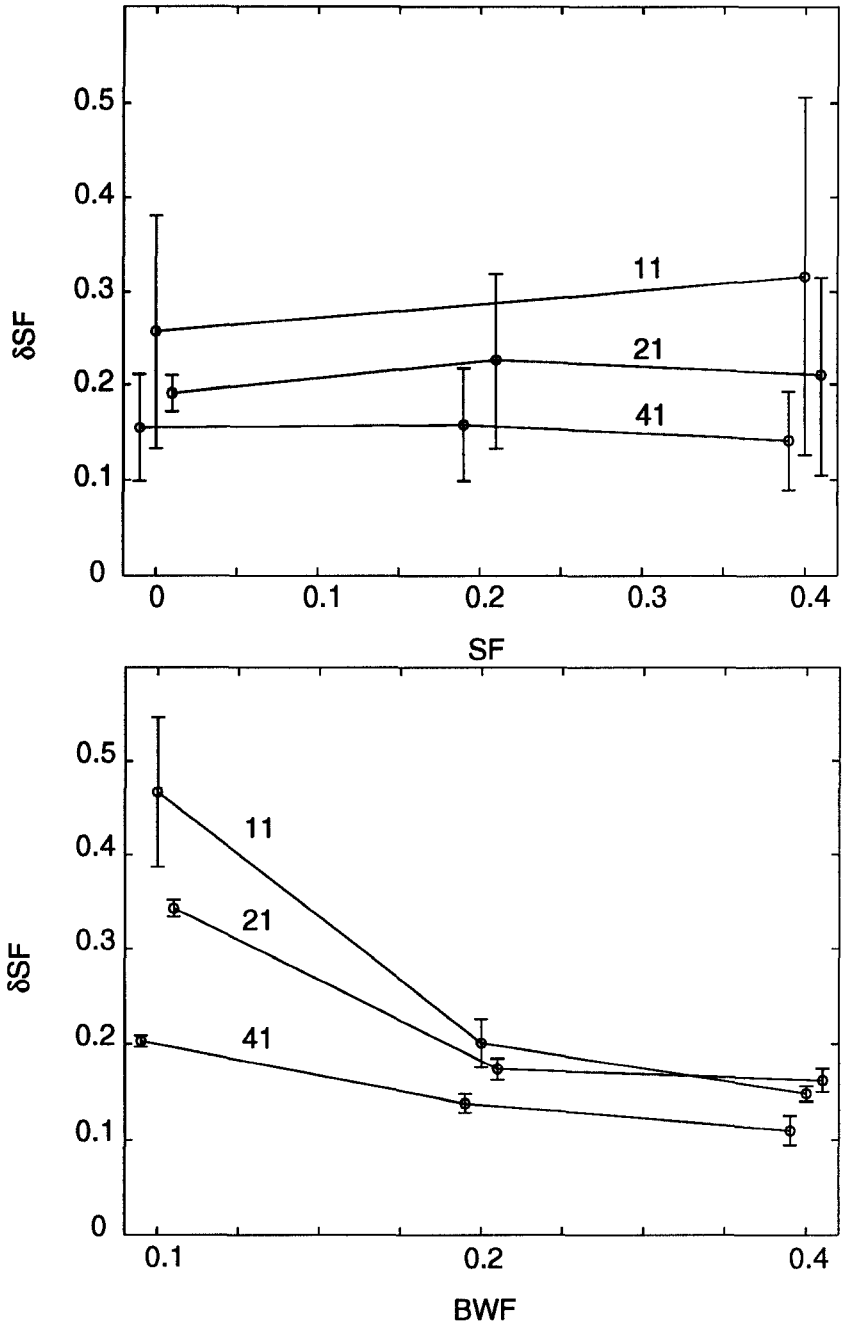


Figure 2.5: Symmetry change detection thresholds for 41, 21, and 11 component complexes, and 15 dB peak level, averaged over four subjects and three BWF's in (a) and three SF's in (b). Large error bars in (a) are due to δSF increase for the narrowest peak (BWF = 0.1) seen in (b). Note that in (b), most δSF changes with spectral density occur at the narrowest peak.

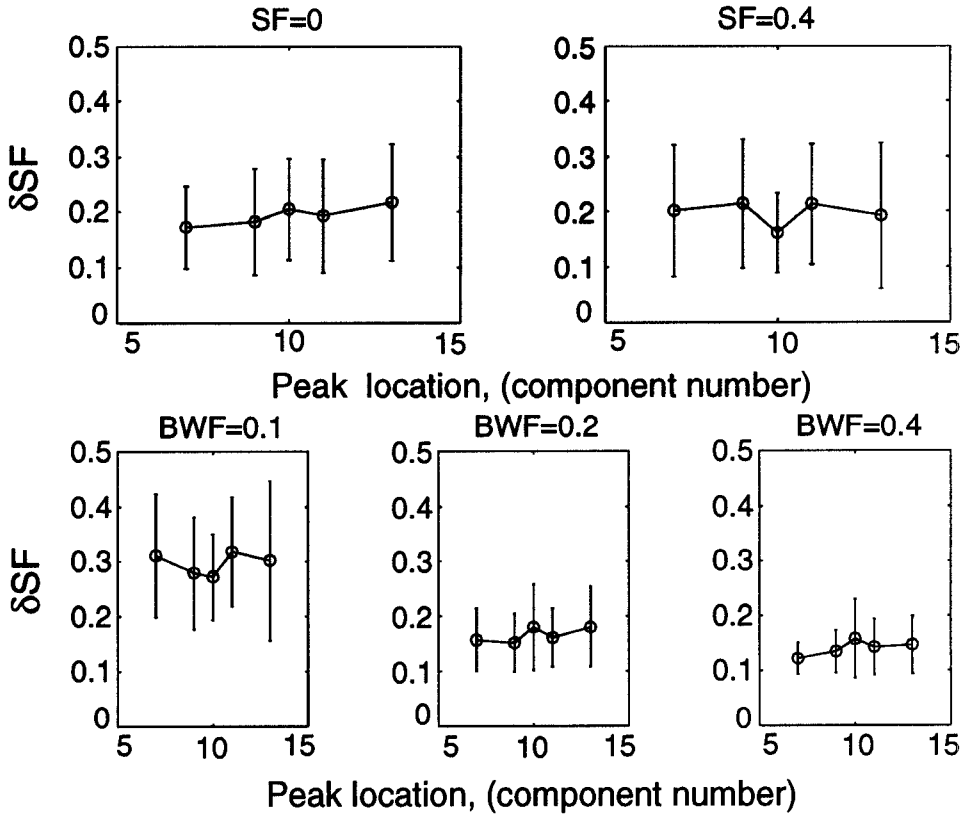


Figure 2.6: δSF thresholds for 21 component complex and 15 dB peak amplitude, are given as a function of the component number of peak's location ($7^{th}, 9^{th}, 10^{th}, 11^{th}, 13^{th}$). Thresholds are averaged over four subjects and: three BWF's in (a) (shown separately for two SF's); and two SF's in (b) (shown separately for three BWF's).

of the BWF's for broader peaks (BWF's=0.2, 0.4), while they increase for the narrowest peak tested (BWF=0.1) (Fig. 2.6(b)).

(2) δSF threshold values are independent of the location of the (largest) peak component. Thus, δSF $\approx$ 0.15 for BWF's=0.2, and 0.4, and δSF $\approx$ 0.3 for BWF=0.1, for all peak locations.

Two subject were tested at two additional off-centered peak locations (5^{th} and 15^{th} component). The thresholds are somewhat lower, presumably reflecting the vicinity of the edges of the profiles.

2.4 Detection of changes in spectral peak bandwidth factor

Experiments described in this section measured detectability of a change in spectral peak shape due only to a change in its bandwidth factor (BWF), while holding the symmetry factor constant. In this sense, these experiments complement those described earlier in Sec. 2.3. For each test, the detection threshold was computed as the relative change in the BWF of the standard, i.e., δBWF/BWF.

2.4.1 Results and discussion

Dependence on symmetry and bandwidth factors of the standard

As in Sec. 2.3.1, a 41 component complex was used and the peak level was kept at 15 dB level above the baseline. Standards of three different bandwidth factors (BWF = 0.1, 0.2, 0.4) and five different symmetry factors (SF = 0, 0.1,

0.15, 0.2, 0.4) were used, i.e., a total of 15 conditions. The average value of the thresholds over three subjects are plotted in Figs. 2.7. The plot in Fig. 2.7(a) is of the average δBWF/BWF as a function of SF. In Fig. 2.7(b), the thresholds are plotted as a function of BWF.

The basic result that emerges from all these tests is that the detection threshold for a BWF change is the same regardless of peak shape (δBWF/BWF $\approx$ 0.22), over the range of peak shapes studied. The corresponding unnormalized rms–thresholds are shown in Appendix A.

Our subjects took longer to train for this task than for the symmetry change detection task. Furthermore, the BWF rms–thresholds are in general higher than the SF rms–thresholds. During the tests, subjects reported listening for several different sound qualities, e.g., pitch and sharpness of sound, in order to recognize the signal. Some of them reported changing their listening strategies depending on the testing conditions.

Dependence on peak levels

The dependence of BWF thresholds on peak levels was examined in three subjects over the following conditions: three SF's (0, 0.2, 0.4), three BWF's (0.1, 0.2, 0.4), and at three peak levels (10 dB, 15 dB, 20 dB). Tests at 15 dB peak level were repeated to account for the fact that two new subjects participated in this sequence of tests. The δBWF/BWF thresholds, first as a function of SF and then as a function of BWF, are given in Figs. 2.8(a) and (b), respectively.

The plots confirm that, at a particular level, the δBWF/BWF threshold is largely independent of peak shape. However, thresholds do vary as a function of peak level, but mostly at lower peak levels. For instance, on average, the rate

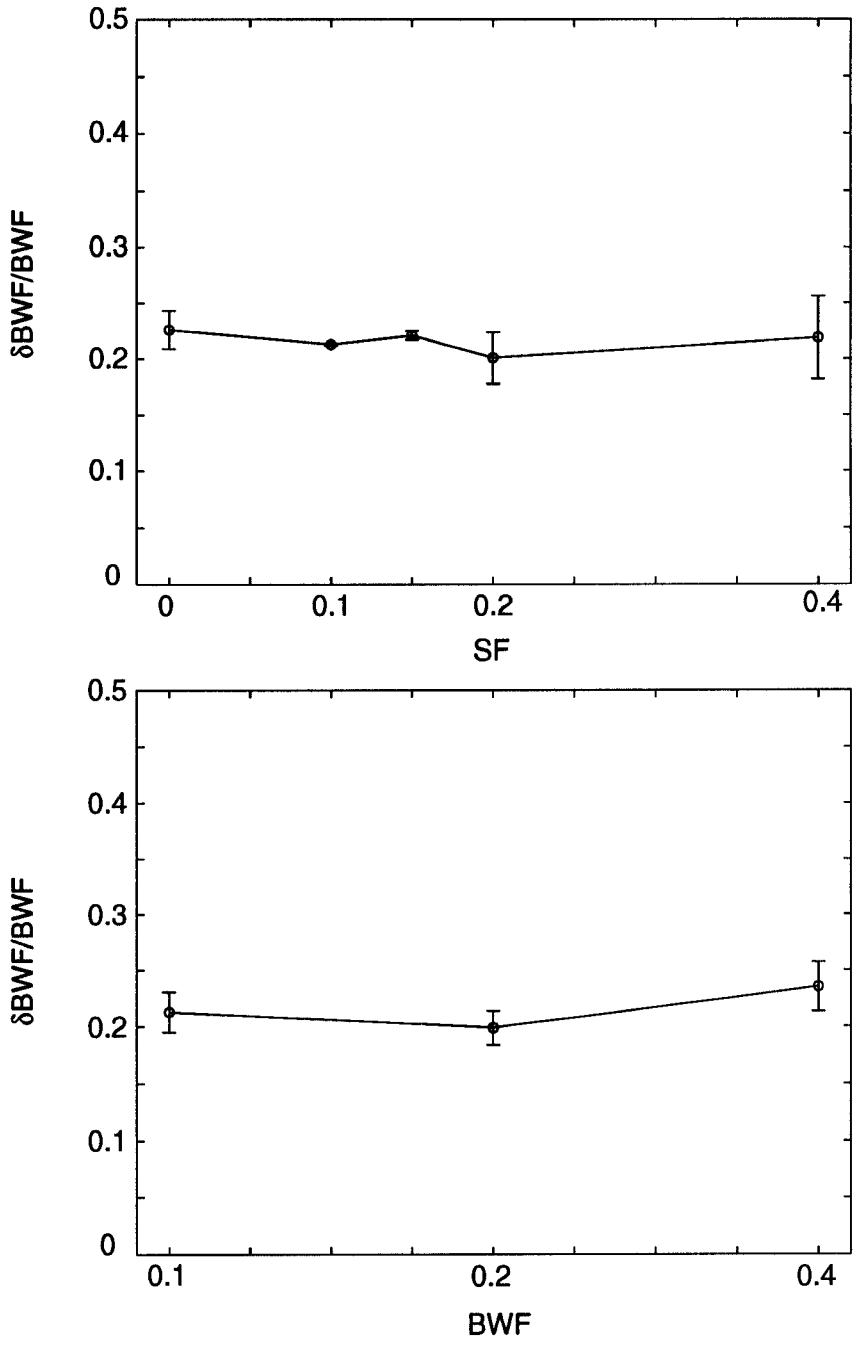


Figure 2.7: Bandwidth change detection δBWF/BWF threshold for 41 frequency components, and 15 dB peak level, averaged over three listeners, and three BWF's (0.1, 0.2, 0.4) in (a), and two SF's (0 and 0.4) in (b). Data are slightly offset along the abscissa for clarity.

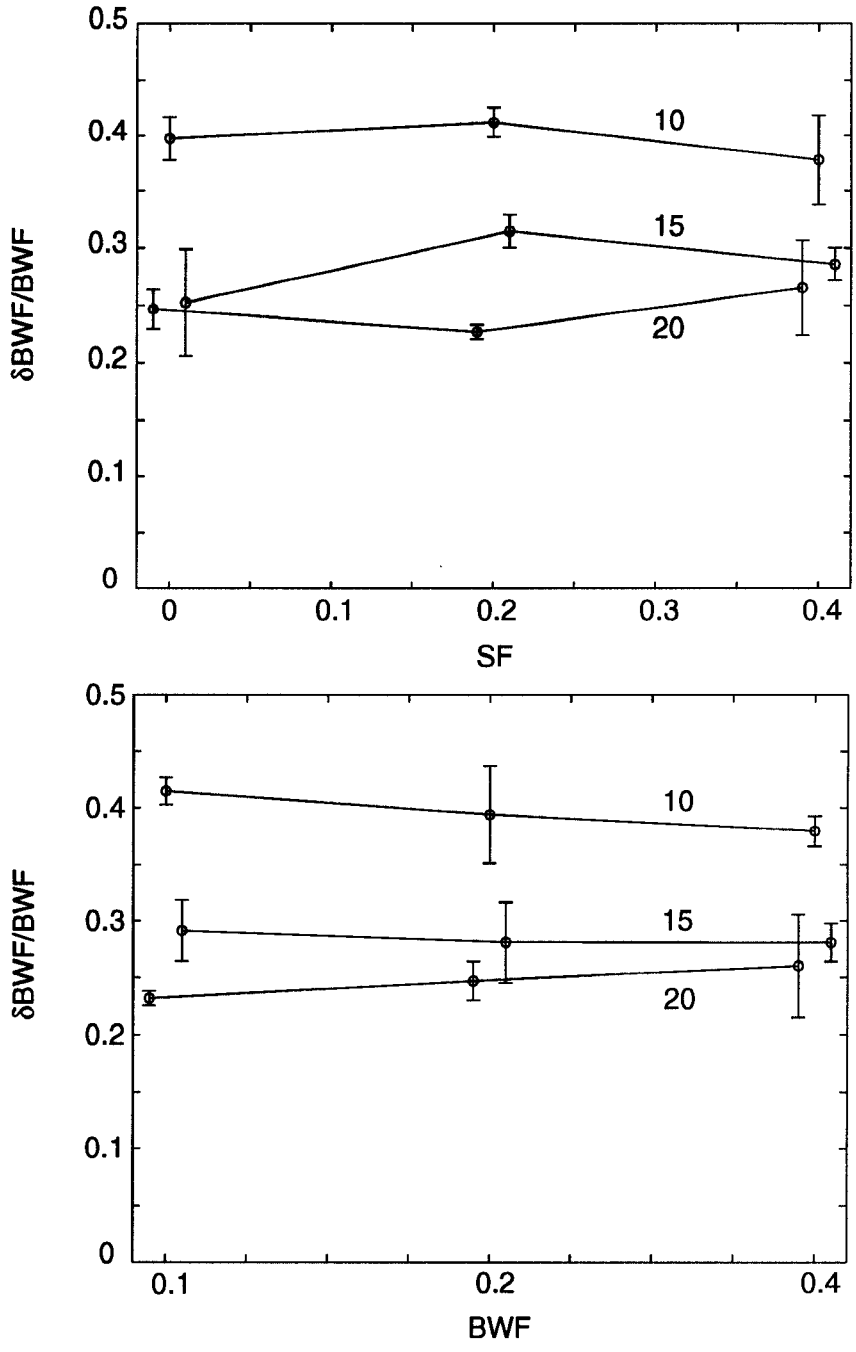


Figure 2.8: Bandwidth change detection δBWF/BWF thresholds for 41 component complex and 3 peak amplitudes: 10 dB, 15 dB, and 20 dB. The thresholds are averages of three subjects and: three BWF's in (a), and three SF's in (b). Points are offset along the abscissa for clarity.

of threshold rise in going from the 20 dB to the 15 dB peaks is less than half of that seen between 15 dB and 10 dB.

Spectral density dependence

Dependence of BWF thresholds on the spectral density was examined for the 15 dB peak level using 11, 21, and 41 component complexes. The average results of three listeners, using two SF's (0 and 0.4) and three BWF's (0.1, 0.2, 0.4), are presented in Figs. 2.9. In Fig. 2.9(a), they are given as a function of SF, and in Fig. 2.9(b) as a function of BWF. The corresponding rms–thresholds are shown in Appendix A.

Once again, δBWF/BWF thresholds remain constant for all conditions tested, i.e., regardless of peak shape and spectral density. The one obvious exception is at the broadest peak for the 11 component case, where the threshold is significantly larger.

Dependence on location of the largest peak component

As in Sec. 2.3.1, a 21 component complex was used and the peak level was kept at 15 dB level above the baseline. Standards of three different bandwidth factors (BWF = 0.1, 0.2, 0.4) and two different symmetry factors (SF = 0, 0.4) were used, i.e., a total of 6 conditions. The average threshold values of two subjects are depicted in Figs. 2.10, as a function of the component number of the location of the (largest) peak component. In Fig. 2.10(a) the thresholds are averaged over two SF's, and in (b) over three BWF's.

The plots confirm the threshold trends established earlier for the centered peaks (Sec. 2.4.1). Namely, the thresholds are, for a particular peak location,

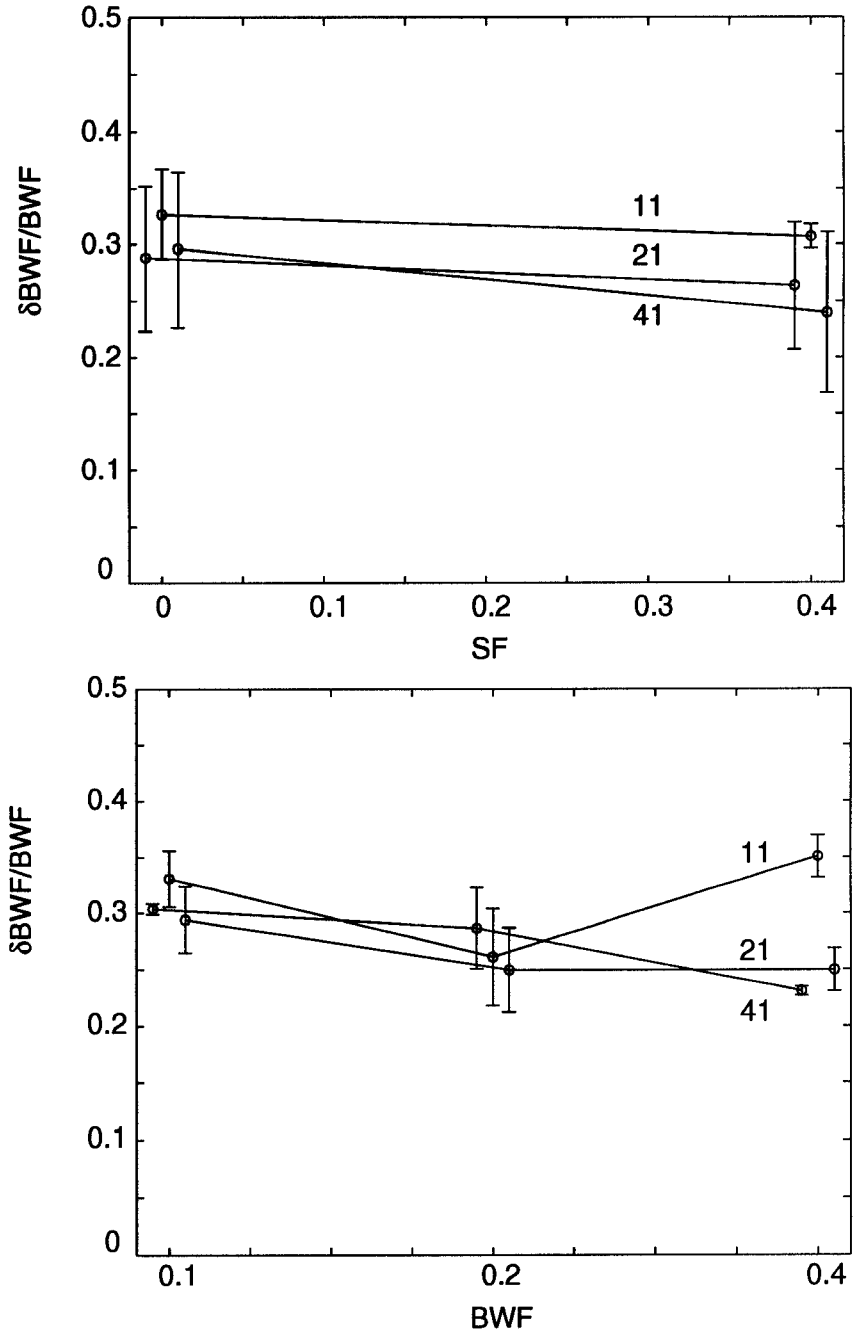


Figure 2.9: δBWF/BWF thresholds for 41, 21, and 11 component complexes, and 15 dB peak level, averaged over three subjects and three BWF's in (a) and two SF's (b). Threshold is independent of spectral density for all but the broadest BWF, where it increases for the 11 component case.

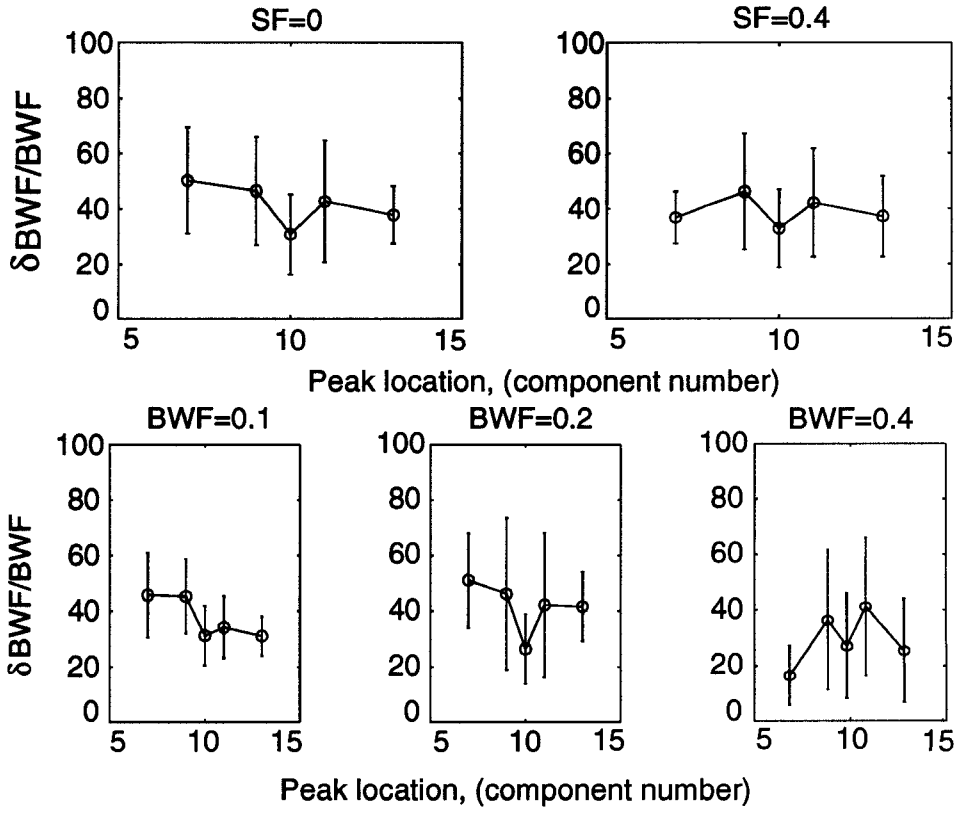


Figure 2.10: δBWF/BWF thresholds for 21 frequency components, and 15 dB peak level, averaged over two listeners, and three BWF's (0.1, 0.2, 0.4) in (a), and two SF's (0, 0.4) in (b).

independent of peak shape (δBWF/BWF $\approx$ 38%). Moreover, they are also independent of the location of the peak component. Large error bars are due to the large differences between the threshold values of the two subjects. Individual results averaged over two SF's are shown in Fig. 2.11 for three different BWF's.

