TECHNICAL RESEARCH REPORT

Representation of Spectral Profiles in the Auditory System Part II: A Ripple Analysis Model

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

T.R. 93-77

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

Based on experimental results presented in [*Vranić-Sowers and Shamma*, 1993], and on further physiological and psychoacoustical evidence, it is argued that the auditory system analyzes a spectral profile along two largely independent dimensions. They correspond to the magnitude and phase of a localized *Fourier* transformation of the profile, closely analogous to the spatial frequency transformations described in the visual system. Within this general framework, a model of profile analysis is proposed in which a spectral profile is assumed to be represented by a weighted sum of sinusoidally modulated spectra (ripples). 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 ripple detection experiments in [*Green*, 1986; *Hillier*, 1991]. Perceptual thresholds are then computed from the filter outputs and compared with thresholds measured for peak profile experiments, and for detection tasks with step, single component increment, and the alternating profiles.

INTRODUCTION

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 [*Durlach, Braida and Ito*, 1986], the maximum difference model [*Bernstein and Green*, 1987], the spectral slope model [*Klatt*, 1982], and the spectral peak model [*Assmann and Summerfield*, 1989]. 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 [*Bernstein and Green*, 1987; *Durlach, Braida and Ito*, 1986], 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 [*Vranić-Sowers and Shamma*, 1993]. 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

[*Campbell and Robson*, 1968; *Levine*, 1985]. It is, however, the elegant anatomical and physiological work of [*Valois and Valois*, 1990] 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 [*Calhoun and Schreiner*, 1993; *Shamma, Versnel and Kowalski*, 1993]. Furthermore, that the ordered mappings of response area bandwidths and asymmetries along the isofrequency planes of AI [*Schreiner and Mendelson*, 1990; *Shamma et al.*, 1993] 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 [*Green*, 1986; *Hillier*, 1991; *Houtgast and Veen*, 1982]. 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 Part I of this paper [*Vranić-Sowers and Shamma*, 1993]. 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. II. Its various parameters are estimated in Sec. III using detection thresholds reported in [*Green*, 1986; *Hillier*, 1991]. The model is used in Secs. IV and V to predict the results of several profile analysis experiments. Finally, the model is discussed in relation to auditory percepts such as timbre and pitch, and in contrast to other profile analysis prediction models.

A. 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., $p(\omega) = \sin(2\pi\Omega\omega + \Phi)$), on the logarithmic frequency axis, ω. A ripple has a ripple frequency Ω (in cycle/octave), and a ripple phase Φ (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.

I. RESTATEMENT OF SPECTRAL PEAK RESULTS

As discussed earlier in [*Vranić-Sowers and Shamma*, 1993] (Sec. V), 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.

A. 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. 10(a) and (b) in [*Vranić-Sowers and Shamma*, 1993]).

Consequently, the δBWF/BWF thresholds from Figs. 6 in [*Vranić-Sowers and Shamma*, 1993] 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 [*Hillier*, 1991]. These are discussed in detail and related to our measurements later in Sec. IV A.

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

As discussed earlier (Sec. V, Figs. 10(c) and (d), in [*Vranić-Sowers and Shamma*, 1993]), 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. 3 in [*Vranić-Sowers and Shamma*, 1993] as follows: *For broad peak profiles (BWF* ≥ 0.2*), subjects detect a* $\delta SF \approx 0.11$*, 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 III in [*Vranić-Sowers and Shamma*, 1993] 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 [*Vranić-Sowers and Shamma*, 1993] (Sec. VI), and will be discussed in relation to the δSF thresholds later in Sec. V.

II. 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.

A. 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. 1(a) and (c)). Therefore,

$$H(\Omega; \Omega_o) = G(\Omega - \Omega_o) + G(\Omega + \Omega_o), \tag{1}$$

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 $g(\omega) = g_o(\omega)\ e^{\mp j2\pi\Omega_o\omega}$, 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. 1(b)), is a Gaussian weighted sinusoid known as the Gabor–function ([*Gabor*, 1946]), 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:

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

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 $p(\omega)$:

$$\begin{aligned} r(\omega_o, \Omega_o, \Phi_o) &= \int h(\omega_o - \omega; \Omega_o, \Phi_o) p(\omega) d\omega \\ &= \int H(\Omega; \Omega_o, \Phi_o, \omega_o) P(\Omega) e^{j2\pi\Omega\omega_o} d\Omega. \end{aligned} \tag{3}$$

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. Further details on the computations of $r(\cdot)$ are found in Appendix I.

In the remainder of this paper, Eq. (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) is to compute the filter outputs only around the center of the profile ($\omega_o = 0$), i.e., to suppress the index ω_o. Finally, all profiles considered will be even or odd symmetric around their center, so that their ripple spectra $P(\Omega)$ have constant phase as a function of Ω. These simplifications lead to:

$$r(\Omega_o) = \int H(\Omega - \Omega_o)\ |P(\Omega)| d\Omega. \tag{4}$$

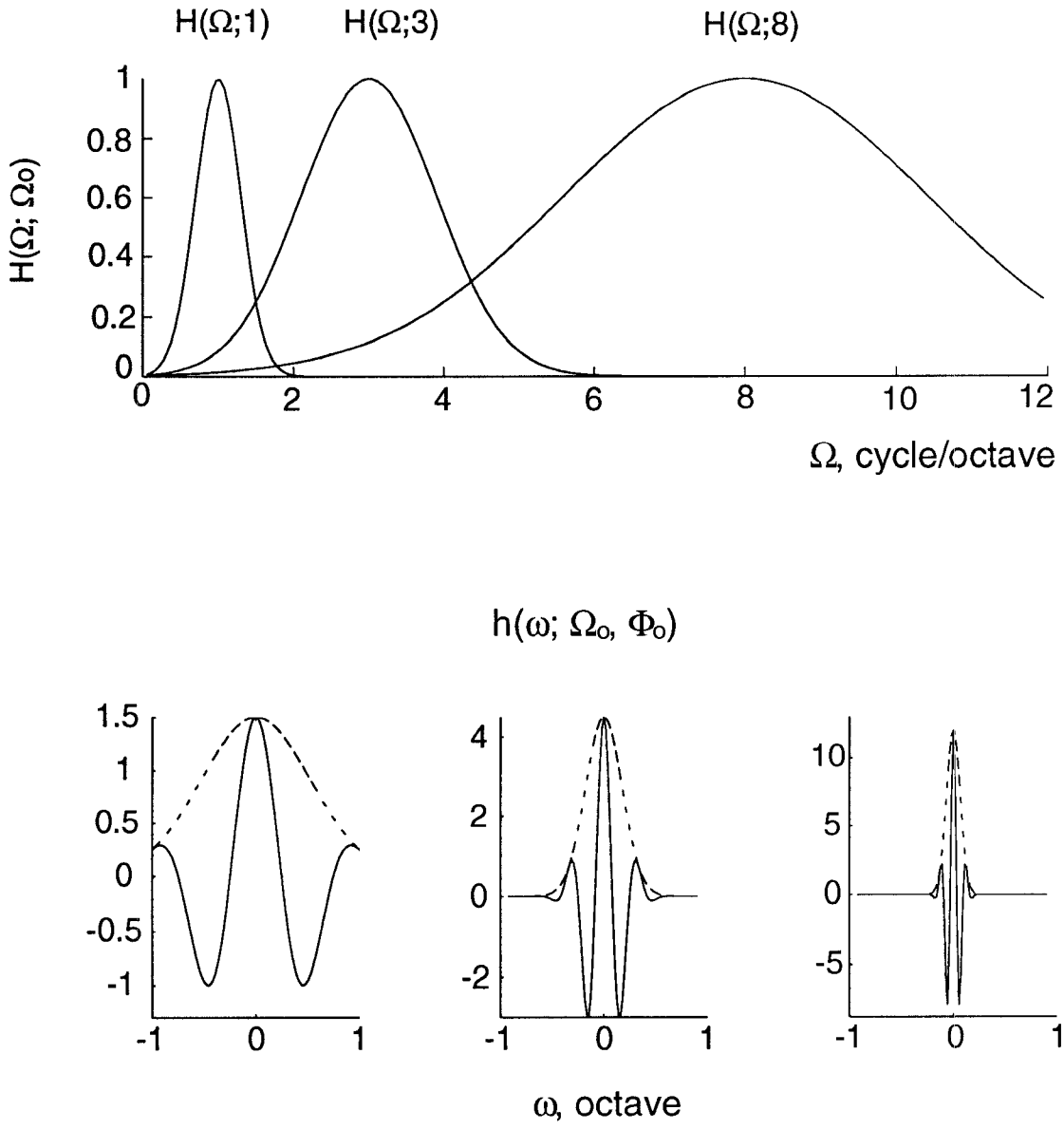

Figure 1: (a) Magnitude of three ripple filters centered at $\Omega_o = 1, 3$, and 8 cycle/octave. (b) Inverse Fourier transforms for the three filters in (a) centered at $\omega_o = 0$, with $\Phi_o = 0$ The inverse is computed as $h(\omega; \Omega_o) = 2\sqrt{2\pi}\sigma(\Omega_o)e^{-\frac{(2\pi\omega\sigma(\Omega_o))^2}{2}}\cos(2\pi\Omega_o\omega)$,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

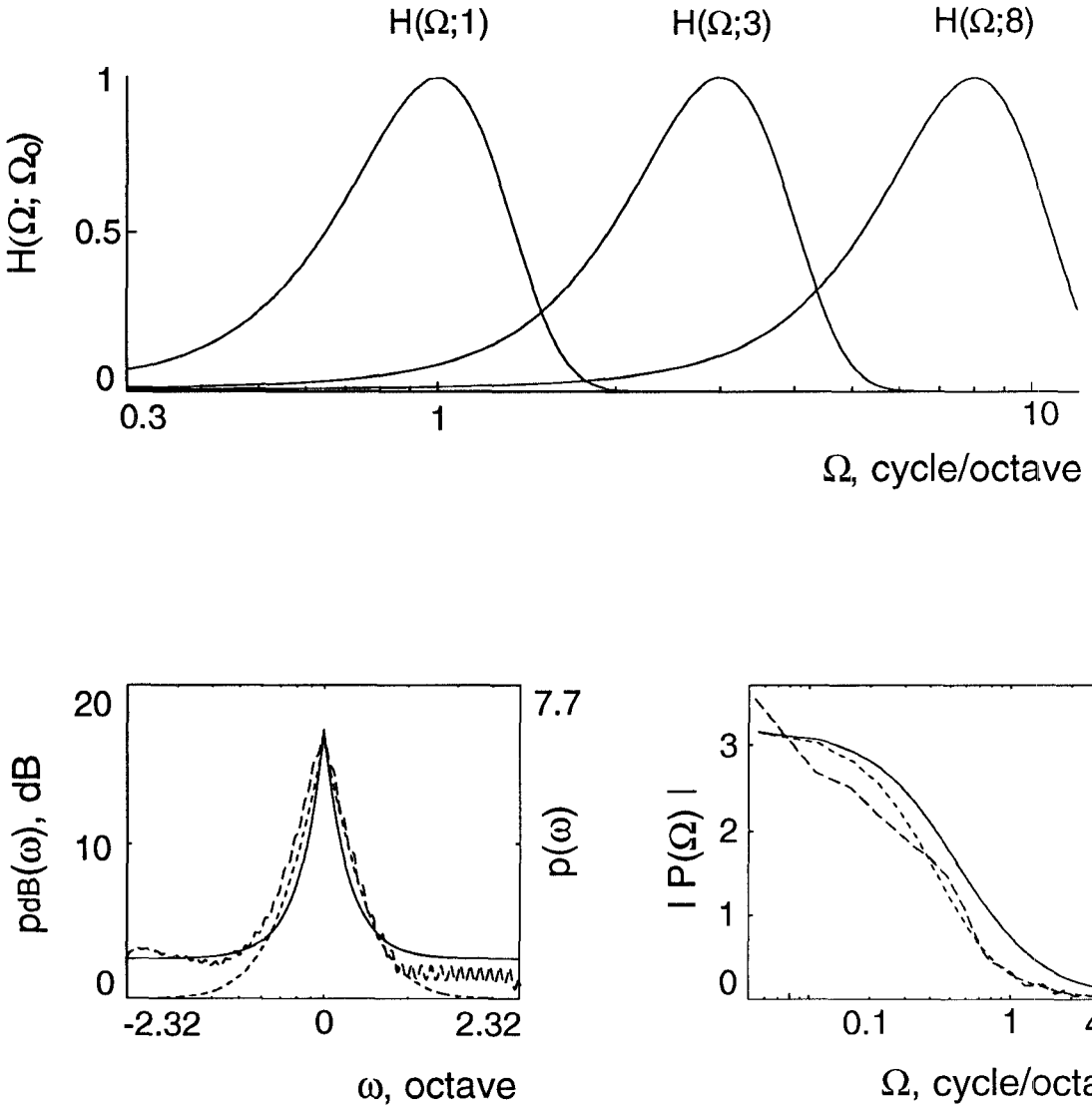

Figure 1: (c) Same as (a) but plotted on a logarithmic Ω axis. (d) Three input representations of a symmetric peak profile with BWF = 0.2 and A_{max} = 15 dB (left), and their corresponding ripple spectra (right). There are little differences between the three representations or their ripple spectra. 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 [*Glasberg and Moore*, 1990] for details).

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. V.

Finally, the filter bank depicted in Figs. 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 [*Hillier*, 1991] and visual gratings [*Valois and Valois*, 1990]), and neurophysiological experiments [*Valois, Albrecht and Thorell*, 1982] in which filter tuning was estimated around various Ω_o's to be approximately 1 octave (measured at the 3 dB points). This corresponds approximately to choosing σ's as $\sigma(\Omega_o) = 0.3\ \Omega_o$.

