I. INTRODUCTION

Computational models of human peripheral frequency selectivity have a long history and a wide range of potential uses. They are a fast way to estimate excitation patterns along the cochlear partition in response to an arbitrary acoustic stimulus and have potential uses in, for example, signal compression algorithms, automatic speech recognition devices and simulations of individual patterns of hearing loss. Historically, the computations have used banks of linear gammatone filters (Patterson et al., 1992) but more recently researchers have attempted to use nonlinear filters (Giguère and Woodland, 1994; Lyon, 1982; Carney, 1993; Goldstein, 1990, 1995; Lopez-Poveda et al., 1998; Meddis et al., 2001). Much of what we know about the nonlinear aspects of peripheral frequency selectivity is derived from animal studies, particularly from direct observation of the response of the cochlear partition to acoustic stimulation (Rhode, 1971; Rhode and Recio, 2000; Robles et al., 1986, 1991; Sellick et al., 1982; Yates et al., 1990; Ruggiero et al., 1992). Nevertheless, this information is consistent with recent psychophysical observations showing changes with level of the filter’s best frequency (BF) (McFadden and Yama, 1983) and bandwidth (BW) (Glasberg and Moore, 1990; Rosen et al., 1998; for a review see Moore and Glasberg, 1987, and Moore, 1998). In this work we present a computational algorithm that was originally designed to accommodate animal observations and demonstrate how it can be adapted to simulate the hearing of a single human listener and the average hearing of a group of listeners.

The new filterbank aims to simulate the basilar membrane (BM) response using an array of point models (Lopez-Poveda et al., 1998; Meddis et al., 2001). It is similar in many respects to gammatone filterbanks that are in current use. The main difference is that each individual gammatone filter is replaced by a dual resonance nonlinear (DRNL) filter unit (to be described in detail later in this work). It has already been shown that the system can be used to model animal observations. The purpose of the present study is to show that, given suitable parameter changes, it can also be used to simulate human psychophysical data. The parameters of the DRNL units vary with respect to position along the cochlear partition but are fixed with respect to the intensity of the stimulus. Nevertheless, it is an emergent property of the system that the effective BW of each unit changes with signal level. The primary aim of the present study is to explore a new methodology for using human psychophysical data to identify parameters of the DRNL units. A second aim of the study is to identify mathematical functions that allow us to specify the parameters of DRNL units at any arbitrary BF.

Plack and Oxenham (2000) have recently published estimates of BM nonlinearity in normal listeners at BFs between 250 Hz and 8 kHz. These data will be used to help define the parameters of the DRNL units. They used a “pulsation threshold” paradigm (Houtgast, 1972), which is based on an auditory illusion whereby an interrupted sound is perceived as being continuous if there is sufficient energy from another sound during the interruptions. Plack and Oxenham presented a stimulus consisting of a signal tone at BF rapidly alternated on three occasions by a low-frequency (LF) tone...
well as the average response of six listeners.

A. The outer/middle ear stage

A sound pressure waveform produces vibration of the tympanic membrane, which, in turn, induces the vibration of the stapes. Each of these two processes of the peripheral hearing is modeled by means of a linear-phase, 512-point, finite impulse response (FIR) filter. The coefficients of each FIR filter were obtained from empirical frequency responses by applying an inverse fast Fourier transform routine with MATLAB (The Mathworks, Inc., ver. 5.3).

The outer-ear (headphone pressure-to-eardrum pressure) frequency response is taken from Pralong and Carlile [1996, Fig. 1(e)] and is shown in Fig. 2(a). It corresponds to a typical human outer-ear pressure-gain function measured at a point close to the tympanic membrane when the stimulus is delivered through a pair of Sennheiser HD-250 Linear headphones. A response measured with a Sony MDR-V6 headset would have been preferred, as this is the model that Plack and Oxenham employed to collect their pulsation-threshold data. Unfortunately, such response is unavailable. However, both the MDR-V6 and HD-250 are circumaural headphones and have a fairly flat frequency response in the range of

III. THE MODEL

The overall structure of the model is shown in Fig. 1. It consists of two main stages: (i) an outer/middle-ear filter function, which transforms a headphone-delivered sound pressure waveform into a stapes velocity waveform, and (ii) a DRNL filter that simulates the BM velocity of vibration in response to stapes velocity.