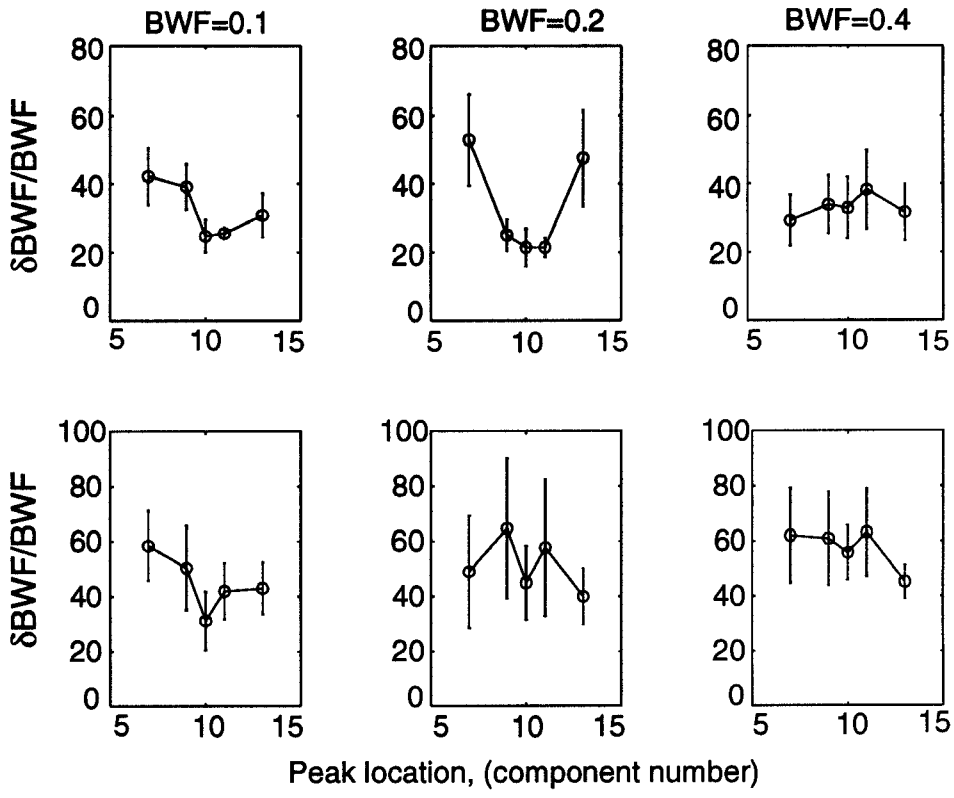


Figure 2.11: Same as in Fig. 2.10(b), but plotted for two individual subjects. Note the different δBWF/BWF scales used.

2.5 Detection of simultaneous changes in both, SF and BWF

In this set of experiments we change simultaneously the SF and BWF of peak profiles, and compare the thresholds to those obtained for pure δSF and δBWF tests. The SF and BWF were changed by decreasing the (absolute value of the) right slope. This tilts the peak profile to the right, and increases, both its SF and BWF. Three subject were tested for 21 and 41 spectral density complexes, for 15 dB peak levels, and for different starting SF's and BWF's. Two out of the three subjects participated also in δSF and δBWF tests reported in Secs. 2.3.1 and 2.4.1.

The threshold results are in Table 2.1(a) and 2.1(b) for the two density cases.

2.5.1 41 spectral sensity case

The δSF thresholds are independent of the starting SF and BWF, except for the increase towards the narrower peaks (Table 2.1(a) and Fig. 2.12(a)). Thus, for BWF's between 0.1 – 0.13, δSF $\approx$ 0.14, while for broader peaks, BWF's between 0.15 – 0.84, δSF $\approx$ 0.11.

The δBWF/BWF thresholds are independent of the starting SF and BWF, for peaks with starting SF between 0 – 0.43 (Table 2.1(b) and Fig. 2.12(b)). The average δBWF/BWF $\approx$ 16.3%, while for highly tilted starting peaks (SF $>$ 0.43), δBWF/BWF $\approx$ 38.4%.

Three conclusions follow from these data:

(1) The δSF threshold trends are similar to those obtained for pure δSF tests

Thresholds for δSF–δBWF test

(41 components)

Starting SF	BWF	δSF	δBWF/BWF (%)
0	0.10	0.13	15.3
0.14	0.12	0.14	19.5
0.33	0.15	0.11	19.9
0.60	0.25	0.11	38.7
0	0.13	0.14	15.8
0	0.20	0.12	14.0
0.11	0.22	0.11	14.5
0.25	0.27	0.12	19.5
0.43	0.35	0.09	19.4
0.67	0.60	0.09	39.7
0	0.40	0.12	13.9

(21 components)

0	0.10	0.14	16.7
0	0.13	0.15	17.1
0	0.20	0.13	15.5
0	0.40	0.15	17.2

Table 2.1: δSF–δBWF/BWF thresholds for various starting SF's and BWF's, for 41 and 21 spectral density cases.

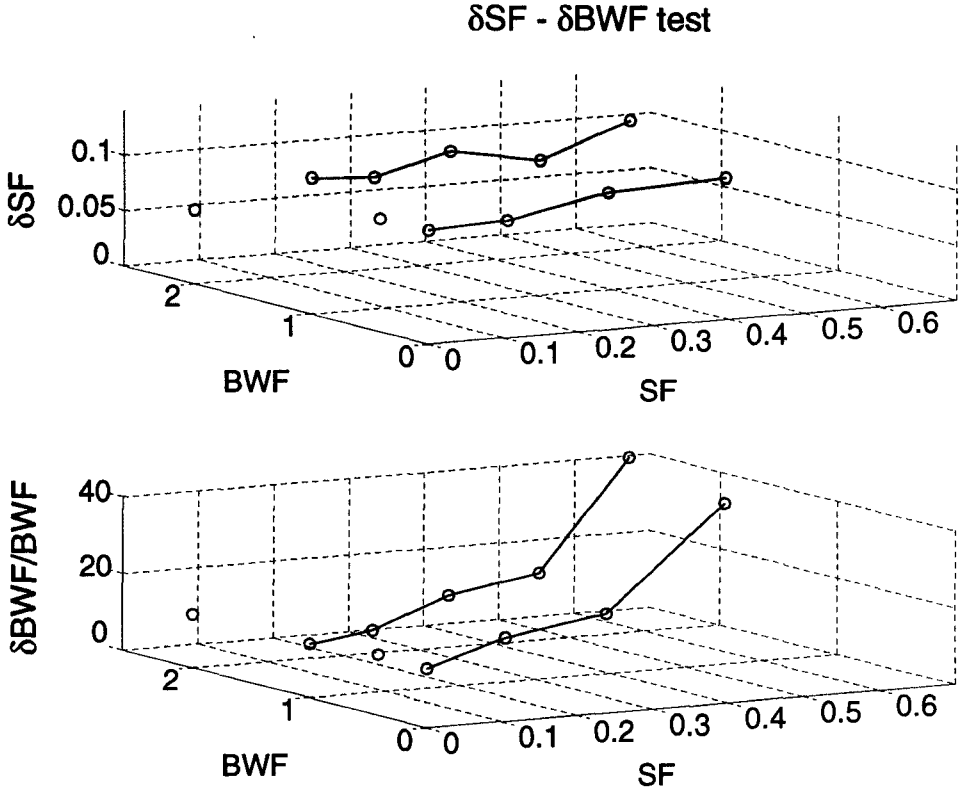


Figure 2.12: (a) δSF and (b) δBWF/BWF thresholds as a function of starting SF's and BWF's. 41 component complexes with peaks at 15 dB peak level were used. Data are averaged over three subjects.

(Sec. 2.3.1). Namely, the thresholds are independent of SF and they increase towards the narrower BWF's.

(2) The δBWF/BWF threshold trends are the same to those observed earlier for pure changes in BWF (Sec. 2.4.1). Namely, the δBWF/BWF thresholds are independent of the starting SF and BWF, for starting SF's between 0 and 0.43. Thresholds increase for larger SF's.

(3) The average threshold values, δSF $\approx$ 0.11 for broader peaks, are the same to those obtained for pure δSF test. The threshold increase for narrower peaks is somewhat smaller here than for δSF case (0.14 versus 0.16). However, the average δBWF/BWF thresholds are significantly smaller than for pure δBWF tests (16.3% versus 22%).

2.5.2 21 spectral density case

The tests were repeated for 21 spectral density case (Table 2.1(b)), over a limited range of SF's and BWF's. The average δBWF/BWF $\approx$ 16.6%, and is independent of the starting BWF. The δSF $\approx$ 0.14 is also independent of the starting BWF, i.e., it does not exhibit a characteristic increase at narrower BWF's.

2.6 Two control experiments for SF and BWF change detection

In this section, we present the results of two control experiments. In the first we randomized the location (frequency) of the peak between signal and standard in

order to minimize or abolish the "pitch" cues that may underlie the detection of SF and BWF changes. In the second experiment we assessed the relative contribution of the change in peak energy to the detection threshold.

2.6.1 Effects of peak frequency randomization

Numerous experimental results have suggested that the detection of spectral shape changes may in some cases be effectively mediated by pitch cues associated with these spectral changes [32, 33, 34, 35]. In order to assess the possible contribution of such a pitch cue in our tests, we measured the effect on thresholds of randomizing peak locations, a procedure which in effect destroys the pitch cue. The change in thresholds was then compared to what would be predicted from the theoretical strength of the pitch cue computed for each test using the so-called Ewaif model (reviewed briefly in Appendix B).

Stimulus

The entire spectral content was randomly shifted in order to prevent listeners from using standard's and signal's complex pitches for spectral shape change detection. Frequency shift was achieved by randomly changing the sampling time in a range of 40 μs to 45 μs, in steps of 0.5 μs. This amounts to shifting the central component from 1000 Hz to 889 Hz, and all the other components accordingly to preserve the frequency spacing.

Two subjects participated in SF and three in BWF change detection series. They were tested at two SF's (0, 0.4) and three BWF's (0.1, 0.2, 0.4) for the 41 spectral density signals, and 15 dB peak level. Thresholds measured are

	BWF					
	0.1		0.2		0.4	
	SF					
δSF test	0	0.4	0	0.4	0	0.4
NR	0.27	0.27	0.13	0.15	0.13	0.11
R	0.36	0.44	0.27	0.23	0.17	0.12
-(NR R)	0.09	0.17	0.14	0.08	0.04	0.01
F_{sta}	1290.52	1327.40	1227.73	1329.10	1223.27	1430.59
F_{sig}	1315.58	1345.59	1258.54	1369.30	1293.42	1495.99
$\Delta F=F_{sta}-F_{sig}$	-25.04	-18.19	-30.83	-40.25	-70.13	-65.39
$\Delta F/F_{sta} \cdot 100$	-1.94	-1.37	-2.51	-3.03	-5.73	-4.57

Table 2.2: (a) Symmetry factor change detection threshold (δSF), for 41 component complex for non-randomized (NR) and randomized (R) spectra. The first two rows are the NR and R δSF thresholds. The third row is the difference of the first two. The forth and fifth rows are the computed Ewaif pitches of standard (F_{sta}) and signal (F_{sig}) at perceptual threshold levels for NR condition, for zero-phase components. The ΔF row is $\Delta F = F_{sta}-F_{sig}$. The last row is the relative pitch difference, $\Delta F/F_{sta}$.

	BWF					
	0.1		0.2		0.4	
	SF					
δBWF/BWF test	0	0.4	0	0.4	0	0.4
NR	0.30	0.31	0.32	0.25	0.23	0.23
R	0.43	0.42	0.62	0.53	0.37	0.36
-(NR R)	0.13	0.11	0.30	0.28	0.14	0.13
F_{sta}	1290.52	1327.40	1227.73	1329.10	1223.27	1430.59
F_{sig}	1243.44	1318.03	1216.00	1377.66	1255.29	1505.61
$\Delta F=F_{sta}-F_{sig}$	38.65	9.66	11.69	-49.96	-26.78	-73.16
$\Delta F/F_{sta} \cdot 100$	2.99	0.73	0.95	-3.76	-2.19	-5.11

Table 2.2: (b) Bandwidth factor change detection threshold (δBWF/BWF), for 41 component complex for non-randomized (NR) and randomized (R) spectra. The table is organized as Table 2.2(a). Note a change in sign of ΔF across various testing conditions, which may explain the change in strategies that our subjects reported in performing this task.

presented in Tables 2.2(a) and 2.2(b), for the SF and BWF change tests, respectively. In each table, the first and second rows contain the detection thresholds for the non-randomized (NR) and randomized (R) peaks. The third row lists their differences (NR–R). The next two rows are the computed Ewaif pitches of standard (F_{sta}) and signal (F_{sig}) at NR thresholds. The ΔF row shows the difference of the previous two. The last row is the relative pitch difference $\Delta F/F_{sta}$. The Ewaif pitches were computed for zero phases, which corresponds to our stimulus condition.

Assessing the data using the Ewaif model

In order to assess the amount of a pitch cue contribution to the detection of changes in our stimulus, the following two arguments were used (see [35] for details):

1) If the detection process relies primarily on a pitch cue (as defined by the Ewaif model), then some minimal pitch difference, ΔF [36], or relative pitch difference, $\Delta F/F_{sta}$ [35], is necessary for detection. Therefore, at perceptual thresholds ΔF or $\Delta F/F_{sta}$ should remain relatively constant.

2) If a threshold deterioration occurs due to the uncertainty in the randomized signal, and not due to the pitch differences across the testing conditions, then it should be uniform across all conditions. Otherwise, the deterioration probably reflects the effective contribution of the pitch cue. This is evaluated by the change in values of the NR–R in Tables 2.2.

Results and discussion

(i) Effects on detection of SF changes (Table 2.2(a))

With respect to the first argument above, it is clear from the ΔF and $\Delta F/F_{sta}$ values in Table 2.2(a) that not all pitch cues are equal at threshold, since both increase approximately 4-fold over the SF's and BWF's tested. However, the rise in δSF for the narrowest peak might be due to decreasing pitch cue. This is further supported by the data with respect to the second argument, namely that the randomization affects only the δSF thresholds of the narrower peaks. Therefore, the evidence here suggest that the pitch cue may be effective only for these peaks.

(ii) Effects on detection of BWF changes (Table 2.2(b))

The ΔF and $\Delta F/F_{sta}$ values vary greatly (approximately 7-fold) across the SF's and BWF's. Note also a change in sign of ΔF across various testing conditions. This strongly suggests that the pitch cue plays a minimal role in this discrimination task. Furthermore, a near uniform increase of the thresholds when the signal is randomized, supports the notion that it is due to an uncertainty effect rather than an abolishment of a pitch cue.

2.6.2 Detecting peak energy change compared to a BWF change

In all the BWF change tests, the peak energy was not equalized as the peak width was altered. So, it could be argued that the δBWF/BWF threshold reflects a change in the energy of the peak, rather than in the BWF per se. One indirect argument against this conclusion is that the rms–thresholds for BWF changes were sometimes 7 dB worse than those for SF changes (e.g., in Appendix A, compare data points at BWF = 0.4 in Figs. A.1(b) and Figs. A.4(b)). If the

tasks were purely based on the total change in energy, then the two thresholds should be comparable.

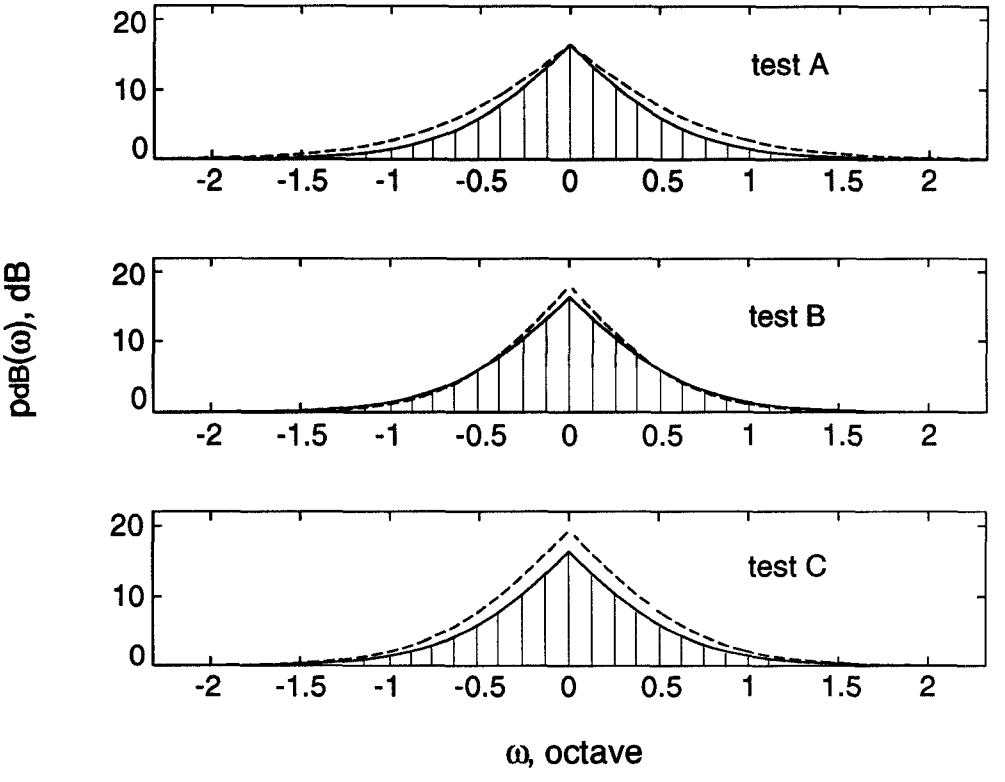


Figure 2.13: (a) Effects of BWF changes in tests A and B (see text), and peak level changes in test C are shown by dashed lines, for a standard peak with BWF = 0.2, SF = 0, and A_{max} = 15dB.

A more direct rebuttal of this hypothesis is provided in Figs. 2.13, where the rms–thresholds of three different tests are compared. In tests A and B, the BWF of the peak was changed in one of two ways: either as usual through a change in the width (Fig. 2.13(a), test A), or through a change in the height of the peak (test B). (In test B, for each peak level of the signal, A_{signal}, the slopes are computed as $R_{signal} = \frac{A_{signal}}{A_{max}} R_{standard}$, and $L_{signal} = -R_{signal}$, where $R_{standard}$ and A_{max} are the right slope and the peak level of the standard.) In test C, the

peak's BWF was kept constant and the rms–threshold is measured for changes in the peak size, and not its shape. In all three tests, 41 component stimuli were used with a starting peak level of 15 dB. Three subjects were tested at two SF (0 and 0.4) and three BWF's (0.1, 0.2, 0.4).

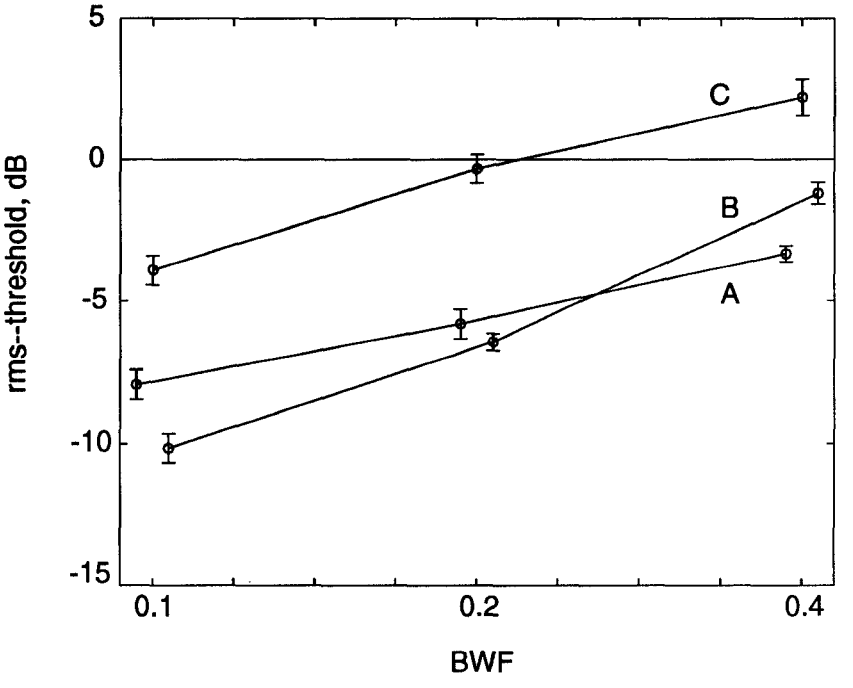


Figure 2.13: (b) The rms–thresholds for the three tests (A, B, and C).

The data in Fig. 2.13(b) reveal that the rms–thresholds (and δBWF/BWF thresholds) are very close for BWF change detection tests (A and B). They are also uniformly and significantly lower (approximately 6 dB) than those due to a change in peak size alone (test C). The conclusions we draw are that (1) a BWF change is a more effective feature to detect than just scaling the peak, and that (2) the relatively small changes in peak energy associated with the BWF tests (as in A and B) are unlikely to contribute significantly to the BWF change detection thresholds.

2.7 Broader interpretations of SF and BWF changes

In all experiments so far, the changes in peak shapes were parameterized in terms of SF and BWF changes. There is, however, an equivalent and more general description of these two manipulations. For instance, a four-fold increase in BWF (from BWF = 0.1 to 0.4) can be viewed as a stretching (or a dilation) of the peak profile along the ω axis, i.e., $p(\omega)$ becomes $p(\alpha \cdot \omega)$ with $\alpha = 1/4$ (see Fig. 2.14(a)). This change in $p(\omega)$ can be equivalently described in the Fourier transform domain of the profile. Namely, if $P(\Omega)$ is the Fourier transform of $p(\omega)$, then dilating the profile by a factor α causes its transform to become $1/\alpha \cdot P(\Omega/\alpha)$ (Fig. 2.14(b)). The computations of the magnitude of the Fourier transform of the peak profile are in Appendix C.1. The units of Ω in the profile transform domain are in terms of the number of cycles per unit distance (octave) along the ω axis. (For instance, a sinusoidal profile with $\Omega_o = 2$ cycle/octave is a profile whose peaks are separated by 1/2 octaves along the ω axis.)

The change in the SF of a peak $p(\omega)$ can be also expressed in terms of a corresponding (though somewhat less intuitive) modification of the peak transform $P(\Omega)$. Specifically, if a small constant phase angle θ_o is added to the phases of all components of the transform $P(\Omega)$, then the corresponding profile $p(\omega)$ becomes tilted in a manner very similar to that caused by a SF change. This is demonstrated in Figs. 2.14(c) and (d) for three SF's and their corresponding θ_o angles: SF = 0.05 (3^o), 0.15 (9^o), and 0.3 (18^o). (The computations are in Appendix C.2).

The above interpretations of the δBWF and δSF imply that these manipu-

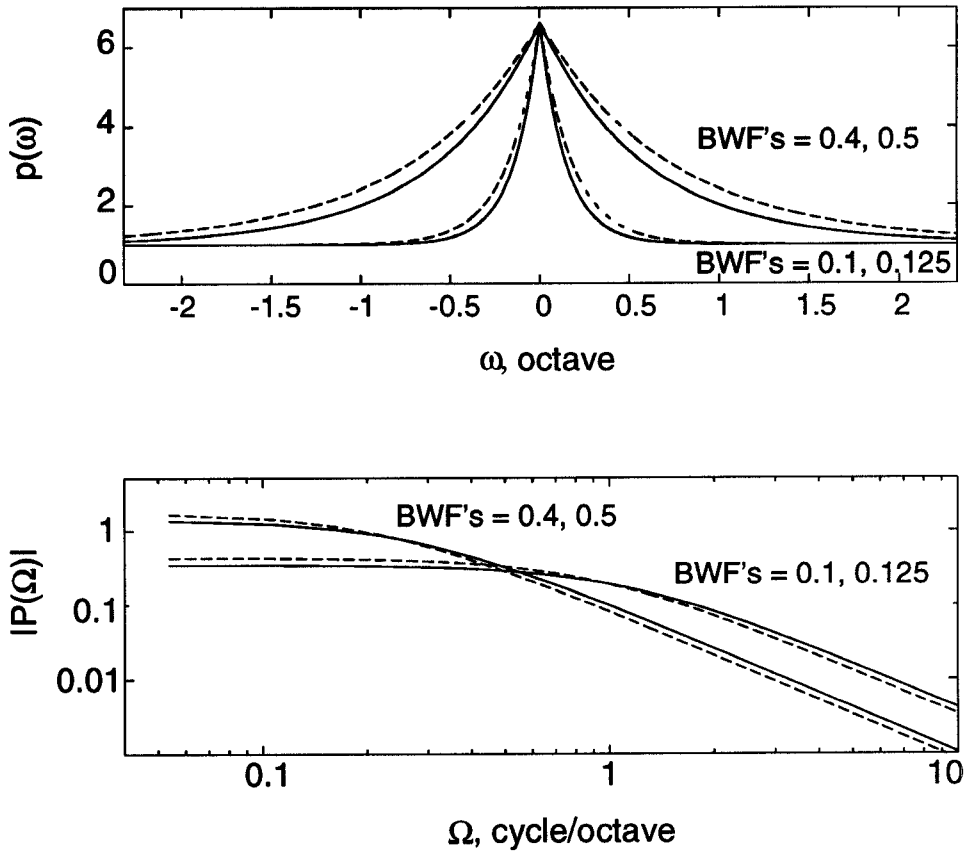


Figure 2.14: (a) Peak profiles with A_{max} = 15 dB, BWF's 0.1 and 0.4 (solid lines), and BWF's at 25% detection threshold (dashed lines). (b) Magnitude of the profiles' Fourier transformations, $|P(\Omega)|$. The effect of the BWF change is a shift in magnitude (and not a change in shape) along the $\log \Omega$ axis.