B. 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 [*Hillier*, 1991; *Shannon*, 1992]. 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 [*Glasberg and Moore*, 1990], and the auditory filter models of [*Chadwick et al.*, 1986; *Hillier*, 1991; *Patterson*, 1986; *Slaney and Lyon*, 1990; *Yang, Wang and Shamma*, 1992].

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 [*Green*, 1986; *Hillier*, 1991; *Houtgast and Veen*, 1982]. 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. 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. VI A.

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.

C. Different types of detection procedures

Figs. 2 illustrate the output of the filter bank (according to Eq. (4)) 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 [*Bernstein and Green*, 1987; *Bernstein, Richards and Green*, 1987], and the ripple detection experiments (called here ripple intensity-difference-limen experiments, or ripple-**idl**) by [*Green*, 1986; *Hillier*, 1991].

Examples of the second category are: The pedestal-type of experiments, such as increments on a single component pedestal [*Green*, 1988] or on a peak profile pedestal (Sec. IV B. in [*Vranić-Sowers and Shamma*, 1993]); The δSF and ripple-**pdl** detection (Secs. II and VI, in [*Vranić-Sowers and Shamma*, 1993]); The δBWF and ripple-**fdl** tests [*Hillier*, 1991; *Vranić-Sowers and Shamma*, 1993].

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

1. Detection of profiles with flat standards

We first use the results of the ripple-**idl** thresholds [*Hillier*, 1991] to determine K (Sec. III A.). 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. IV).

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. 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 $r(\cdot)$ [*Vranić-Sowers and Shamma*, 1993]. 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. While other features of $r(\cdot)$ are theoretically equivalent, our choice is motivated by the fact that the ripple transforms of arbitrary

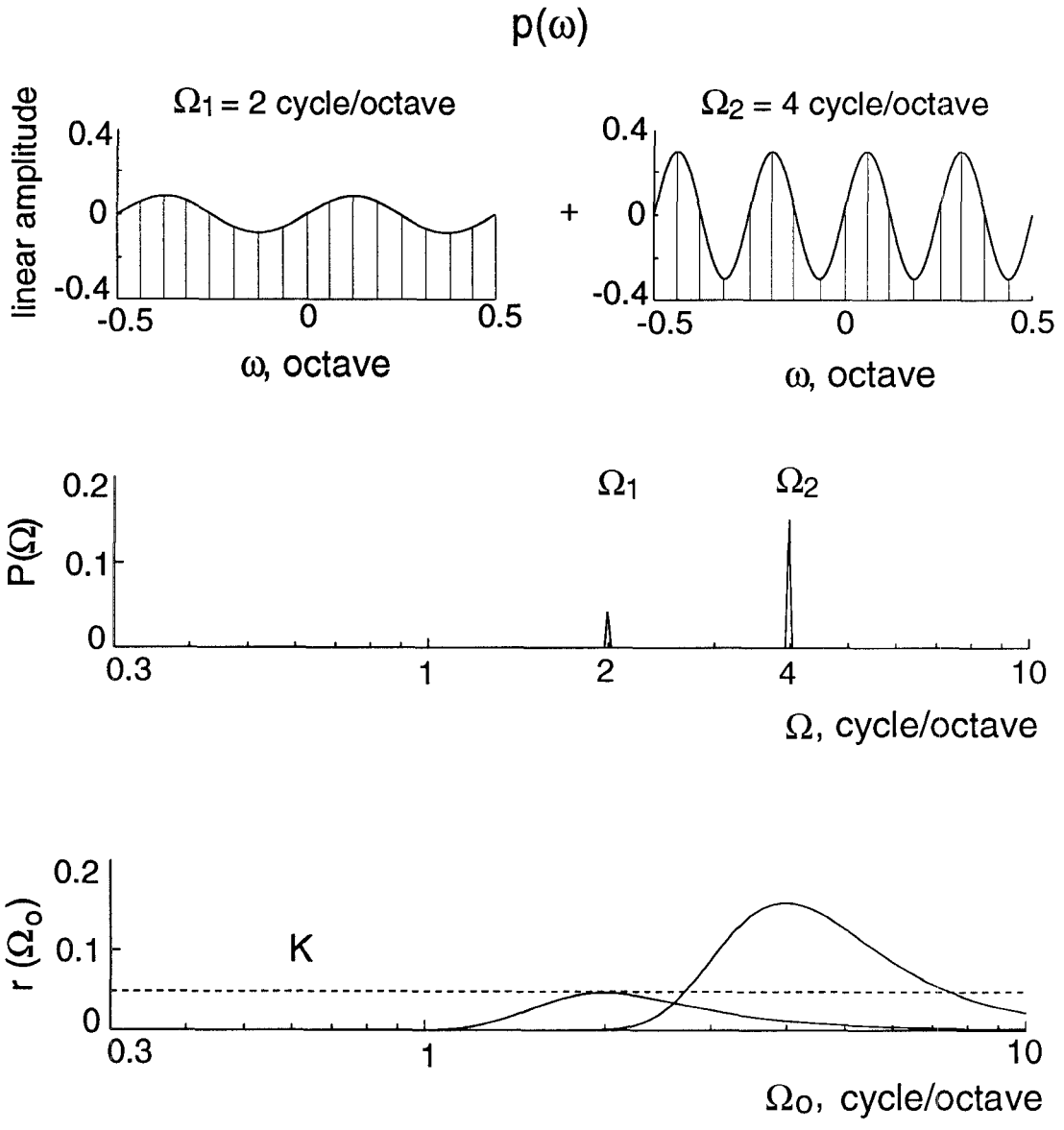

Figure 2: (a) Two single ripple profiles with ripple frequencies $\Omega_1 = 2$ and $\Omega_2 = 4$ cycle/octave, and amplitudes 0.1 and 0.3, respectively. The Ω_1 ripple is at its just detectable level ([*Hillier*, 1991]). (b) Ripple spectrum of the two ripples. (c) Ripple transform for the profiles in (a) (according to Eq. (4)). The $K = 0.05$ value corresponds to the detection threshold of the Ω_1 ripple at the output $r(\cdot)$.

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 [*Hillier*, 1991] (Fig. 3 and Sec. III B).