A. The outer/middle ear stage

A typical human headphone-to-eardrum sound pressure gain (from Pralong and Carlile, 1996, Fig. 1E) is assumed to be a typical human outer-ear pressure-gain function measured at a point close to the tympanic membrane when the stimulus is delivered through a pair of Sennheiser HD-250 Linear headphones. A response measured with a Sony MDR-V6 headset would have been preferred, as this is the model that Plack and Oxenham employed to collect their pulsation-threshold data. Unfortunately, such response is unavailable. However, both the MDR-V6 and HD-250 are circumaural headphones and have a fairly flat frequency response in the range of

III. THE MODEL

The overall structure of the model is shown in Fig. 1. It consists of two main stages: (i) an outer/middle-ear filter function, which transforms a headphone-delivered sound pressure waveform into a stapes velocity waveform, and (ii) a DRNL filter that simulates the BM velocity of vibration in response to stapes velocity.

A. The outer/middle ear stage

A sound pressure waveform produces vibration of the tympanic membrane, which, in turn, induces the vibration of the stapes. Each of these two processes of the peripheral hearing is modeled by means of a linear-phase, 512-point, finite impulse response (FIR) filter. The coefficients of each FIR filter were obtained from empirical frequency responses by applying an inverse fast Fourier transform routine with MATLAB (The Mathworks, Inc., ver. 5.3).

The outer-ear (headphone pressure-to-eardrum pressure) frequency response is taken from Pralong and Carlile [1996, Fig. 1(e)] and is shown in Fig. 2(a). It corresponds to a typical human outer-ear pressure-gain function measured at a point close to the tympanic membrane when the stimulus is delivered through a pair of Sennheiser HD-250 Linear headphones. A response measured with a Sony MDR-V6 headset would have been preferred, as this is the model that Plack and Oxenham employed to collect their pulsation-threshold data. Unfortunately, such response is unavailable. However, both the MDR-V6 and HD-250 are circumaural headphones and have a fairly flat frequency response in the range of

III. THE MODEL

The overall structure of the model is shown in Fig. 1. It consists of two main stages: (i) an outer/middle-ear filter function, which transforms a headphone-delivered sound pressure waveform into a stapes velocity waveform, and (ii) a DRNL filter that simulates the BM velocity of vibration in response to stapes velocity.

A. The outer/middle ear stage

A sound pressure waveform produces vibration of the tympanic membrane, which, in turn, induces the vibration of the stapes. Each of these two processes of the peripheral hearing is modeled by means of a linear-phase, 512-point, finite impulse response (FIR) filter. The coefficients of each FIR filter were obtained from empirical frequency responses by applying an inverse fast Fourier transform routine with MATLAB (The Mathworks, Inc., ver. 5.3).

The outer-ear (headphone pressure-to-eardrum pressure) frequency response is taken from Pralong and Carlile [1996, Fig. 1(e)] and is shown in Fig. 2(a). It corresponds to a typical human outer-ear pressure-gain function measured at a point close to the tympanic membrane when the stimulus is delivered through a pair of Sennheiser HD-250 Linear headphones. A response measured with a Sony MDR-V6 headset would have been preferred, as this is the model that Plack and Oxenham employed to collect their pulsation-threshold data. Unfortunately, such response is unavailable. However, both the MDR-V6 and HD-250 are circumaural headphones and have a fairly flat frequency response in the range of

III. THE MODEL

The overall structure of the model is shown in Fig. 1. It consists of two main stages: (i) an outer/middle-ear filter function, which transforms a headphone-delivered sound pressure waveform into a stapes velocity waveform, and (ii) a DRNL filter that simulates the BM velocity of vibration in response to stapes velocity.

A. The outer/middle ear stage

A sound pressure waveform produces vibration of the tympanic membrane, which, in turn, induces the vibration of the stapes. Each of these two processes of the peripheral hearing is modeled by means of a linear-phase, 512-point, finite impulse response (FIR) filter. The coefficients of each FIR filter were obtained from empirical frequency responses by applying an inverse fast Fourier transform routine with MATLAB (The Mathworks, Inc., ver. 5.3).

The outer-ear (headphone pressure-to-eardrum pressure) frequency response is taken from Pralong and Carlile [1996, Fig. 1(e)] and is shown in Fig. 2(a). It corresponds to a typical human outer-ear pressure-gain function measured at a point close to the tympanic membrane when the stimulus is delivered through a pair of Sennheiser HD-250 Linear headphones. A response measured with a Sony MDR-V6 headset would have been preferred, as this is the model that Plack and Oxenham employed to collect their pulsation-threshold data. Unfortunately, such response is unavailable. However, both the MDR-V6 and HD-250 are circumaural headphones and have a fairly flat frequency response in the range of
interest (100–8000 Hz). Hence, it is reasonable to assume that the outer-ear response of Fig. 2(a) is a good approximation to that measured with the MDR-V6 headset.