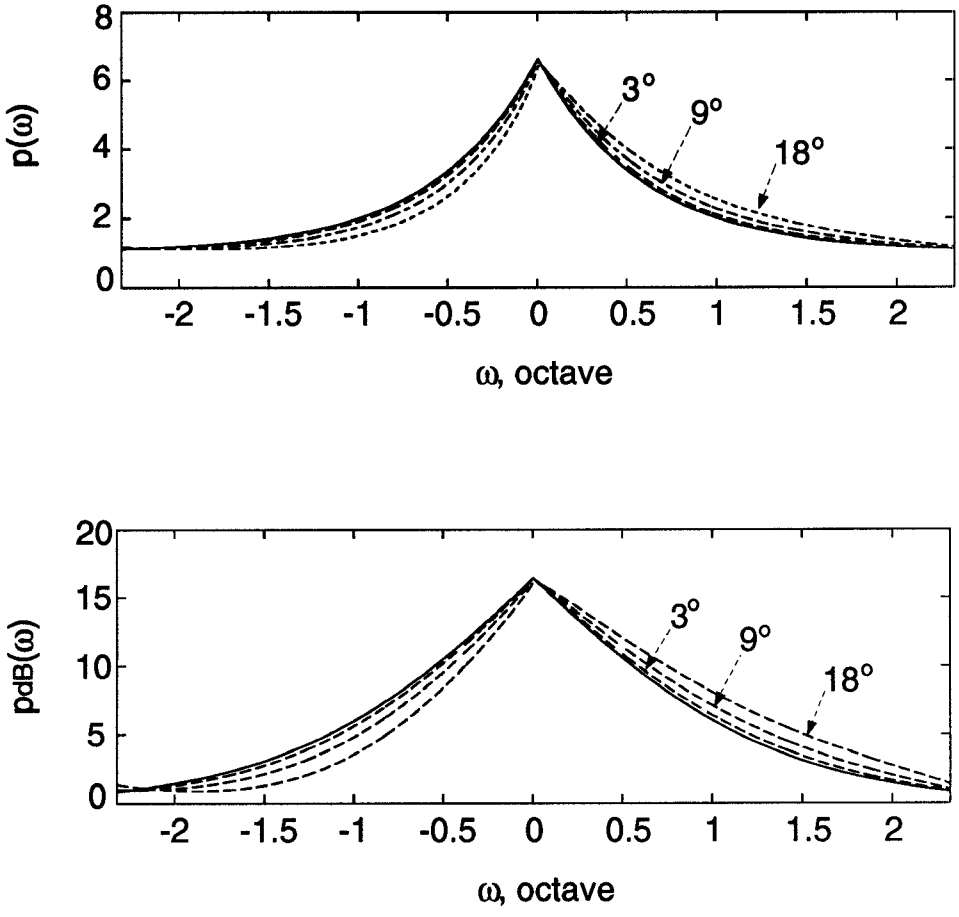


Figure 2.14: (c) and (d) The effects on changes in the symmetries of a peak profile (BWF = 0.4 and SF = 0) due to adding constant phases (3^o, 9^o, and 18^o) to its Fourier transform.

lations can be readily applied to any arbitrary spectral profile. The sensitivity measurements can then be directly compared across different profiles. Specifically, we shall be interested in comparing the dilation (δBWF/BWF) and phase-shift (δSF) thresholds of the peaks to those of sinusoidally modulated spectra, or ripples, which are the basis functions of the Fourier transform. Dilating a rippled spectrum simply changes its ripple (or envelope) frequency, and shifting the spectrum along the ω axis changes its phase (Fig. 2.15). While ripple frequency-difference-limen thresholds were measured previously [37, 2], no ripple phase sensitives have been reported in the literature. The experiments described below provide these measurements.

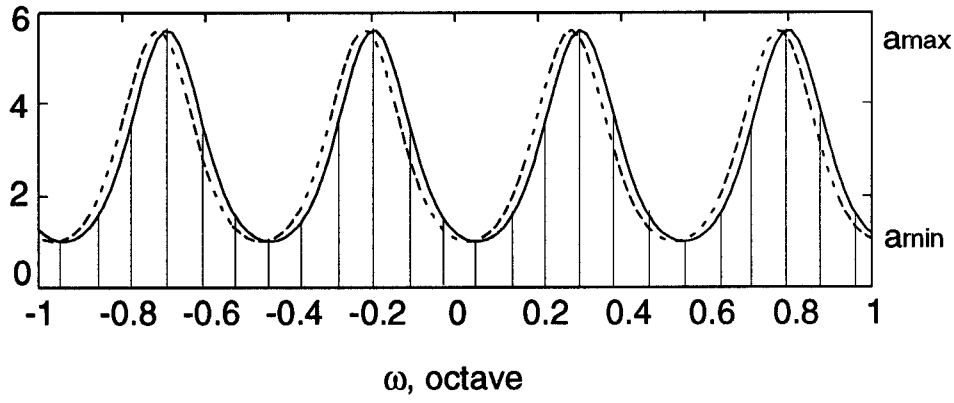


Figure 2.15: A sinusoidal ripple profile with ripple frequency of 2 cycle/octave, and 15 dB peak-to-valley amplitude (computed as $20\log_{10}(a_{max}/a_{min})$). Its 16° phase shifted version is shown in dashed lines.

2.8 Phase difference limen experiments

Sensitivity to ripple phase changes was measured in sinusoidally modulated profiles on a dB amplitude scale (Fig. 2.15), and the thresholds, termed phase-difference-limen (**pdl**), are reported in units of degrees.

2.8.1 Stimulus

For all testing conditions, the number of frequency components was 161 (34 per octave), and the frequency components were equally spaced on a logarithmic scale between 0.2–5 kHz. The starting ripple phase was kept constant at zero degrees for the data reported here. Other starting phases were also tested, and results were very similar. The peak-to-valley ratio was defined as $20 \log \frac{a_{max}}{a_{min}}$, where a_{max} and a_{min} are the peak and valley amplitudes of the sinusoid (see Fig. 2.15). The ripple frequency (Ω) was fixed over a set of trials at 0.25, 0.5, 1, 2, or 4 cycle/octave, for 15 dB and 25 dB peak-to-valley ratios. One of the two subjects also completed the test for 8 cycle/octave, and for 35 dB peak-to-valley ratio, while the other was tested at 2 cycle/octave and 20 dB and 35 dB levels.

The overall intensity was varied across and within the trials over a 20 dB range in 1 dB steps.

2.8.2 Results

The average data for two subjects are presented in Fig. 2.16(a) as a function of ripple frequency, for two levels. The results show that thresholds are constant below about 2 cycle/octave at both levels tested, achieving a minimum of

about 6° for the larger level. Phase sensitivity decreases with increasing ripple frequencies beyond 2 cycle/octave.

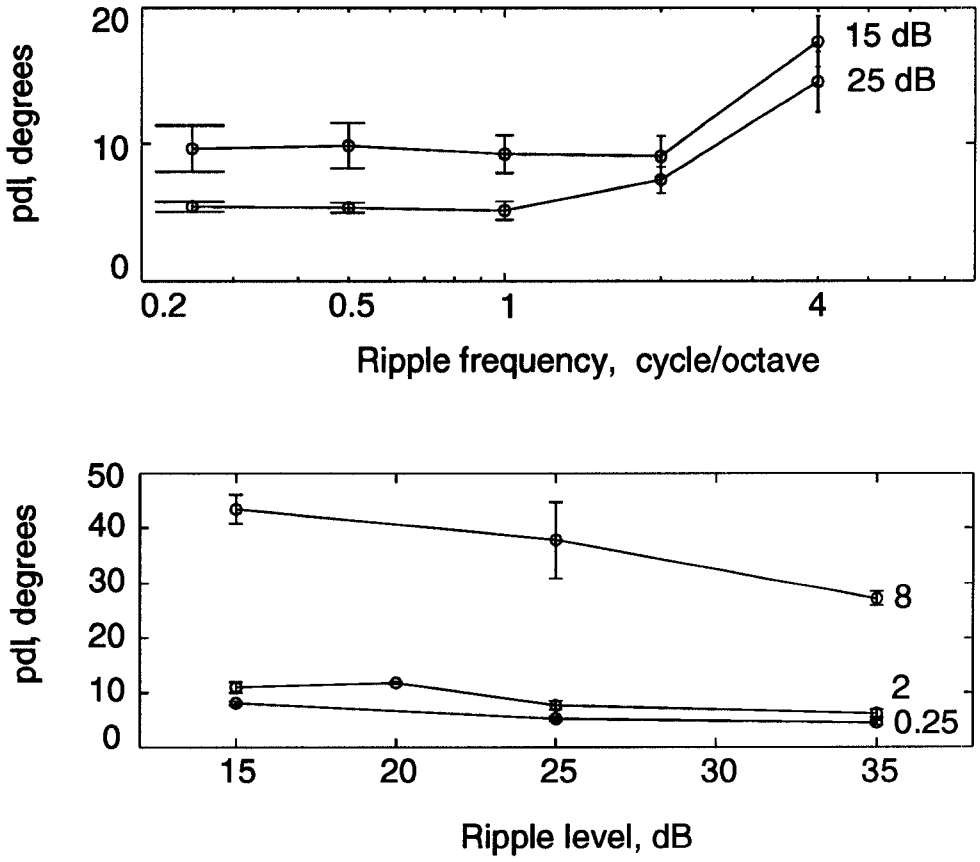


Figure 2.16: (a) Phase difference limen threshold (**pdl**) as a function of ripple frequency, for 15 dB and 25 dB peak-to-valley amplitudes (or ripple levels), averaged over 2 subjects. (b) Individual **pdl** thresholds at three ripple frequencies as a function of ripple level (subject 1 was tested at 0.25 and 8 cycle/octave, and subject 2 at 2 cycle/octave).

Figure 2.16(b) depicts the data for individual subjects as a function of ripple level. Thresholds saturate with increasing levels at all ripple frequencies tested.

2.8.3 Discussion

There are two important characteristics of the data in Fig. 2.16(a). The first is that for lower ripple frequencies (Ω), subjects detect a constant phase shift and not a constant displacement of the peaks, as is probably the case for $\Omega > 2$ cycle/octave. The second is that the lowest detectable phase shift (6^o) is very close to the phase shift implied by the δSF thresholds ($= 0.11$) measured for the peaks (Figs. 2.3). The correspondence between these two thresholds confirms the association made between them as explained in Sec. 2.7. It also suggests that this threshold is independent of the particular spectral shape used. The implications of this finding are discussed in more detail in Chapter 3.

2.9 Summary of results

The experiments described here measured subjects' ability to detect changes in the symmetry and bandwidth factors of spectral peaks under various conditions. The choice of these spectral features was inspired by the physiological finding that the primary auditory cortex encodes explicitly the locally averaged gradient of the acoustic spectrum. In the case of spectral peaks, the local gradient is directly related to the symmetry of the peak. Since the shape of a peak can be effectively described by its symmetry and bandwidth, our goal was to examine the perceptual sensitivity of, and interdependence between, these features.

The basic result that emerges is that **thresholds to changes in SF and BWF are (with one exception) approximately constant regardless of peak shape parameters tested.** Thus, for the detection of SF changes, δSF

thresholds are near 0.11 for all SF's and almost all BWF's (Figs. 2.3). The exception occurs towards the narrowest peak (BWF = 0.1) where (1) the detection threshold increases gradually to 0.16 (Fig. 2.3(b)), and (2) pitch cues associated with this detection task become more effective (Sec. 2.6.1(i), Table 2.2(a)). For the detection of BWF changes, all δBWF/BWF thresholds remain constant at around 0.22 regardless of the peak shape (Figs. 2.7).

Also measured were the effects of three additional manipulations that did not change the shape of the peak: (1) change in the peak level; (2) spectral density of the complex; and (3) location of the largest peak component. For the first, all thresholds maintain the same trends regardless of peak level. Their absolute values, however, slightly improve for higher peak levels (Figs. 2.4 and 2.8). For lower peaks, the deterioration in δBWF/BWF thresholds accelerates with decreasing peak levels. It is possible that the uniform rise in threshold is mediated by increased masking effects of the base upon the smaller peak. For the second manipulation, δSF thresholds increase gradually with decreasing densities only at the narrowest peak (Fig. 2.5(b)), whereas δBWF/BWF thresholds deteriorate only for the lowest density (11 components) at the broadest peak (Fig. 2.9(b)). The thresholds are independent of the location of the maximal peak component (Secs. 2.3.1 and 2.4.1).

The experiments reported in Sec. 2.5 investigated how simultaneous changes in SF and BWF combine into a single detection cue. The trends in SF and BWF changes are the same as for the pure δSF and δBWF tests, for the same range of starting SF's and BWF's (from 0 to 0.4). However, the threshold values are δSF $\approx$ 0.11 and δBWF/BWF $\approx$ 16%, i.e., they are approximately the same as those for pure δSF tests and much smaller than those for pure δBWF test.

Although limited to only one way of changing both the SF and BWF, this set of experiments indicates that the overall detection occurs when one of the cues, in this case a change in SF, reaches its thresholds. The thresholds increase for more tilted peaks (SF>0.4).

Finally, rms–threshold values for SF and BWF detection tasks are comparable to other profile detection tasks (see Appendix A). Furthermore, they are significantly lower than rms–thresholds of changes that do not affect peak shape (Fig. 2.13(b)).

More generally, it is conjectured that for an arbitrary spectral profile, changing the SF and BWF, corresponds to changing the phase and the magnitude of the Fourier transform of the profile (Sec. 2.7). A fundamental conclusion from this analysis is that the detection of peak shape changes can be parametrized along two sensitive and largely independent axes: peak SF and BWF. This result lends support to the notion that the underlying physiological representation of these two features of a peak may be separated along orthogonal dimensions. For instance, one conjecture might be that the SF is mapped explicitly by the gradient map found in AI [9]. Then, this map is duplicated more than once, each at a different scale of local averaging of bandwidth, in essence providing the BWF dimension. While a physiological substrate for such a multiscale representation is yet unavailable in AI, maps of gradually changing tuning in the response areas of cells along the isofrequency planes in AI are in harmony with this view [10].

2.10 Profile analysis models

The choice of a threshold measure implies an underlying profile analysis model. Here we first apply to our data, two profile analysis models which have been shown to perform well in a variety of detection tasks. Both models presume that profile changes are conveyed by independent channels distributed across the spectrum. The first is the channel model for discrimination of broadband spectra proposed by [25], which basically combines information from all independent channels. The other is the maximum difference model described in [24], which is based on detecting the largest difference between any pair of components in the signal, i.e., it uses only two channels in computing the thresholds. We examine how these two models predict the detectability of peak shape changes by monitoring the constancy of the index d' (see Sec. 1.2) computed at perceptual thresholds at various SF and BWF combinations.

2.10.1 The channel model

This model is described in detail in [23, 25]. It consists of N noisy channels whose variances (σ) are assumed to be constant. Some interdependence between the channel outputs is introduced because of the level randomization in the experiments. The uniform roving level distribution over a 20 dB range ($\sigma_R = 5.6dB$) is approximated by a normal distribution of $\sigma_R = 5dB$. Furthermore, it is assumed that the channel variances are such that $\sigma_R \cdot N \gg \sigma$. The level difference between the standard's and signal's i^{th} component is defined as $\Lambda_i = 20\ \log((p_i)_{signal}/(p_i)_{standard})$. These assumptions lead to $d' = \sqrt{(\sum \Lambda_i^2 - (\sum \Lambda_i)^2)}/\sigma$. The numerator (or $d'\sigma$) was computed at perceptual

thresholds for different testing conditions (Tables 2.3) and at the limits of the error bars, in order to determine its sensitivity to threshold changes.

For δSF tests, the stimuli are "balanced" (see [25]), in that $\sum \Lambda_i \approx 0$, or at least $(\sum \Lambda_i)^2 \ll \sum \Lambda_i^2$. The channel model predicts reasonably well the average thresholds as a function of peak's starting symmetry (SF) (Table 2.3(a)). It however fails to predict the δSF threshold trends as a function of BWF. For instance, to maintain a constant $d'\sigma$, the average δSF at BWF = 0.2 (21 component stimulus) needs to be larger by 21% ($\approx$ 0.20). A similar decrease in threshold is necessary at BWF = 0.4 for the 41 component stimulus (27%, or to 0.08).

For δBWF tests, all, with one exception, $d'\sigma$'s are comparable when considering the significant overlap due to the error bars (Table 2.3(b)). The only stimulus for which the model clearly fails is the broadest symmetric peak (SF = 0, BWF = 0.4) for both spectral densities.

The model also fails to account for the detection thresholds measured in the control experiment (C) described in Sec. 2.6.2. Specifically, it predicts higher than perceptual thresholds for the narrowest peaks (Table 2.3(c)).

Finally, the $d'\sigma$ for the phase data (Table 2.3(d)) increases with increasing threshold values at higher ripple frequencies. This is true for both 15 dB and 25 dB levels. The model therefore predicts a constant **pdl** instead of the increasing thresholds seen at higher ripple frequencies. For instance, a $d'\sigma = 10.51dB$ for 8 cycle/octave and 15 dB level stimulus, would predict a 9^o threshold, compared to the 49.58^o perceptual value. Note that the stimuli here are "balanced" as in the δSF tests.

In summary, the channel model predicts reasonably well most of the threshold

$d'\sigma$ for δSF test (21 components)

SF	BWF 0.1	BWF 0.2	BWF 0.4	average
0	4.46 ±0.13	3.15 ±0.19	4.36 ±0.52	3.99
0.4	3.68 ±0.07	3.58 ±0.13	4.84 ±0.30	4.03
average	4.07	3.37	4.60	
δSF threshold	0.34 ±0.01	0.17 ±0.01	0.17 ±0.02	

$d'\sigma$ for δSF test (41 components)

SF	BWF 0.1	BWF 0.2	BWF 0.4	average
0	2.96 ±0.19	2.88 ±0.13	3.97 ±0.18	3.27
0.4	3.36 ±0.22	3.22 ±0.15	4.33 ±0.20	3.64
average	3.16	3.05	4.15	
δSF threshold	0.16 ±0.01	0.11 ±0.005	0.11 ±0.005	

Table 2.3: $d'\sigma$ values for the "independent channel model" (Sec. 2.10.1). (a) $d'\sigma$ for δSF tests for 21 and 41 component spectra, evaluated at threshold and error bar limit values (in brackets) which are given at the bottom of each table. (For example, for BWF = 0.1 and SF = 0, δSF = 0.34 with error bar limit of $\pm$ 0.01, and the corresponding $d'\sigma = 4.46 \pm 0.13$.) Thresholds are from Figs. 2.5(b) and 2.3(b) for 21 and 41 density cases.

	$d'\sigma$ for δBWF test (21 components)			
	BWF			
SF	0.1	0.2	0.4	average
0	3.04 ±0.27	2.68 ±0.46	1.98 ±0.14	2.57
0.4	3.03 ±0.27	2.80 ±0.49	2.52 ±0.17	2.78
average	3.03	2.74	2.25	
δBWF/BWF threshold	30 ±3%	25 ±5%	25 ±2%	

	$d'\sigma$ for δBWF test (41 components)			
	BWF			
SF	0.1	0.2	0.4	average
0	3.11 ±0.24	3.02 ±0.19	2.58 ±0.21	2.90
0.4	3.12 ±0.25	3.16 ±0.20	3.27 ±0.25	3.18
average	3.11	3.09	2.92	
δBWF/BWF threshold	21.5 ±1.8%	20.0 ±1.4%	23.6 ±2.1%	

Table 2.3: (b) $d'\sigma$ values for δBWF tests. The table is organized as Table 2.3(a), with threshold values from Figs. 2.9(b) and 2.7(b) for the two density tests.

	$d'\sigma$ for control experiment C			
	BWF			
SF	0.1	0.2	0.4	average
0	4.22 ±0.86	6.40 ±1.86	5.46 ±1.28	5.36
δA_{max} threshold	2.79 ±0.55	3.49 ±0.98	3.44 ±0.81	
0.4	4.41 ±0.80	6.51 ±1.02	6.62 ±1.29	5.85
δA_{max} threshold	2.91 ±0.51	3.53 ±0.53	3.63 ±0.69	
average	4.31	6.45	6.04	

Table 2.3: (c) Similar to Table 2.3(a) for control experiment C (Fig. 2.13(b)). Thresholds are given separately for the two SF's.

	$d'\sigma$ for phase–ripple experiment					
	ripple frequency (cycle/octave)					
	0.25	0.5	1	2	4	8
$d'\sigma$	9.59	10.44	8.89	9.47	19.0	49.58
pdl, 15 dB	8.09°	9.08°	7.53°	8.1°	16.22°	43.47°
$d'\sigma$	10.67	10.20	10.08	14.32	28.39	
pdl, 25 dB	5.39°	5.31°	5.11°	7.34°	14.53°	

Table 2.3: (d) $d'\sigma$ for **pdl** tests with spectral sinusoids, for 15 dB and 25 dB peak levels and thresholds from Fig. 2.16(a).

trends measured in our experiments. However, we can discern no simple pattern to the failures since they occur at various BWF's, and are apparently unrelated to the number of stimulus components (for the two cases tested). It is possible that some of the simplifying assumptions are invalid, for instance the constant σ over all channels or the actual number of independent channels used.

2.10.2 The maximum difference model

This model is based on the detection of the maximum level difference between only two spectral regions. The model was derived from experimental results with flat standards, and was defined accordingly for such tests. It predicts well the thresholds in a number of profile analysis tasks [24]. In order to apply the model to the peak stimuli, the computational scheme was slightly modified. For instance, we define the level difference between the i^{th} and j^{th} frequency component as $\Lambda_{i,j} = 20\ \log((p_i)_{signal}/(p_i)_{standard}) - 20\ \log((p_j)_{signal}/(p_j)_{standard})$. Also, contrary to the assumption of the original model [24], we take the σ's to be constant for all channels, and hence the largest d' is defined by the largest $\Lambda_{i,j}$, or Λ (i.e., $d' = \max_{i,j} d'_{i,j}$, where $d'_{i,j} = \Lambda_{i,j}/\sigma_{i,j}$).

The computed Λ's for the δSF tests are shown in Table 2.4(a). As a function of a peak's BWF, the trends are well predicted for the larger BWF's (0.2 and 0.4). For the narrow peaks (BWF = 0.1), the model predicts smaller threshold than is observed. It also predicts a slight dependence of the thresholds on SF where none exists in the data. Note that including a variable σ would probably worsen the predictions. This is because for broader peaks Λ occurs closer to the edges of the profile where σ is larger. Consequently d' becomes smaller than indicated by the table.

Λ for δSF test (21 components)				
	BWF			
SF	0.1	0.2	0.4	average
0	5.12	2.53	2.58	3.41
0.4	4.35	3.29	3.31	3.65
average	4.73	2.91	2.94	
δSF threshold	0.34	0.17	0.17	

Λ for δSF test (41 components)				
	BWF			
SF	0.1	0.2	0.4	average
0	2.39	1.65	1.66	1.90
0.4	3.07	2.04	2.09	2.40
average	2.73	1.84	1.87	
δSF threshold	0.16	0.11	0.11	

Table 2.4: Maximal difference levels, Λ (dB), for the "maximal difference model" (Sec. 2.10.2). Tables are organized as Tables 2.3 (with the same threshold values as in Tables 2.3). (a)

For δBWF tests, Λ is approximately constant for all SF's and BWF's except for the narrowest peak for 21 component stimulus (Table 2.4(b)). Therefore, with the assumption of constant σ's across all spectral regions, the model predicts well the perceptual trends. However, increasing σ's towards the edges (as in [24]) would cause the d' to decrease with increasing BWF, predicting erroneously higher thresholds for these conditions.

Predicted threshold trends for the control experiment are consistent (Table 2.4(c)) with the experimental ones, since Λ values appear scattered around 2.7 dB for all conditions. They are, however, larger than those of the δBWF tests, indicating that the model accounts well for the trends in the thresholds, but not for their absolute values.

The maximal difference for the ripple-phase data is roughly constant at lower ripple frequencies, and follows the threshold increase at higher ripple frequencies (Table 2.4(d)). This is similar to the $d'\sigma$ trend in the channel model. Thus, the maximal difference model predicts constant thresholds at all ripple frequencies contrary to the observed data.

In summary, the picture that emerges from applying the maximum difference model to our data is a mixed one. For instance, the model clearly accounts for several of the observed trends in our data, especially the δBWF tests. Using variable σ's may extend the applicability of the model to a bigger portion of the tests, but it clearly destroys the good predictions of the δBWF tests. The pattern of prediction errors is somewhat more structured than for the channel model, in that the model seems to fail mostly for the narrowest peaks. It also fails to predict the δSF independence of SF.

A for δBWF test (21 components)

SF	BWF 0.1	0.2	0.4	average
0	1.85	1.64	1.68	1.72
0.4	1.86	1.66	1.68	1.73
average	1.85	1.65	1.68	
δBWF/BWF threshold	30%	25%	25%	

A for δBWF test (41 components)

SF	BWF 0.1	0.2	0.4	average
0	1.43	1.37	1.59	1.46
0.4	1.45	1.37	1.59	1.47
average	1.44	1.37	1.59	
δBWF/BWF threshold	21.5%	20%	23.6%	

Table 2.4: (b)

	A for control experiment C			
		BWF		
SF	0.1	0.2	0.4	average
0	2.42	3.03	2.61	2.69
δA_{max} threshold	2.79	3.49	3.44	
0.4	2.53	3.08	2.83	2.81
δA_{max} threshold	2.91	3.53	3.63	
average	2.47	3.05	2.72	

Table 2.4: (c)

	Λ for phase–ripple experiment					
	ripple frequency (cycle/octave)					
	0.25	0.5	1	2	4	8
Λ	2.12	2.38	1.97	2.12	4.23	11.11
pdl, 15 dB	8.09°	9.08°	7.53°	8.1°	16.22°	43.47°
Λ	2.35	2.32	2.23	3.16	6.32	
pdl, 25 dB	5.39°	5.31°	5.11°	7.34°	14.53°	

Table 2.4: (d)

2.10.3 Discussion

Both the channel and the maximum difference models described above account partially for the data. And it is possible that one or both of models can be made to account fully for the data with enough parameter adjustments. So it is difficult to pass judgement on these models on these grounds.

