TECHNICAL RESEARCH REPORT

Representation of Spectral Profiles in the Auditory System Part I: Detection of Spectral Peak Shapes and Ripple Phases

by S. Vranic-Sowers and S.A. Shamma

T.R. 92-112r1

The Institute for Systems Research is supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

Abstract

This paper explores the question of how spectral profiles are represented in the auditory system. Using profile analysis methods, listeners' sensitivities to changes in spectral peak shapes and ripple phases were measured. Peak shapes were uniquely described in terms of two parameters: a symmetry factor (SF) which roughly measures the local evenness or oddness of a peak, and a bandwidth factor (BWF) which reflects the tuning or sharpness of a peak. Thresholds to changes in these parameters (defined as δSF and δBWF/BWF) were measured together with the effects of several manipulations such as using different peak levels, varying spectral component densities, and randomizing the frequencies of the peaks. The basic result that emerges is that δSF and δBWF/BWF *thresholds are largely constant regardless of the standard's peak shape.* The only exception occurs for the narrowest peaks (smallest BWF's) where δSF thresholds rise. A fundamental conclusion arising from these data is that peak profiles are represented along two sensitive and largely independent axes: peak bandwidth and symmetry factors. More generally, it is conjectured that for an arbitrary spectral profile these two axes correspond to the magnitude and phase of a *Fourier* transformation of the profile. In this light, the last set of experiments measured listeners' sensitivity to ripple phase changes in sinusoidal ripple stimuli. The thresholds obtained are similar in value and trends to δSF thresholds.

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 [*Shamma et al.*, 1993]. 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. Several other response features have also been mapped in the AI, including FM directional sensitivity [*Shamma et al.*, 1993] and response area bandwidth and tuning [*Schreiner and Mendelson*, 1990].

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, and peak frequency randomization.

The experiments reported here are similar in methodology to previously reported profile analysis experiments ([*Bernstein and Green*, 1987; *Bernstein, Richards and Green*, 1987; *Green, Mason and Kidd*, 1984; *Kidd, Mason and Green*, 1984]). They also share the same overall goals of the phonetic distance measure experiments described in [*Assmann and Summerfield*, 1989] and [*Klatt*, 1982]. 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. V, 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 this part of the paper, threshold measurements for all above mentioned manipulations are presented and critically interpreted in the context of two profile analysis models [*Bernstein and Green*, 1987; *Durlach, Braida and Ito*, 1986]. In Part II [*Vranić-Sowers and Shamma*, 1993], these and other data from several profile analysis experiments [*Bernstein and Green*, 1987; *Green*, 1986; *Hillier*, 1991] are integrated within a "ripple analysis model" based on the idea that the auditory system internally represents a spectral profile in terms of its Fourier transform.

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. II) and bandwidth (Sec. III) of peak profiles, for different peak shapes, levels, and spectral densities. Two control experiments are described in Sec. IV in which the relevance of pitch cues and peak energy changes in the above discrimination tasks are evaluated. In Sec. V, the results are briefly discussed within a general theoretical framework and further experiments with rippled spectra are performed (Sec. VI). We end with a general discussion of the results in relation to other profile analysis experiments.

I. GENERAL PROCEDURE

A. Methods

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. 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.

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. VI.

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 ([*Levitt*, 1971]). 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 ([*Green*, 1988]).

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.

B. Spectral peak stimulus parameters

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. 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(a)):

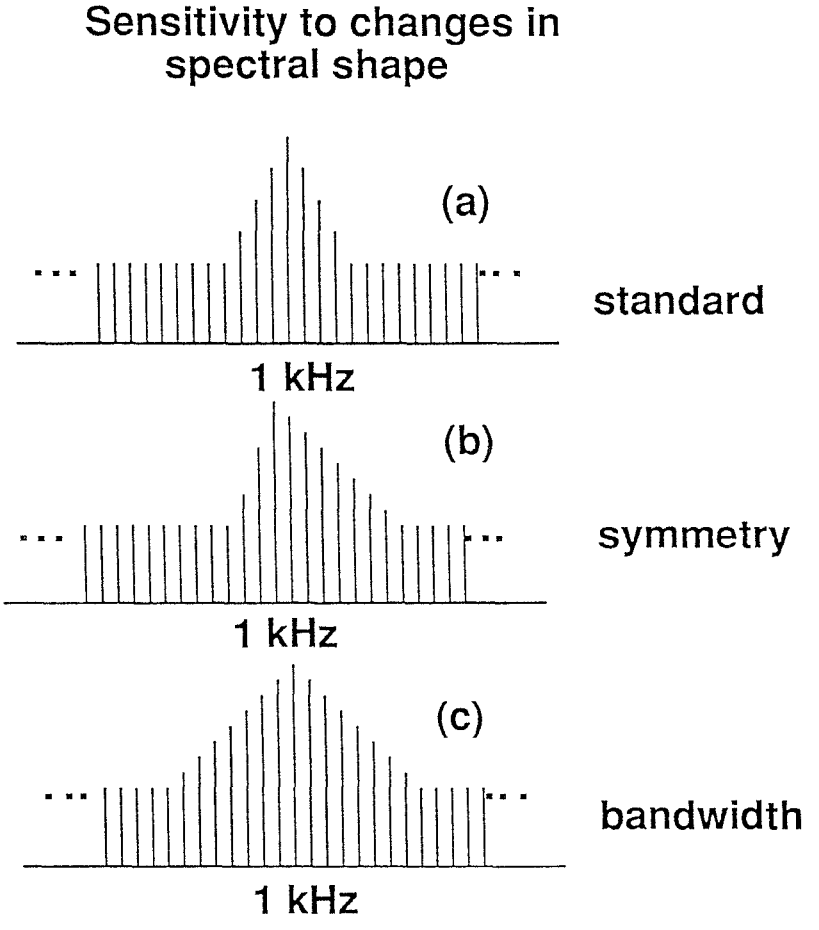

Figure 1: (a) Complex waveform consists of a flat base and a peak added to it. Peak takes different symmetries (b) and bandwidths (c).

- 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 $A_{max} = 20 \log_{10}(\frac{a_{max}}{b})$.

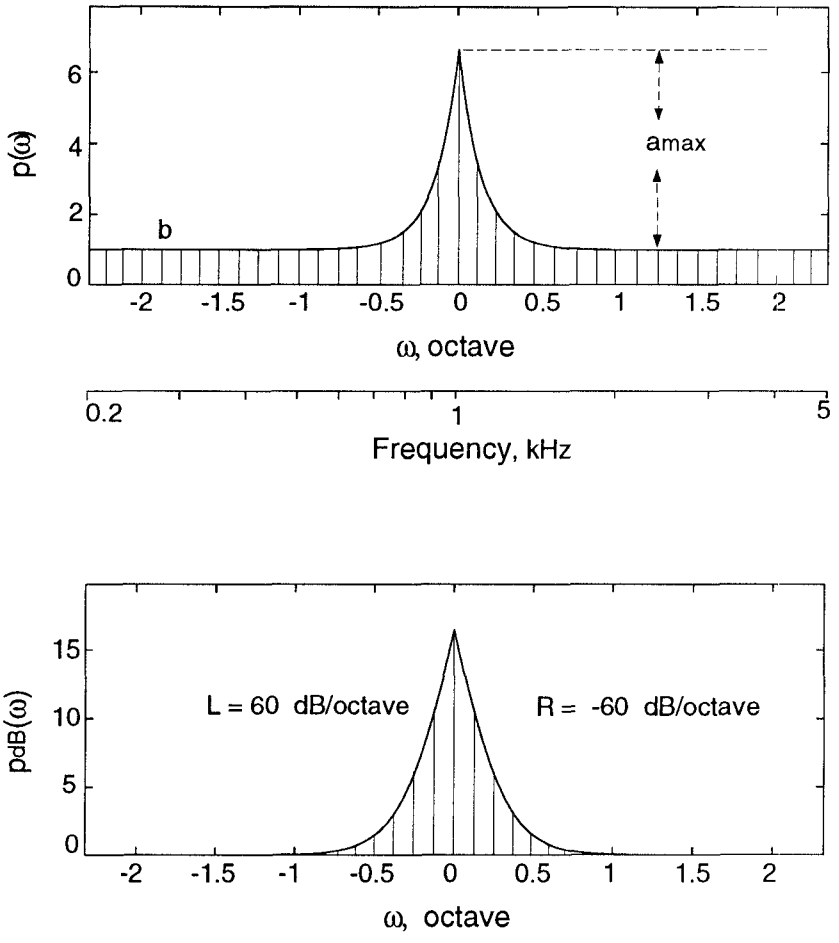

Figure 2: (a) Peak profile plotted on a linear (top) and logarithmic (bottom) amplitude scale. Peak level (A_{max}) is 15 dB, and BWF = 0.1 and SF = 0.

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

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

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(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(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(a) has SF = 0 and BWF = 0.1 octave. Peaks with various other SF's and BWF's are shown in Fig. 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 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

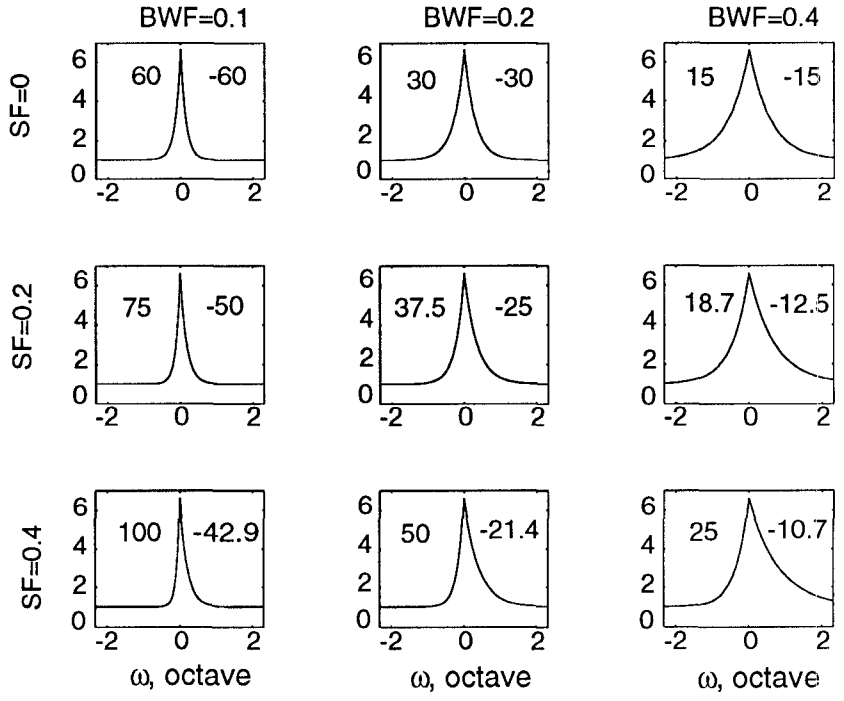

Figure 2: (b) Envelopes of various peak profiles plotted on a linear amplitude axis. Columns share the same BWF's, and rows share the same SF's. Corresponding left and right slope values (in dB/octave) are shown for each case.

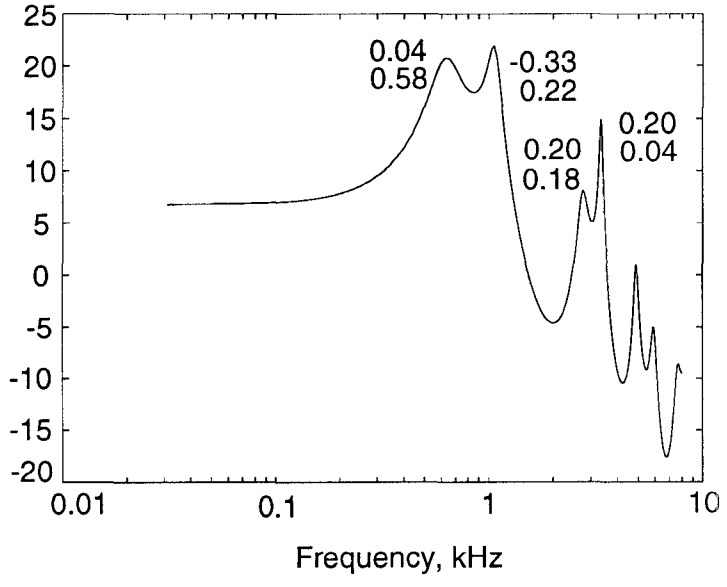

Figure 2: (c) SF's and BWF's for the spectral peaks of a naturally spoken vowel "aw".

also correspond to those that might be computed from the spectral envelope of speech sounds, as shown in Fig. 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 with the peak always centered at 1 kHz ($\omega = 0$ octaves). The waveform was turned on 10 ms following the onset in order to suppress the large amplitudes due to zero phases. No other phase conditions were tested since numerous previous experiments have shown that phase effects on signal detection are minimal ([*Bernstein, Richards and Green*, 1987; *Green and Mason*, 1985]).