The middle-ear response (stapes velocity as a function of stimulus frequency) is shown in Fig. 2(b) for a stimulus level at the eardrum of 0 dB SPL. The data is derived from stapes displacement measurements in cadavers by Goode et al. (1994, Fig. 1) after sound pressure stimulation near the tympanic membrane. Consistent with the observations of Goode et al., peak stapes velocity is assumed to increase linearly with stimulus pressure. The range of empirical data points has been extrapolated from 400–6500 Hz to 100–10 000 Hz (see Fig. 2(b)) in order to be able to evaluate the model over a wider frequency range. The extrapolation is consistent with the measurements of Kringlebotn and Gundersen (1985).

The same outer- and middle-ear filters have been used throughout the modeling work described next.

B. The DRNL filter

Stapes motion transmits energy to the intracochlear fluid, which induces, in turn, motion of the BM. This process is modeled by a DRNL filter (Meddis et al., 2001) which simulates the velocity of vibration of a given site along the BM in response to a given stapes velocity waveform. Its structure and parameters are shown in Fig. 3(a). The input signal follows two independent paths, one linear and one nonlinear. In the linear path, a gain, \( g \), is applied and then the signal is filtered through a cascade of two or three, see later in this work, first-order gammatone (GT) filters (parameters: \( CF_{lin} \) and \( BW_{lin} \)) followed by a cascade of four second-order low-pass filters. In the nonlinear path, the input signal is filtered through a cascade of three first-order GT filters (parameters: \( CF_{nl} \) and \( BW_{nl} \)) followed by a nonlinear gain (see later in this work), followed by another cascade of three GT filters having the same parameters (\( CF_{nl} \) and \( BW_{nl} \)). During parameter estimation, the \( CF_{nl} \) is set to the frequency of the probe signal being studied and is not a free parameter. However, the \( CF \) of the linear path (\( CF_{lin} \)) is different and typically below \( CF_{nl} \) (see later in this work).

The nonlinear gain function is

\[
y(t) = \text{sign}(x(t)) \cdot \min[a|x(t)|, b|x(t)|^c],
\]

where \( x(t) \) and \( y(t) \) are the input and the output signals of the nonlinearity, respectively, and \( a, b, \) and \( c \) are parameters of the model. The details of the time-domain digital implementation of the DRNL filter are given in the Appendix.
The output from the DRNL filter is the sum of the output from the linear and the nonlinear paths. For consistency with its original description by Meddis et al., it is assumed to represent BM velocity.

Although the DRNL unit is simple in construction, the effects of a change in signal level on its properties are not immediately apparent. To help visualize these, two extra figures are supplied. Figure 3(b) shows separately the filter output as a function of signal frequency for the linear and the nonlinear path for a 30 dB SPL input signal. These functions are generated using the parameters to be used later for the 1-kHz site (subject YO, Table I). The combined output of the DRNL unit is shown as the thick continuous line and is simply the sum of the two filter functions. Note that the nonlinear filter function dominates the output and is the main determinant of the shape of the summed output function. Figure 3(c) shows the same set of functions but this time for an 85 dB SPL signal. The linear filter function has grown considerably while the nonlinear filter function has grown very little. This is because the nonlinearity is compressive. As a consequence, the output from the linear function dominates the aggregate output. The thick line representing the summed output of the DRNL unit is now a much wider filter function than that shown in Fig. 3(b). In addition, we can see that the BF of the DRNL unit has shifted to a lower frequency.

At very low signal levels, the DRNL unit operates linearly. This is because the nonlinear path is linear for low signal inputs [see Eq. (1)]. At very high signal levels, the DRNL also operates essentially as a linear filter. This is because the linear filter comes to dominate the output. These properties of the DRNL filter are consistent with the data of Plack and Oxenham (2000) which often show a linear response at low signal levels, followed by a compressive nonlinearity and then followed by a return to linearity at 80 dB SPL. A similar return to linearity is also sometimes found in observations of animal BM response, although its significance is disputed (Ruggero et al., 1996, Fig. 1; Johnstone et al., 1986, Fig. 5; Rhode and Cooper, 1996, Fig. 7). Note that the effect of signal level on filter width, BF and nonlinearity are emergent properties of a level-independent set of model parameters.

III. MODELING HUMAN BM NONLINEARITY FOR NORMAL-HEARING SUBJECTS

A. Method

The model was tuned to simulate human BM nonlinearity as estimated by psychoacoustical experiments of pulsation threshold (Plack and Oxenham, 2000). Plack and Oxenham presented subjects with a stimulus consisting of a number of interleaved segments of a signal tone and masker tone. Each individual segment was ramped up and down (2-ms raised-cosine ramps) and the frequency of the masker was 0.6 times the frequency of the signal. They measured signal frequencies of 250, 500, 1000, 2000, 4000, and 8000 Hz. For any given signal level, the task was to measure the masker threshold level at which the subject reports the signal to sound “pulsating” as opposed to continuous. Background noise was used to prevent off-frequency listening. Figure 4 symbols show their experimental findings for one subject (YO, filled circles) together with the average response of the six subjects considered in their experiment (open circles).