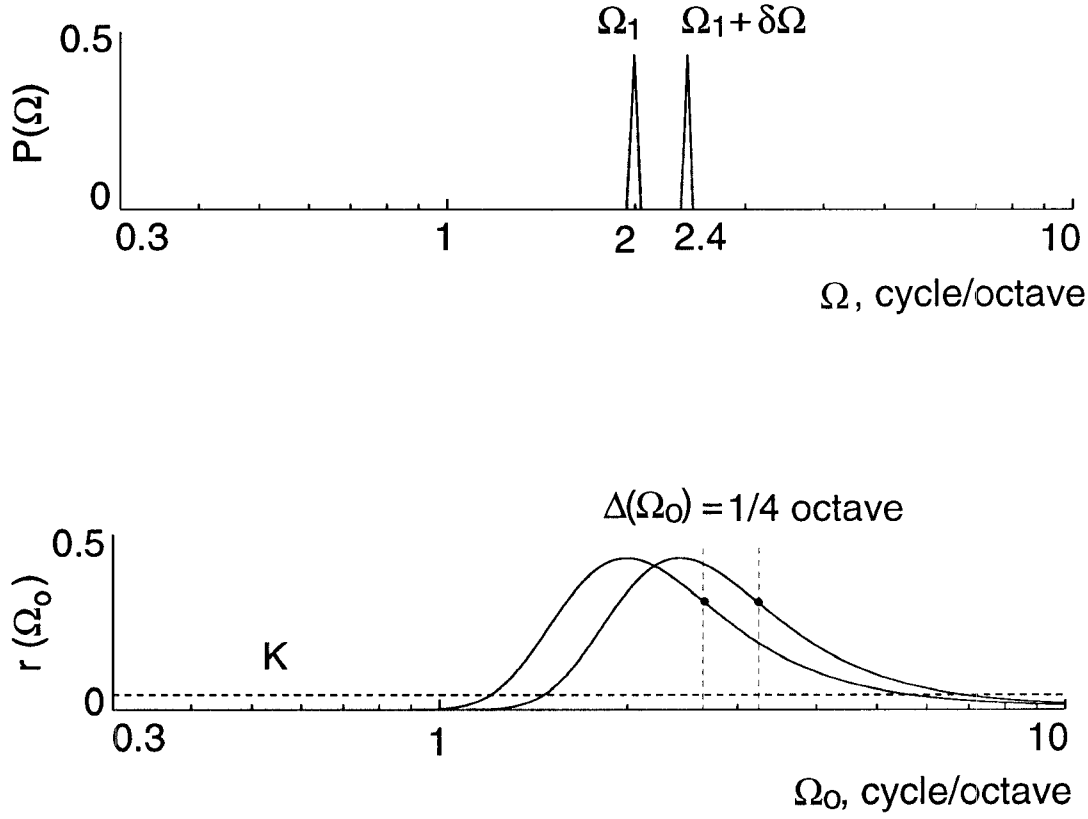

Figure 3: A two ripple spectrum and its ripple transform. The ripple frequencies are $\delta\Omega = 20\%$ apart (**fdl**–threshold, [*Hillier*, 1991]), with $\Omega_1 = 2$ and $\Omega_1 + \delta\Omega = 2.4$ cycle/octave. The amplitudes are well above K. The (vertical) dashed lines mark the locations of the steepest lowpass edges of $r(\cdot)$.

The filter output, $r(\cdot)$, as computed from Eq. (4) above, reflects only the magnitude of ripple transform as a function of Ω (see Appendix I). Therefore, it is insensitive to the ripple phase and cannot be used to predict δSF and ripple-**pdl** thresholds. Instead, these are briefly related to the model in Sec. V (and can be explicitly computed as in Appendix I).

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. IV E, the model predictions for the two pedestal-type experiments mentioned above will be discussed.

3. 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 [*Durlach, Braida and Ito*, 1986] and the more specific model of [*Bernstein and Green*, 1987] 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. IV B.

III. DETERMINING THE PERCEPTUAL THRESHOLD PARAMETERS OF THE RIPPLE ANALYSIS MODEL

In this section, the results of the ripple-**idl** and -**fdl** experiments by [*Green*, 1986; *Hillier*, 1991] 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 ($\approx$ 1 octave). The filter outputs $r(\cdot)$ are computed as in Eq. (4).

A. Ripple idl's

Fig. 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 [*Hillier*, 1991] and reproduced in Fig. 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' threshold 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 as shown in Fig. 4(b). Note that this is equivalent to assuming a variable σ_n, analogous to the model of [*Bernstein and Green*, 1987].

B. Ripple fdl's

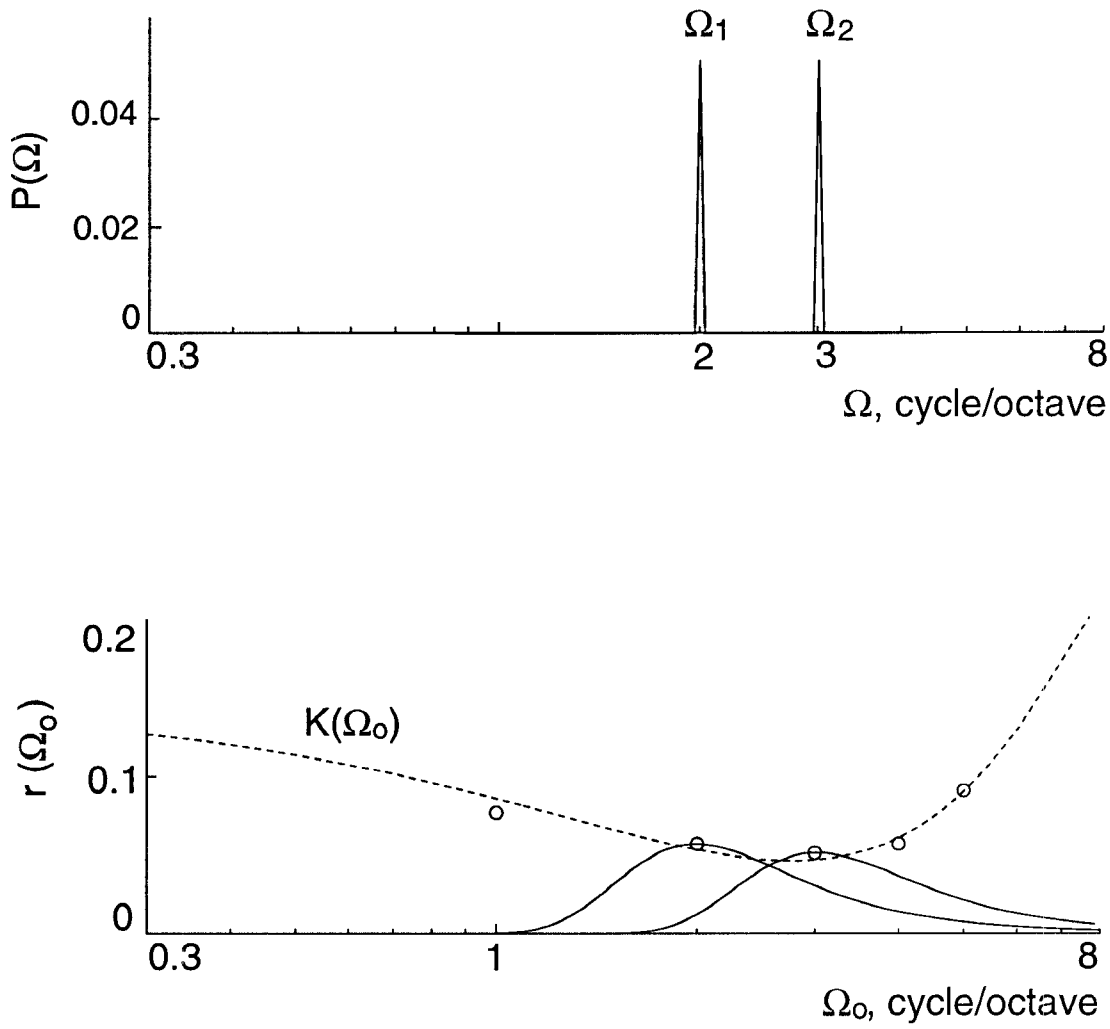

Figure 4: (a) A two ripple spectrum and (b) its ripple transform. Both ripples are at their **idl**–threshold values. The dashed line is a polynomial approximation to the measured data points (denoted by circles) reproduced from Fig. 3.27 in [*Hillier*, 1991]. The detection threshold $K(\Omega_o)$ reflects the shape of the perceptual threshold, $\mathbf{idl}(\Omega_o)$.