C. Spectral peak threshold measures

Threshold measures reported here were derived from the threshold estimate 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 I). 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 paper the presentation and interpretations of the first type of threshold. The rms–thresholds for all tests are compiled in Appendix I, mostly to facilitate comparisons with results from other profile analysis experiments previously reported.

II. 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).

A. Results and discussion

1. 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. 3. In Fig. 3(a) the data are averaged over the four BWF's and plotted against SF. In Fig. 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. 3(a)). However, there is a slight consistent decrease in threshold as a function of BWF (Fig. 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 I. 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. IV A.

2. Dependence on peak amplitudes

In order to determine how the detection threshold depended on peak levels, the tests described in Sec. II A.1 were repeated at two other peak levels: 10 dB 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. 4. As in Figs. 3, data are averaged over the BWF's in Fig. 4(a) and over the SF's in Fig. 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 I). Note that on average, the two subjects here exhibited uniformly higher thresholds than the earlier five in Sec. II A.1.

(2) δSF thresholds as a function of BWF (Fig. 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. 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.

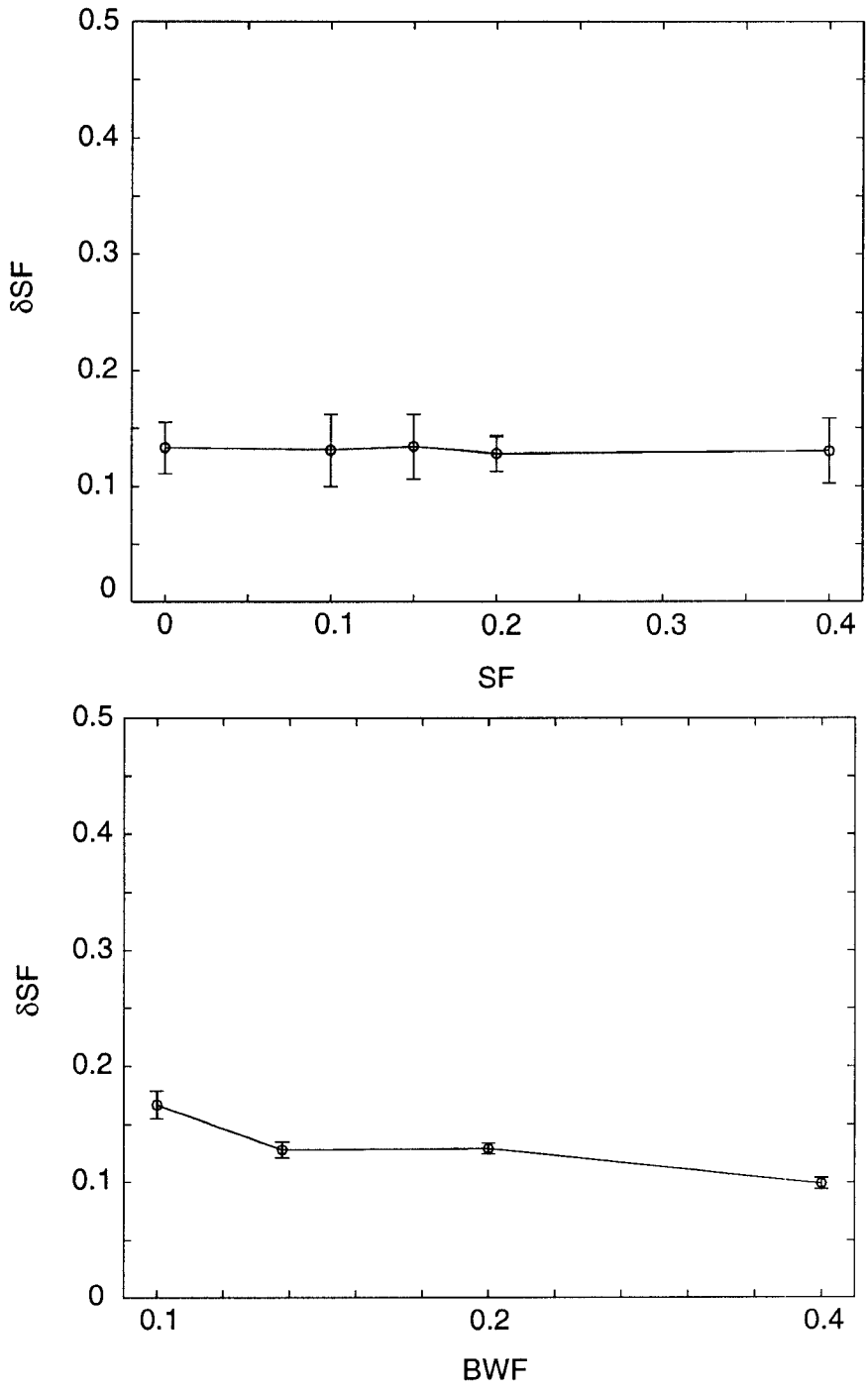

Figure 3: Symmetry change detection δSF thresholds for 41 component complex and 15 dB peak amplitude, averaged over five subjects and: four BWF's in (a), and five SF's in (b). The δSF threshold measure is defined as the change in SF between the signal at threshold and the standard. In (b), the δSF increases for the narrowest BWF. The error bars are the standard deviations of the means.

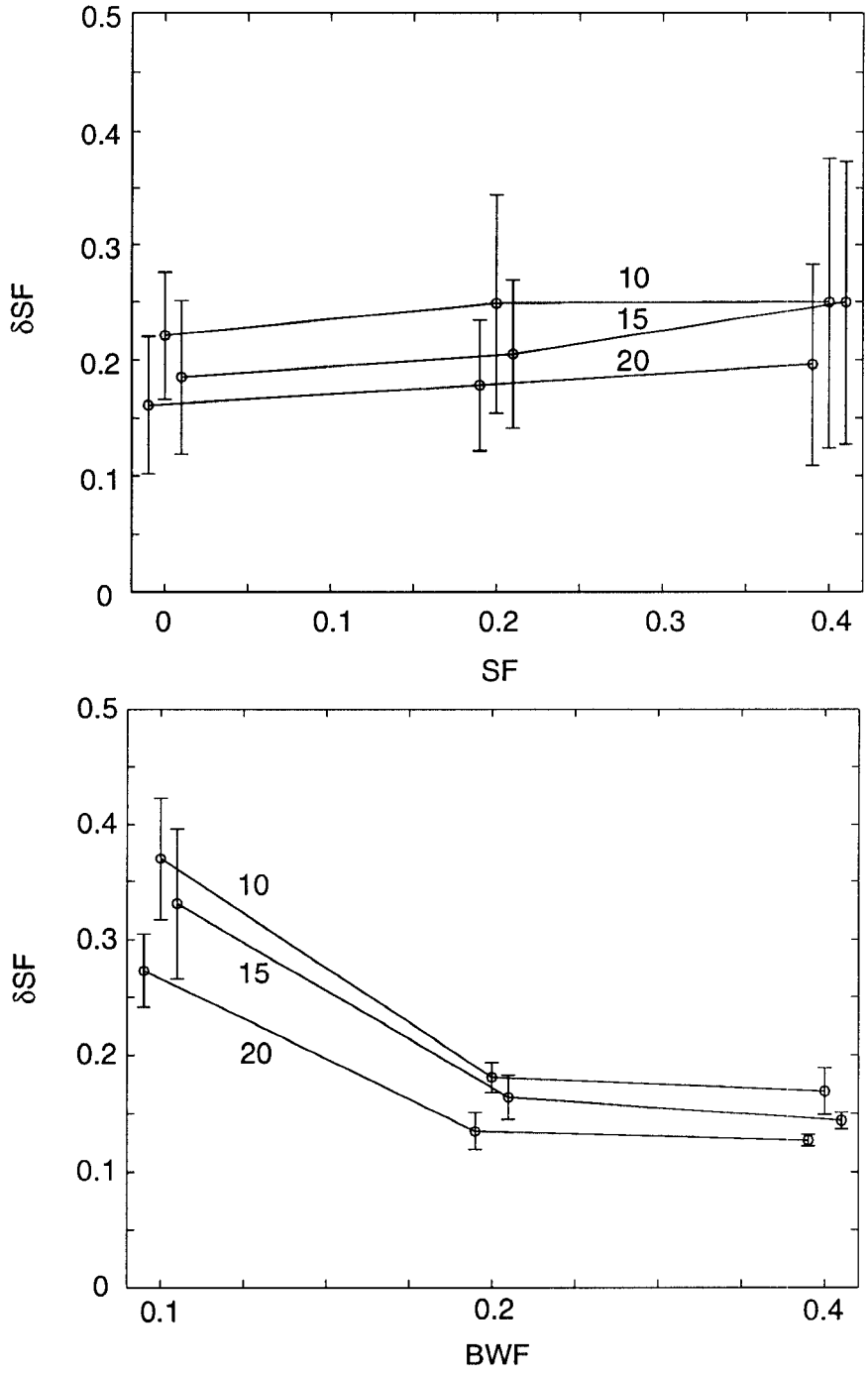

Figure 4: Symmetry change detection δSF thresholds for 41 component complex and 3 peak amplitudes: 10 dB, 15 dB, and 20 dB, relative to baseline. The data are averages of three subjects and: three BWF's in (a), and three SF's in (b). The values along the ordinates are defined as in Fig. 3. The large 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.

3. 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 15 dB above the baseline.

Once again, all δSF values and trends described earlier largely hold regardless of spectral densities (Figs. 5). The most prominent change in δSF thresholds occurs as a function of spectral density at the narrowest peak (Fig. 5(b)). The threshold here deteriorates rapidly as the spectral density decreases and, as in Figs. 4, it is largely this accelerated rise that is responsible for the upward shifts in the mean δSF in Fig. 5(a).

Note that the rms–threshold plots in Appendix I 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.

III. 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. II. For each test, the detection threshold was computed as the relative change in the BWF of the standard, i.e., δBWF/BWF.

A. Results and discussion

1. Dependence on symmetry and bandwidth factors of the standard

As in Sec. II A.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. 6. The plot in Fig. 6(a) is of the average δBWF/BWF as a function of SF. In Fig. 6(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 I.