Here, it is assumed that pulsation threshold occurs when the signal and the masker produce the same amount of excitation at the place of the BM of maximum excitation to the signal frequency. For this reason, the paradigm for evaluating the model was a simplified version of that used by Plack and Oxenham. Instead of using a pulsating stimulus, the signal tone and the masker tone were passed independently through the model. Signal and masker frequencies were the same as those used by Plack and Oxenham. Both tones had a duration

<table>
<thead>
<tr>
<th>Signal frequency (Hz)</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. cascaded GT filters</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CF_{lin} (Hz)</td>
<td>235</td>
<td>460</td>
<td>945</td>
<td>1895</td>
<td>3900</td>
<td>7450</td>
</tr>
<tr>
<td>BW_{lin} (Hz)</td>
<td>115</td>
<td>150</td>
<td>240</td>
<td>390</td>
<td>620</td>
<td>1550</td>
</tr>
<tr>
<td>g</td>
<td>1400</td>
<td>800</td>
<td>520</td>
<td>400</td>
<td>270</td>
<td>250</td>
</tr>
<tr>
<td>LP_{lin} cutoff (Hz)</td>
<td>CF_{lin}</td>
<td>CF_{lin}</td>
<td>CF_{lin}</td>
<td>CF_{lin}</td>
<td>CF_{lin}</td>
<td>CF_{lin}</td>
</tr>
<tr>
<td>No. cascaded LP filters</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>DRNL filter BF (Hz)</td>
<td>260</td>
<td>508</td>
<td>1002</td>
<td>2006</td>
<td>3978</td>
<td>7720</td>
</tr>
<tr>
<td>DRNL filter BW_{sum} (Hz)</td>
<td>47</td>
<td>70</td>
<td>118</td>
<td>210</td>
<td>415</td>
<td>755</td>
</tr>
</tbody>
</table>
of 84 ms and were ramped up and down with 2-ms raised-cosine ramps. At each frequency, the signal level was varied within the ranges used by Plack and Oxenham. The sampling frequency \( f_s \) was 64 000 Hz.

At each BF, the model’s peak response to the signal \( O_S \) and the masker \( O_M \) was measured during the last half of the stimulus duration in order to avoid any effects that may occur at the onset. For each signal level \( L_S \), the task was to find out a masker level \( L_M \) such that the ratio \( O_S / O_M \) was equal to (or just exceeded) a value of one.

**B. Parameter optimization**

Two sets of model parameters were chosen so as to optimize the match between the model results and the two sets of experimental data shown in Fig. 4. However, the model contains a large number of parameters. Not all of them were allowed to vary. Some parameters were fixed by introducing constraints to the search space. These constraints are based on our current understanding of the cochlear response as explained later in this work.

CF\(_{nl}\) was fixed at the probe signal frequency. Consistent with Meddis et al. (2001), the cut-off frequencies of the low-pass filters in the linear and the nonlinear paths (LP\(_{lin}\) and LP\(_{nl}\)) were fixed equal to the CF of the GT filters in their respective paths (CF\(_{lin}\) and CF\(_{nl}\)).

The compression exponent, \( c \), was also fixed at 0.25 across BF. This may appear to be surprising in view of the fact that the slopes of the functions in Fig. 4 change with signal frequency. However, the slope of the psychophysical functions in Fig. 4 is only indirectly related to the compression function in Eq. (1). The psychophysical functions, in the model at least, are based on the summed activity of two components, only one of which is nonlinear. The linearity of one component dilutes the nonlinearity of the other in the output of the DRNL unit. We had previously found (Meddis et al., 2001) that a fixed value of the compression exponent, \( c \), was consistent with the variation observed in the slopes of the input/output functions in the animal data. By adopting the same strategy on this occasion, it was possible to produce a quantitative match to the data while benefitting from the need to estimate one parameter less.

Parameter \( a \) of the nonlinearity is responsible for the “sensitivity” at the tip of filter. It, therefore, determines the BM response at the absolute hearing threshold (AHT). There is evidence that AHT occurs at a BM velocity in the region of \( 5 \times 10^{-5} \) m/s (Ruggero et al., 1997). For this reason, \( a \) was fixed so that the DRNL filter output peak velocity is \( 5 \times 10^{-5} \) m/s at the subject’s hearing threshold.