Both models, however, have been reported to raise fundamental questions when applied to a different stimulus – the sinusoidal ripple [24, 37]. The maximum difference model predicts well the detectable "amplitude of the ripple" [24]. The model also correctly predicts that thresholds are independent of the number of ripple cycles since they are estimated from a single pair of channels. This, however, runs directly counter to the premise of the channel model – that more independent channels of information must lead to lower thresholds [37]. The success of the maximum difference model with rippled (and alternating) profiles therefore raises the question: Why is the additional information provided by other independent channels not used?

One way to resolve this difficulty is to change the definition of the independent channels. For instance, if one thinks of the maximum level difference (from a pair of channels) as the amplitude of the ripple, then adding more ripple cycles (and hence more channels) does not add new information. Thus, an alternative definition (or model) of the channels is that they sense the amplitude (and perhaps the phase) of ripples of different frequencies. Such a channel does not measure the energy difference at one point in the spectrum, rather it conveys information about a more structured spectral pattern (e.g., the ripple).

This "ripple analysis model" implies that the detection strategy of the spectral profile is not applied to the profile directly, but instead to some transfor-

mation of the profile. Such an approach is not unusual – afterall, the spectral profile itself is a transformation from the time-domain of the signal, and the classical channel model is applied to a (Fourier) transformation of the signal (the spectrum). An elaboration of this idea is presented in the next chapter.

Chapter 3

A new ripple analysis model

3.1 Motivation

In characterizing the perception of spectral profiles, a basic objective is to select a model representation upon which an appropriate metric can be defined. Several such models have been proposed to account for data from a wide range of psychoacoustical tests, including profile analysis experiments and discriminations of simultaneous vowels. Examples are the independent channel model [25], the maximum difference model [24], the spectral slope model [29], and the spectral peak model [12]. Despite their unique characteristics, all models share the same fundamental starting point, that the spectral profile is represented by the acoustic spectrum on a logarithmic frequency axis. Relative to this profile, various operations are then defined to predict the measured perceptual thresholds.

However, based on experiments with peak profiles, and analysis of the results using two models [25, 24], it was suggested that a more parsimonious explanation of the results emerges if the spectral profile is assumed to be represented in the

auditory system by its Fourier transformation (Sec. 2.7). Specifically, such a representation implies that an arbitrary complex profile is analyzed into a collection of elementary sinusoidal profiles of different frequencies (called ripples). This is accomplished via a bank of filters, each independently tuned to a different ripple frequency (and possibly to a specific phase of the ripple). Detection thresholds would then be computed from this representation of the profile.

The idea that the brain analyzes and perceives its sensory patterns in this manner is relatively common, especially in the vision literature where it has been variously called multi-resolution or multi-scale representation, and spatial frequency analysis [15, 16]. It is, however, the elegant anatomical and physiological work of [38] that has provided the most immediate inspiration to pursue this type of model for the auditory system.

Recent neurophysiological recordings in the auditory cortex have confirmed that cortical cells are indeed tuned to specific ripple frequencies and phases [17, 18]. Furthermore, that the ordered mappings of response area bandwidths and asymmetries along the isofrequency planes of AI [9, 10] are just a different manifestation of these two ripple response properties. These findings support the notion that auditory cortical cells are able to perform the necessary ripple analysis.

Essential to the development of an auditory ripple analysis model, however, are basic psychoacoustical sensitivity measures using simple rippled spectra. A few such experiments have already been performed [37, 39, 2]. Using these data, and the conceptual framework outlined above, an explicit computational model is developed to interpret the results of various profile analysis experiments, including the peak profiles described in Sec. 2.7.

In the following sections, we first restate the results of the peak profile experiments in the context of the ripple analysis view. The basic outlines of the computational model are presented in Sec. 3.3. Its various parameters are estimated in Sec. 3.4 using detection thresholds reported in [2, 37]. The model is discussed in relation to auditory percepts such as timbre and pitch in Secs. 3.5.2 and 3.5.3.

3.1.1 Terminology and notation

Several terms will be frequently used in this paper to describe the ripple analysis representation of profiles. These are:

Ripple: refers to a sinusoidal spectral profile (e.g., $p(\omega) = \sin(2\pi\Omega\omega + \theta)$), on the logarithmic frequency axis, ω. A ripple has a ripple frequency Ω (in cycle/octave), and a ripple phase θ (in radians or degrees).

Ripple spectrum, $P(\Omega)$: refers strictly to the Fourier transform of the profile $p(\omega)$.

Ripple analysis filter, $H(\Omega; \Omega_o, \Phi_o)$: refers to a bandpass filter that is tuned around a characteristic ripple frequency (Ω_o) and phase (Φ_o).

Ripple transform, $r(\cdot)$: refers to the output of a bank of ripple analysis filters.

3.2 Restatement of spectral peak results

As discussed earlier in Sec. 2.7, changing the bandwidth factor (BWF) and symmetry factor (SF) of a peak profile is equivalent to a shift in the magnitude

and the phase of its ripple spectrum. In this section, the δSF and δBWF/BWF detection thresholds are restated in the context of the ripple analysis model.

3.2.1 δBWF as a shift in the magnitude of the peak's ripple spectrum

A BWF change represents a dilation of the profile $p(\omega)$ along the ω axis. Thus, changing the BWF by a factor α, changes the profile to $p(\alpha\omega)$, and its ripple spectrum from $P(\Omega)$ to $1/\alpha\ P(\Omega/\alpha)$. Therefore, on a logarithmic Ω axis, the dilation causes $P(\Omega)$ to translate by $\log_2 \alpha$ octaves without changing its shape (see Figs. 2.14(a) and (b)).

Consequently, the δBWF/BWF thresholds from Figs. 2.7 can be restated as follows: **Regardless of the shape of the peak, subjects detect approximately a 22% change in BWF, or a profile dilation factor $\alpha \approx 0.8$. This corresponds to 0.29 octave translation of the ripple spectrum.**

For a single ripple profile, a dilation simply changes the frequency of the ripple. Thus, a dilation threshold in this case measures sensitivity to ripple frequency changes, or the so-called frequency-difference-limen (**fdl**). Such measurements were recently reported in a Ph.D. Dissertation by [2]. These are discussed in detail and related to our measurements later in Secs. 3.4.2 and 4.1.

3.2.2 δSF as a shift in the phase of the peak's ripple spectrum

As discussed earlier (Sec. 2.7, Figs. 2.14(c) and (d)), small δSF's (or tilts) in the peak profile $p(\omega)$ can be equivalently produced by adding a constant phase angle to all components of its ripple spectrum $P(\Omega)$, leaving the magnitude of the ripple spectrum unaltered.

Therefore, δSF thresholds are in effect a measure of the subjects' ripple phase sensitivity, and as such, one may restate the results from Figs. 2.3 as follows: **For broad peak profiles (BWF $\geq$ 0.2), subjects detect a δSF $\approx$ 0.11, which corresponds approximately to a 6^o shift in the phase of the ripple spectrum of the profile. Thresholds increase to about 0.16 or 10^o for the narrowest peak (BWF = 0.1).** (See Appendix C for computational details).

For single ripple profiles, this task is exactly equivalent to measuring the sensitivity to ripple phase shifts, termed here phase-difference-limen (**pdl**). Such measurements are reported in (Sec. 2.8), and will be discussed in relation to the δSF thresholds later in Sec. 4.4.

3.3 General description of the ripple analysis model

The ripple analysis model can be conceptually divided into two stages: (1) A ripple transformation stage modeled as a filter bank which simply converts the input spectral profile into its corresponding ripple transform; (2) A detection

model which operates on the magnitude or phase of the ripple transform, or on selected features of it such as its maxima and edges.

3.3.1 Computing the ripple transform of a spectral profile

This stage consists of a bank of ripple selective filters analogous to the frequency selective filters of the cochlea. Each filter $H(\Omega, \Omega_o)$ is assumed to be Gaussian shaped and centered around a particular ripple frequency Ω_o (Figs. 3.1(a) and (c)). Therefore,

$$H(\Omega; \Omega_o) = G(\Omega - \Omega_o) + G(\Omega + \Omega_o), \tag{3.1}$$

where, $G(\Omega \pm \Omega_o) = e^{-\frac{(\Omega \pm \Omega_o)^2}{2\sigma^2}}$, and σ determines the width of the ripple filter centered at Ω_o. The inverse transform (impulse response) of $G(\Omega \pm \Omega_o)$ is $g(\omega) = g_o(\omega)\ e^{\mp \jmath 2\pi\Omega_o\omega}$, where $g_o(\omega) = \sqrt{2\pi}\sigma\ e^{-\frac{(2\pi\omega\sigma)^2}{2}}$. The inverse transform of the $H(\Omega; \Omega_o)$ filter, $h(\cdot)$ (Fig. 3.1(b)), is a Gaussian weighted sinusoid known as the Gabor–like function [40] (they are also a subset of the so-called localized trigonometric basis functions, [41]), given by:

$$h(\omega; \Omega_o) = 2\ g_o(\omega) \cos(2\pi\Omega_o\omega).$$

In general, the weighted sinusoid can have an arbitrary phase, Φ_o, such that:

$$h(\omega; \Omega_o, \Phi_o) = 2\ g_o(\omega) \cos(2\pi\Omega_o\omega - \Phi_o). \tag{3.2}$$

Therefore the output of any given filter can be computed by convolving its impulse response $h(\omega; \Omega_o, \Phi_o)$ with the input profile $p(\omega)$:

$$r(\omega_o, \Omega_o, \Phi_o) = \int h(\omega_o - \omega; \Omega_o, \Phi_o) p(\omega) d\omega$$

$$= \int H(\Omega;\Omega_o,\Phi_o,\omega_o)P(\Omega)e^{\jmath 2\pi\Omega\omega_o}d\Omega. \tag{3.3}$$

Note that each ripple filter is indexed by three parameters $(\Omega_o, \Phi_o, \omega_o)$ reflecting the ripple frequency and phase selectivity of the filter, and the location along the tonotopic axis around which the analysis is performed.

In the remainder of this paper, Eq. 3.3 will be simplified to focus on the magnitude (rather than the phase) of the ripple filter outputs. This is justified for profile manipulations which affect only the magnitude of the ripple transform, e.g., changing the BWF of peak profiles. Another simplification of Eq. 3.3 is to compute the filter outputs only around the center of the profile ($\omega_o = 0$), i.e., to suppress the index ω_o. The impulse response of a ripple filter centered at $\omega_o = 0$ is given in Eq. 3.2, and its localized Fourier transform is:

$$H(\Omega;\Omega_o,\Phi_o) = (G(\Omega-\Omega_o)+G(\Omega+\Omega_o))\cos(\Phi_o)-\jmath\,(G(\Omega-\Omega_o)-G(\Omega+\Omega_o))\sin(\Phi_o)$$

$$= H(\Omega;\Omega_o)\cos(\Phi_o) + \hat{H}(\Omega;\Omega_o)\sin(\Phi_o),$$

where,

$$\hat{H}(\Omega;\Omega_o) = \mathcal{H}H(\Omega;\Omega_o),$$

and $\mathcal{H}$ is a Hilbert transformer defined as $\mathcal{H} = -\jmath\ sign(\Omega)$. Note that the filter is real, stable, and causal [42]. Since $G(\Omega \pm \Omega_o) = e^{-\frac{(\Omega\pm\Omega_o)^2}{2\sigma^2}}$, $H(\Omega;\Omega_o)$ is pure real and $\hat{H}(\Omega;\Omega_o)$ is pure imaginary. Then:

$$H(\Omega;\Omega_o,\Phi_o) = H(\Omega;\Omega_o)\cos(\Phi_o) - \jmath\ sign(\Omega)H(\Omega;\Omega_o)\sin(\Phi_o)$$

$$= H(\Omega;\Omega_o)\ e^{-\jmath\ sign(\Omega)\Phi_o}.$$

The ripple transform response to an arbitrary (real) input profile with ripple spectrum $P(\Omega) = |P(\Omega)|e^{\jmath\theta(\Omega)}$, is:

$$r(\Omega_o,\Phi_o) = \int_{-\infty}^{\infty} H(\Omega;\Omega_o,\Phi_o)P(\Omega)d\Omega = \int_{-\infty}^{\infty} H(\Omega;\Omega_o)|P(\Omega)|e^{\jmath(-sign(\Omega)\Phi_o+\theta(\Omega))}d\Omega$$

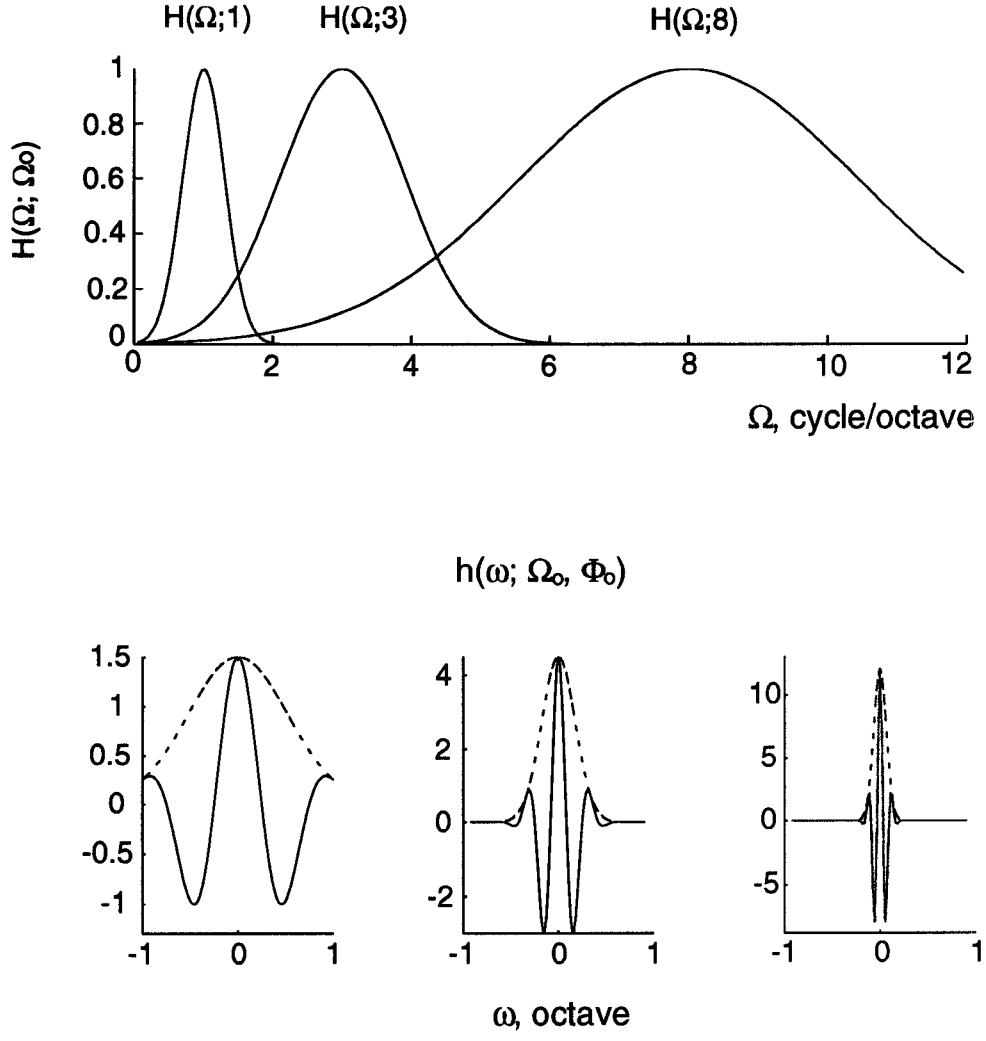


Figure 3.1: (a) Magnitude of three ripple filters centered at $\Omega_o = 1, 3$, and 8 cycle/octave. (b) Inverse Fourier transforms for the three filters in (a) centered at $\omega_o = 0$, with $\Phi_o = 0$ The inverse is computed as $h(\omega; \Omega_o) = 2\sqrt{2\pi}\sigma(\Omega_o)e^{-\frac{(2\pi\omega\sigma(\Omega_o))^2}{2}}\cos(2\pi\Omega_o\omega)$, for $\sigma(\Omega_o) = 0.3\ \Omega_o$.

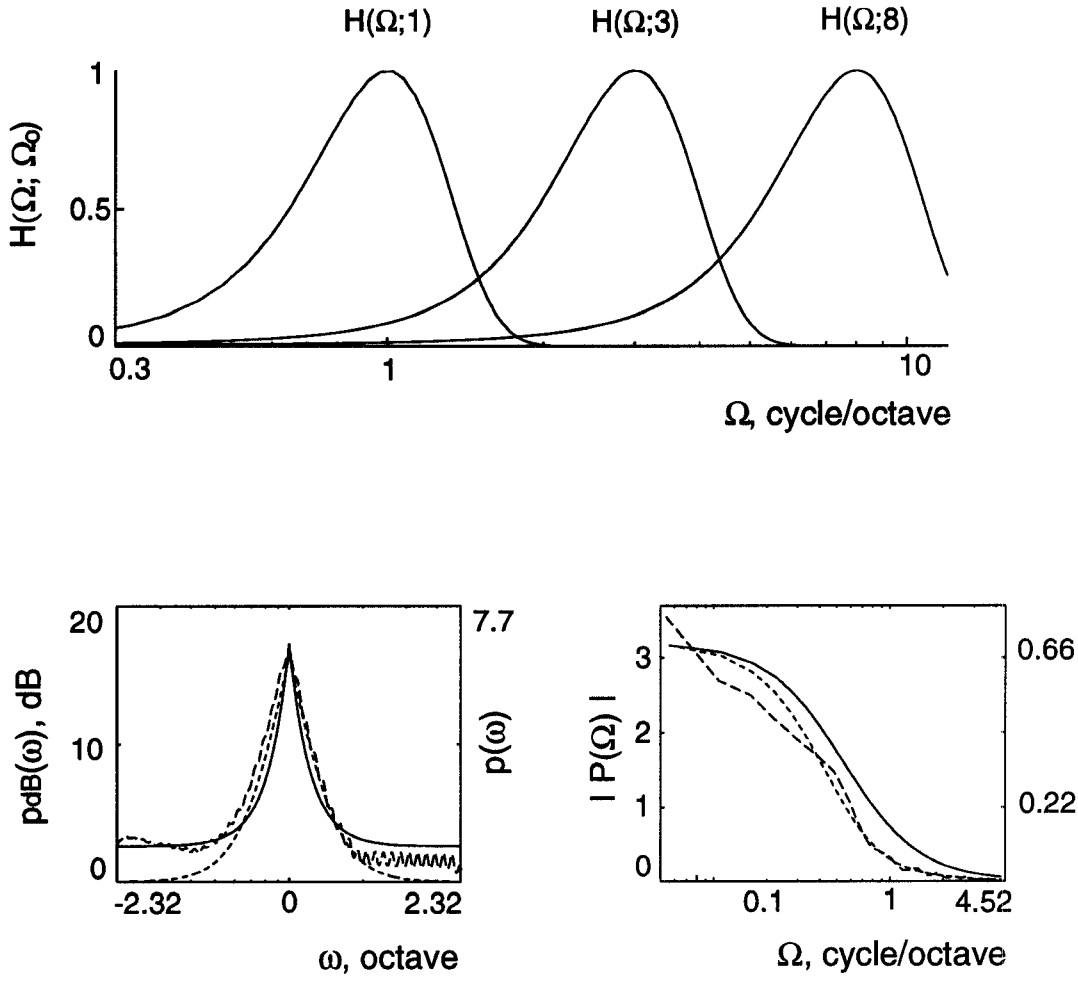


Figure 3.1: (c) Same as (a) but plotted on a logarithmic Ω axis. (d) Three input representations of a symmetric peak profile with BWF = 0.2 and A_{max} = 15 dB (left), and their corresponding ripple spectra (right). There are little differences between the three representations or their ripple spectra. The solid line is the normalized linear representation of the peak (right ordinate). The dotted line is the same peak profile represented on a logarithmic amplitude scale. The dashed line depicts the output of the excitation pattern model (no corrections were applied in the model, and the base was 0 dB amplitude; see [1] for details).

$$= 2\int_0^\infty H(\Omega;\Omega_o)|P(\Omega)|\cos(\Phi_o - \theta(\Omega))d\Omega.$$

There are two important observations to make here:

1) The ripple transform is a (sinusoidal) function of the characteristic phase of the filter. Thus, there is a particular Φ_o for which the response will be a maximum. Its value depends on the $\theta(\Omega)$ of the input profile.

2) If the input profile is dilated by factor α, i.e., $P(\Omega) \longrightarrow 1/\alpha P(\Omega/\alpha)$, then $r(\Omega_o, \Phi_o)$ becomes:

$$r(\Omega_o, \Phi_o) = 2\int_0^\infty H(\Omega;\Omega_o)|P(\Omega/\alpha)|\cos(\Phi_o - \theta(\Omega/\alpha))d\Omega/\alpha.$$

Evaluating $r(\cdot)$ at Ω_o/α $(= \Omega'_o)$, and letting $\Omega/\alpha = \Omega'$, we get:

$$r(\Omega'_o, \Phi_o) = 2\int_0^\infty H(\Omega';\Omega'_o)|P(\Omega')|\cos(\Phi_o - \theta(\Omega'))d\Omega', \tag{3.4}$$

which is identical to $r(\Omega_o, \Phi_o)$ prior to dilation. Therefore, a dilation simply translates the filter outputs by $\log_2 \alpha$ octaves, preserving their shape against a $\log_2 \Omega_o$ axis. Furthermore, the response as a function of Φ_o remains unaltered.

Depending on the input profile and the type of manipulations applied to it, it is possible to simplify the above expression for $r(\cdot)$. Majority of profiles considered here will be even or odd symmetric around their center, so that their ripple spectra $P(\Omega)$ have constant phase as a function of Ω. In this case, $\theta(\Omega) = sign(\Omega)\theta_o$, and the response becomes:

$$r(\Omega_o, \Phi_o) = \cos(\Phi_o - \theta_o)\int_{-\infty}^\infty H(\Omega;\Omega_o)|P(\Omega)|d\Omega.$$

Depending on the phase (asymmetry) of the input profile θ_o, the maximal response is at $\Phi_o = \theta_o$:

$$r(\Omega_o) = \int_{-\infty}^\infty H(\Omega;\Omega_o)|P(\Omega)|\ d\Omega. \tag{3.5}$$

This is the expression used in computing the ripple transforms of various input profiles.

Once again, for a pure dilation manipulation of an input profile (e.g., δBWF test), the ripple filter response $r(\Omega_o)$ translates along the $\log \Omega_o$ axis, but remains unchanged along the Φ_o axis (i.e., it is largest at $\Phi_o = \theta_o$ as before).

The two profile experiments that do not satisfy the above assumptions are the detection of ripple phase shifts, and detection of SF changes of the peak profiles. Both tasks are discussed separately in Sec. 4.4.

Finally, the filter bank depicted in Figs. 3.1(a) and (c) is assumed to be a constant Q bank, i.e., its filters have constant widths on a $\log \Omega$ axis, or equivalently have linearly increasing widths (σ's) with Ω_o. This choice is primarily justified by adaptation experiments (both with ripples [2] and visual gratings [38]), and neurophysiological experiments [43] in which filter tuning was estimated around various Ω_o's to be approximately 1 octave (measured at the half amplitude point). This corresponds approximately to choosing σ's as $\sigma(\Omega_o) = 0.3\ \Omega_o$.

The representation of the input spectral profile

Another important issue to consider here is how the auditory system represents the input spectral profile. For example, is the profile represented on a linear or logarithmic (dB) amplitude scale? There are only a few psychoacoustical studies that explicitly contrast these two representations, and their results are somewhat contradictory [2], [44]. The logarithmic amplitude spectrum has been the most widely assumed representation in profile analysis experiments. It is a scale-normalized spectrum in that it preserves only level-independent features of the spectrum. A linear, but similarly level-tolerant representation, would result

if the amplitude spectrum is normalized by the total energy or average power or some other level-sensitive quantity such as the height of the base profile. Other representations can be used that range from a simple scale-normalized power spectrum, to more complicated biologically and psychoacoustically inspired representations such as the excitation pattern model [1], and the auditory filter models of [45, 2, 46, 47, 48].

In general, an inappropriate profile representation distorts its intended ripple spectrum $P(\Omega)$. In some cases, the distortions are rather small, as with low amplitude ripples where linear and logarithmic ripples look very similar, and their perceptual thresholds are closely matched [39, 37, 2]. Similarly, peak profiles exhibit similar ripple spectra regardless of whether they are represented on a logarithmic or linear scale, or via a more complicated model (Fig. 3.1(d)). In other cases, the distortion is large but inconsequential in the context of the ripple analysis model as will be discussed in more detail in Sec. 3.5.1.

In summary, for most profiles considered here, the choice of the input representation does not alter the basic conclusions. Therefore, $p(\omega)$ is taken to be the linear amplitude spectrum of the stimulus, normalized by the amplitude of the base, i.e., $p(\omega)$ is the profile amplitude relative to its base. For the case of spectral peaks, where the whole amplitude is denoted by $p(\omega)$ (Eq. 2.1), the input representation is actually $p(\omega)/b$.