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

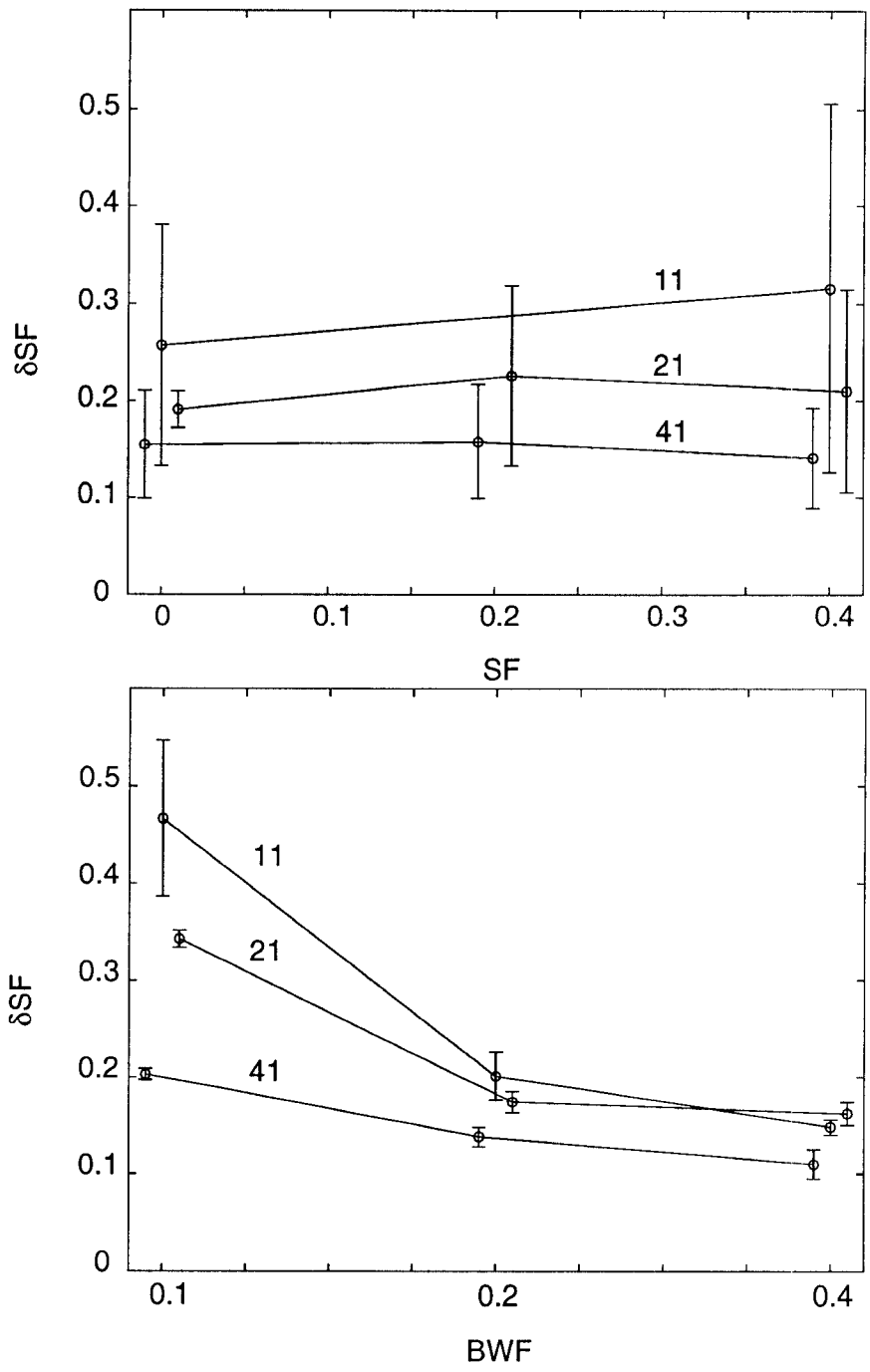

Figure 5: Symmetry change detection thresholds for 41, 21, and 11 component complexes, and 15 dB peak level, averaged over four subjects and three BWF's in (a) and three SF's in (b). Large error bars in (a) are due to δSF increase for the narrowest peak (BWF = 0.1) seen in (b). Note that in (b), most δSF changes with spectral density occur at the narrowest peak.

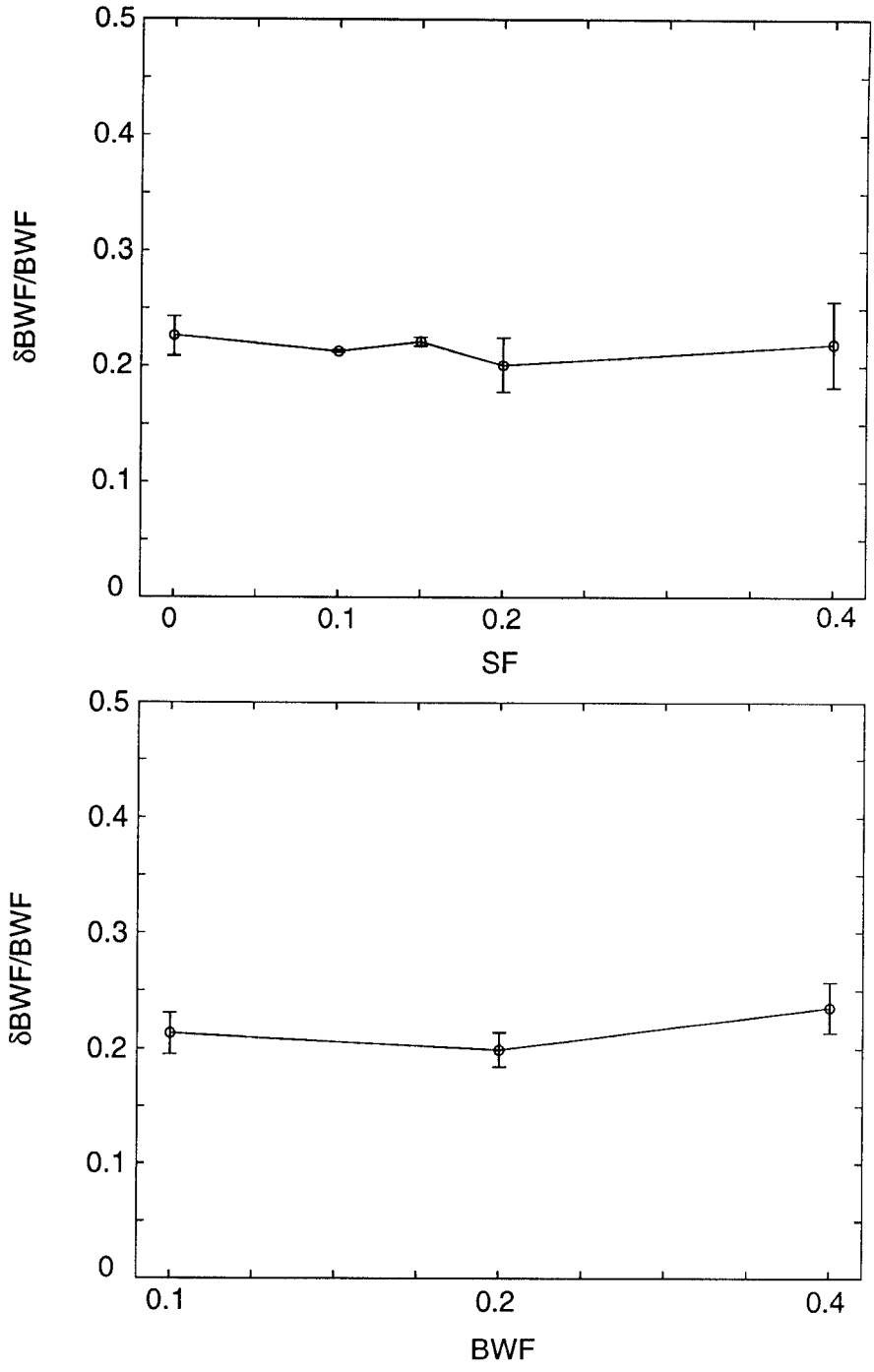

Figure 6: Bandwidth change detection δBWF/BWF threshold for 41 frequency components, and 15 dB peak level, averaged over three listeners, and three BWF's (0.1, 0.2, 0.4) in (a), and two SF's (0 and 0.4) in (b). Data are slightly offset along the abscissa for clarity.

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.

2. 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. 7(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 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.

3. 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. 8. In Fig. 8(a), they are given as a function of SF, and in Fig. 8(b) as a function of BWF. The corresponding rms–thresholds are shown in Appendix I.

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.

IV. 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.

A. 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 ([*Berg, Nguyen and Green*, 1992; *Feth, O'Malley and Ramsey*, 1982; *Richards, Onsan and Green*, 1989; *Stover and Feth*, 1983]). 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

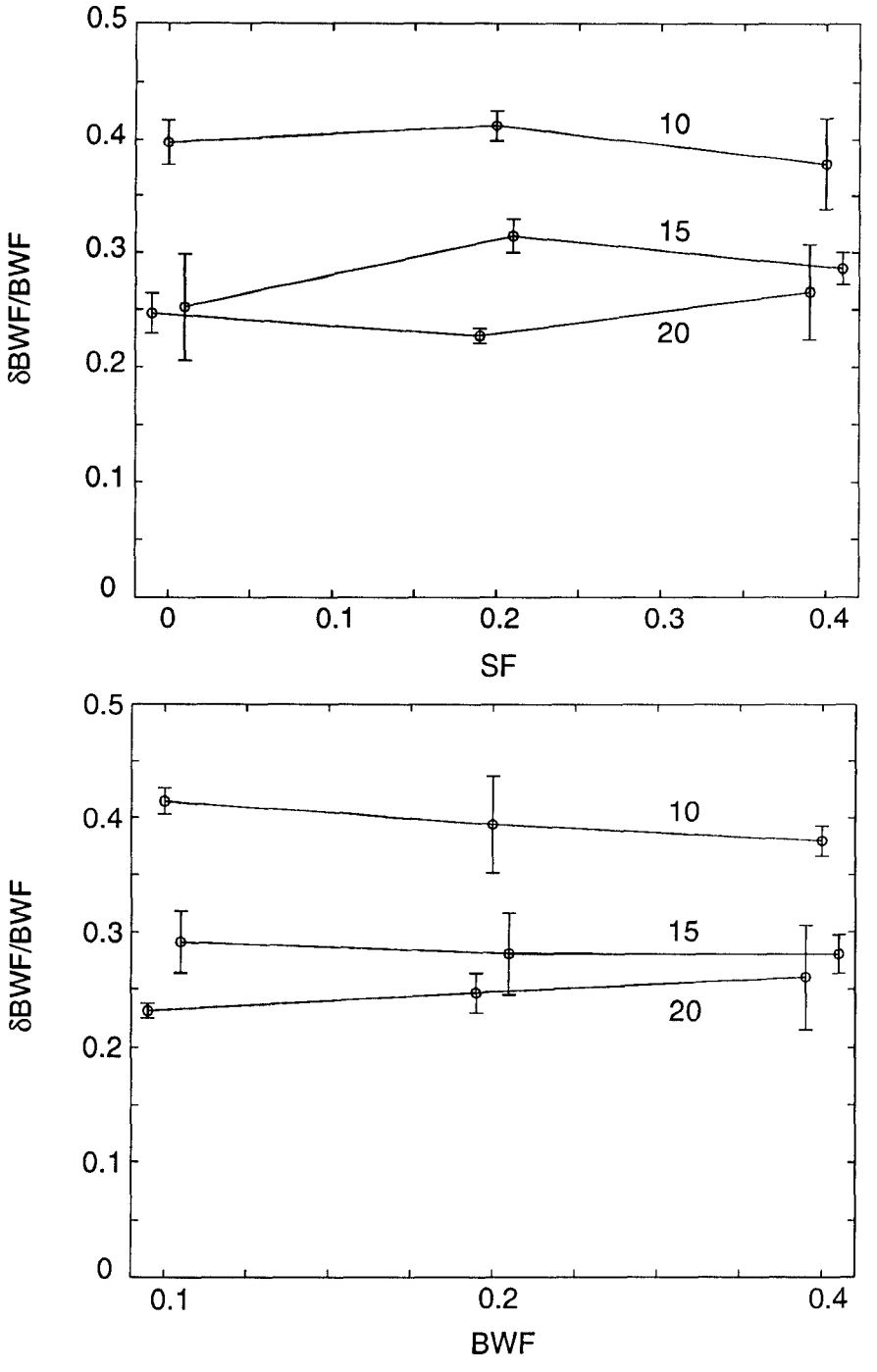

Figure 7: Bandwidth change detection δBWF/BWF thresholds for 41 component complex and 3 peak amplitudes: 10 dB, 15 dB, and 20 dB. The thresholds are averages of three subjects and: three BWF's in (a), and three SF's in (b). Points are offset along the abscissa for clarity.