Parameter CF\(_{lin}\) determines the BF of the DRNL filter at high levels, when the linear filter path dominates the DRNL output (see Fig. 3). There is physiological (e.g., Rhode and Recio, 2000) and psychophysical (McFadden and Yama, 1983) evidence of a shift in BF as a function of level. For this reason, an important initial constraint was set on CF\(_{lin}\) by requiring it to be lower than CF\(_{nl}\). The shift found by McFadden and Yama from 65 to 95 dB SPL gives a ratio BF\(_{65}\)/BF\(_{95}\) of approximately 1.1 to 1.4. This does not imply directly that the ratio CF\(_{nl}/CF_{lin}\) must be set within that range in the model. Because the DRNL filter is the sum of two components, each of which consists of a number of cascaded gammatone filters, followed by low-pass filters, its BF is not equal to either CF\(_{lin}\) or CF\(_{nl}\). However, by trial and error, we
Another important constraint was set by requiring the DRNL filter functions near threshold to have 3-dB-down bandwidths (BW_{nl}) that are consistent with the human psychophysical observations specified in the formula given by Glasberg and Moore (1990) (see later in this work). This criterion constrains BW_{nl} as the nonlinear filter path dominates the DRNL filter function at threshold. It does not mean, however, that BW_{nl} is set equal to the value given by Glasberg and Moore. Since the DRNL filter function at threshold is the result of a number of cascaded GT filters, the effective DRNL filter BW is somewhat lower than BW_{nl}, the BW of the individual GT filters. As a result, BW_{nl} must be greater than the human psychophysical bandwidths (see Tables I and II).

A weaker constraint was introduced on BW_{lin} so that the effective DRNL filter BW at high levels (when the linear filter dominates the DRNL filter output) is larger than at threshold, in agreement with psychophysical human filter functions. It was decided not to set any stronger numerical constraints on BW_{lin} as it deeply influences DRNL filter output to the masker and hence the psychophysical data functions in Fig. 4. These functions can be characterized (from left to right) as an initial linear, a compressed, and a final linear section. When concentrating on these separate features, the modeler was able to use the following rules when varying the remaining parameters (b, g, and BW_{lin}) to optimize the fit. The gain b was varied to set the height of the compressed section in the functions of Fig. 4. Finally, once parameter a has been fixed (see earlier in this work), the gain g together with BW_{lin} were set so that the height of the initial linearity could be reproduced and so that O_S (the response to the masker) is approximately equal to O_M (the response to the signal) at 85 dB SPL, in agreement with the psychophysical data of Fig. 4.

With the above mentioned constraints, the variable parameters were varied manually to optimize the fit by minimizing the Euclidean distance between the psychophysical data and the model results. Parameters to fit the data of subject YO (Table I) were found in the first place. These were then adjusted to find optimum parameters to fit the average pulsation thresholds in the Plack and Oxenham study (Table II). It is important to notice that across BFs the experimental average pulsation thresholds are 8 to 9 dB higher than those of subject YO (in Fig. 4, compare open and closed symbols). Based on our main assumption that pulsation threshold occurs when the masker and the signal produce equal levels of BM excitation (see also Houtgast, 1972, and Plack and Oxenham, 2000), this result implies that the auditory filters corresponding to the average response are necessarily steeper in their low-frequency tail (at 0.6×BF) than those of subject YO. In the model, this effect can be achieved in two ways: either by reducing BW_{lin} or by increasing the number of cascaded gammatone filters in the linear path. The necessary adjustments to BW_{lin} would require its value to be lower than BW_{nl}, hence violating one of the constraints set previously. It was, therefore, decided to allow the number of cascaded gammatone filters in the linear path to vary from two (for subject YO) to three (for the average response) in order to account for the “vertical” variability in pulsation thresholds. The issue is further discussed in Sec. IV.

### C. Creating a human filterbank

A computational model like the one presented here is expected to be used as part of larger models of the auditory periphery. This sort of application usually requires a filterbank rather than a discrete number of filters.
D. Results

1. Input/output curves

Figure 4 compares the experimental data (Plack and Oxenham, 2000, Fig. 2) and the model results for both data sets (subject YO and average responses). For the most part, the fit is quantitatively very good in both cases. At BF=250 Hz, the model masking function is mainly linear, while at 1000 Hz, the model shows a three stage function (linear, compressed, then linear again). At 8000 Hz, the model replicates the almost wholly compressed nature of the function except for the return to linearity at high signal levels. The only significant failures occur at BF=2000 Hz, where the model fails to simulate the linearity at 30 dB SPL for subject YO, and at BF=8000 Hz, where the model fails to simulate the linearity at 80–85 dB SPL. The failure at 2000 Hz is linked to one of the parameter constraints that fixes the tip sensitivity parameter, \( a \), to agree with the YO’s or the average hearing threshold. It may also be a consequence of the fact that the outer-ear or the cadaver-based, middle-ear function do not correspond to subject YO’s. The failure at 8000 Hz is linked to the deep notch in the outer-ear response [see Fig. 2(a)]. Pralong and Carlile (1996) pointed out that the greatest variability in outer-ear responses from trial to trial occurs precisely at around 8000 Hz. Their result is consistent with the findings of Kulkarni and Colburn (2000) and Möller et al. (1995). This suggests that at this particular frequency the outer-ear function shown in Fig. 2(a) is inappropriate to model the response of the subjects with the headphones used by Plack and Oxenham. Although not shown here, we did find an almost perfect match to the 8-kHz data assuming a flat outer-ear response at this frequency.