3.3.2 Different types of detection procedures

Figures 3.2 illustrate the output of the filter bank (according to Eq. 3.5) for a two ripple profile. The input ripple spectrum $P(\Omega)$ is a pair of impulses, and $r(\cdot)$ therefore reflects the filter shapes. Since the model is linear, the output

amplitude is proportional to the amplitude of the ripple. For a ripple that is just detectable against a flat base (e.g., at Ω_1), one can define a critical level (K) of the output pattern at perceptual threshold. Therefore, for a loud ripple (e.g., at Ω_2) $r(\cdot)$ exceeds K significantly.

In order to account for perceptual data based on such outputs, two types of tests will be distinguished. The first are those in which the profile is to be detected against a flat standard, i.e., the task is to detect the existence of the profile. In this case, a small portion near the maximum of $r(\cdot)$ exceeds K at threshold, and hence other details of $r(\cdot)$ do not play a role in the detection. The second type of experiments are those for which the standard is not flat. Instead, the subject is to detect a change in some parameter of an audible profile, for instance in the phase or frequency of a ripple. In this case, $r(\cdot)$ is well above K and a change in the profile affects a global change in $r(\cdot)$. Consequently, the change in shape or some feature of the $r(\cdot)$ pattern must be taken into account.

Almost all previously reported profile detection experiments fall in the first category. They include those described by [24, 27], and the ripple detection experiments (called here ripple intensity-difference-limen experiments, or ripple-**idl**) by [2, 37].

Examples of the second category are: The pedestal-type of experiments, such as increments on a single component pedestal [23] or on a peak profile pedestal (Sec. 2.6.2); The δSF and ripple-**pdl** detection (Secs. 2.3 and 2.8); The δBWF and ripple-**fdl** tests (Sec. 2.4 and [2]).

The detection procedures for these two types of experiments are slightly different:

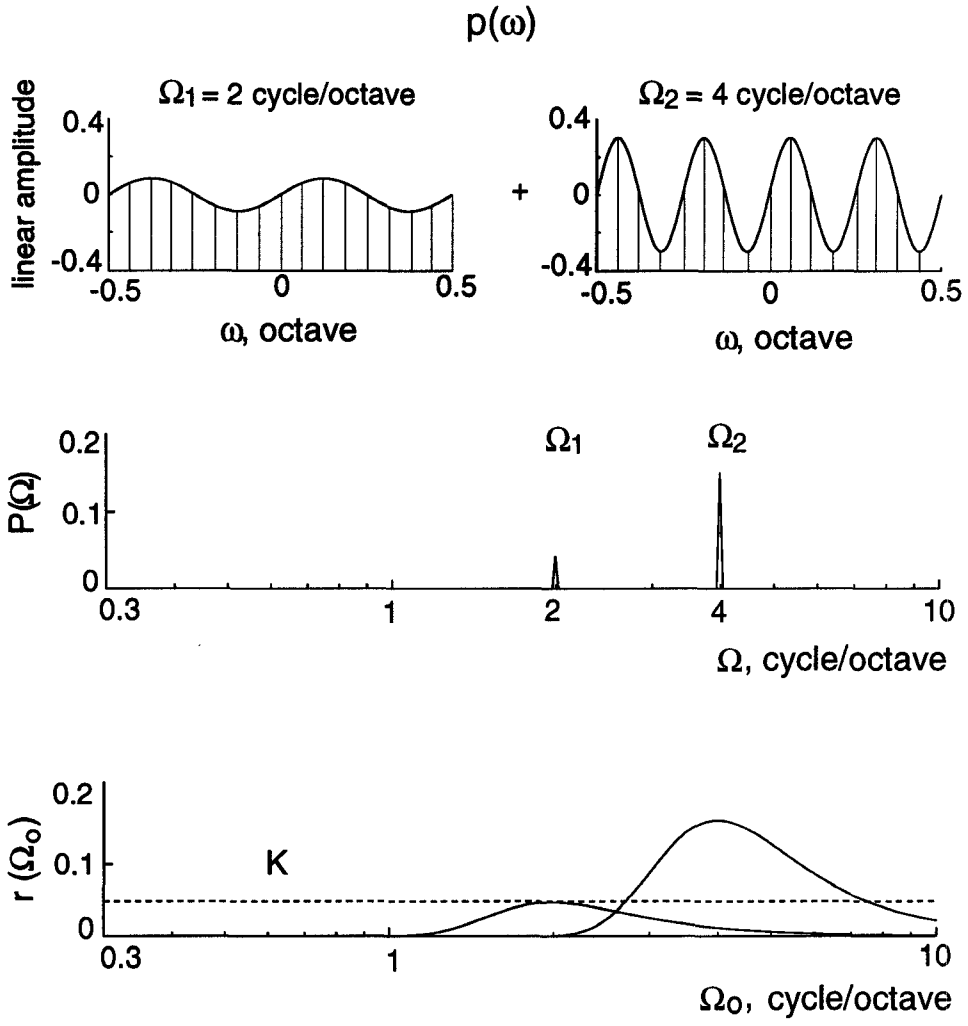


Figure 3.2: (a) Two single ripple profiles with ripple frequencies $\Omega_1 = 2$ and $\Omega_2 = 4$ cycle/octave, and amplitudes 0.1 and 0.3, respectively. The Ω_1 ripple is at its just detectable level [2]. (b) Ripple spectrum of the two ripples. (c) Ripple transform for the profiles in (a) (according to Eq. 3.5). The $K = 0.05$ value corresponds to the detection threshold of the Ω_1 ripple at the output $r(\cdot)$.

Detection of profiles with flat standards

We first use the results of the ripple-**idl** thresholds [2] to determine K (Sec. 3.4.1). For other profiles with flat standards, the $r(\cdot)$ at perceptual thresholds are computed and are compared to K in order to evaluate the model performance (Sec. 4.2).

Detection of profiles with non–flat standards

The procedure is slightly more elaborate for the second type of experiments. It is best illustrated by the ripple-**fdl** example in Fig. 3.3. Here, the filter outputs are well above K. A change of the ripple frequency (from Ω_1 to $\Omega_1 + \delta\Omega$) simply translates or shifts $r(\cdot)$ along the logarithmic Ω_o axis. In such an experiment, it is presumed that the subject detects the amount of shift in $r(\cdot)$ (Sec. 2.7). Since all $r(\cdot)$ points translate the same amount, the shift can be measured anywhere on the pattern, e.g., at its maximum, or at its right or left edges. We shall always detect and measure the shift in the steepest lowpass edge in $r(\cdot)$. These points are marked by the two dashed lines in $r(\cdot)$ of Fig. 3.3. While other features of $r(\cdot)$ are theoretically equivalent, our choice is motivated by the fact that the ripple transforms of arbitrary complex profiles are necessarily bandlimited (i.e., have lowpass edges), but may not always exhibit clear maxima. Therefore, to estimate detection thresholds in non-flat standard tasks, the shift in $r(\cdot)$ is measured and compared to a just detectable change Δ which is determined from the ripple-**fdl** measurements in [2] (Fig. 3.3 and Sec. 3.4.2).

The filter output, $r(\cdot)$, as computed from Eq. 3.5 above, reflects only the magnitude of ripple transform as a function of Ω (see Sec. 3.3.1). Therefore, it is insensitive to the ripple phase and cannot be used to predict δSF and ripple-**pdl**

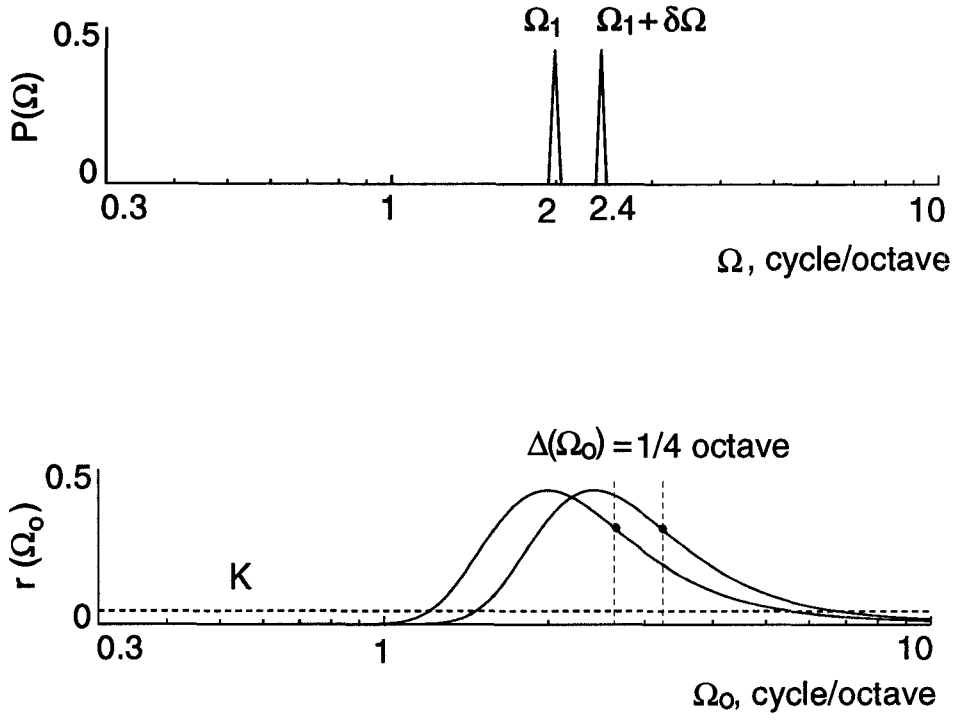


Figure 3.3: A two ripple spectrum and its ripple transform. The ripple frequencies are $\delta\Omega = 20\%$ apart (**fdl**–threshold, [2]), with $\Omega_1 = 2$ and $\Omega_1 + \delta\Omega = 2.4$ cycle/octave. The amplitudes are well above K. The (vertical) dashed lines mark the locations of the steepest lowpass edges of $r(\cdot)$.

thresholds. Instead, these are briefly related to the model in Sec. 4.4 (and can be explicitly computed as in Sec. 3.3.1).

Finally, in the pedestal-type experiments mentioned above, an audible profile is increased in amplitude until the change is detected. In the linear model, the corresponding $r(\cdot)$ output becomes proportionately larger. Threshold is simply taken to be the percentage increase in $r(\cdot)$ needed for detection. In Sec. 4.3, the model predictions for the two pedestal-type experiments mentioned above will be discussed.

Stochastic detection procedures

The filter bank of the ripple analysis model can be viewed as a set of independent channels conveying information about the ripple spectrum of the profile ($P(\Omega)$). In this sense, it is analogous to the classical view of the critical band channels operating upon the spectral profile. Hence, the independent channel model of [25] and the more specific model of [24] can be formally adapted and applied to the outputs of the ripple filter bank. This strategy is not pursued here because of the lack of sufficient data on such parameters of the channels as their variances. While one may postulate various hypotheses to explain threshold changes with level or ripple frequency, it is best at present to adopt simpler approaches as discussed below in Sec. 3.4.

3.4 Determining the perceptual threshold parameters of the ripple analysis model

In this section, the results of the ripple-**idl** and -**fdl** experiments by [37, 2] are used to compute the model's perceptual thresholds K and Δ. The model consists of a bank of equal amplitude constant Q filters $H(\Omega, \Omega_o)$ with parameter $\sigma(\Omega_o) = 0.3\ \Omega_o$, implying a constant width along the logarithmic Ω axis (≈ 1 octave). The filter outputs $r(\cdot)$ are computed as in Eq. 3.5.

3.4.1 Ripple idl's

Figure 3.4(a) illustrates $r(\cdot)$ for two just detectable ripples Ω_1 and Ω_2. The amplitudes of the ripples (0.05) are derived from the **idl** curve reported in [2] and reproduced in Fig. 3.4(b). Since the two ripples are simple dilations of each other, their ripple spectra are related by a translation, and the constant Q filter bank preserves this relationship in $r(\cdot)$. To detect them, it is assumed that $r(\cdot)$ must exceed K. In a stochastic detection procedure, d' index is defined as K/σ_n, where σ_n^2 is the variance of the noisy channel located at the maximum of $r(\cdot)$.

To obtain the same **idl** results from the model, the simplest approach is to assume that K directly reflects the **idl** curve, i.e., $K(\Omega_o) = \mathbf{idl}(\Omega_o)$. Alternatively, one may assume K to be constant and instead weight the input profile or the filter heights by the inverse of the **idl** bowl. While all these assumptions are equivalent with respect to the single ripple **idl**'s, they generally have different consequences for arbitrary input profiles. In the absence of additional supporting data, we shall adopt the simplest approach, taking K to be a function of Ω_o

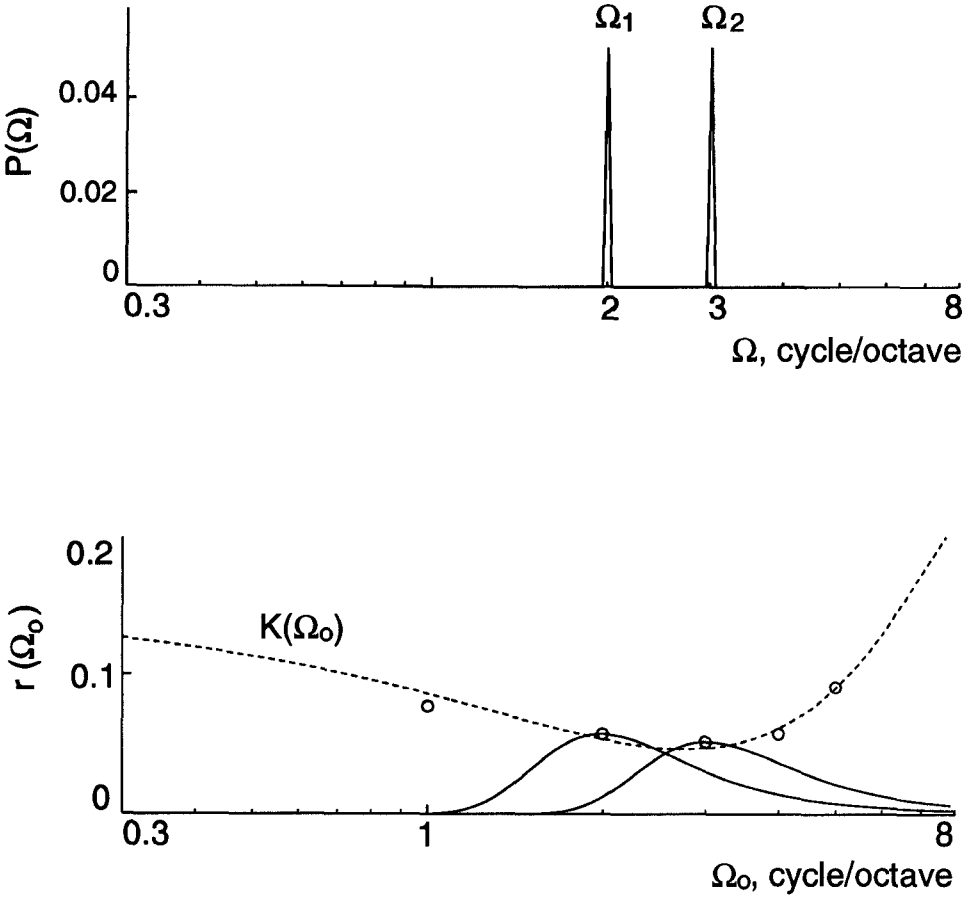


Figure 3.4: (a) A two ripple spectrum and (b) its ripple transform. Both ripples are at their **idl**–threshold values. The dashed line is a polynomial approximation to the measured data points (denoted by circles) reproduced from Fig. 3.27 in [2]. The detection threshold $K(\Omega_o)$ reflects the shape of the perceptual threshold, $\mathbf{idl}(\Omega_o)$.

as shown in Fig. 3.4(b). Note that this is equivalent to assuming a variable σ_n, analogous to the model of [24].

3.4.2 Ripple fdl's

In **fdl** tests, input ripples are large enough such that $r(\cdot)$ is well above the perceptual threshold $K(\cdot)$. As demonstrated earlier (Sec. 3.3.2, Fig. 3.3), a shift in the ripple frequency from Ω_1 to $\Omega_1 + \delta\Omega$ causes the corresponding $r(\cdot)$ to translate by the same amount along the $\log \Omega_o$ axis. This is a direct consequence of the dilation relation among the filters, i.e., the choice of the filter widths (σ) to be a linear function of Ω_o. No other parameters in the model affect this shift.

Measured **fdl**'s are reproduced from [2] in Fig. 3.5. To obtain the same results from the model, the minimal detectable shifts along the Ω_o axis, $\Delta(\Omega_o)$, are assumed to be equal to the **fdl** curve (the ordinate on the right in Fig. 3.5), i.e.:

$$\Delta(\Omega_o) = \log_2(1 + \mathbf{fdl}(\Omega_o)) \quad \text{(octave)}. \tag{3.6}$$

In the intermediate range of ripple frequencies (0.7–6 cycle/octave), the **fdl**'s are constant at approximately 20%, or $\Delta(\Omega_o) \approx \log_2(1.2) \approx 0.26$ octave. The **fdl**'s rise outside of this range.

More elaborate schemes can be used to incorporate the **fdl** curve into the model. For instance, it can be partially encoded into the observed shifts of the output patterns by adding a constant to the σ, e.g., $\sigma(\Omega_o) = 0.3\ \Omega_o + 0.05$. This increases the filter widths substantially only in the low Ω_o region ($\leq$ 1 cycle/octave), in effect increasing the **fdl**'s in this range as observed in the data. Similarly, the **fdl** increase in the high Ω_o region ($\geq$ 6 cycle/octave) may be related

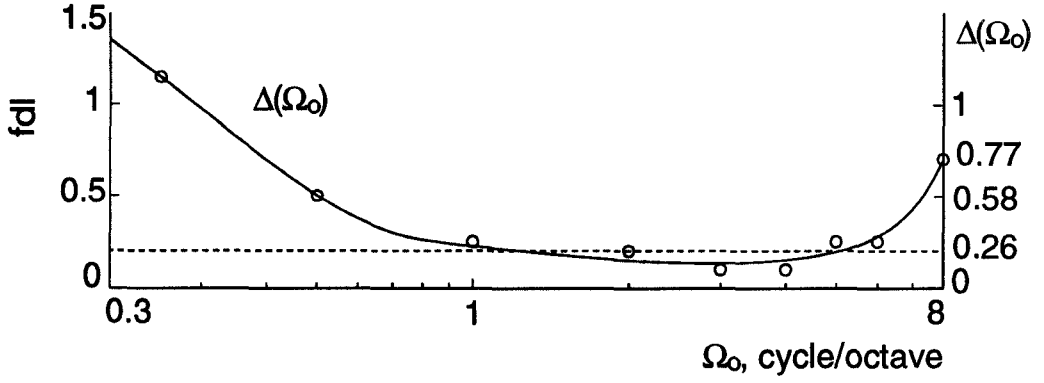


Figure 3.5: Interpolated **fdl**–thresholds (denoted by circles) reproduced from Fig. 3.30 in [2]. The corresponding $\Delta(\Omega_o)$ values (Eq. 3.6) are shown on the right scale.

to the increasing **idl**'s there, and hence can be accounted for by introducing level-sensitive procedures for the detection of shifts in $r(\cdot)$.

In the absence of further definitive data in favor of any of the above approaches, we shall adopt the simplest and assume Δ (in octaves) to be a function of Ω_o in the shape of the **fdl** curve as in Fig. 3.5 (Eq. 3.6).

3.4.3 Summary of the computational steps

The ripple analysis model consists of the following computational steps:

(1) Compute $P(\Omega)$, the ripple spectrum of the input profile $p(\omega)$: $P(\Omega) = \int p(\omega) e^{-j2\pi\Omega\omega} d\omega$.

(2) Compute $r(\Omega_o)$, the output of the filter bank using Eq. 3.5. The width of the filter $H(\Omega, \Omega_o)$ centered at Ω_o is determined from $\sigma(\Omega_o) = 0.3\ \Omega_o$.

(3) For flat standard profile experiments, compare $r(\Omega_o)$ to the perceptual threshold curve $K(\Omega_o)$ as given in Fig. 3.4(b).

(4) For **fdl**-type tests, compute $r(\Omega_o)$ and compare its shifts (at the location of the maximal lowpass slope) to $\Delta(\Omega_o)$ (defined by Eq. 3.6 and Fig. 3.5). For pedestal-type experiments, thresholds are computed from the percentage scale change of $r(\cdot)$.

3.5 Discussion

3.5.1 Summary of the ripple analysis model and underlying assumptions

A simplified ripple analysis model is presented to integrate findings from various profile analysis experiments. The basic operation implied by the model is a transformation of the profile into its ripple transform domain. Various manipulations on the profile are then interpreted and detected in this domain. Two sets of assumptions underlie the model: the nature and linearity of the input representation, and the parameters of the ripple analysis filters.

The representation and linearity of the input profiles

It is assumed in this model that the auditory system analyzes the amplitude spectrum on a linear, rather than on a logarithmic, scale. Neither is known to be the true auditory representation, and other representations such as the power spectrum or some output of a cochlear filter model might be more appropriate. The effects of using a distorted representation are minimal in the cases examined in this paper, because it usually creates distortion components of smaller

amplitudes that, for **idl** tests, are effectively sub-threshold at their corresponding filters. For **fdl**-type tests where the input is large, the pattern is dilated as a whole, and hence all distortion components are shifted consistently with the fundamental.

The exact nature of the input profile is more consequential in cases where metrics between different complex profiles are considered (see discussion later) or when profiles are added linearly. This brings up a fundamental assumption of the ripple analysis model, that the auditory system analyzes linearly a profile in terms of ripples. How does the cochlea with all of its nonlinearities preserve the principle of superposition of spectral ripples? And if not, in what form is the linearity preserved so as to permit this type of ripple analysis? Hillier ([2], Sec. 4.4) attempted to address this issue using adaptive experiments. Recent models of cochlear processing have also tackled this question [49]. However, the validation of linearity hypothesis must await direct tests from psychoacoustical and neurophysiological experiments in search of systematic interactions among a small number of simultaneously presented ripples.

The parameters of the ripple analysis filters

The filters determine the shape of the ripple transform $r(\Omega_o)$, and hence the interpretation of the results. The choices made here regarding the parameters and shape of these filters satisfy two basic experimental findings, reported in [2] (Secs. 4.4 and 4.5) and in psychophysical experiments in vision using analogous stimuli (summarized in [38]). These are: (1) the filters are roughly of a constant Q factor, and (2) their width is approximately 1 octave (i.e., $\sigma(\Omega_o) = 0.3\ \Omega_o$). It might be argued that other details of the filter shapes (such as their heights and

form) can be inferred from the **idl** and **fdl** measurements. Such an inference, however, as discussed in Sec. 3.4, is uncertain since other parameters can be readily adjusted with similar effects on the model outputs. To avoid making such specific commitments in the model, the filters are defined in as general a form as possible.

3.5.2 The complete model and its relevance to timbre perception

The simplified ripple analysis model ignores the explicit depiction of two axes of the complete model: The tonotopic axis (ω) and the ripple phase axis (Φ). In its full version (Eqs. 3.1, 3.2, and 3.3), a profile would be analyzed by a bank of filters at each point along the profile. Similarly, each filter of the bank would be repeated many times, each selective to one phase of the ripple. As such, a spectral profile would be expanded into a three-dimensional representation.

An important aspect of the complete representation is that it is local with respect to the profile. This is emphasized by the preservation of the tonotopic axis in the output, i.e., the filter outputs change as a function of location along the tonotopic axis. The above statements are exactly analogous to the locality in time of a spectrogram of running speech. They simply indicate that the profile is (in effect) first windowed prior to the application of the ripple transformation. Computationally, the locality of the ripple analysis is implied by the relatively broad bandwidths of the ripple filters, or equivalently, the limited extent of the impulse response of the filters (Fig. 3.1(b)).

Since changes along any of the three axes are perceptually detectable, com-

paring two arbitrary profiles must be done along all three dimensions of the ripple representation. Such a metric might be considerably simpler in this domain (e.g., simple Eucledian distance), than metrics based on the profiles in the usual spectral domain, since the model transformations imply many of the "conditioning" operations often required in these metrics. For example, the metric suggested by [12] applies (among other things) a second derivative upon the profile, effectively de-emphasizing the slow variations (or equivalently, the low frequency ripples) of the profile. Such an operation is implied in the model by the highpass edge of the **idl** curve.

3.5.3 Relation to rippled noise stimuli and the pitch of complex sounds

A different rippled spectrum stimulus that has been widely used in studies of pitch perception is the so-called rippled noise [50]. It has a sinusoidal spectral envelope defined against a linear, rather than a logarithmic, frequency axis, i.e., is similar to a harmonic spectrum. On a logarithmic frequency axis, however, a harmonic spectrum appears to have an exponentially increasing ripple frequency.

The representation of harmonic spectra in the ripple analysis model leads to many interesting hypotheses regarding the encoding of complex pitch in the auditory system. Of immediate relevance to the focus of this paper, however, is the interpretation of the "dominance region" in pitch models. Specifically, it has long been known that the $2^{nd}, 3^{rd}$, and 4^{th} harmonics in a series are dominant in conveying the pitch of the complex [51]. From a computational point of view, pitch models have taken this phenomenon into account by emphasizing (or

weighting more heavily) these regions of the spectral profile prior to estimating the pitch strength and value [52].

The dominance region can be viewed in the context of the ripple analysis model as a correlate of the ripple **idl** sensitivity curve (Fig. 3.4(b)), which has its lowest thresholds for ripples around 2 cycle/octave. In a harmonic spectrum defined against a logarithmic axis, the spectral profile around the $2^{nd} - 4^{th}$ harmonics has this same ripple frequency. Thus, the emergence of the ripple **idl** curve may share the same origins as those responsible for the dominance region, namely a combination of suppressive and other interactions at relatively peripheral stages of the auditory system [53, 52].