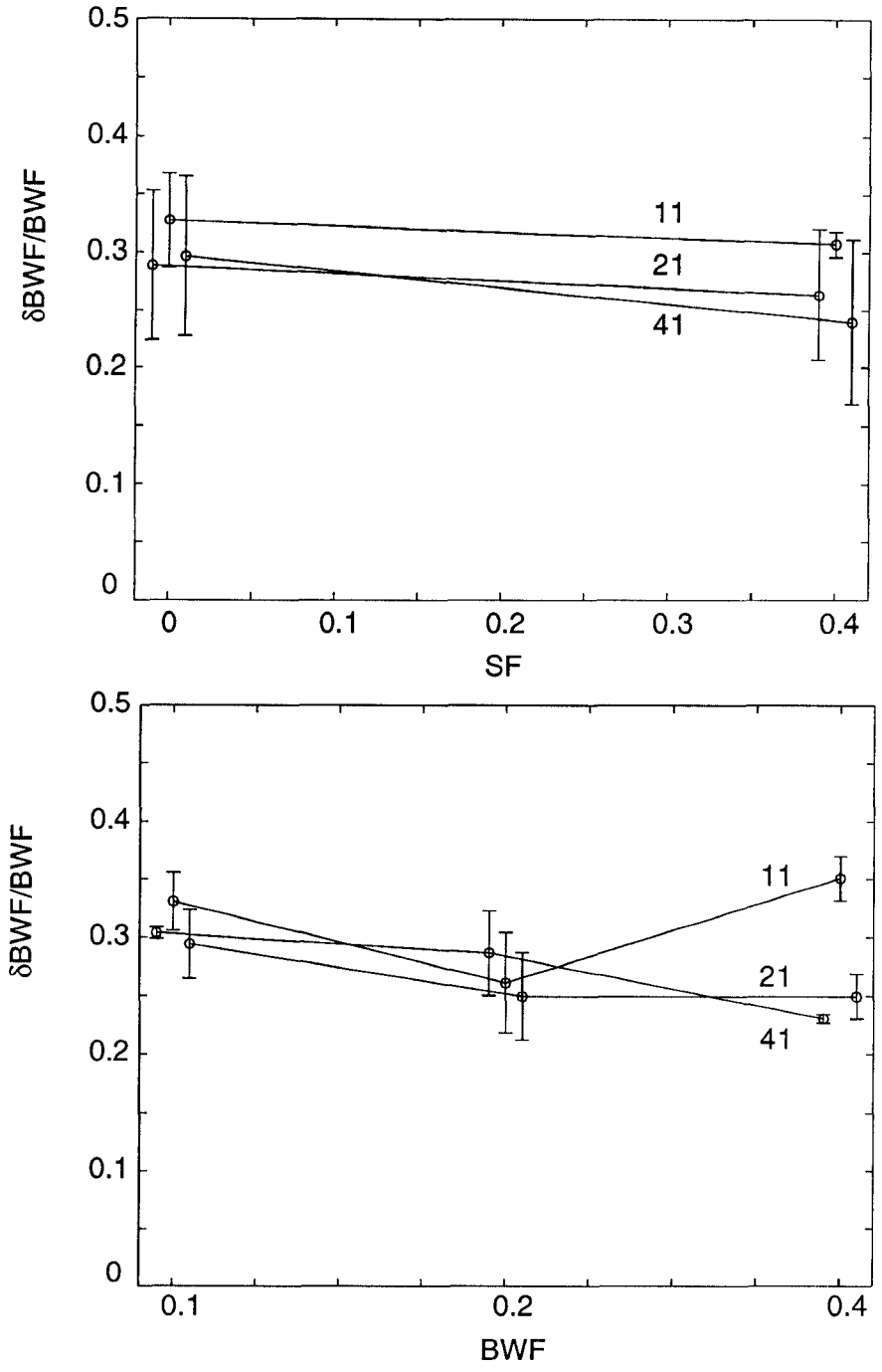

Figure 8: δBWF/BWF thresholds for 41, 21, and 11 component complexes, and 15 dB peak level, averaged over three subjects and three BWF's in (a) and two SF's (b). Threshold is independent of spectral density for all but the broadest BWF, where it increases for the 11 component case.

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 II).

1. 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 presented in Tables I(a) and I(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.

2. 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 [*Richards, Onsan and Green*, 1989] 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 ([*Feth and Stover*, 1987]), or relative pitch difference, $\Delta F/F_{sta}$ ([*Richards, Onsan and Green*, 1989]), 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 I.

3. Results and discussion

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

With respect to the first argument above, it is clear from the ΔF and $\Delta F/F_{sta}$ values in Table I(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

	BWF					
	0.1		0.2		0.1	
	SF					
δSF test	0	0.1	0	0.1	0	0.1
NR	0.27	0.27	0.13	0.15	0.13	0.11
R	0.36	0.11	0.27	0.23	0.17	0.12
-(NR R)	0.09	0.17	0.11	0.08	0.01	0.01
F_{sta}	1290.52	1327.10	1227.73	1329.10	1223.27	1130.59
F_{sig}	1315.58	1345.59	1258.51	1369.30	1293.12	1195.99
$\Delta F=F_{sta}-F_{sig}$	-25.01	-18.19	-30.83	-10.25	-70.13	-65.39
$\Delta F/F_{sta} \cdot 100$	-1.91	-1.37	-2.51	-3.03	-5.73	-1.57

Table 1: (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, $\Delta F/F_{sta}$.

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 I(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.

B. Detecting peak energy change compared to a BWF change

	BWF					
	0.1		0.2		0.1	
	SF					
δBWF/BWF test	0	0.4	0	0.1	0	0.1
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	1130.59
F_{sig}	1213.41	1318.03	1216.00	1377.66	1255.29	1505.61
$\Delta F = F_{sta} - F_{sig}$	38.65	9.66	11.69	-19.96	-26.78	-73.16
$\Delta F / F_{sta} \cdot 100$	2.99	0.73	0.95	-3.76	-2.19	-5.11

Table 1: (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 I(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.

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 I, compare data points at BWF = 0.4 in Figs. A3(b) and A6(b)). If the tasks were purely based on the total change in energy, then the two thresholds should be comparable.

A more direct rebuttal of this hypothesis is provided in Figs. 9, 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. 9(a), test A), or through a change in the height only (test B)[1] of the peak. In test

[1]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.

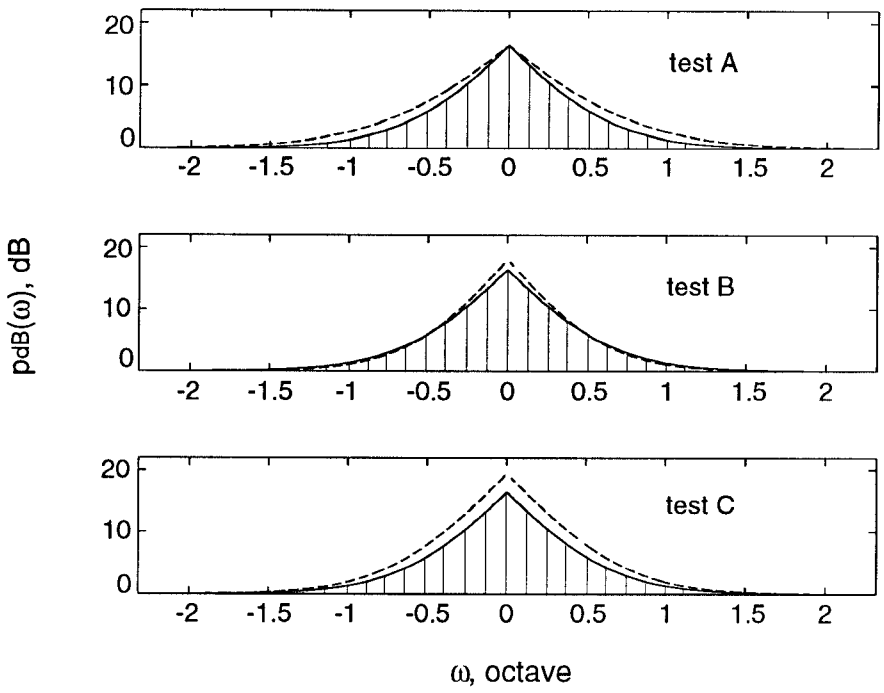

Figure 9: (a) Effects of BWF changes in tests A and B (see text), and peak level changes in test C are shown by dashed lines, for a standard peak with BWF = 0.2, SF = 0, and A_{max} = 15dB.

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).

The data in Fig. 9(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.

V. 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. 10(a)). This change in $p(\omega)$

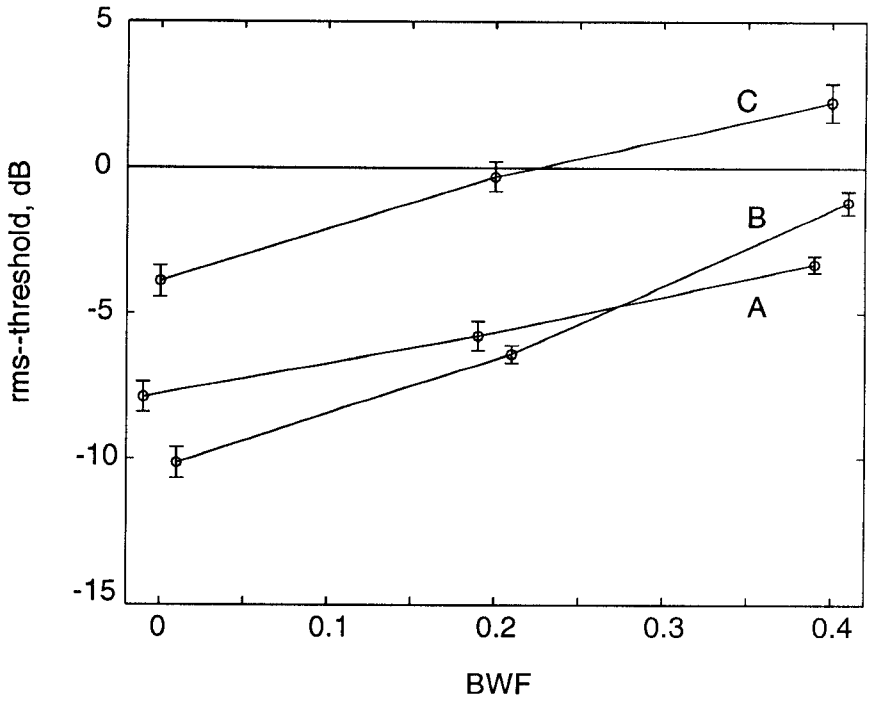

Figure 9: (b) The rms–thresholds for the three tests (A, B, and C).

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. 10(b))[2].

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. 10(c) and (d) for three SF's and their corresponding θ_o angles: SF = 0.05 (3°), 0.15 (9°), and 0.3 (18°). (The computations are in Appendix III).

[2]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 sinusoidal profile whose peaks are separated by 1/2 octaves along the ω axis.
The magnitude of the Fourier transform of the peak profile (Fig. 10(b)) is for $\Omega > 0$ computed as:

$$|P(\Omega)| = \frac{a_{max}}{b} \frac{20 \text{ BWF}}{3(\ln 10)\ M} \cdot$$

$$|1 + j\ 2\pi\Omega\ 20/(3\ \ln 10)\ \text{SF BWF} + (\pi\Omega\ 20/(3\ \ln 10)\ \text{BWF})^2(1 - \text{SF}^2)|^{-1},$$

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

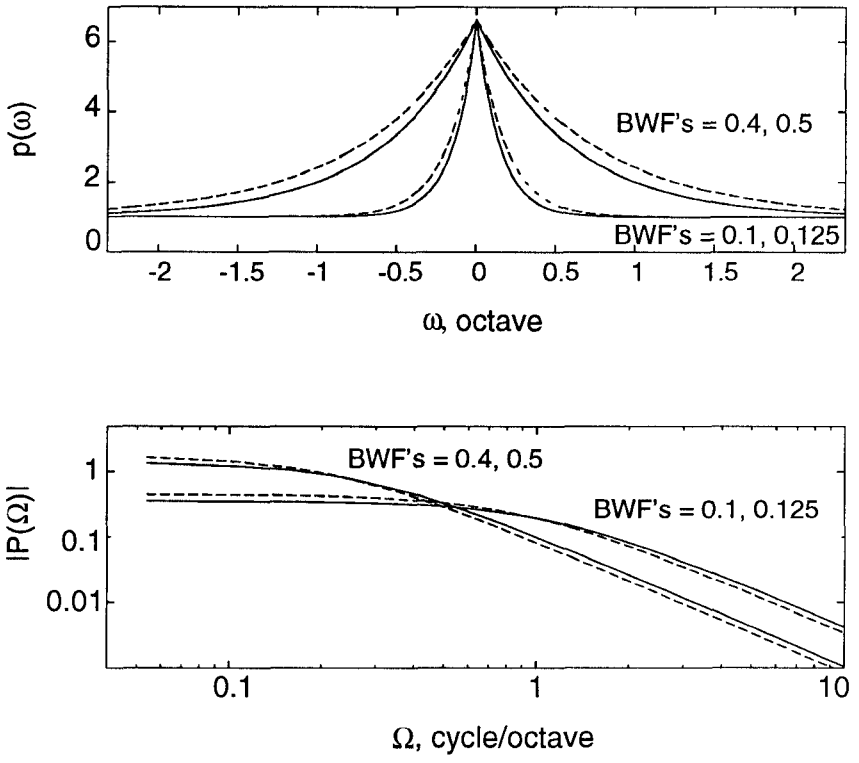

Figure 10: (a) Peak profiles with A_{max} = 15 dB, BWF's 0.1 and 0.4 (solid lines), and BWF's at 25% detection threshold (dashed lines). (b) Magnitude of the profiles' Fourier transformations, $|P(\Omega)|$. The effect of the BWF change is a shift in magnitude (and not a change in shape) along the $\log \Omega$ axis.