The dashed thin line in Fig. 4 shows the fit to the average data using the filterbank parameters calculated by linear regression. The overall behavior is maintained. The filterbank response is within one standard deviation of the average response for the most cases. The fit, however, gets worse, as would be expected from using a set of parameters that is not optimum for the specific data set.

2. Thresholds and filter bandwidths

Figure 6(a) shows a good match between the experimental and the model thresholds (assumed to occur at an output velocity of \( 5 \times 10^{-5} \) m/s) both for subject YO and the average data. This justifies the values selected for parameter \( a \).

The figure also shows the thresholds estimated with the average-filterbank parameters. The larger deviations between experimental and filterbank values are less than 5 dB. They

### Table III. Regression-line coefficients \( p_0 \) and \( m \) for creating the filterbank assuming a relationship of the form: \( \log_{10}(\text{parameter}) = p_0 + m \log_{10}(\text{BF}) \), with BF expressed in Hz.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Subject YO</th>
<th>Average response</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{CF}_{lin} )</td>
<td>-0.102 05</td>
<td>-0.067 62</td>
</tr>
<tr>
<td>( \text{BW}_{lin} )</td>
<td>0.184 58</td>
<td>0.037 28</td>
</tr>
<tr>
<td>( \text{CF}_{nl} )</td>
<td>-0.059 12</td>
<td>-0.052 52</td>
</tr>
<tr>
<td>( \text{BW}_{nl} )</td>
<td>-0.037 39</td>
<td>-0.031 93</td>
</tr>
<tr>
<td>( a )</td>
<td>0.788 80</td>
<td>1.402 98</td>
</tr>
<tr>
<td>( b )</td>
<td>1.439 70</td>
<td>1.619 12</td>
</tr>
<tr>
<td>( c )</td>
<td>-0.602 06</td>
<td>-0.602 06</td>
</tr>
<tr>
<td>( g )</td>
<td>4.305 06</td>
<td>4.204 05</td>
</tr>
<tr>
<td>( \text{LP}_{lin\text{ cutoff}} )</td>
<td>-0.102 05</td>
<td>-0.067 62</td>
</tr>
<tr>
<td>( \text{LP}_{nl\text{ cutoff}} )</td>
<td>-0.059 12</td>
<td>-0.052 52</td>
</tr>
</tbody>
</table>
occur at 0.5, 2.0, and 8.0 kHz and correspond to the largest deviations of the optimum values of parameter \( a \) over its regression line [Fig. 5(b)].

Figure 6(b) shows the approximate \( BW_{3dB} \) of the modeled filters compared with the human as calculated by the Glasberg and Moore (1990) formula:

\[
BW_{3dB} = 0.89 \times ERB = 0.89 \times [24.7 \times (4.37BF + 1)],
\]

where BF is expressed in kHz. The modeled \( BW_{3dB} \) were calculated directly from the isoresponse curves and not by applying the critical band paradigm used by Glasberg and Moore. For this purpose, the isoresponse curves were evaluated at a number of stimulus frequencies around the filter’s BF enough to determine the 3-dB-down cutoff frequencies with reasonable accuracy (±2 Hz). It can be seen that the DRNL filter BWs are a good match to the psychophysical data. The average filterbank \( BW_{3dB} \) match the human bandwidths except for 8 kHz, where the filterbank \( BW_{3dB} \) is around 150 Hz lower.

3. Isointensity response

Figure 7(a) shows the iso-intensity response of the model with the set of parameters given in Table II, optimized to fit the average pulsation threshold data (the response with the subject YO parameters of Table I is similar). The curves correspond to stimulus levels from 30 to 70 dB SPL in steps of 10 dB and have been normalized to give a gain of 0 dB at their peak value. The psychophysical filtershapes derived by Baker et al. (1998) are shown in Fig. 7(b) for comparison. Finally, Fig. 7(c) shows the filter shapes of the average filterbank (Table III).