3.5.4 Relation to visual processing

An appealing aspect of the ripple analysis model is that it shares identically the conceptual framework of spatial frequency analysis that has long been prevalent in visual processing. While having its critics (see reviews in [38]), this approach has been supported by substantial anatomical, neurophysiological, and psychophysical evidence, elegantly detailed in [38]. Interestingly, in the vision community, the idea that the brain performs a local Fourier transformation is motivated by its similarity to the cochlear transformations of the auditory system!

From the perspective of the auditory system, however, the cochlea simply transforms sound into an input spectral profile. The auditory nervous system then treats this profile the same way the visual system treats its retinal image. The notion of a Fourier transformation of a spectrum is common in engineering

speech applications, and is known as convolutional homomorphic processing. It involves computing Fourier-like coefficients of the spectral profile, known as cepstral coefficients [54]. While quite different in details, the cepstral coefficients encode roughly similar types of information about the shape of the spectrum as the ripple transform, and have been found especially useful in automatic speech recognition systems.

Finally, the correspondence between auditory ripple analysis and visual spatial frequency analysis goes deeper than a mere analogy. As evidence to this claim, consider the closely matched values of the filter parameters and detection thresholds measured in the visual system, e.g., roughly constant Q and 1 octave wide filters, with constant 6^{o} phase sensitivity increasing at higher ripple frequencies (Table 6.1, and Figs. 6.11 and 8.3 in [38]). These remarkable equivalences may simply reflect modality-independent limitations imposed by identically structured sensory areas in the central nervous system. For instance, the resolution of the analysis filters may simply be dictated by developmental rules limiting the minimum divergence or convergence of dendritic fields along the sensory epithelium (be it auditory or visual). Clearly, exploring further equivalences (and differences) between similarly defined psychophysical measures, e.g., **fdl**'s for ripples versus gratings (which apparently have not been reported in the literature), would shed considerable light on the underlying functional organization of both systems.

Chapter 4

Predictions of the ripple analysis model for various input profiles

In the first part of this chapter, the ripple analysis model developed in Chapter 3, is used to account for the perceptual thresholds measured in a wide range of profile analysis experiments. Prediction thresholds are computed from the filter outputs and compared with thresholds measured perceptually.

First, the δBWF/BWF thresholds of spectral peaks are considered. Then, detection thresholds for three profiles against a flat standard: the alternating, the step, and the single component increment profiles; are computed and compared to the perceptual thresholds reported in [35]. Finally, two pedestal-type experiments are compared. In Sec. 4.4 we discuss the results of the ripple phase sensitivity for **pdl** and δSF experiments.

Based on these results a prediction model applicable to a wide range of stimuli is proposed in Sec. 4.5.

4.1 Predicting the δBWF/BWF thresholds for peak profiles

As explained earlier in Sec. 3.2.1, a change in the BWF of a spectral peak is equivalent to a dilation of the peak, which in turn causes its ripple transform $r(\cdot)$ to translate laterally. This is similar to a change in the frequency of a ripple profile, i.e., an **fdl**-type experiment. In Figs. 4.1 the ripple profile and corresponding $r(\cdot)$ for peaks with BWF = 0.1 and 0.4 are shown (solid lines) together with their 20% (or $\alpha \approx 0.8$) dilated versions (dashed lines). Since the location of the maximal lowpass slopes for these profiles fall within the range $0.7 - 6$ cycle/octave, then the model predicts approximately constant dilation thresholds of 0.8 (20%), or $\Delta(\Omega_o) \approx 1/4$ octave shift. Note that if $r(\cdot)$ in $0.7 - 6$ cycle/octave region is well above $K(\cdot)$, then thresholds will be determined by $r(\cdot)$ shifts in this (most sensitive) range. These estimates compare well with the δBWF/BWF thresholds measured at approximately 0.22 (Sec. 3.2.1). Finally, the model also predicts that δBWF/BWF thresholds are independent of SF (Fig. 2.7(a)), since the symmetry (SF) of a peak does not affect the magnitude of its ripple transform.

Two conclusions can be drawn from the above analysis. First, the constancy of the δBWF/BWF thresholds in the peak profiles is directly related to the constancy of the **fdl**'s in the $0.7-6$ cycle/octave range. If broader or narrower peaks are used such that their $r(\cdot)$ lie outside of this range, then the model predicts that δBWF/BWF thresholds would also rise reflecting the **fdl** increase. The second conclusion is that the above trends in dilation thresholds hold **regardless of the exact details of the profile shape** since it is the shift in the profile

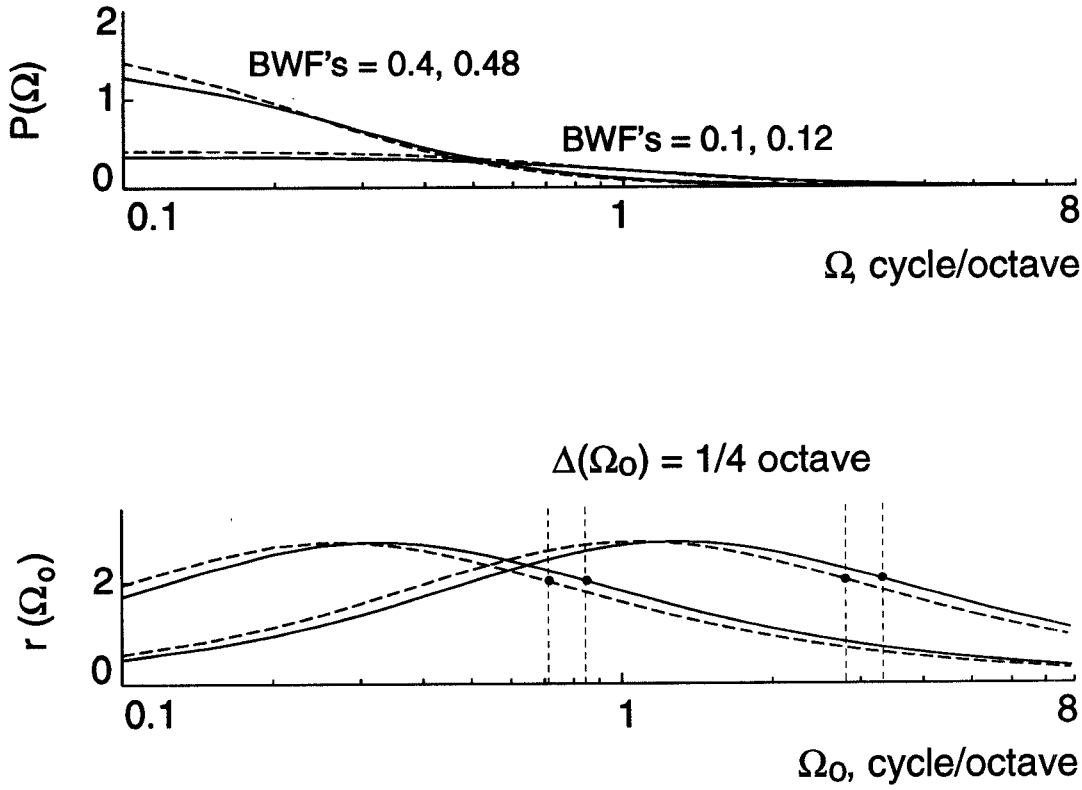


Figure 4.1: (a) Ripple spectra of symmetric peak profiles with BWF's = 0.1 and 0.4 (solid lines), and the corresponding profiles at δBWF/BWF = 20% perceptual thresholds (dashed lines). (b) The ripple transforms of the peak profiles in (a). The locations of the steepest lowpass edges are, for both cases, 1/4 octaves apart.

that is being detected. Clearly, this only holds if $r(\cdot)$ is well above the perceptual threshold $K(\cdot)$.

4.2 Predicting detection thresholds for profiles with flat standards

4.2.1 Alternating profile

The alternating profile [24] consists of 21 uniformly distributed components (0.2–5 kHz) that alternate above and below a flat (unit) base (Figs. 4.2).

Thresholds for detection of such a profile are reported at –21.7 dB ($= 20\log(\delta a) = 20\log 0.08$), where δa is the amplitude of an alternating component relative to the unit base. Such a profile can be considered approximately a ripple at the highest possible frequency representable by this complex, i.e., at 10 cycles per 4.64 octaves, or approximately 2.15 cycle/octave. The amplitude δa of the just detectable ripple at this frequency can be predicted from the **idl** curve as $\delta a = 2K(2.15 \text{ cycle/octave}) \approx 0.1$ (or –20.0 dB), which is close to the measured amplitude (Fig. 4.2(c)).

4.2.2 Step profile

The task in this experiment is to detect the presence of a (linear) step in a 21 component flat standard [24] (Fig. 4.3(a)).

For a step-up that is centralized (located at 1 kHz), threshold is reached at –23.1 dB ($= 20\log \delta a = 20\log 0.07$), where δa is the height of the step (relative

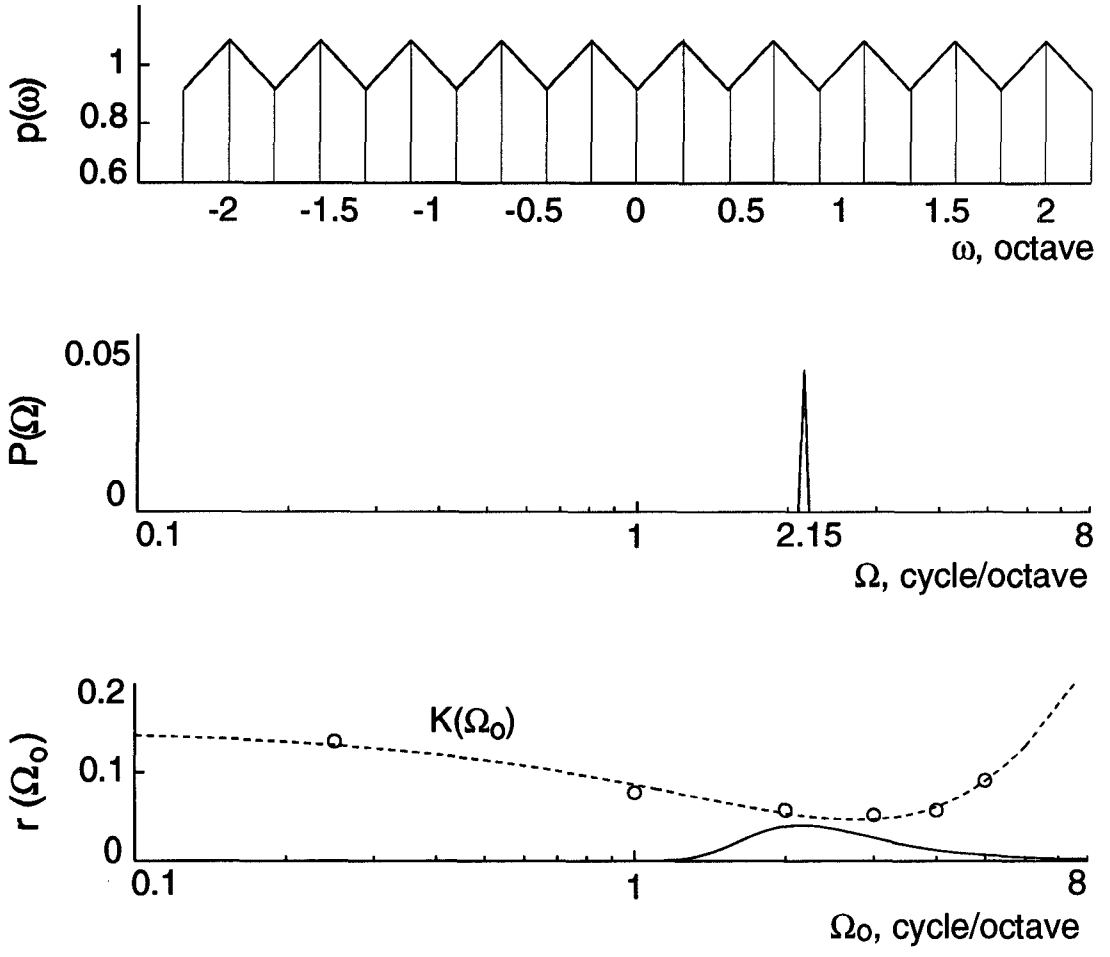


Figure 4.2: (a) The alternating profile at threshold amplitude (0.08, or –21.7 dB). (b) Ripple spectrum of the alternating profile in (a). (c) Ripple transform of the ripple spectrum in (b). The detection threshold $K(\Omega_o)$ in (c) is reached near 2.2 cycle/octave.

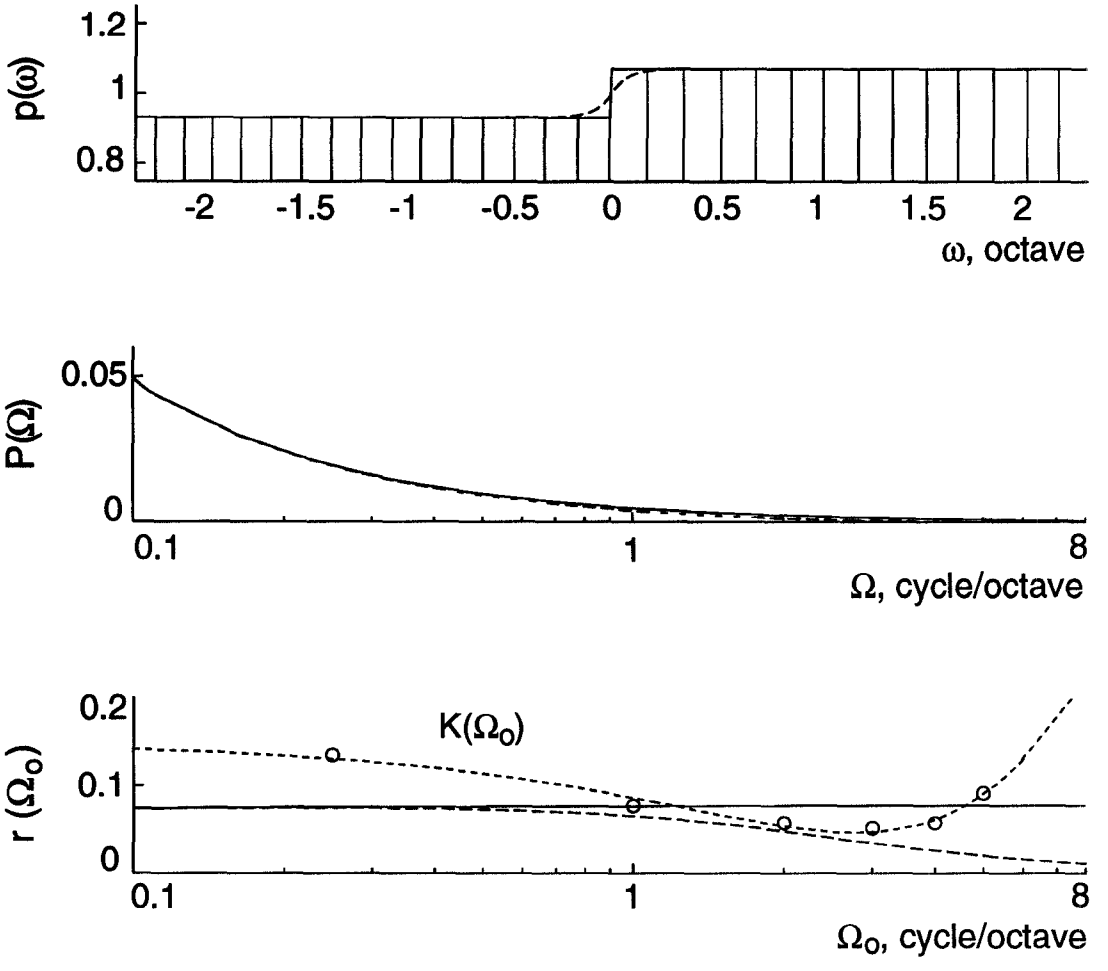


Figure 4.3: (a) Profile of a step function (solid line) at threshold amplitude (0.07, or -23.1 dB), and its smoothed version (dashed line). (The smoothed version is obtained by convolving the step with the narrow symmetric peak profile of BWF = 0.1 and $A_{max} = -30$ dB). The ripple spectra and ripple transforms are in (b) and (c), respectively. The ripple transform of the profile is above the detection threshold, $K(\Omega_o)$, between 1.3 - 4.4 cycle/octave, while its smoothed version reaches the threshold near 2 cycle/octave.

to the unit base). Figure 4.3(b) illustrates the ripple spectrum of this (idealized) profile ($P(\Omega)$). The corresponding model output $r(\cdot)$ is a constant because of the dilation relation of the filters (Fig. 4.3(c), solid line). Namely, for a constant Q factor filter ($\sigma(\Omega_o) = \sigma_{rel}\ \Omega_o$, and in particular $\sigma_{rel} = 0.3$), and for a step profile input (with magnitude of ripple spectrum $|P(\Omega)| = 1/|\Omega|$), filter outputs $r(\cdot)$ are independent of Ω_o, and for nonzero ripple frequencies (Ω larger than some $\Omega_1 > 0$):

$$r(\cdot) = 2 \int_{\Omega_1}^{\infty} \frac{1}{\Omega} e^{-\frac{(\Omega-\Omega_o)^2}{2\sigma^2(\Omega_o)}} d\Omega = 2 \int_{\Omega_1}^{\infty} \frac{1}{\Omega} e^{-\frac{(1-\Omega/\Omega_o)^2}{2\sigma_{rel}^2}} d\Omega = 2 \int_{\Omega_1}^{\infty} \frac{1}{\Omega'} e^{-\frac{(1-\Omega')^2}{2\sigma_{rel}^2}} d\Omega',$$

where, $\Omega' = \Omega/\Omega_o$, and $r(\cdot)$ is a function of σ_{rel} only. The predicted threshold is smaller than measured (0.05 versus 0.07, or –26 dB versus –23.1 dB). However, a more realistic representation of the step is with a gradual (or ramped) transition because of cochlear filter smoothing (dashed lines in Figs. 4.3). The smoothing of the ideal profile lowers the $P(\Omega)$ (Fig. 4.3(b)), and the corresponding $r(\cdot)$ is more lowpass filtered (Fig. 4.3(c)) and just detectable near 2 cycle/octave, or at –24.8 dB.

Since the phase of the ripple spectrum does not play a role here, predicted thresholds remain the same for the reversed (step-down) profile, as is indeed measured. Finally, the simplified model cannot account for the rise in thresholds as the step is moved towards the edges of the spectrum [24]. It may be possible, however, to account for this trend by including the effects of the base edge in $p(\omega)$, and using the complete model (i.e., Eq. 3.3).

4.2.3 Single component increment profile

In this experiment, a single component in the profile is incremented relative to the base [23] (Fig. 4.4(a)). The threshold is approximately -20.1 dB ($= 20 \log \delta a = 20 \log 0.09$), where δa is the height of component relative to the (unit) base. In order to apply this profile to the ripple analysis model, it is assumed that the finite bandwidth of the cochlear filters broadens the impulse-like profile, making it appear as a narrow peak profile (e.g., with BWF = 0.1 and SF = 0) with same height as before (= 0.09). Figure 4.4(c) illustrates that for such a peak the corresponding output $r(\cdot)$ reaches perceptual threshold $K(\Omega_o)$ near $\Omega_o = 2.3$ cycle/octave, or at -25.5 dB. Note that approximating the increment by a peak with slightly different BWF's causes correspondingly small shifts in the broad $r(\cdot)$, without affecting the estimated thresholds significantly.

4.3 Comparing detection thresholds for two pedestal experiments

A clearly audible profile serves as the standard (the pedestal) in these experiments. The profile is then increased in amplitude until the change is detected. Results from two such profile experiments are compared here: The first is the control experiment C reported in Sec. 2.6.2 (Fig. 2.13(a)) in which the pedestal is a peak profile. The second employs a single component pedestal (Fig. 5.4 in [23]). In both cases, the pedestal profiles produce $r(\Omega_o)$ outputs that are far above the perceptual threshold curve $K(\Omega_o)$. Threshold is defined here as the relative change (i.e., in dB or percentage) of the model outputs $r(\cdot)$ needed for

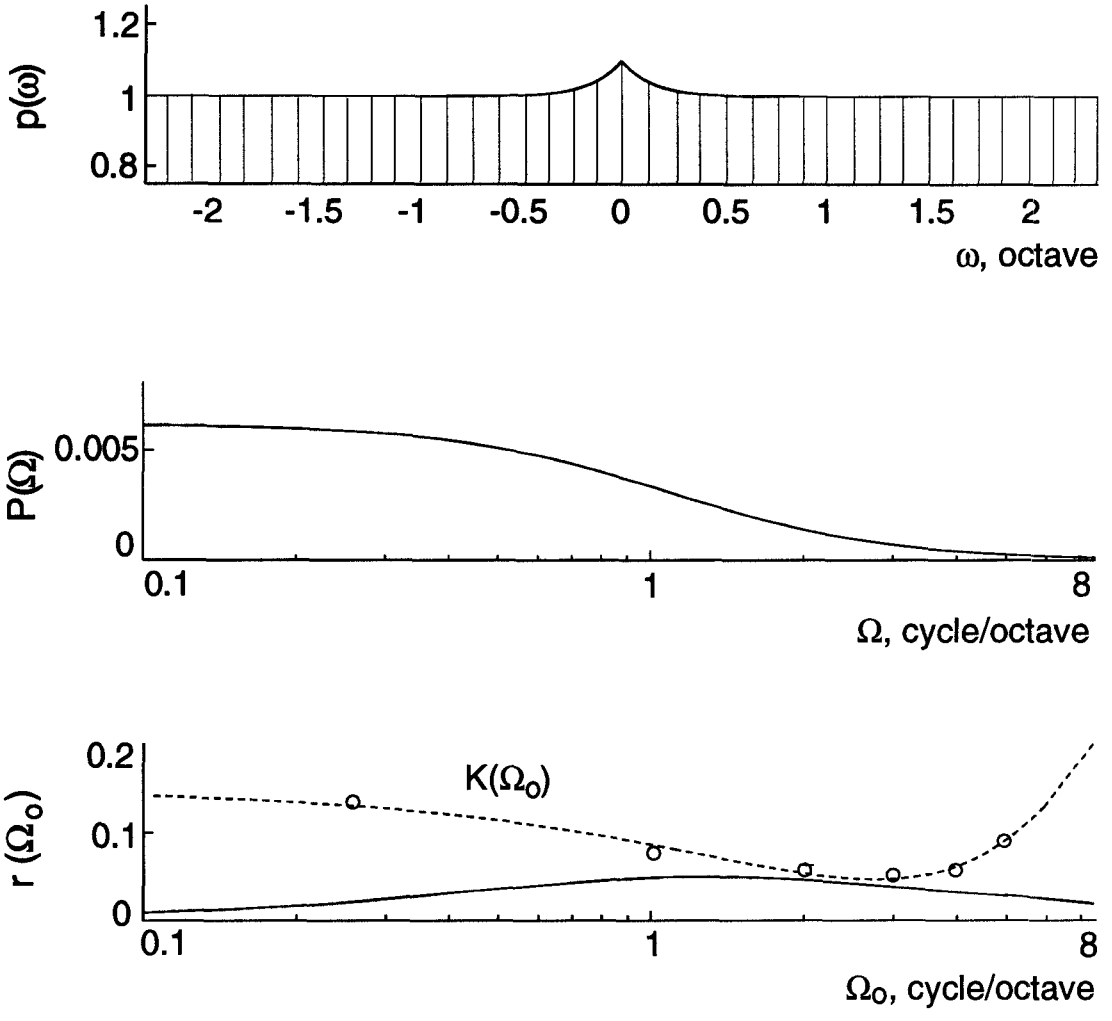


Figure 4.4: (a) Profile of a single increment on a flat base is at its perceptual threshold (0.09, or –20.1 dB). The ripple spectrum and ripple transform are shown in (b) and (c), respectively. The single increment is approximated with the BWF = 0.1 symmetric peak of –20.1 dB amplitude (see text). The detection threshold is reached approximately at 2.3 cycle/octave.

detection. Because of the linearity of the model, this relative change is directly related to the relative change in the input profiles, and hence thresholds can be directly compared without reference to the model.

Threshold for the narrowest symmetric peak profile (BWF = 0.1 and SF = 0), expressed as a peak level difference between the signal and standard, is 2.8 dB (Fig. 2.13(b) and Table 2.3(c)). This is equivalently expressed as $20\log(\delta a/a) = 20\log 0.26 \approx -12$ dB where a is the pedestal (or peak) height and δa is the detectable increment. Comparable detection thresholds (≈ -14 dB) were obtained for a single component pedestal of roughly an equivalent height (16 dB) and a complex of similar density (Fig. 5.4. in [23]).

Therefore, the model predicts that regardless of the exact shape of the profile, pedestal experiments with "loud" standard profiles, and reasonably dense complexs should exhibit similar thresholds as computed above. What is not accounted for here is the dependence of thresholds on such parameters as the density of the complex and the height of the pedestal at low levels. Such variabilities presumably involve considering the effects of masking and adjusting detection strategies for $r(\cdot)$ near perceptual threshold levels.

4.4 Sensitivity to ripple phase shifts

So far, only the magnitude of the outputs of the ripple analysis model has been considered. The other major representational axis is the phase axis. Two sets of data are relevant to this issue: (1) **pdl** thresholds for single ripple stimuli, and (2) δSF thresholds for peak profiles.