In **fdl** tests, input ripples are large enough such that $r(\cdot)$ is well above the perceptual threshold $K(\cdot)$. As demonstrated earlier (Sec. II C.2, Fig. 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 [*Hillier*, 1991] in Fig. 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. 5), i.e.:

$$\Delta(\Omega_o) = \log_2(1 + \mathbf{fdl}(\Omega_o)) \quad \text{(octave)}. \tag{5}$$

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.

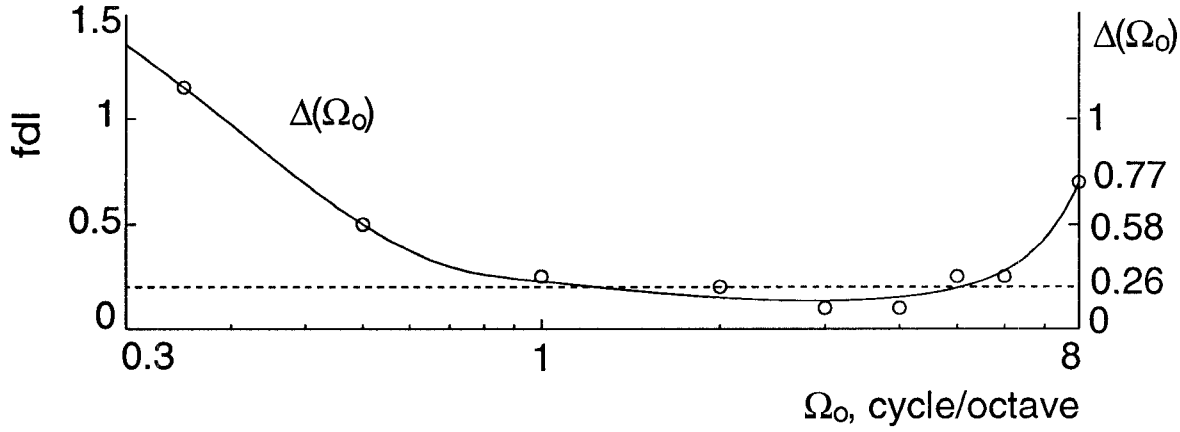

Figure 5: Interpolated **fdl**–thresholds (denoted by circles) reproduced from Fig. 3.30 in [*Hillier*, 1991]. The corresponding $\Delta(\Omega_o)$ values (Eq. (5)) are shown on the right scale.

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 (≤ 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 (≥ 6 cycle/octave) may be related 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. 5 (Eq. (5)).

C. Summary of the ripple analysis model

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. (4). 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. 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. (5) and Fig. 5). For *pedestal-type experiments*, thresholds are computed from the percentage scale change of $r(\cdot)$).

IV. PREDICTIONS OF THE RIPPLE ANALYSIS MODEL FOR VARIOUS INPUT PROFILES

In this section, the model is used to account for the perceptual thresholds measured in a wide range of profile analysis experiments. 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 [*Richards, Onsan and Green*, 1989]. Finally, two pedestal-type experiments are compared.

A. Predicting the δBWF/BWF thresholds for peak profiles

As explained earlier in Sec. I A, 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. 6 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[1]. These estimates compare well with the δBWF/BWF thresholds measured at approximately 0.22 (Sec. I A). Finally, the model also predicts that δBWF/BWF thresholds are independent of SF (Fig. 6(a) in [*Vranić-Sowers and Shamma*, 1993]), 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

[1]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.

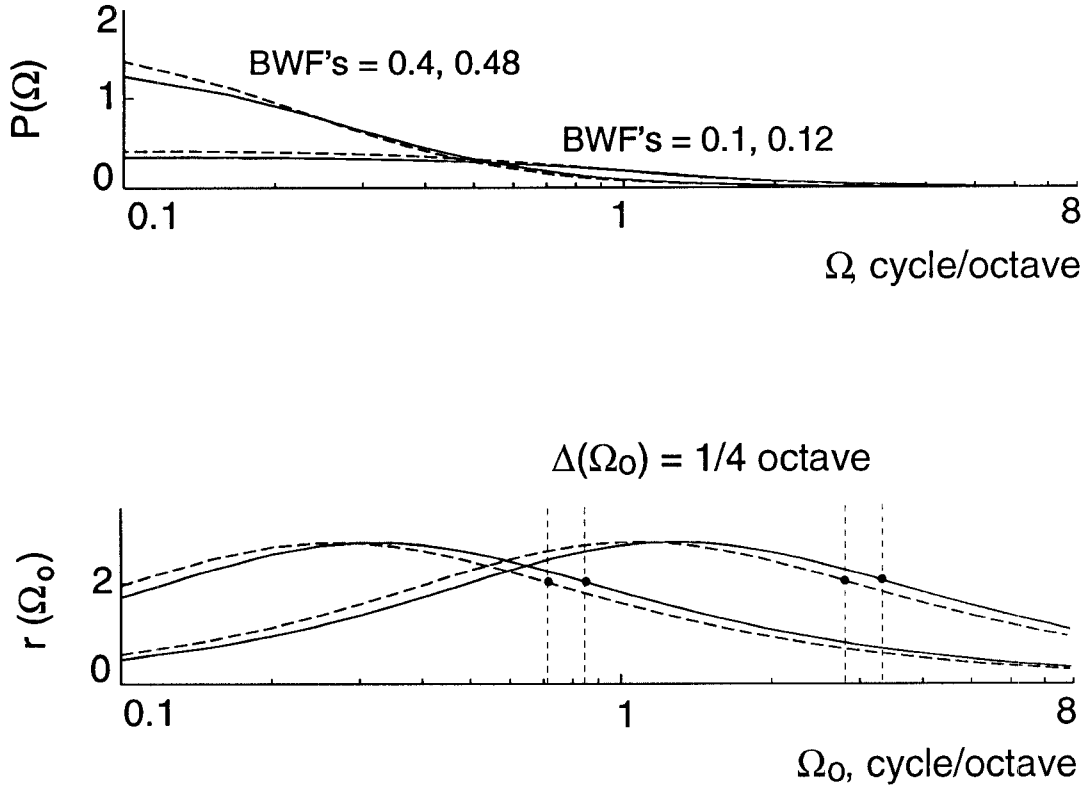

Figure 6: (a) Ripple spectra of symmetric peak profiles with BWF's = 0.1 and 0.4 (solid lines), and the corresponding profiles at δBWF/BWF = 20% perceptual thresholds (dashed lines). (b) The ripple transforms of the peak profiles in (a). The locations of the steepest lowpass edges are, for both cases, 1/4 octaves apart.

the above trends in dilation thresholds hold *regardless of the exact details of the profile shape* since it is the shift in the profile that is being detected. Clearly, this only holds if $r(\cdot)$ is well above the perceptual threshold $K(\cdot)$.