The model curves show distinct notches below BF that are not reported in the psychophysical literature. In the model, the notches are the result of phase cancellation between the outputs of the linear and nonlinear paths. Similar notches have been consistently observed in physiological measurements of the cochlear partition in chinchilla (e.g., Rhode and Recio, 2000) and may be the explanation for “Nelson’s notches” (Kiang et al., 1986, Fig. 5) occasionally reported in auditory nerve rate/intensity functions. If human listeners monitor a number of adjacent filters simultaneously and pay attention to the filter with the best signal/noise ratio, it is unlikely that these notches will be observed physiologically as they offer a relatively poor signal to noise ratio.

The model filter function gets wider as the level of the stimulus increases. This is an important property of the DRNL filter consistent with the physiological (e.g., Rhode and Recio, 2000) and psychophysical behavior (e.g., Baker et al., 1998; Rosen et al., 1998; Glasberg and Moore, 2000).

Although not seen in Fig. 7, from 65 to 95 dB SPL there is a shift in the model BF towards lower BFs such that the ratio \( BF_{65}/BF_{95} \) ranges from about 1.10 (at BF=4 kHz) to 1.24 (at BF=8 kHz). The shifts are even larger (1.15 to 1.52) with YO’s optimum parameters. Such shifts are consistent with the physiology (e.g., Rhode and Recio, 2000) and have also been observed psychophysically when adequate simultaneous masking paradigms are employed (e.g., McFadden and Yama, 1983).

IV. DISCUSSION

The modeling exercise had two main aims. The first aim was to show that DRNL filters could be used to simulate psychophysical estimates of filter width and compression. The second aim was to show that the DRNL parameter changes are orderly with respect to BF. The close qualitative
and quantitative fits of the modeling results to the psychophysical data in Fig. 4 indicate that the first aim has been met with respect to both a single listener, YO, and the average response of six listeners. Similarly, the good fit of the regression functions in Table III suggests that the relationship between the parameters and BF is orderly enough to allow a filterbank to be generated for any point along the basilar membrane. This appears to be the case at least for the region representing a range of BFs between 250 Hz and 8 kHz.

The fitting process depends critically on three important assumptions. First, it assumes that the DRNL architecture gives an adequate account of the response of the cochlear partition to acoustic stimulation. Arguments in favor of this assumption are presented in Meddis et al. (2001). Second, it assumes that the pulsation threshold really does represent a condition in which the excitation of the masker, $O_M$, is equal to that generated by the probe signal, $O_S$. Arguments in favor of this assumption are presented by Houtgast (1972) and Plack and Oxenham (2000). However, the data of Plack and Oxenham also show that subjects with similar absolute hearing thresholds have pulsation thresholds that differ by as much as 20 dB (cf. subjects AO and YO in the original study). If the assumption was strictly correct, it would imply that their filters would be as much as 20 dB different at the masker frequency. Such intersubject variability seems rather large. Moore (1998, Fig. 3.19) gives an example of intersubject variability at BF=1 kHz, where the filter shapes of four subjects at 600 Hz ($=0.6\times$BF) differ by no more than 10 dB. Therefore, it is possible that the pulsation threshold occurs at ratios $O_S/O_M$ less than one for some subjects. The topic needs further investigation.

The third assumption is that listeners are able to attend selectively to the output of a single filter. This assumption is made in many psychophysical experiments and has proved to be of pragmatic value. To reduce the risk associated with this assumption, Plack and Oxenham (2000) took steps to minimize off-frequency listening.

The study does have some technical weaknesses that will need to be addressed in future studies. The outer ear function used in the model was based on measurements made on different subjects wearing different headphones from those in the Plack and Oxenham study. Furthermore, the middle-ear function used in the model was based on measurements made in cadavers. Ideally, we should have used measurements of the stapes response to headphone-delivered acoustic stimulation for the same subjects, with the same headphones as used in the psychophysical study. This is, of course, not practical and it is not immediately clear how this function is to be best estimated. Although the subjects’ audiograms were available, they could not be used as a substitute for an outer/middle ear function because this would imply that no processes subsequent to the outer/middle ear contributes to the audiogram.

A second weakness of the model is that it does not address the issue how the filter BF changes with signal level. Animal studies show that BF does shift in this way. If this is also the case with human listeners, then it follows that listeners must redirect their attention to a different site when the probe signal level changes. We have made no attempt here to model this effect. In effect, the procedure previously described assumes that adjacent filters have very similar characteristics. This is despite that fact that the study as a whole shows that the characteristics of the filters are changing along the cochlear partition. Future studies will need to use a fine grain filterbank with facilities for taking the output from the filter whose current BF corresponds to the probe signal frequency.