The above interpretations of the δBWF and δSF imply that these manipulations 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. 11). While ripple frequency-difference-limen thresholds were measured previously ([*Green*, 1986; *Hillier*, 1991]), no ripple phase sensitives have been reported in the literature. The experiments described below provide these measurements.

VI. PHASE DIFFERENCE LIMEN EXPERIMENTS

Sensitivity to ripple phase changes was measured in sinusoidally modulated profiles on a dB amplitude scale (Fig. 11), and the thresholds, termed phase-difference-limen (**pdl**), are reported in units of degrees.

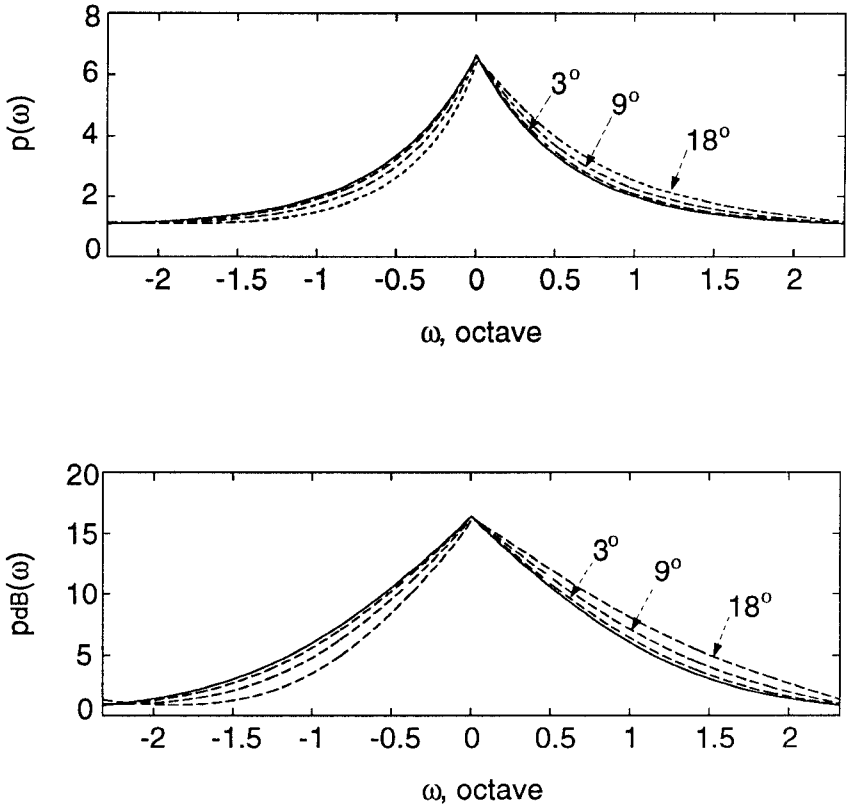

Figure 10: (c) and (d) The effects on changes in the symmetries of a peak profile (BWF = 0.4 and SF = 0) due to adding constant phases (3^o, 9^o, and 18^o) to its Fourier transform.

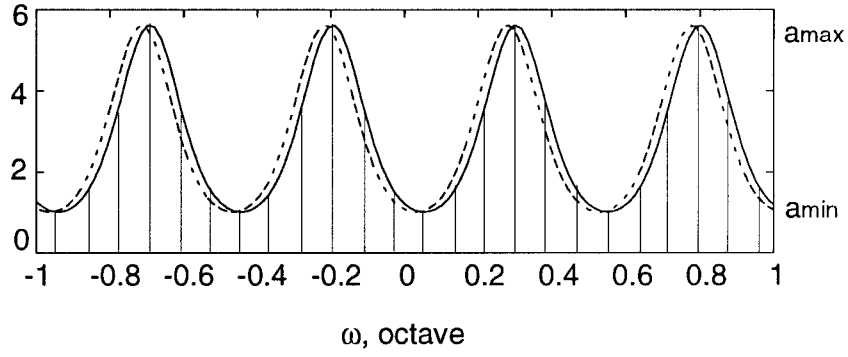

Figure 11: A sinusoidal ripple profile with ripple frequency of 2 cycle/octave, and 15 dB peak-to-valley amplitude (computed as $20\log_{10}(a_{max}/a_{min})$). Its 16^o phase shifted version is shown in dashed lines.

A. 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. 11). 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.

B. Results

The average data for two subjects are presented in Fig. 12(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^o for the larger level. Phase sensitivity decreases with increasing ripple frequencies beyond 2 cycle/octave.

Figure 12(b) depicts the data for individual subjects as a function of ripple level. Thresholds saturate with increasing levels at all ripple frequencies tested.

C. Discussion

There are two important characteristics of the data in Fig. 12(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. 3). The correspondence between these two thresholds confirms the association made between them as explained in Sec. V. 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 Part II of this paper.

VII. GENERAL DISCUSSION

A. Summary of basic 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

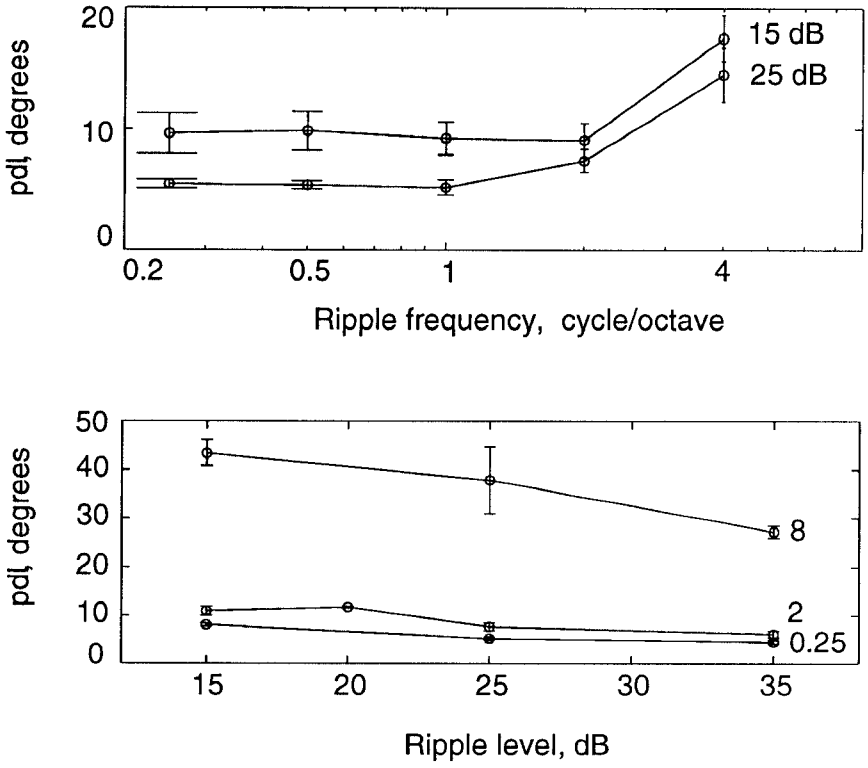

Figure 12: (a) Phase difference limen threshold (**pdl**) as a function of ripple frequency, for 15 dB and 25 dB peak-to-valley amplitudes (or ripple levels), averaged over 2 subjects. (b) Individual **pdl** thresholds at three ripple frequencies as a function of ripple level (subject 1 was tested at 0.25 and 8 cycle/octave, and subject 2 at 2 cycle/octave).

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. 3). The exception occurs towards the narrowest peak (BWF = 0.1) where (1) the detection threshold increases gradually to 0.16 (Fig. 3(b)), and (2) pitch cues associated with this detection task become more effective (Sec. IV A.3 (i), Table I(a)). For the detection of BWF changes, all δBWF/BWF thresholds remain constant at around 0.22 regardless of the peak shape (Figs. 6).

Also measured were the effects of two additional manipulations that did not change the shape of the peak: (1) change in the peak level and (2) spectral density of the complex. For the first, all thresholds maintain the same trends regardless of peak level. Their absolute values, however, slightly improve for higher peak levels (Figs. 4 and 7). 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. 5(b)), whereas δBWF/BWF thresholds deteriorate only for the lowest density (11 components) at the broadest peak (Fig. 8(b)).

Finally, rms–threshold values for SF and BWF detection tasks are comparable to other profile detection tasks (see Appendix I). Furthermore, they are significantly lower than rms–thresholds of changes that do not affect peak shape (Fig. 9(b)).

In summary, a fundamental conclusion from these data 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 ([*Shamma et al.*, 1993]). 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 ([*Schreiner and Mendelson*, 1990]).

B. Profile analysis models

The choice of a threshold measure implies an underlying profile analysis model. Such a model based on the δSF and δBWF/BWF measures is described in detail in Part II of this paper. Here we apply a two alternative profile analysis models which have been shown to perform well in a variety of detection tasks. Both 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 [*Durlach, Braida and Ito*, 1986], which basically combines information from all independent channels. The other is the *maximum difference* model described in [*Bernstein and Green*, 1987], 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' computed at perceptual thresholds at various SF and BWF combinations.

1. The channel model

This model is described in detail in [*Durlach, Braida and Ito*, 1986; *Green*, 1988]. 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 ($\sigma_R = 5.6$) is approximated by a normal distribution of $\sigma_R = 5$. 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 II) 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 [*Durlach, Braida and Ito*, 1986]), 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 II(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% (≈ 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 II(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. IV B. Specifically, it predicts higher than perceptual thresholds for the narrowest peaks (Table II(c)).

Finally, the $d'\sigma$ for the phase data (Table II(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 $d'\sigma = 10.51$ for 8 cycle/octave and 15 dB level stimulus, would predict a 9° threshold, compared to the 49.58° perceptual value.