B. Predicting detection thresholds for an alternating profile

The alternating profile [*Bernstein and Green*, 1987] consists of 21 uniformly distributed components (0.2–5 kHz) that alternate above and below a flat (unit) base (Figs. 7).

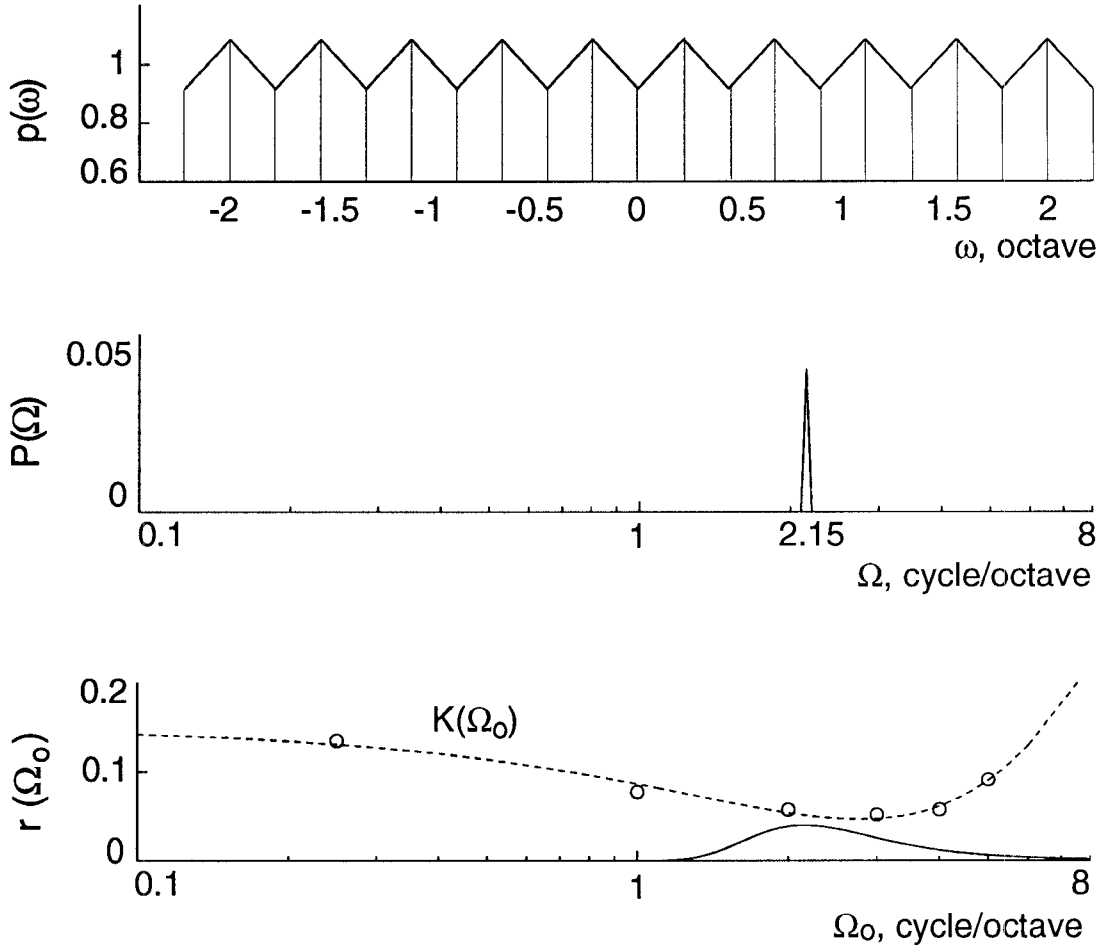

Figure 7: (a) The alternating profile at threshold amplitude (0.08, or –21.7 dB). (b) Ripple spectrum of the alternating profile in (a). (c) Ripple transform of the ripple spectrum in (b). The detection threshold $K(\Omega_o)$ in (c) is reached near 2.2 cycle/octave.

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. 7(c)).

C. Predicting detection thresholds for a step profile

The task in this experiment is to detect the presence of a (linear) step in a 21 component flat standard [*Bernstein and Green*, 1987] (Fig. 8(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 to the unit base). Figure 8(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[2] (Fig. 8(c), solid line). 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. 8). The smoothing of the ideal profile lowers the $P(\Omega)$ (Fig. 8(b)), and the corresponding $r(\cdot)$ is more lowpass filtered (Fig. 8(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 [*Bernstein and Green*, 1987]. 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)).

D. Predicting detection thresholds for a single component increment profile

In this experiment, a single component in the profile is incremented relative to the base [*Green*, 1988] (Fig. 9(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 9(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

[2]For a constant Q factor filter ($\sigma(\Omega_o) = \sigma_{rel}\ \Omega_o$, and in particular $\sigma_{rel} = 0.3$). For a step profile input (with magnitude of ripple spectrum $|P(\Omega)| = 1/\Omega$), filter outputs $r(\cdot)$ are independent of Ω_o.

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

where, $\Omega' = \Omega/\Omega_o$, and $r(\cdot)$ is a function of σ_{rel} only.