Ripple phase is explicitly included in the model in Eqs. 3.2 and 3.3, and

Sec. 3.3.1, where each ripple filter is assumed to be selective both to a ripple frequency and phase (Ω_o and Φ_o). In all profiles discussed in the previous section, profile manipulations were such that the phase of its ripple spectrum was not affected, hence it was justifyably ignored in the analysis. In the following, no explicit computations of the response phase are carried out. Instead, we compare the phase thresholds obtained from two experiments, in which the magnitude of the ripple spectrum is held constant while its phase is changed.

The most direct measurement of the phase sensitivity is provided by the ripple-**pdl** measurements in Sec. 2.8 (Figs. 2.16 and Table 2.3(d)). Results from these experiments reveal that, for low ripple frequencies (< 2 cycle/octave) subjects detect a constant phase shift of approximately 6^o. The amount of positional shift (which occurs along the tonotopic axis) decreases with increasing frequencies of the ripple. Eventually, at some high ripple frequency (above about 2 cycle/octave), the subjects switch from a constant phase detection mode to a constant positional shift of the ripple peaks along the tonotopic axis. The constant positional shift can be estimated from the slope of the detection threshold curves for ripples > 2 cycle/octave (Fig. 2.16(a) and Table 2.3(d)) to be 3.8^ooctave, or approximately 0.73% in Hz (which corresponds to 0.011 octaves). Viewed in the context of the ripple analysis model, these data suggest that only the lower ripple frequency filters exhibit fine ripple phase selectivity. Thresholds also increase for low amplitude ripples.

Thresholds measured with single ripple stimuli are supported well by δSF measurements in peak profiles (Sec. 2.3 and Appendix C.2). As discussed earlier (Sec. 3.2.2), δSF changes in the peak profiles can be equivalently described as constant phase shifts in the ripple spectrum of the peaks. Thus, consistent

with single ripple data, subjects detect a 0.11 change in SF, which is equivalent to a 6^o phase shift, the same threshold obtained from lower frequency ripples. Furthermore, this threshold increases to 10^o for the narrowest peak (BWF = 0.1), reflecting its higher ripple frequency spectrum.

These results suggest a general prediction of the model: For arbitrary input profiles with strong components at lower ripple frequencies, subjects should exhibit similar ripple phase sensitivities as those observed for broader peaks and single lower frequency ripples. For instance skewing the profile of a vowel or a musical note by adding a constant phase shift to their ripple transforms would be detectable at approximately 6^o.

4.5 Predictions of the model for any arbitrary profile

The computations outlined in this paper served to illustrate the competence of the ripple analysis model in accounting for many previous profile analysis measurements. However, other models such as the maximum difference model [24] and the independent channel models can account for a significant portion of the same data [25]. It is, therefore, important to come up with specific tests that can distinguish these models. Two such tests follow from the fundamental predictions which emerge from the ripple analysis model:

Given any arbitrary spectral profile whose ripple transform ($r(\Omega_o)$) is large relative to the perceptual threshold $K(\Omega_o)$, then:

(1) the dilation threshold is constant, and is due to the most sen-

sitive ripple component in the transform (e.g., 20–30% if it contains ripples in the range 0.7–6 cycle/octave).

(2) the ripple phase threshold for the profile is constant at approximately 6^{o}, if $r(\Omega_o)$ contains at least some large low frequency ripples (< 2 cycle/octave).

Both of these predictions are unintuitive and hence their future confirmation reflects well on the model. Both tests, however, are not of the **idl**-type the maximum difference model was developed to predict. A more direct **idl**-type test might be to measure the detection threshold of a profile composed of several ripples that do not appreciably interact within the same filter (e.g., separated by more than an octave). For different phases of the ripples, the shape of the profile changes and so too the predicted thresholds of the maximum difference model. In contrast, the ripple analysis model predicts that detection thresholds are independent of the relative phases of the ripples. A specific example of such a profile is the square wave, which is composed of a large fundamental ripple and smaller odd harmonics. The maximum difference model predicts the threshold based on the amplitude of the square wave, whereas the ripple analysis model predicts it based on the amplitude of the fundamental (largest) ripple component. The two amplitudes differ by a small (hopefully measurable) factor of $4/\pi$.

Chapter 5

Conclusion and further research

5.1 Summary and conclusions

First part of this dissertation discusses the results of psychoacoustical experiments on human subjects (Chapter 1). The experiments were motivated by the recent findings in physiological experiments in the ferret primary auditory cortex [9]. These indicate the existence of a systematic variation of inhibitory response patterns along the isofrequency plane. Responses to spectrally shaped noise bursts, show that neurons with inhibition to frequencies below (above) their best frequency (BF) respond optimally to noise bursts with least spectral energy below (above) BF.

Psychoacoustical measurements were carried out in order to explore the perceptual implications of these findings. The complex stimuli used were multi-component spectral peaks of different symmetries (SF) and bandwidths (BWF). Sensitivity to changes in spectral peak shape was tested for a number of different conditions: various starting peak shapes, different peak amplitudes, spectral densities, and peak locations. The effect of profile frequency randomization is

also discussed.

The main results indicate that the detection of SF and BWF changes is largely independent of the initial spectral shape (Sec. 2.9). Moreover, the detection threshold for simultaneous changes in SF and BWF is the same as for the pure change in SF (Sec. 2.5). More generally, it is conjectured that for an arbitrary spectral profile, these two manipulations perceptually correspond to changes in the phase and the magnitude of the Fourier transform of the profile (Sec. 2.7). As such, the human auditory system independently detects changes in the SF and BWF of the peak profiles (Sec. 3.2.2). Moreover, the experiments in which we changed both the SF and BWF, indicate the detection occurs when the detection os reached in ne of the two cues. The thresholds obtained for **pdl**-experiments with single ripple stimuli are similar in value (and trends) to the thresholds for changes in the symmetry factor (Sec. 2.8).

Within this general framework, a model of profile analysis is proposed. The first stage of the model presents the transfer function of the auditory cortex as a bank of bandpass filters, each tuned to a particular ripple frequency and ripple phase. The filters convert the input spectral profile into its corresponding ripple transform, basically performing a windowed Fourier transformation (Sec. 3.3).

The second part is the detection model, which operates on the magnitude or phase of the ripple transform (Sec. 3.3.2). Two different types of detection models are derived for two types of profile analysis experiments. In one, where the task is to detect the existence of the profile against a flat standard, the ripple transform of the profile is compared to detection thresholds obtained from the modulation transfer function, i.e. from the **idl**-experiments (with single ripple stimuli). The model accounts well for the perceptual thresholds in all three

profiles considered: with (centered) step, (centered) single component increment, and the alternating profiles (Sec. 4.2). In order to account for the increase in perceptual thresholds as the step is moved towards the edges of the spectrum, a complete model (possibly with the effects of the base edges in the input profile) should be used.

Another detection model is defined for experiments with non-flat standards, for which the task is to detect a change in some parameter of an audible profile. Depending on the parameter in question, the change in the ripple transform is compared to detection thresholds obtained, either from **fdl**-experiments for δBWF test (Sec. 4.1), or from **pdl**-experiments for δSF test (Sec. 4.4). The model predicts well the perceptual thresholds in these tests.

For pedestal-type experiments (e.g., with increments on a single component pedestal or on a peak profile pedestal) the detection thresholds are defined as the relative change in the input profile (i.e., without computing the ripple transform) (Sec. 4.3).

Finally, two fundamental predictions are hypothesized for any arbitrary spectral profile whose ripple transform is large relative to the perceptual threshold (4.5). They are:

(1) the dilation threshold is constant, and is due to the most sensitive ripple component in the transform (e.g., 20–30% if it contains ripples in the range 0.7–6 cycle/octave).

(2) the ripple phase threshold for the profile is constant at approximately 6^o, if $r(\Omega_o)$ contains at least some large low frequency ripples (< 2 cycle/octave).

5.2 Further research

It is of the immediate importance to this work to verify the predictions of the two hypotheses from above, for any arbitrary profile. This requires extensive experimental work since it involves testing subject's sensitivity to dilation or phase addition manipulations of several profiles (e.g., vowels) (Secs. 4.4 and 3.5). Preliminary results with double peaks in which the SF or BWF was changed in either one of them indicate similar threshold trends as observed for single peaks. To complete these studies, measurements need to be collected with double peak spectra in which the whole spectrum is dilated or tilted. Next would be to test the constancy of **pdl**-thresholds for various starting phases in two or more ripples, as outlined in Sec. 3.5.

The fundamental question, however, is how simultaneous variations of both magnitude and phase are combined in the perception of the spectrum. Preliminary results based on analogous experiments with single peaks (Sec. 2.5), indicate that the detection occurs when the threshold is reached independently in either one of the two parameters. Work is currently under way to design such tests with single ripples, in which both the ripple frequency and phase are varied simultaneously.

Based on new neurophysiological evidence it will be possible to modify the filters and detection criteria to reflect in a greater manner their biological nature (Secs. 3.3.2 and 3.4). This in turn might help to extend the applicability of the model to account for such perceptual dependencies as dependence on amplitude and spectral density.

The ultimate goal, however, is to define a perceptual metric which would be

used to compare any arbitrary profiles. Such a metric would be based on the three-dimensional representation of the profile at the auditory cortex level (i.e., along the frequency, ripple-frequency, and ripple-phase axes).

Appendix A

Detection thresholds measured in root-mean-square units

We have collected in this Appendix all rms–thresholds obtained in the peak SF and BWF change detection tasks. The rms–threshold is derived from SF and BWF changes based on the assumptions that all changes in the spectral components (which are largest around the peak) contribute equally to the detection process. The rms–threshold is defined as $20 \log \sqrt{\sum_{i=1}^{n} (\Delta p_i / p_i)^2}$, where Δp_i is the change in the amplitude of the i^{th} component at threshold, p_i is the amplitude of the i^{th} component in the standard, and n is the number of components.

This measure is closely related to that used in most profile analysis experiments previously reported. Specifically, for the case of a flat standard of amplitude p, the measure usually used is: $20 \log(\frac{\text{rms}_{signal}}{p})$. This measure converts to our units if we add a constant $10 \log n$, which accounts for the number of signal components n. Thus, the threshold for the single increment detection task ($n = 1$) is the same under both units. For other commonly studied detection tasks in profile analysis: 21 components step function at 1 kHz, alternating am-

plitudes spectrum, and ripple signals; the threshold values reported are: –23.33 dB, –23.07 dB, and –21.58 dB, respectively (see [35, 37]). Computed in our unit, the thresholds are: –10.11 dB, –10.06 dB, and –8.36 dB, respectively.

A.1 Detection of changes in spectral peak symmetry

Stimuli and testing conditions are described in Secs. 2.2.2 and 2.3.

(i) Dependence on symmetry and bandwidth factors of the standard

Threshold trends are as described earlier in Sec. 2.3.1 in that detection of a change in peak symmetry is independent of the peak shape of the standard (Figs. A.1). The average detection threshold was $\approx$ –8.5 dB. This value is comparable to that measured in other profile analysis experiments.

(ii) Dependence on peak amplitudes

Data are averaged and presented as in Sec. 2.3.1. The same trends established earlier for the 15 dB case hold also for the other two levels. However, unlike the δSF thresholds (Figs. 2.4), the average rms–threshold monotonically increases with peak levels (Figs. A.2). The increase is small, being of the order 0.25 dB per 1 dB change in peak level.

(iii) Dependence on spectral density

rms–Threshold increases with increasing spectral density, from 41 to 11 components tests (Figs. A.3). Note that the 11 component thresholds are lower than those for the 41 component signal. Some of this difference is probably due to masking effects among the 41 closely spaced components ([24]). Another possi-

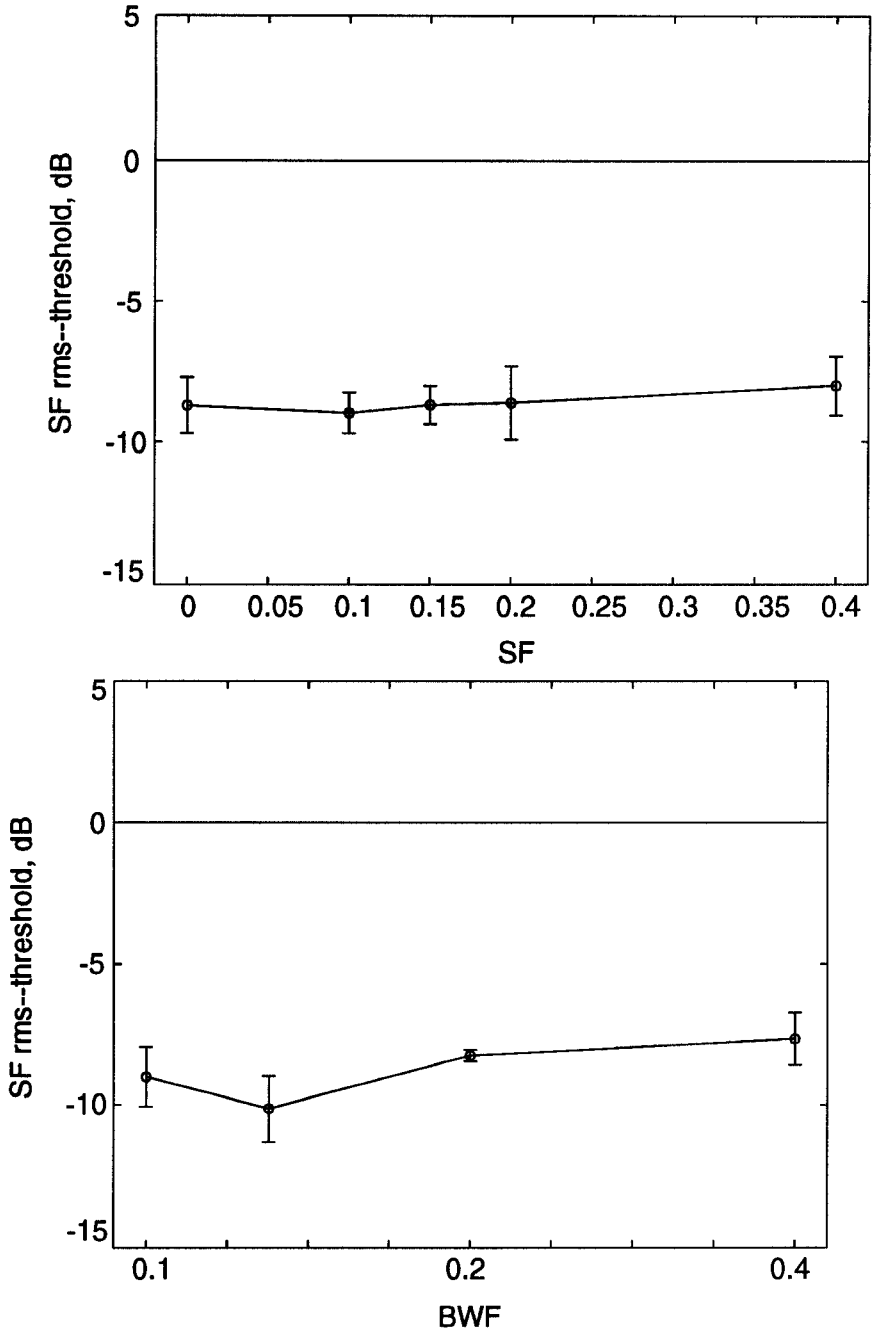


Figure A.1: Symmetry change detection rms–thresholds for 41 component complex and 15 dB peak amplitude, averaged over five subjects and: four BWF's in (a), and five SF's and (b). The rms–threshold is defined as: 20 $\log\sqrt{\sum_{i=1}^{n}(\Delta p_i/p_i)^2}$, where Δp_i is the change in the amplitude of the i^{th} component at threshold, p_i is the amplitude of the i^{th} component in the standard, and $n = 41$. rms–Threshold is independent of SF and BWF. The error bars are the standard deviations of the means.

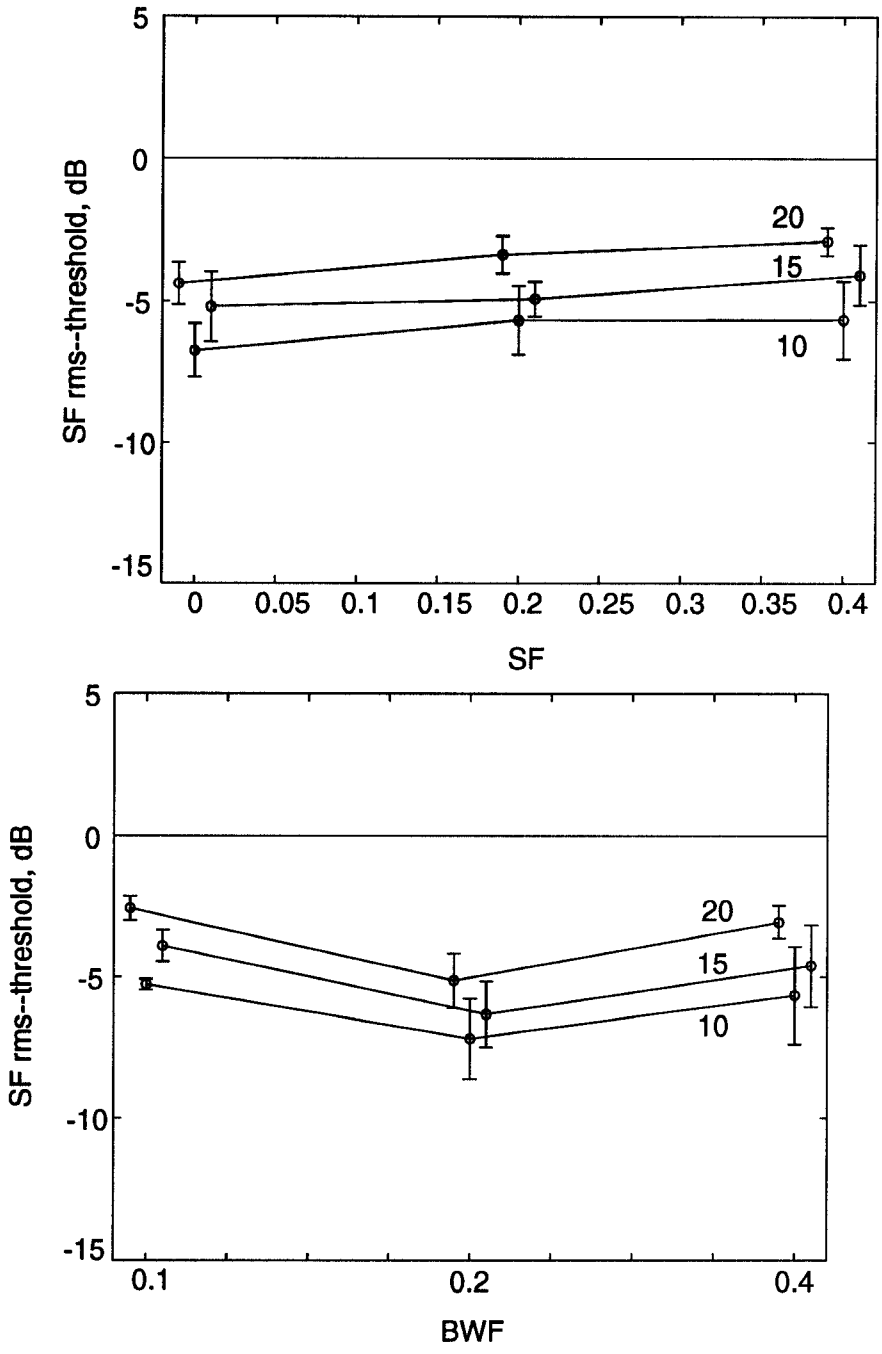


Figure A.2: Symmetry change detection rms–thresholds for 41 component complex and 3 peak amplitudes: 10 dB, 15 dB, and 20 dB, relative to baseline. The data are averages of three subjects and: three BWF's in (a), and three SF's and (b). The values along the ordinates are defined as in Fig. A.1. Points are slightly offset for clarity reason.

ble source is the large frequency spacing among the 11 components which may cause the task to be perceived as amplitude changes in several smaller peaks rather than the detection of symmetry in a single (broader) peak.

A.2 Detection of changes in spectral peak bandwidth factor

Stimuli and tests are described in Secs. 2.2.2 and 2.4 of the text.

(i) Dependence on symmetry and bandwidth factors of the standard

The data are averaged and presented as described in Sec. 2.4.1. Detection thresholds are independent of SF for all BWF's. However, they increase monotonically with standard's BWF. This trend is more clearly depicted in Fig. A.4(b), where the rms–thresholds are averaged over the five SF's and then plotted against BWF. The functional form of this dependence, which best approximates the experimental data points in the least square error sense, is:

$$\text{threshold(dB)} = -6.85\text{(dB)} + 3.3\text{(dB/octave)}\ \log_2(10\ \text{BWF})\ \text{(octave)}.$$

(ii) Dependence on peak amplitudes

Data are averaged and presented as in Sec. 2.4.1. Mean rms–thresholds tend to increase with peak level in a manner similar to that seen earlier in the SF change detection task.

(iii) Dependence on spectral density

The rms–thresholds increase monotonically with BWF, and with spectral density (Fig. A.6).

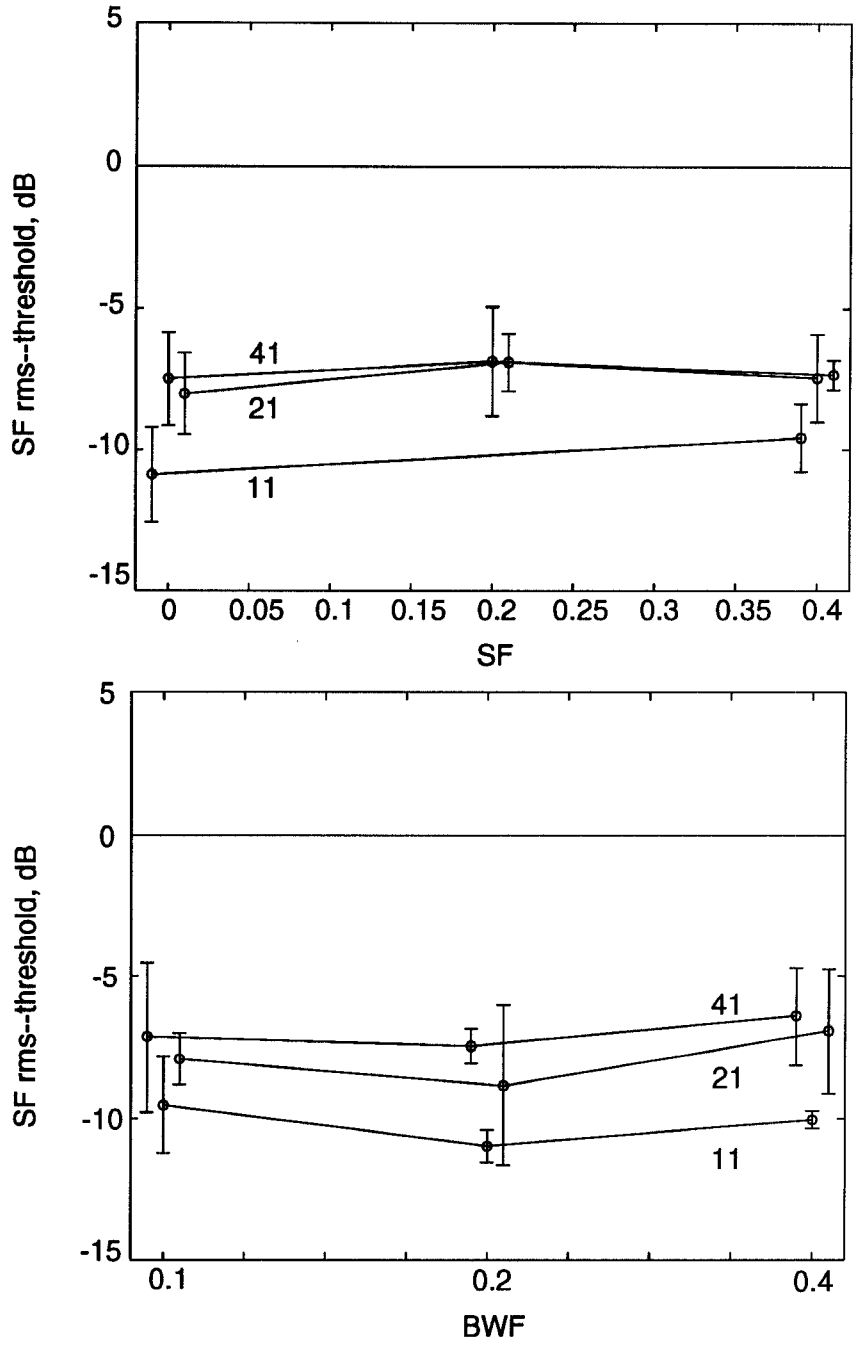


Figure A.3: Symmetry change detection thresholds for 41, 21, and 11 component complexes, averaged over four subjects and three BWF's (a) and three SF's (b). rms–Threshold increases with increasing spectral density, from 41 to 11 component tests.