	$d'\sigma$ for δSF test (21 components)			
	BWF			
SF	0.1	0.2	0.1	average
0	4.46 ±0.13	3.15 ±0.19	1.36 ±0.52	3.99
0.1	3.68 ±0.07	3.58 ±0.13	1.81 ±0.30	4.03
average	4.07	3.37	1.60	
δSF threshold	0.31 ±0.01	0.17 ±0.01	0.17 ±0.02	

	$d'\sigma$ for δSF test (11 components)			
	BWF			
SF	0.1	0.2	0.1	average
0	2.96 ±0.19	2.88 ±0.13	3.97 ±0.18	3.27
0.1	3.36 ±0.22	3.22 ±0.15	1.33 ±0.20	3.61
average	3.16	3.05	1.15	
δSF threshold	0.16 ±0.01	0.11 ±0.005	0.11 ±0.005	

Table 2: $d'\sigma$ values for the "independent channel model" (Sec. VII B.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. 5(b) and 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	3.01 ±0.27	2.68 ±0.16	1.98 ±0.11	2.57
0.1	3.03 ±0.27	2.80 ±0.19	2.52 ±0.17	2.78
average	3.03	2.71	2.25	
δBWF/BWF threshold	30 ±3%	25 ±5%	25 ±2%	

	$d'\sigma$ for δBWF test (11 components)			
	BWF			
SF	0.1	0.2	0.1	average
0	3.11 ±0.21	3.02 ±0.19	2.58 ±0.21	2.90
0.1	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 ±1.1%	23.6 ±2.1%	

Table 2: (b) $d'\sigma$ values for δBWF tests. The table is organized as Table II(a), with threshold values from Figs. 8(b) and 6(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	5.16 ±1.28	5.36
$\delta\Lambda_{max}$ threshold	2.79 ±0.55	3.49 ±0.98	3.14 ±0.81	
0.4	4.41 ±0.80	6.51 ±1.02	6.62 ±1.29	5.85
$\delta\Lambda_{max}$ threshold	2.91 ±0.51	3.53 ±0.53	3.63 ±0.69	
average	1.31	6.45	6.01	

Table 2: (c) Similar to Table II (a) for control experiment C (Fig. 9(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	10.11	8.89	9.17	19.0	19.58
pdl, 15 dB	8.09°	9.08°	7.53°	8.1°	16.22°	13.17°
$d'\sigma$	10.67	10.20	10.08	11.32	28.39	
pdl, 25 dB	5.39°	5.31°	5.11°	7.31°	11.53°	

Table 2: (d) $d'\sigma$ for **pdl** tests with spectral sinusoids, for 15 dB and 25 dB peak levels and thresholds from Fig. 12(a).

Note that the stimuli here are "balanced" as in the δSF tests.

In summary, the channel model predicts reasonably well most of the threshold 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. 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 [*Bernstein and Green*, 1987]. 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 ([*Bernstein and Green*, 1987]), 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.

The computed Λ's for the δSF tests are shown in Table III(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.

For δBWF tests, Λ is approximately constant for all SF's and BWF's except for the narrowest peak for 21 component stimulus (Table III(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 [*Bernstein and Green*, 1987]) 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 III(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 III(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.

A for δSF test (21 components)

SF	BWF 0.1	0.2	0.4	average
0	5.12	2.53	2.58	3.11
0.1	4.35	3.29	3.31	3.65
average	1.73	2.91	2.91	
δSF threshold	0.31	0.17	0.17	

A for δSF test (11 components)

SF	BWF 0.1	0.2	0.4	average
0	2.39	1.65	1.66	1.90
0.1	3.07	2.01	2.09	2.40
average	2.73	1.81	1.87	
δSF threshold	0.16	0.11	0.11	

Table 3: Maximal difference levels, Λ (dB), for the "maximal difference model" (Sec. VII B.2). Tables are organized as Tables II (with the same threshold values as in Tables II). Table 3 (a)

A for δBWF test (21 components)

SF	BWF 0.1	0.2	0.1	average
0	1.85	1.61	1.68	1.72
0.1	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 (11 components)

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

Table 3: (b)

A for control experiment C				
	BWF			
SF	0.1	0.2	0.4	average
0	2.12	3.03	2.61	2.69
δA_{max} threshold	2.79	3.49	3.11	
0.1	2.53	3.08	2.83	2.81
δA_{max} threshold	2.91	3.53	3.63	
average	2.17	3.05	2.72	

Table 3: (c)

A for phase ripple experiment						
	ripple frequency (cycle/octave)					
	0.25	0.5	1	2	1	8
A	2.12	2.38	1.97	2.12	1.23	11.11
pdl, 15 dB	8.09°	9.08°	7.53°	8.1°	16.22°	13.17°
A	2.35	2.32	2.23	3.16	6.32	
pdl, 25 dB	5.39°	5.31°	5.11°	7.34°	14.53°	

Table 3: (d)

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.

C. Ripple analysis model

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 [*Bernstein and Green*, 1987; *Green*, 1986]. The maximum difference model predicts well the detectable amplitude of the ripple [*Bernstein and Green*, 1987]. 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 [*Green*, 1986]. 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 transformation 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 Part II of this paper.

ACKNOWLEDGMENTS

This work was supported by grants from the Air Force Office of Scientific Research, The Office of Naval Research, and the National Science Foundation (NSFD CDR8803012).

The authors are members of the Institute for Systems Research. We also thank Daniel Naar from Apple Corporation for his helpful comments concerning the manuscript.

Appendix I: Detection thresholds measured in rms 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 $20 \log \sqrt{\sum_{i=1}^{n} (\Delta a_i / a_i)^2}$, where Δa_i is the change in the amplitude of the i^{th} component at threshold, a_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 a, the measure usually used is: $20 \log(\frac{\text{rms}_{signal}}{a})$. 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 amplitudes spectrum, and ripple signals; the threshold values reported are: –23.33 dB, –23.07 dB, and –21.58 dB, respectively (see [*Green*, 1986; *Richards, Onsan and Green*, 1989]). Computed in our unit, the thresholds are: –10.11 dB, –10.06 dB, and –8.36 dB, respectively.

In order to facilitate the comparison with corresponding figures in the paper using different threshold measures, we shall use the same figure numbers as before except for an additional prefix (A).

1. Detection of changes in spectral peak symmetry

Stimuli and testing conditions are described in Sec. II.

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

Threshold trends are as described earlier in Sec. II in that detection of a change in peak symmetry is independent of the peak shape of the standard (Figs. A3). 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. II. The same trends established earlier for the 15 dB case hold also for the other two levels. However, unlike the δSF thresholds (Figs. 4), the average rms–threshold monotonically *increases* with peak levels (Figs. A4). 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. A5). 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 ([*Bernstein and Green*, 1987]). Another possible source

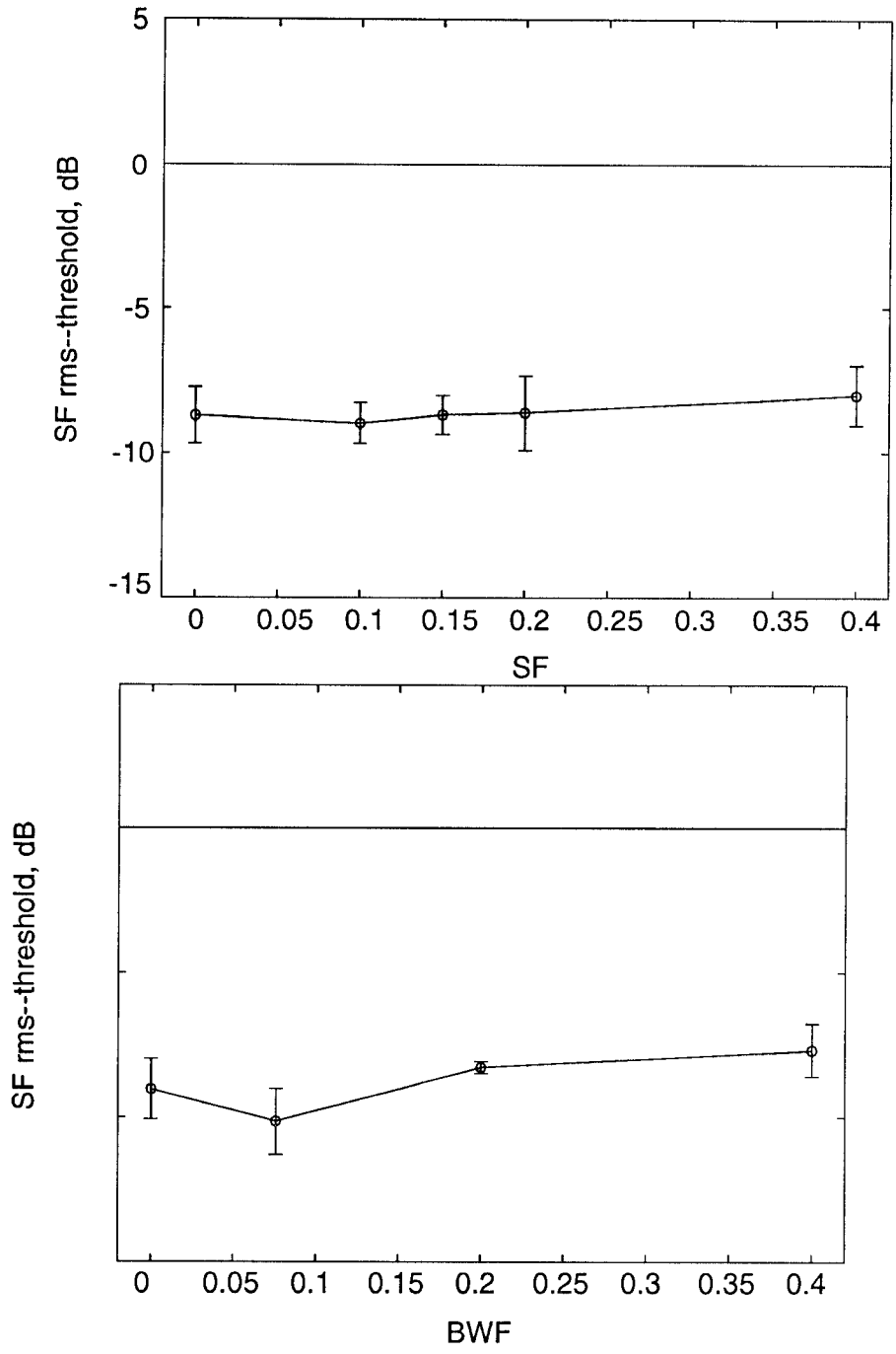

Figure 3: A3: Symmetry change detection rms–thresholds for 41 component complex and 15 dB peak amplitude, averaged over five subjects and: four BWF's in (a), and five SF's and (b). The rms–threshold is defined as: 20 $\log \sqrt{\sum_{i=1}^{n}(\Delta a_i/a_i)^2}$, where Δa_i is the change in the amplitude of the i^{th} component at threshold, a_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.