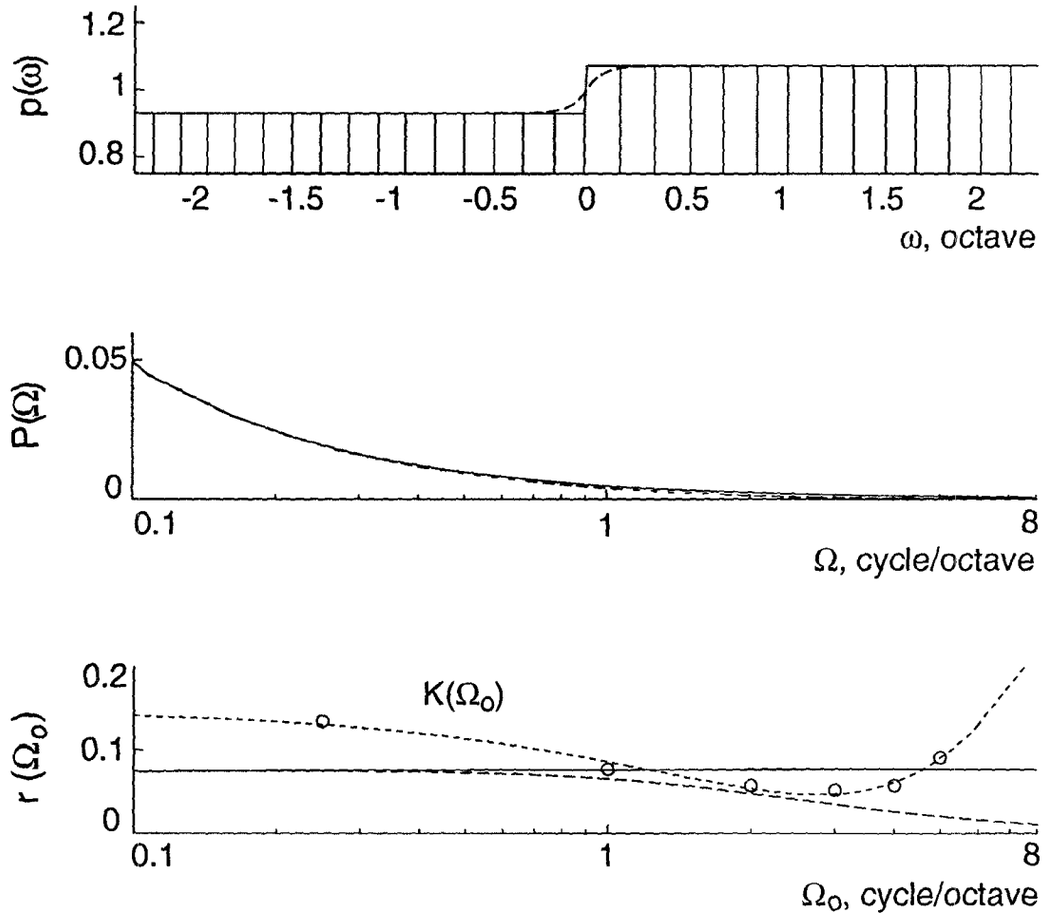

Figure 8: (a) Profile of a step function (solid line) at threshold amplitude (0.07, or –23.1 dB), and its smoothed version (dashed line). (The smoothed version is obtained by convolving the step with the narrow symmetric peak profile of BWF = 0.1 and A_{max} = −30 dB). The ripple spectra and ripple transforms are in (b) and (c), respectively. The magnitude of the ripple spectra is computed as: $|P(\Omega)| = 2a/|j\ 2\pi\Omega\ M|$, where a is the amplitude at threshold. 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.

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.

E. 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 [*Vranić-Sowers and Shamma*, 1993] (Sec. IV B., Fig. 9(a)) in which the pedestal is a peak profile. The second employs a single component pedestal (Fig. 5.4 in [*Green*, 1988]). 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 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. 9(b) and Table II(c) in [*Vranić-Sowers and Shamma*, 1993]). 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 [*Green*, 1988]).

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.

V. 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. (2) and (3), and Appendix I, 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

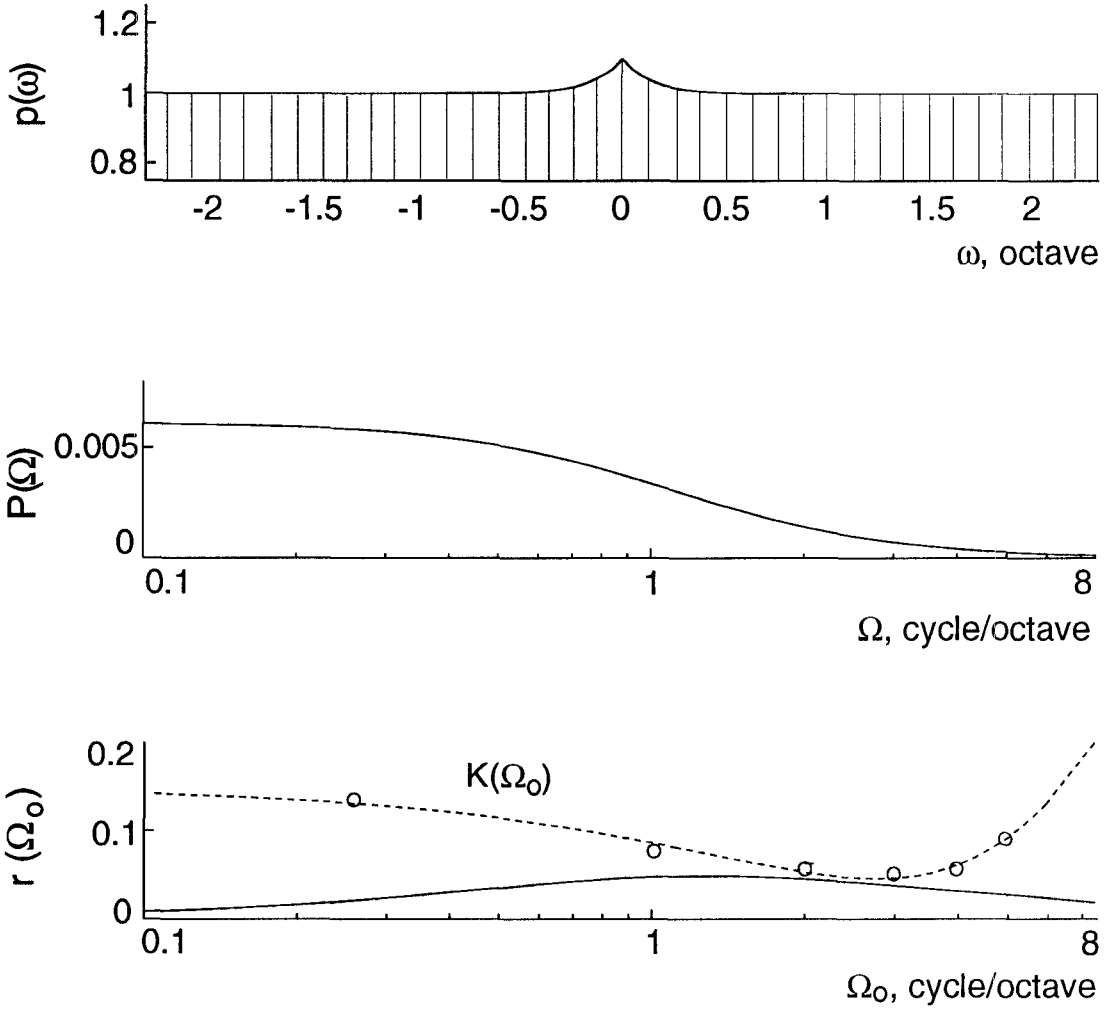

Figure 9: (a) Profile of a single increment on a flat base is at its perceptual threshold (0.09, or –20.1 dB). The ripple spectrum and ripple transform are shown in (b) and (c), respectively. The single increment is approximated with the BWF = 0.1 symmetric peak of –20.1 dB amplitude (see text). The detection threshold is reached approximately at 2.3 cycle/octave.

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. VI (Figs. 12 and Table II(d)) of [*Vranić-Sowers and Shamma*, 1993]. Results from these experiments reveal that for low frequency ripples (< 2 cycle/octave), subjects exhibit a constant phase shift of approximately 6^o. Thresholds increase for low amplitude ripples. They also increase linearly against the logarithmic ripple frequency axis above about 2 cycle/octave, i.e., subjects switch from a constant phase detection mode to a constant positional shift of the ripple peaks along the tonotopic axis[3]. 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 measured with single ripple stimuli are supported well by δSF measurements in peak profiles (Sec. II and Appendix III, in [*Vranić-Sowers and Shamma*, 1993]). As discussed earlier (Sec. I B), δSF changes in the peak profiles can be equivalently described as constant phase shifts in the ripple transform 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.