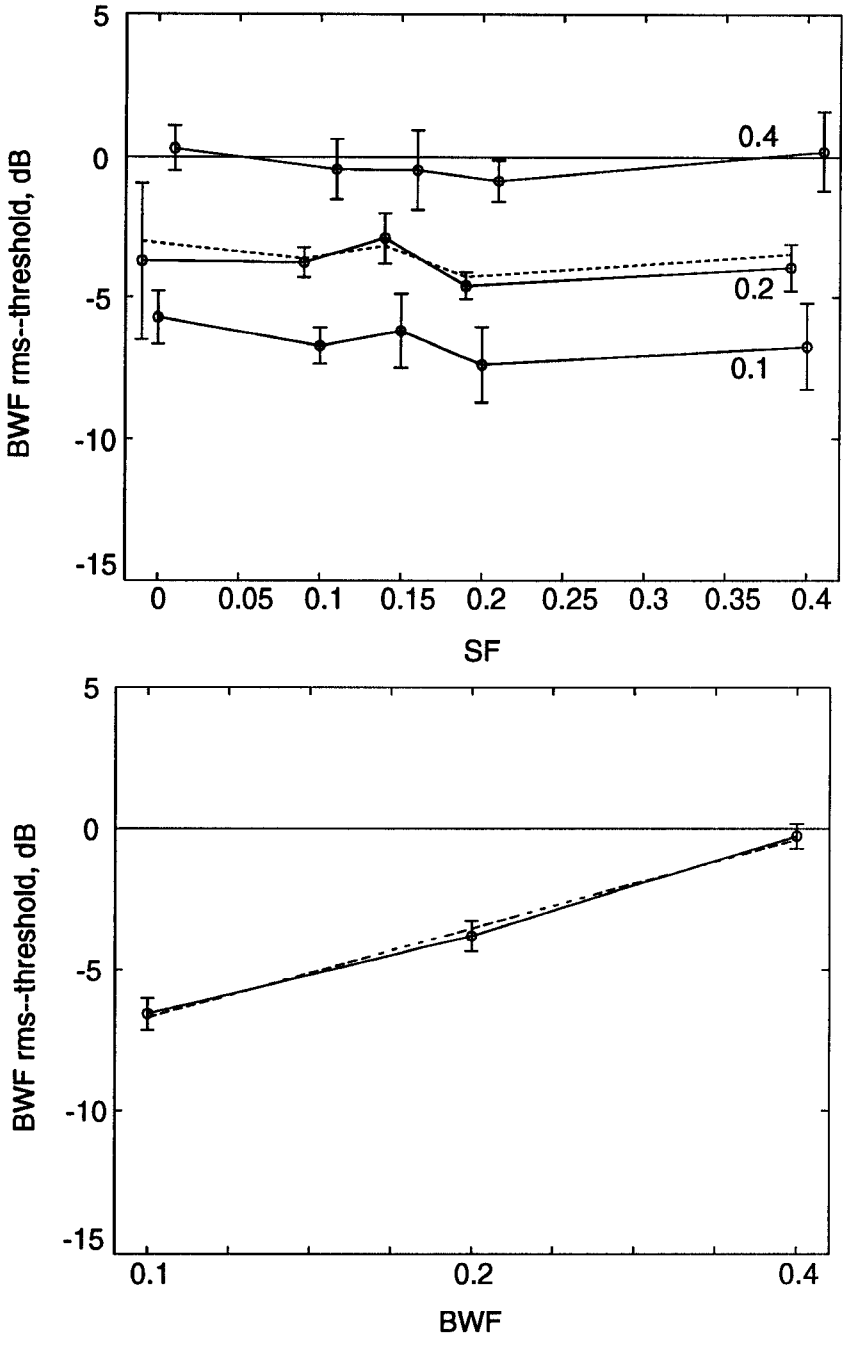


Figure A.4: (a) Bandwidth change detection rms–threshold for 41 frequency components, 15 dB peak level, and three BWF's: 0.1, 0.2, and 0.4, averaged for three listeners. Thresholds monotonically increase with BWF, and the form of this dependence, averaged over five SF's, is depicted in (b). The dotted line in (b) is the least square error linear approximation of this dependences: threshold (dB) = -6.85 + 3.3 $\log_2$ (10 BWF). Data are slightly offset for clarity.

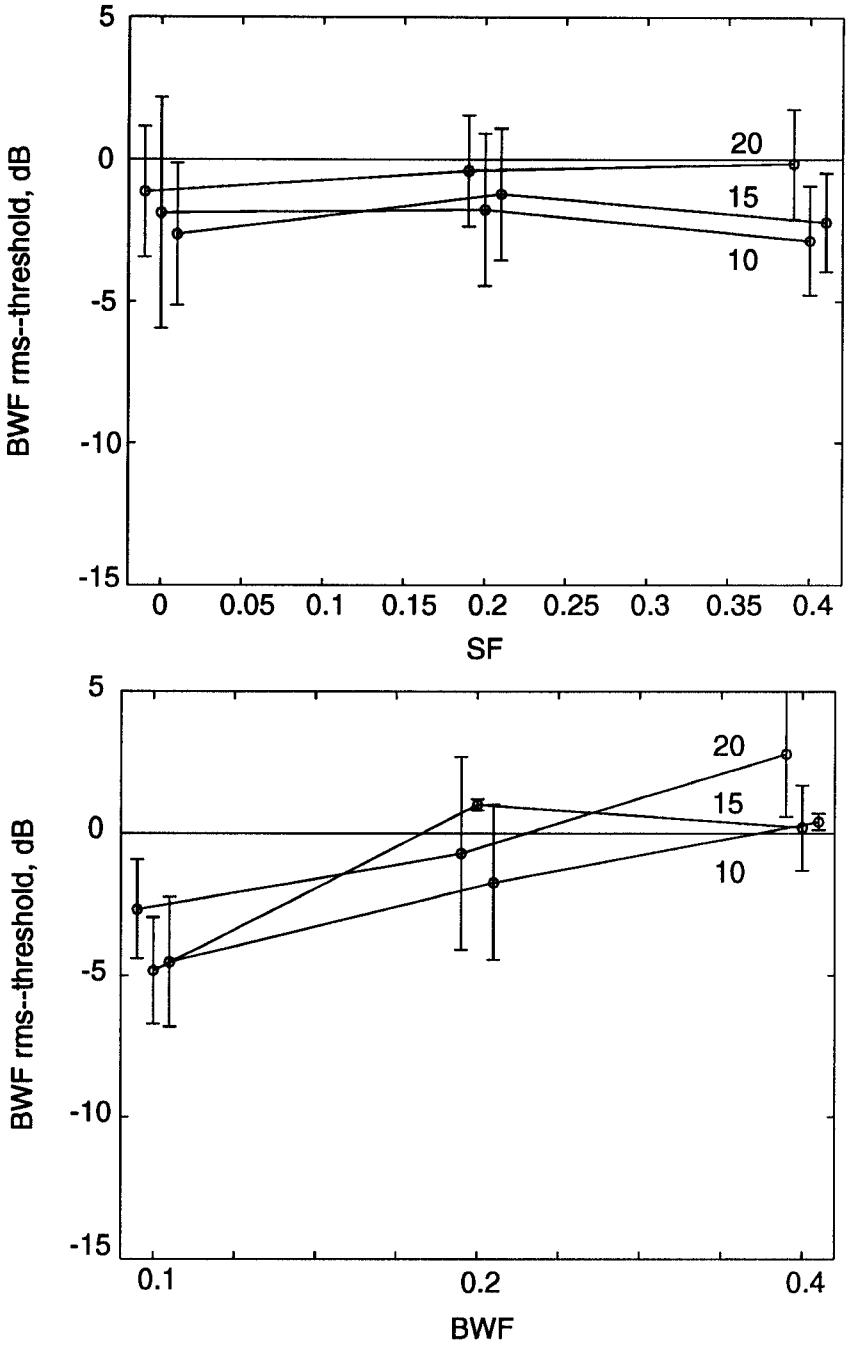


Figure A.5: Bandwidth change detection rms–thresholds for 41 component complex and 3 peak amplitudes: 10 dB, 15 dB, and 20 dB. The thresholds are averages of three subjects and: three BWF's in (a), and three SF's in (b). Points are offset for clarity.

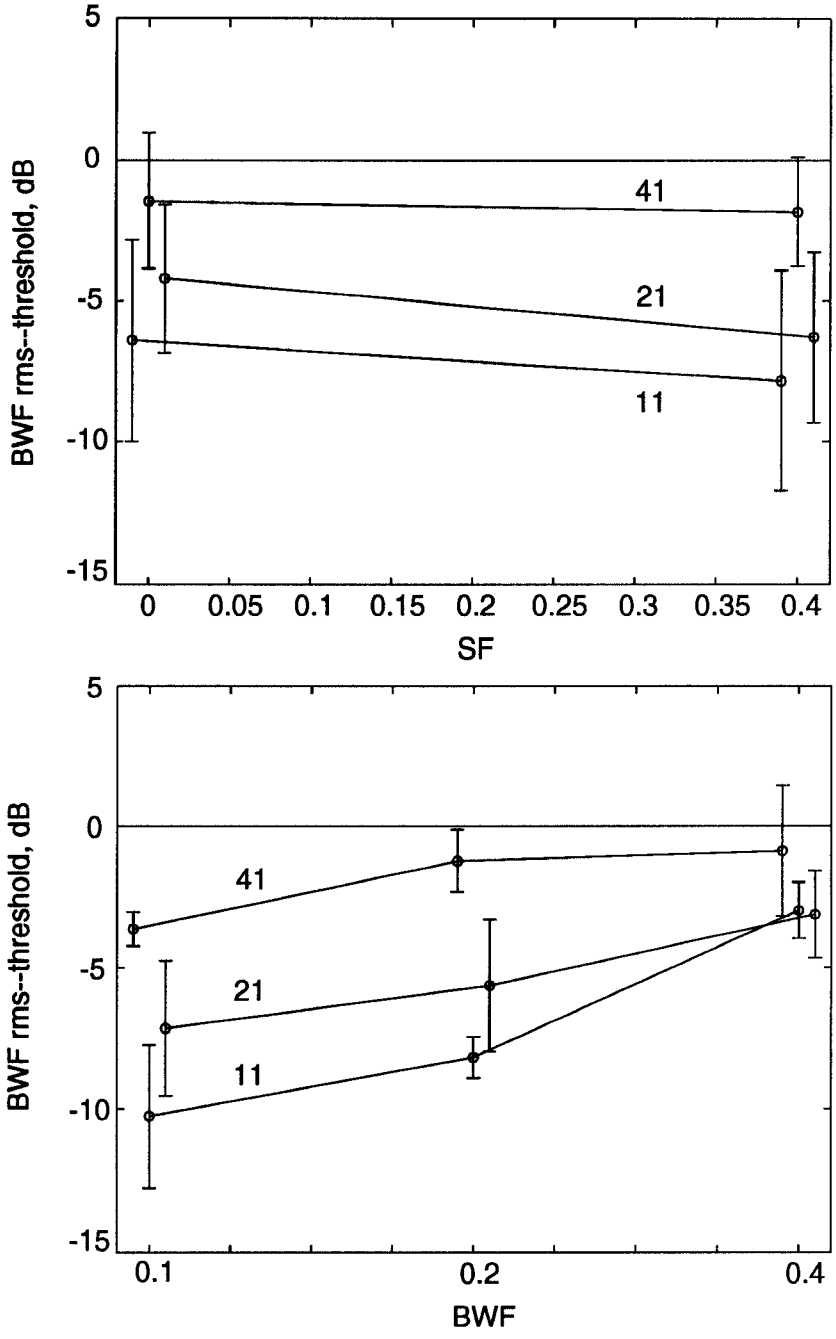


Figure A.6: Bandwidth change detection thresholds for 41, 21, and 11 component complexes, averaged over three subjects and three BWF's (a) and two SF's (b). rms–Thresholds are in general higher for 41 than for 21 and 11 component cases.

Appendix B

Brief review of the Ewaif model

An analytic function $m(t)$ with envelope $e(t)$ and phase $\phi(t)$ is related to a real waveform $p(t)$ with Hilbert transform $\hat{p}(t) = p(t) * \frac{1}{\pi t}$ ([54, 55, 56]), as:

$$m(t) = e(t)\ e^{\jmath\phi(t)} = p(t) + \jmath\hat{p}(t),$$

where

$$e(t) = \sqrt{p^2(t) + \hat{p}^2(t)},$$

and

$$\phi(t) = \arctan\frac{\hat{p}(t)}{p(t)}.$$

The equivalent pitch of a complex sound is defined by the Ewaif model (Envelope Weighted Average Instantaneous Frequency) [57] as:

$$\text{Ewaif} = \frac{\int_0^T e(t)\ \text{inst}f(t)dt}{\int_0^T e(t)\ dt},$$

where T is a stimulus duration, and $\text{inst}f(t)$ is an instantaneous frequency of $p(t)$, defined as $\text{inst}f(t) = \frac{1}{2\pi}\frac{d\phi}{dt}$.

For our n component stimulus:

$$p(t) = \sum_{i=1}^{i=n} p_i \cos(2\pi f_i + \varphi_i),$$

the instantaneous envelope and frequency are:

$$e^2(t) = \sum_{i=1}^{n} p_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} p_i p_j \cos(2\pi(f_i - f_j)t + \varphi_i - \varphi_j)$$

and

$$\text{inst}f(t) = \frac{1}{e^2(t)} \sum_{i=1}^{n} p_i^2 \, f_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} p_i \, p_j(f_i + f_j) \, \cos(2\pi(f_i - f_j)t + \varphi_i - \varphi_j).$$

Appendix C

On the transform of the peak profile

C.1 Fourier transform of the peak profile

The Fourier transform of the peak profile is for $\Omega > 0$ cycle/octave, computed as:

$$P(\Omega) = |P(\Omega)|e^{j\theta} = \frac{a_{max}}{b}\frac{20\ \mathrm{BWF}}{3\ln 10\ M}\cdot \frac{1}{(1 + \jmath\ 2\pi\Omega\ 20/(3\ln 10)\ \mathrm{SF}\ \mathrm{BWF} + (\pi\Omega\ 20/(3\ln 10)\ \mathrm{BWF})^2(1-\mathrm{SF}^2))}, \tag{C.1}$$

and $P(0) = 1 + \frac{a_{max}}{b}\frac{20\ \mathrm{BWF}}{3\ln 10\ M}$.

Define the gain as $G = \frac{a_{max}}{b}\frac{20\ \mathrm{BWF}}{3\ln 10\ M}$, and let $c = 20\pi/(3\ln 10) = 9.1$. The magnitude and phase of $P(\Omega)$ are:

$$|P(\Omega)| = \frac{G}{\sqrt{(1 + \Omega^2\ (\mathrm{c\ BWF})^2(1-\mathrm{SF}^2))^2 + \Omega^2(2\mathrm{c\ SF\ BWF})^2}}$$

$$\theta(\Omega) = -\arctan\frac{2\ \Omega\ \mathrm{c\ BWF\ SF}}{1 + \Omega^2(\mathrm{c\ BWF})^2(1-\mathrm{SF}^2)}.$$

The δBWF task corresponds to shifting the ripple spectrum and ripple transform of the profile along the log Ω_o axis (Eq. 3.4). The ripple transform remains unchanged along the phase axis, as explained in Sec. 3.3.1.

C.2 Adding a constant phase to the Fourier transform of the profile

Consider the profile $p(\omega)$ whose Fourier transform is $P(\Omega)$:

$$p(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\Omega) e^{j2\pi\Omega\omega} d(2\pi\Omega).$$

Adding a constant phase angle θ_o to all the transform components changes the profile to:

$$p_{\theta_o}(\omega) = \int_{-\infty}^{0} P(\Omega) e^{j\theta_o} e^{j2\pi\Omega\omega} d\Omega + \int_{0}^{\infty} P(\Omega) e^{-j\theta_o} e^{j2\pi\Omega\omega} d\Omega,$$

where the integral is split to emphasize that the phase function (added to negative frequencies and subtracted from positive frequencies) must be odd as a function of Ω in order for p_{θ_o} to remain real. This expression can be simplified further by substituting $e^{\pm j\theta_o} = \cos(\theta_o) \pm j \sin(\theta_o)$, and collecting terms:

$$p_{\theta_o}(\omega) = \cos(\theta_o) \int_{-\infty}^{\infty} P(\Omega) e^{j2\pi\Omega\omega} d\Omega - \sin(\theta_o) \int_{-\infty}^{\infty} jP(\Omega) \cdot sign(\Omega) \cdot e^{j2\pi\Omega\omega} d\Omega.$$

Therefore,

$$p_{\theta_o}(\omega) = \cos(\theta_o) p(\omega) + \sin(\theta_o) \mathcal{H}(p(\omega)),$$

where $\mathcal{H}(p(\omega))$ is the so-called Hilbert transform of $p(\omega)$. A simpler expression can be used for the case of small θ_o ($\cos(\theta_o) \approx 1$ and $\sin(\theta_o) \approx \theta_o$):

$$p_{\theta_o}(\omega) = p(\omega) + \theta_o \mathcal{H}(p(\omega)).$$

This is the expression used in computing the profiles in Figs. 2.14 (c) and (d).

Bibliography

[1] B.R. Glasberg and B.C.J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138, 1990.

[2] David A. Hillier. *Auditory processing of sinusoidal spectral envelopes.* PhD thesis, Washington University, Sever Institute of Technology, 1991.

[3] R. Real and T. Imig. Tonotopic organization of auditory cortex in the cat. *Journal of Comp. Neurol.*, 192:265–291, 1980.

[4] M. Merzenich, P. Knight, and G. Roth. Representation of cochlea within primary auditory cortex in the cat. *Journal of Neurophysiology*, 28:231–249, 1975.

[5] M. Merzenich, H. Kaas, and G. L. Roth. Auditory cortex in the grey squirrel: tonotopic organization and architectonic fields. *Journal of Comp. Neurol.*, 166:387–402, 1976.

[6] J. Kelly, P. Judge, and D. Phillips. Representation of the cochlea in primary auditory cortex of the ferret *mustela putorius. Hearing Research*, 24:111–115, 1986.

[7] J. Middlebrooks, P. Dykes, and M. Merzenich. Binaural response-specific bands in primary auditory cortex (ai) of the cat: Topographical organization orthogonal to isofrequency contours. *Brain Res.*, 181:31–48, 1980.

[8] T. Imig and H. Adrian. Binaural columns in the primary field (ai) of cat auditory cortex. *Brain Res.*, 138:241–257, 1977.

[9] Shihab A. Shamma, James W. Fleshman, Philip W. Wiser, and Huib Versnel. Organization of response areas in ferret primary auditory cortex. *Journal of Neurophysiology*, 69(2):367–383, February 1993.

[10] C. E. Schreiner and J. Mendelson. Functional topography of cat primary auditory cortex: Distribution of integrated excitation. *Journal of Neurophysiology*, 64(5):1442–1459, November 1990.

[11] Sarah L. Pallas, Laurie S. Carman, and Mriganka Sur. Visual inputs and information processing in sensory cortex: An *in vivo* developmental study. In Frank H. Eeckman, editor, *Neural Systems: Analysis and modeling*, pages 167–178. Kluwer Academic Publishers, Boston, MA, 1993.

[12] P. F. Assmann and A. Q. Summerfield. Modeling the perception of concurrent vowels: vowels with the same fundamental frequency. *Journal of the Acoustical Society of America*, 85(1):327–338, January 1989.

[13] S. A. Shamma, S. Vranić, and P. Wiser. Spectral gradient columns in primary auditory cortex: physiological and psychoacoustical correlates. In *Advances in the Biosciences, Vol. 83.* 9th International Symposium on Hearing, 1992.

[14] S. Vranić-Sowers, H. Versnel, and S.A. Shamma. Single and double spectral peaks: Psychoacoustical and physiological results. 16th A.R.O. Midwinter Meeting, February 1993.

[15] F.W. Campbell and J.G. Robson. Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197:551–566, 1968.

[16] M. Levine. *Vision in Man and Machine.* McGraw-Hill, New York, 1985.

[17] S.A. Shamma, H. Versnel, and N.A. Kowalski. Organization of primary auditory cortex evident in responses to rippled complex sound stimuli. Neurosc. Meeting, November 1993.

[18] B.M. Calhoun and C.E. Schreiner. Spatial frequency filters in cat auditory cortex. Neurosc. Meeting, November 1993.

[19] David M. Green and J.A. Swets. *Signal Detection Theory and Psychophysics.* John Wiley and Sons, Inc., New York, NY, 1966.

[20] J.A. Swets. *Signal Detection and Recognition by Human Observers.* John Wiley and Sons, Inc., New York, NY, 1964.

[21] W. Levitt. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49:467–477, 1971.

[22] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of mathematical statistics*, 22:400–407, 1951.

[23] David M. Green. *Profile Analysis: Auditory Intensity Discrimination.* Oxford Press, New York, NY, 1988.

[24] L.R. Bernstein and D.M. Green. Detection of simple and complex changes of spectral shape. *Journal of the Acoustical Society of America*, 82(5):1587–1592, 1987.

[25] N.I. Durlach, L.D. Braida, and Y. Ito. Towards a model for discrimination of broadband signals. *Journal of the Acoustical Society of America*, 80(1):63–72, 1986.

[26] D.M. Green, C.R. Mason, and G. Kidd. Profile analysis: Critical bands and duration. *Journal of the Acoustical Society of America*, 75:1163–1167, 1984.

[27] L.R. Bernstein, V.M. Richards, and D.M. Green. *Auditory Processing of Complex Sounds: The detection of spectral shape change.* Lawrence Erlbaum Associates, Inc., New Jersey, 1987.

[28] G.Jr. Kidd, C.R. Mason, and D.M. Green. Auditory profile analysis of irregular sound spectra. *Journal of the Acoustical Society of America*, 79(4):1045–1053, 1986.

[29] D.H. Klatt. Prediction of perceived phonetic distance from critical-band spectra: A first step. In *Proc. ICASSP*, volume 2, pages 1278–1281, 1982.

[30] W.L. Hays. *Statistics.* Holt, Rinehart, and Winston, Inc., New York, NY, 1988.

[31] D.M. Green and C.R. Mason. Auditory profile analysis: Frequency, phase, and weber's law. *Journal of the Acoustical Society of America*, 77(3):1155–1161, 1985.

[32] L.L. Feth, H. O'Malley, and J. Jr. Ramsey. Pitch od unresolved two-component complex tones. *Journal of the Acoustical Society of America*, 72:1403–1412, 1982.

[33] L.J. Stover and L.L. Feth. Pitch od narrow-band signals. *Journal of the Acoustical Society of America*, 73:1701–1707, 1983.

[34] B.G. Berg, Q.T. Nguyen, and D.M. Green. Discrimination of narrow-band spectra. i: Spectral weights and pitch cues. *Journal of the Acoustical Society of America*, 92:1911–1918, 1992.

[35] V.M. Richards, Z.A. Onsan, and D.M. Green. Auditory profile analysis: Potential pitch cues. *Hearing Research*, 39:27–36, 1989.

[36] L.L. Feth and L.J. Stover. *Auditory Processing of Complex Sounds: Demodulation processes in auditory perception.* Lawrence Erlbaum Associates, Inc., New Jersey, 1987.

[37] David M. Green. 'Frequency' and the detection of spectral shape change. In B. J. C. Moore and R. D. Patterson, editors, *Auditory Frequency Selectivity*, pages 351–359. Plenum Press, 1986.

[38] Russell L. De Valois and Karen K. De Valois. *Spatial Vision.* Oxford Science Publications, New York, NY, 1990.

[39] T. Houtgast and T.M. van Veen. On the just-detectable modulation of the spectral envelope on a log-f scale. In *Journal of the Acoustical Society of America Suppl. 1*, volume 71, 1982.

[40] D. Gabor. Theory of communications. *J. Inst. Elec. Eng. (London)*, 93:429–457, 1946.

[41] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38, March 1992.

[42] L. R. Rabiner and R. W. Schafer. *Digital processing of speech signals.* Prentice-Hall Inc., 1978.

[43] R.L. De Valois, Albrecht, and Thorell. *Vision Research*, 22:545–559, 1982.

[44] R. Shannon, 1992. Personal correspondance.

[45] R. D. Patterson. Auditory filters and excitation patterns as representations of frequency resolution. In Brian C. J. Moore and R.D. Patterson, editors, *Auditory Frequency Selectivity*, chapter 3, pages 123–177. Plenum Press Inc., Cambridge, 1986.

[46] R. Chadwick, K. Morrish, S.A. Shamma, and J. Rinzel. *Parameter sensitivity in a mathematical model of basilar membrane mechanics.* Springer-Verlag, New York, 1986.

[47] M. Slaney and R. F. Lyon. Perceptual pitch detector. In *Proc. ICASSP*, volume 1, Albuquerque, NM, 1990.

[48] Xiaowei Yang, Kuansan Wang, and Shihab A. Shamma. Auditory representations of acoustic signals. *IEEE Transactions on Information Theory, Special Issue on Wavelet Transforms and Multiresolution Signal Analysis*, 38(2):824–839, March 1992.

[49] Kuansan Wang and Shihab A. Shamma. Self-normalization and noise-robustness in early auditory representations. *IEEE Transactions on Speech and Audio Processing*, October 1993. accepted for publication.

[50] W.A. Yost, R. Hill, and T. Perez-Falcon. Pitch and pitch discrimination in brodband signals with rippled power spectra. *Journal of the Acoustical Society of America*, 63(4), 1978.

[51] Reinier Plomp. *Aspects of tone sensation: a psychophysical study*, chapter 7, pages 111–142. Academic Press, New York, NY, 1976.

[52] W.A. Yost and R. Hill. Models of the pitch and pitch strength of ripple noise. *Journal of the Acoustical Society of America*, 66(2), 1979.

[53] F.A. Bilsen, J.H. ten Kate, T.J.F. Buunen, and J. Raatgever. Response of single units in the cochlear nucleus of the cat to cosine noise. *Journal of Acoustic Society of America*, 53:858–866, 1975.

[54] A. V. Oppenheim and R. W. Schafer. *Digital Signal Processing.* Prentice-Hall Inc., Eaglewood Cliffs, New Jersey, 1975.

[55] Athanasios Papoulis. *The Fourier Integral and its Applications.* McGraw Hill, New York, 1962.

[56] H. B. Voelcker. Towards a unified theory of modulation. Part I: Phase-envelope relationship. 54(3), 1966.

[57] L.L. Feth. Frequency discrimination of complex periodic tones. *Percept. and Psychophys.*, 15:375–378, 1974.