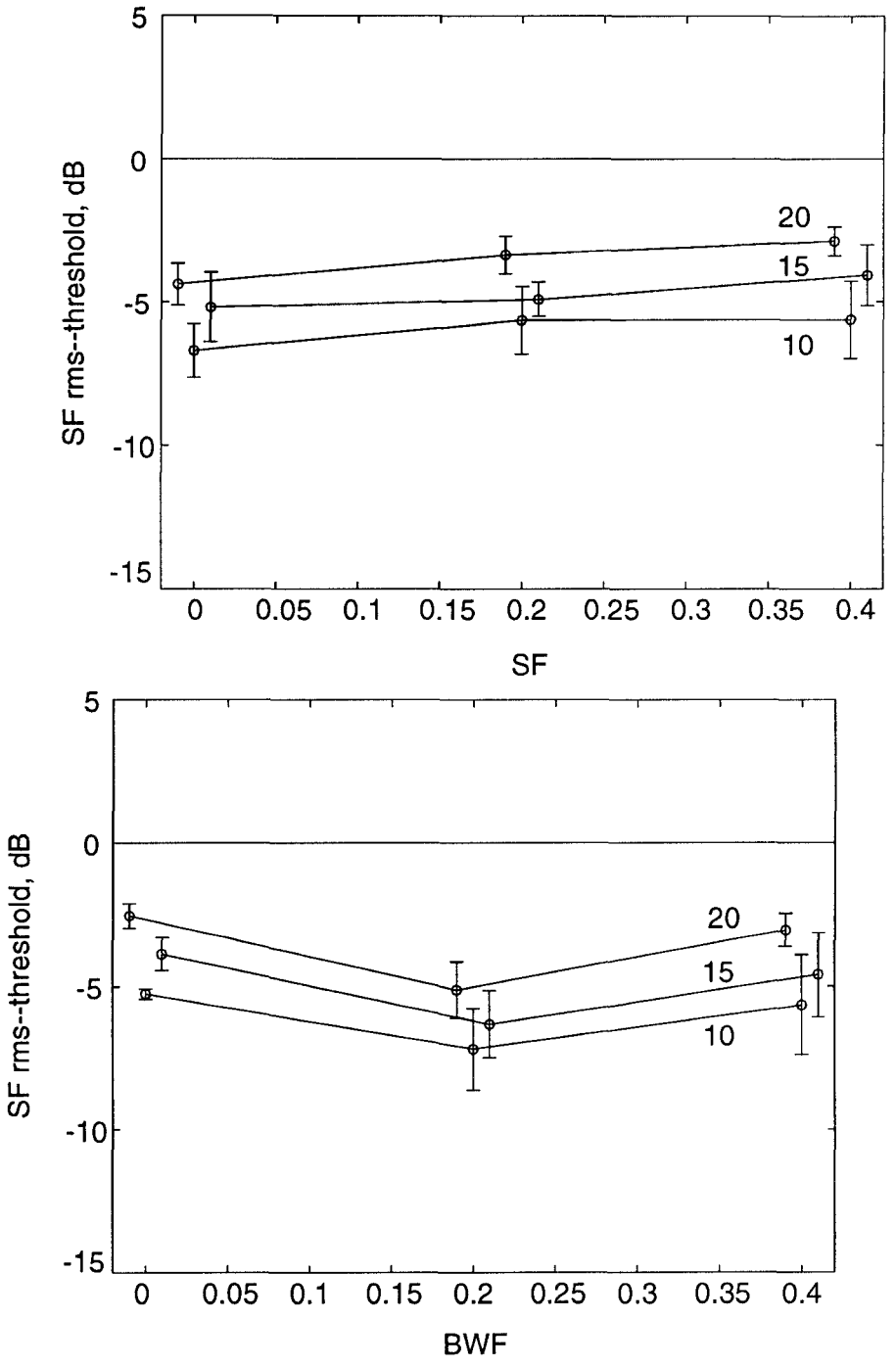

Figure 4: A4: Symmetry change detection rms–thresholds for 41 component complex and 3 peak amplitudes: 10 dB, 15 dB, and 20 dB, relative to baseline. The data are averages of three subjects and: three BWF's in (a), and three SF's and (b). The values along the ordinates are defined as in Fig. 3. Points are slightly offset for clarity reason.

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.

2. Detection of changes in spectral peak bandwidth factor

Stimuli and paradigm are described in Sec. III of the text.

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

The data are averaged and presented as described in Sec. III. 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. A6(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. III. 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. A8(a)).

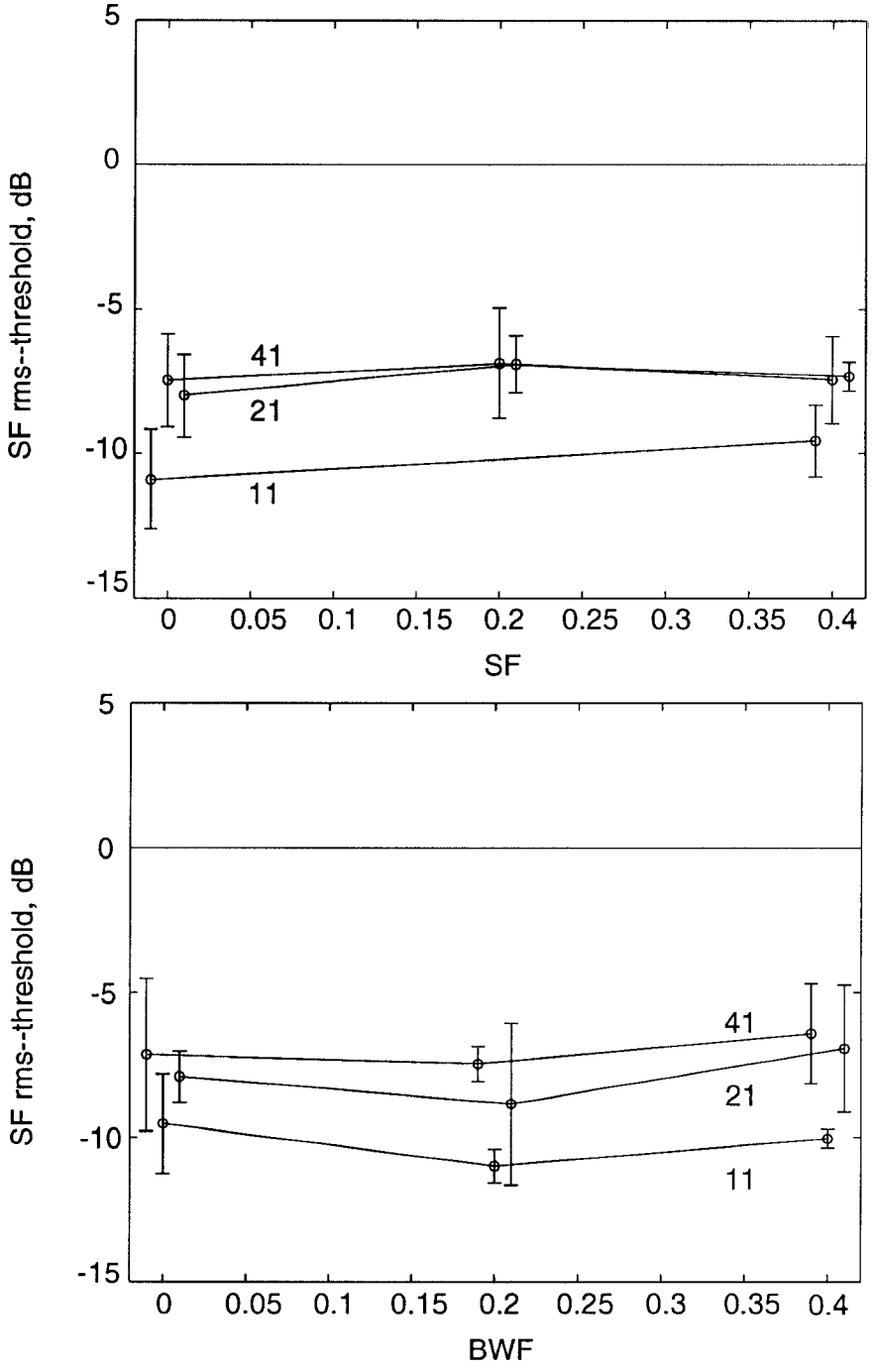

Figure 5: A5: Symmetry change detection thresholds for 41, 21, and 11 component complexes, averaged over four subjects and three BWF's (a) and three SF's (b). rms–Threshold increases with increasing spectral density, from 41 to 11 component tests.

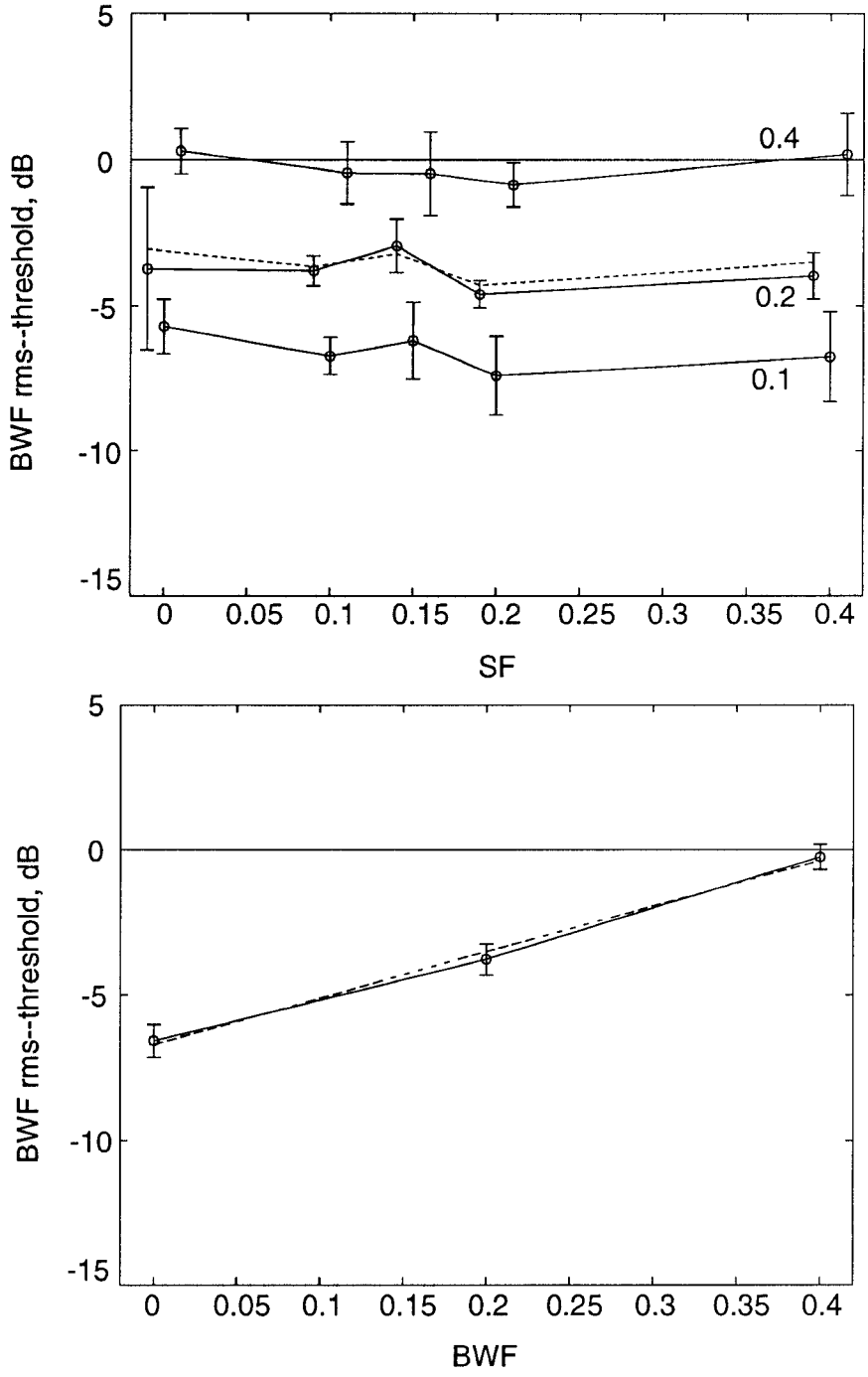

Figure 6: A6: (a) Bandwidth change detection rms–threshold for 41 frequency components, 15 dB peak level, and three BWF's: 0.1, 0.2, and 0.4, averaged for three listeners. Thresholds monotonically increase with BWF, and the form of this dependence, averaged over five SF's, is depicted in (b). The dotted line in (b) is the least square error linear approximation of this dependences: threshold (dB) = $-6.85 + 3.3 \log_2 (10\ \mathrm{BWF})$. Data are slightly offset for clarity.