VI. DISCUSSION

A. 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.

[3]To see this argument more clearly, consider the case of a sinusoidal profile. Then, detecting a 6^o phase shift corresponds to detecting a certain horizontal shift of the pattern. For higher frequencies of the ripple, the horizontal (positional) shift corresponding to this phase sensitivity decreases. Eventually, at some high ripple frequency, the system reaches its detection limit and the observed phase shift at threshold begins to increase. The constant positional shift can be estimated from the slope of the detection threshold curves for ripples > 2 cycle/octave (Fig. 12(a) and Table II(d) in [*Vranić-Sowers and Shamma*, 1993]) to be 3.8^ooctave, or approximately 0.73% in Hz (which corresponds to 0.011 octaves).

1. 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 ([*Hillier*, 1991], Sec. 4.4) attempted to address this issue using adaptive experiments. Recent models of cochlear processing have also tackled this question [*Wang and Shamma*, 1993]. 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.

2. 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 [*Hillier*, 1991] (Secs. 4.4 and 4.5) and in psychophysical experiments in vision using analogous stimuli (summarized in [*Valois and Valois*, 1990]). 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. III, 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.

B. Distinguishing the ripple analysis model from other profile analysis models

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 [*Bernstein and Green*, 1987] and the independent channel models can account for a significant portion of the

same data [*Durlach, Braida and Ito*, 1986]. 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 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*).

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 $4/\pi$.

C. 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. (1), (2), and (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. 1(b)).

Since changes along any of the three axes are perceptually detectable, comparing two arbitrary profiles must be done along all three dimensions of the ripple representa-

tion. Such a metric might be considerably simpler in this domain (e.g., simple Euclidian 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 [*Assmann and Summerfield*, 1989] 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.

D. 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* [*Yost, Hill and Perez-Falcon*, 1978]. 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 [*Wilson*, 1970]. 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 [*Yost and Hill*, 1979].

The dominance region can be viewed in the context of the ripple analysis model as a correlate of the ripple **idl** sensitivity curve (Fig. 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 [*Bilsen et al.*, 1975; *Yost and Hill*, 1979].

E. 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 [*Valois and Valois*, 1990]), this approach has been supported by substantial anatomical, neurophysiological, and psychophysical evidence, elegantly detailed in [*Valois and Valois*, 1990]. 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 [*Oppenheim and Schafer*, 1990]. 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 [*Valois and Valois*, 1990]). 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.

Appendix I: Ripple filter response

Consider the impulse response of the ripple filter $h(\omega; \Omega_o, \Phi_o) = 2\, g_o(\omega) \cos(2\pi\Omega_o\omega - \Phi_o)$, centered at $\omega_o = 0$. Its Fourier transform is given by:

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

where,

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

and $\mathcal{H}$ is a Hilbert transformer defined as $\mathcal{H} = -\jmath\, sign(\Omega)$. Since $G(\Omega \pm \Omega_o) = e^{-\frac{(\Omega \pm \Omega_o)^2}{2\sigma^2}}$ (see text), $H(\Omega; \Omega_o)$ 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:

$$r(\Omega_o, \Phi_o) = \int_{-\infty}^{\infty} H(\Omega; \Omega_o, \Phi_o) P(\Omega) d\Omega = \int_{-\infty}^{\infty} H(\Omega; \Omega_o) |P(\Omega)| e^{\jmath(-sign(\Omega)\Phi_o + \theta(\Omega))} d\Omega$$
$$= 2 \int_0^{\infty} H(\Omega; \Omega_o) |P(\Omega)| \cos(\Phi_o - \theta(\Omega)) d\Omega.$$

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:

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

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

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

which is identical to $r(\Omega_o, \Phi_o)$ prior to dilation. Therefore, a dilation simply translates the filter outputs by $\log \alpha$ octaves, preserving their shape against a $\log \Omega_o$ 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 $r(\cdot)$. For example, in the case when $\theta(\Omega) = sign(\Omega)\theta_o$ (e.g., for symmetric or antisymmetric profiles), 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 (computed in Eq. (4)) is at $\Phi_o = \theta_o$:

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

Once again, for a pure dilation manipulation of an input profile (e.g., δBWF test for peaks at any SF such that $\theta(\Omega) = sign(\Omega)\theta_o$), 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).

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.

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.

Bilsen, F. A., 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, *J. Acoust. Soc. Am.*, *58*, 858–866, 1975.

Calhoun, B. M. and C. E. Schreiner, Spatial frequency filters in cat auditory cortex , *Neurosc. Meeting, November*, *19*, 581.8, 1993.

Campbell, F. W. and J. G. Robson, Application of Fourier Analysis to the visibility of gratings , *J. Physiol.*, *197*, 551–566, 1968.

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

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.

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

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

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.

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

Houtgast, T. and T. M. van Veen, On the just-detectable modulation of the spectral envelope on a log-f scale, *J. Acoust. Soc. Am. Suppl. 1*, *71*, S37, 1982.

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

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

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

Patterson, R. D., Auditory filters and excitation patterns as representations of frequency resolution, in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson, pp. 123–177, Plenum Press, Cambridge, 1986.

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.

Shamma, S. A., H. Versnel and N. A. Kowalski, Organization of Primary Auditory Cortex evident in responses to rippled complex sound stimuli, *Neurosc. Meeting, November*, *19*, 581.9, 1993.

Shannon, R., Personal correspondance, 1992.

Slaney, M. and R. F. Lyon, Perceptual Pitch Detector, *Proc. ICASSP*, *1*, 357–360, 1990.

Valois, R. L. D., Albrecht and Thorell, *Vision Research* , *22*, 545–559, 1982.

Valois, R. L. D. and K. K. D. Valois, in *Spatial Vision*, Oxford University Press, New York, 1990.

Vranić-Sowers, S. and S. A. Shamma, Representation of spectral profiles in the auditory system. Part I: Detection of spectral peak shapes and ripple phases, *J. Acoust. Soc. Am. (submitted)*, , 1993.

Wang, K. and S. A. Shamma, Self-Normalization and Noise Robustness in Early Auditory Represenatation, *IEEE Trans. Acoust. Speech Sig. Proc. (submitted)*, , 1993.

Wilson, J. P., An Auditory afterimage, in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. Smoorenburg, pp. 303–315, Sitthoff, Leiden, The Netherlands, 1970.

Yang, X., K. Wang and S. A. Shamma, Auditory Representations of Acoustic Signals, *IEEE Trans. on Information Theory*, *38*, 824–839, 1992.

Yost, W. A. and R. Hill, Models of the pitch and pitch strength of ripple noise, *J. Acoust. Soc. Am.*, *66(2)*, 400–410, 1979.

Yost, W. A., R. Hill and T. Perez-Falcon, Pitch and pitch discrimination in brodband signals with rippled power spectra, *J. Acoust. Soc. Am.*, *63(4)*, 1166–1173, 1978.