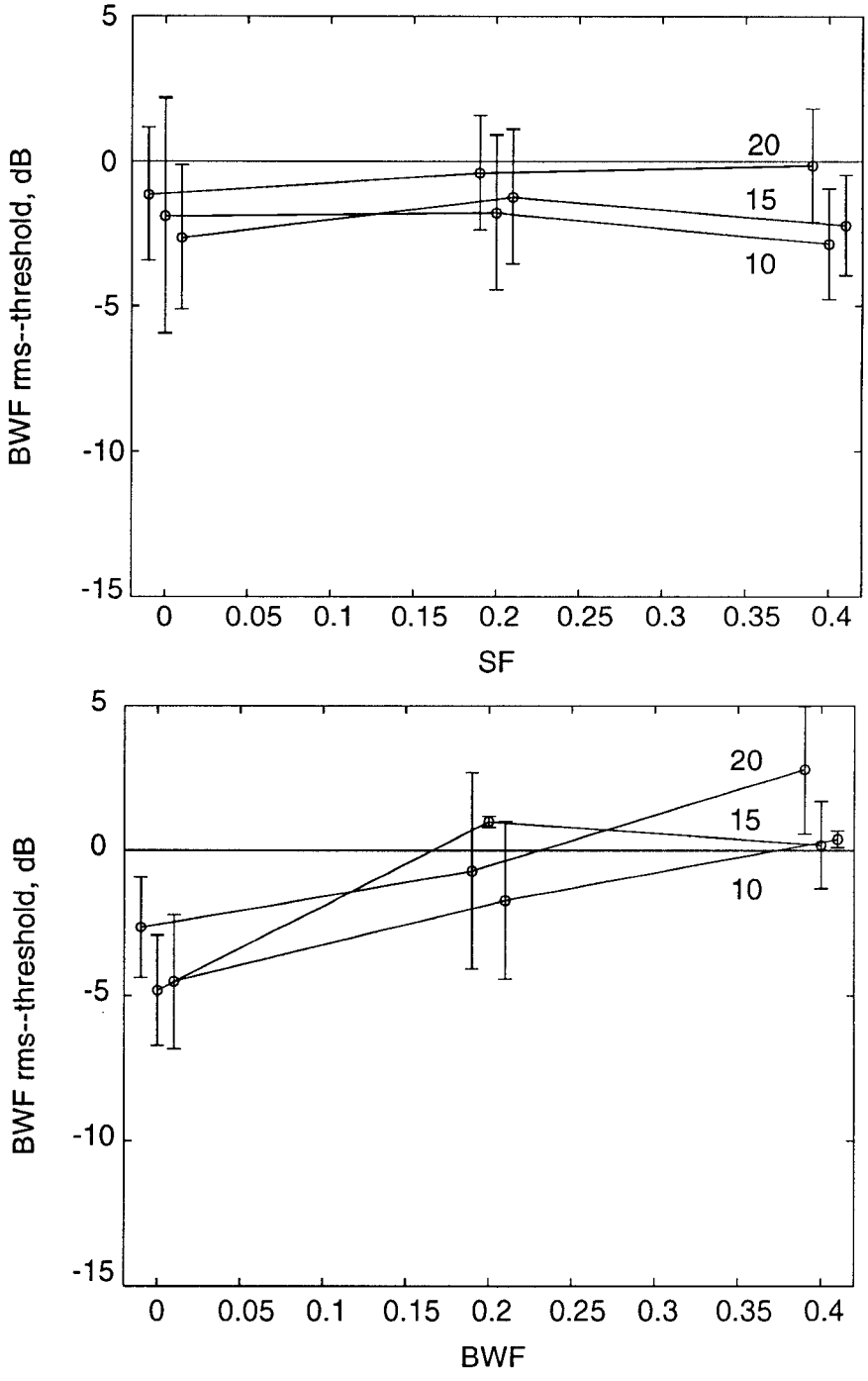

Figure 7: A7: Bandwidth change detection rms–thresholds for 41 component complex and 3 peak amplitudes: 10 dB, 15 dB, and 20 dB. The thresholds are averages of three subjects and: three BWF's in (a), and three SF's in (b). Points are offset for clarity.

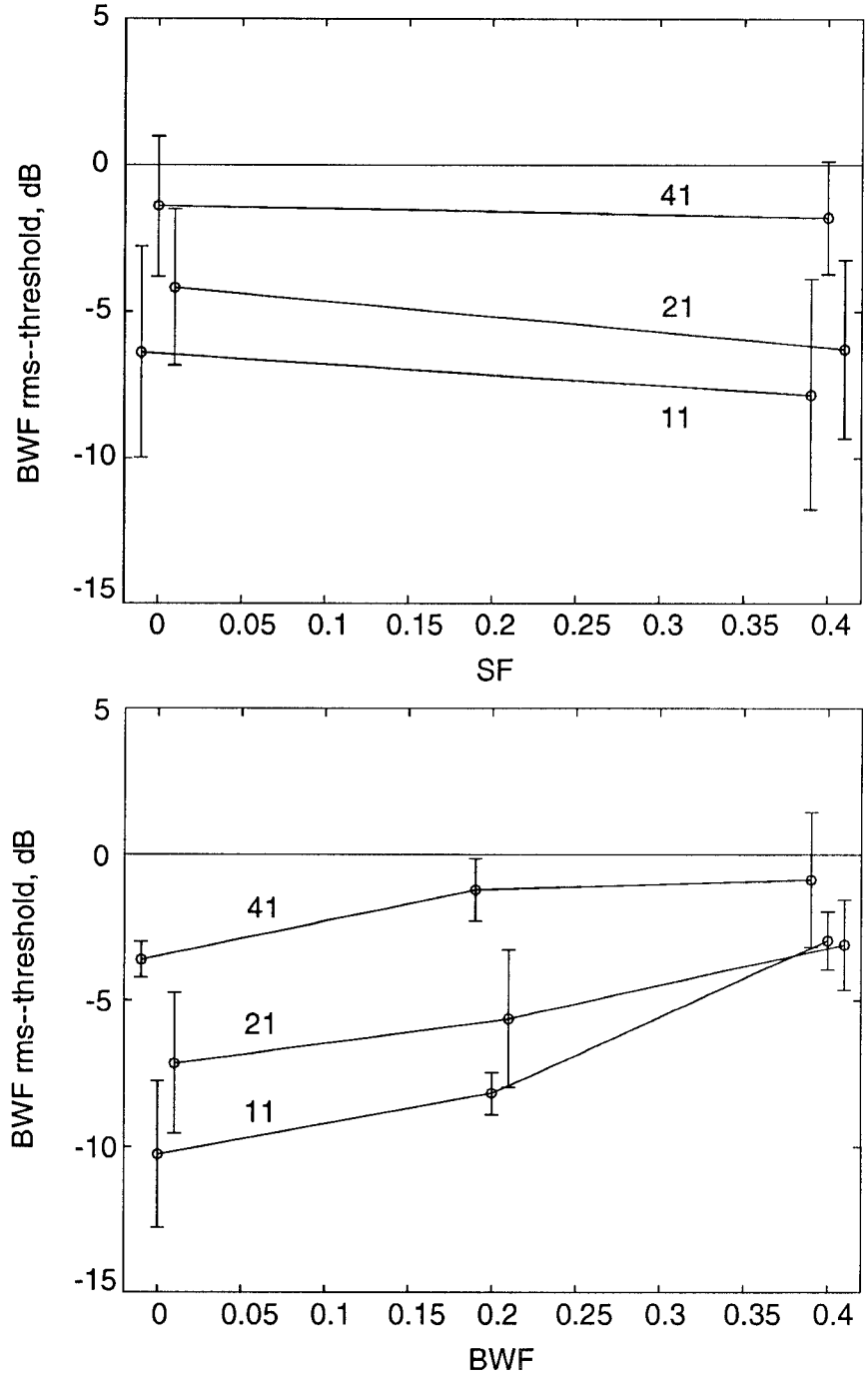

Figure 8: A8: Bandwidth change detection thresholds for 41, 21, and 11 component complexes, averaged over three subjects and three BWF's (a) and two SF's (b). rms–Thresholds are in general higher for 41 than for 21 and 11 component cases.

Appendix II: Brief review of the Ewaif model

An analytic function $m(t)$ with envelope $|m(t)|$ and phase $\phi(t)$ is related to a real waveform $s(t)$ with Hilbert transform $\hat{s}(t)$ ([*Oppenheim and Schafer*, 1990; *Papoulis*, 1962; *Voelcker*, 1966]), as:

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

where

$$|m(t)| = \sqrt{s^2(t) + \hat{s}^2(t)},$$

and

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

The equivalent pitch of a complex sound is defined by the Ewaif model as:

$$\text{Ewaif} = \frac{\int_0^T |m(t)|\ \text{instf}(t)dt}{\int_0^T |m(t)|dt},$$

where T is a stimulus duration, and instf(t) is an instantaneous frequency of $s(t)$, defined as instf$(t) = \frac{1}{2\pi}\frac{d\phi}{dt}$.
For our stimulus of n components, with k^{th} component amplitude, frequency and phase denoted as: a_k, f_k, and φ_k, respectively, the above definition translates to:

$$\text{Ewaif} = \frac{\int_0^T (\sum_{k=1}^{n} a_k^2 f_k + \sum_{k=1}^{n-1}\sum_{j=k+1}^{n} a_k a_j (f_k + f_j)\ \cos(2\pi(f_k - f_j)t + \varphi_k - \varphi_j))\ e(t)dt}{\int_0^T e(t)dt}$$

where

$$e(t)^2 = \sum_{k=1}^{n} a_k^2 + 2\sum_{k=1}^{n-1}\sum_{j=k+1}^{n} a_k a_j \cos(2\pi(f_k - f_j)t + \varphi_k - \varphi_j).$$

Appendix III: 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 $p(\omega)$. This is the expression used in computing the profiles in Figs. 11. A simpler expression can be used for the case of small θ_o ($\cos(\theta_o) \approx 1$ and $\sin(\theta_o) \approx \theta_o$):

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

REFERENCES

Assmann, P. F. and Q. Summerfield, Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency, *J. Acoust. Soc. Am.*, *85*, 327–338, 1989.

Berg, B. G., Q. T. Nguyen and D. M. Green, Discrimination of narrow-band spectra. I: Spectral weights and pitch cues, *J. Acoust. Soc. Am.*, *92*, 1911–1918, 1992.

Bernstein, L. R. and D. M. Green, Detection of simple and complex changes of spectral shape, *J. Acoust. Soc. Am.*, *82(5)*, 1587–1592, 1987.

Bernstein, L. R., V. M. Richards and D. M. Green, in *Auditory Processing of Complex Sounds: The detection of spectral shape change*, pp. 6–15, Lawrence Erlbaum Associates, Inc., New Jersey, 1987.

Durlach, N. I., L. D. Braida and Y. Ito, Towards a model for discrimination of broad-band signals, *J. Acoust. Soc. Am.*, *80 (1)*, 63–72, 1986.

Feth, L. L., H. O'Malley and J. J. Ramsey, Pitch od unresolved two-component complex tones, *J. Acoust. Soc. Am.*, *72*, 1403–1412, 1982.

Feth, L. L. and L. J. Stover, in *Auditory Processing of Complex Sounds: Demodulation processes in auditory perception*, pp. 76–86, Lawrence Erlbaum Associates, Inc., New Jersey, 1987.

Green, D. M., 'Frequency' and the detection of spectral shape change, in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson, pp. 351–359, Plenum Press, Cambridge, 1986.

Green, D. M., in *Profile Analysis*, Oxford Press, New York, 1988.

Green, D. M. and C. R. Mason, Auditory Profile Analysis: Frequency, Phase, and Weber's Law, *J. Acoust. Soc. Am.*, *77*, 1155–1161, 1985.

Green, D. M., C. R. Mason and G. Kidd, Profile Analysis: Critical Bands and Duration, *J. Acoust. Soc. Am.*, *75*, 1163–1167, 1984.

Hillier, D. A., Auditory Processing of Sinusoidal Spectral Envelopes, Ph.D. Dissertation, The Washington University and Severn Institute, 1991.

Kidd, G. J., C. R. Mason and D. M. Green, Auditory profile analysis of irregular sound spectra, *J. Acoust. Soc. Am.*, *75*, 1163–1167, 1984.

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

Levitt, W., Transformed Up-Down Methods in Psychoacoustics, *J. Acoust. Soc. Am.*, *49*, 467–477, 1971.

Oppenheim, A. and R. Schafer, in *Digital Signal Processing*, Prentice-Hall, New Jersey, 1990.

Papoulis, A., in *The Fourier Integral and its Applications*, McGraw-Hill, New York , 1962.

Richards, V. M., Z. A. Onsan and D. M. Green, Auditory profile analysis: Potential pitch cues, 1989.

Schreiner, C. and J. Mendelson, Functional topography of cat primary auditory cortex: Distribution of integrated excitation., *J. Neurophysiol.*, *64(5)*, 1442–1459, 1990.

Shamma, S. A., J. W. Fleshman, P. R. Wiser and H. Versnel, Organization of Response Areas in Ferret Primary Auditory Cortex, *J. Neurophysiol.*, *69* , 367–383, 1993.

Stover, L. J. and L. L. Feth, Pitch od narrow-band signals, *J. Acoust. Soc. Am.*, *73*, 1701–1707, 1983.

Voelcker, H. B., Towards a Unified Theory of Modulation Part I: Phase-Envelope Relationship, *Proc. of IEEE*, *54*(3), 340–353, 1966.

Vranić-Sowers, S. and S. A. Shamma, Representation of spectral profiles in the auditory system. Part II: A ripple analysis model, *J. Acoust. Soc. Am. (submitted)*, , 1993.