The DRNL model is a point model in that there are no connections between the individual filter units. As a consequence, distortion products generated in one DRNL unit do not propagate to other units. This is unlikely to give rise to fatal difficulties when modeling the response to two tones at well-spaced frequencies, as is the case the Plack and Oxenham data. However, one of the attractions of a nonlinear filterbank is its potential for modeling such phenomena as two-tone suppression and combination tones. In this respect, while the DRNL units of the filterbank proposed earlier are

![Graph](image-url)
able to generate "local" suppression and distortion products (Meddis et al., 2001), a more complete model is needed to simulate effects that are not local to the probe tone. For instance, suppression has been shown to occur for suppressor tones with frequencies well below the probe (Abbas and Sachs, 1976; Duijnhuis, 1980). Models alternative to the DRNL filter exist that claim to be able to account for this phenomenon (Goldstein, 1990, 1995). Future work is required to address this issue as far as the DRNL is concerned, as well as to make a detailed comparison concerning the relative merits of the various nonlinear filters that have been published recently.

The modeling study above was careful to use the data of a single subject (at least for the compression data). This is because DRNL units may prove useful for modeling the impaired hearing of individuals. If this proves to be feasible on a routine basis, it could be used to optimize the characteristics of hearing aids before supplying them to patients. Figure 8 shows an example of how the model can be tuned to fit hearing-impaired data for three different subjects (Oxenham and Plack, 1997, Fig. 5). In this case, the signal frequency was 2 kHz and the masker frequency 1 kHz. The model was evaluated using the same paradigm as described earlier. The parameters were identical to those in Table I (2 kHz) for normal-hearing subject YO, except for the gains (\(a, b,\) and \(g\)), which were reduced to model impaired hearing. For subject JK, with severe hearing loss, it was also necessary to reduce BW\(_{in}\) (see Fig. 8 caption). Further studies will be required to evaluate the potential of this methodology.

ACKNOWLEDGMENTS

The authors acknowledge the advice and support of Chris Plack. The authors also thank two anonymous reviewers for constructive comments on an earlier version of this article. The development of the DRNL filter has progressed over many years and has involved contributions from a number of collaborators. The authors would like to thank, in particular, Michael Hewitt, Trevor Shackleton, and Mike Stone for substantial assistance at various times. Author EALP's work was supported by the Junta de Comunidades de Castilla La Mancha (Consejería de Sanidad, Ref. 2001-01044) and by the Universidad de Castilla–La Mancha.

APPENDIX: DIGITAL IMPLEMENTATION OF THE DRNL FILTER

The DRNL filter was implemented digitally in the time domain by implementing each of its filters and gains as a digital component. The implementation of each building component was done as follows.

1. The gammatone filters

The GT filter has an impulse response of the form (Patterson et al., 1992)

\[
\begin{align}
&h(t) = kt^{a-1} \exp(-2\pi B t) \cos(2\pi f_c t + \varphi) \quad (t \geq 0), \\
&h(t) = 0 \quad (t < 0),
\end{align}
\]

where \(n\) is the order of the filter, \(B\) is its bandwidth, \(f_c\) is its center frequency, \(\varphi\) is its phase, and \(k\) is a gain.

The DRNL filter uses several cascades of first-order \((n = 1)\) GT filters only (see Fig. 3 and main text). They were implemented digitally as an infinite impulse response filter as follows (M. Stone, Exp. Psychology Lab., Cambridge, UK, personal communication, see also Slaney, 1993):

\[
y[i] = a_0 \cdot x[i] + a_1 \cdot x[i-1] - b_1 \cdot y[i-1] - b_2 \cdot y[i-2],
\]

where \([x]\) refers to the \(i\)th sample of the digital signal, \(x\) and \(y\) are the input and output signals to/from the filter, respectively, and the coefficients are calculated as follows:

\[
a_0 = \frac{1 + b_1 \cos \theta - j b_1 \sin \theta + b_2 \cos(2\theta) - j b_2 \sin(2\theta)}{1 + \alpha \cos \theta - j \alpha \sin \theta},
\]

\[
a_1 = \alpha \cdot a_0,
\]

\[
b_1 = 2\alpha,
\]

\[
b_2 = \exp(-2\phi),
\]

where

\[
\theta = 2\pi f_c dt,
\]

\[
\phi = 2\pi B dt,
\]

\[
\alpha = -\exp(-\phi)\cos \theta,
\]

and \(j = \sqrt{-1}\), and \(dt\) is the sampling period of the digital signal.
2. The low-pass filters

The DRNL filter implementation includes several cascades of second-order Butterworth lowpass filters (see Fig. 3 and main text). These were implemented digitally as follows:

\[ y[i] = C \cdot x[i] + 2 \cdot C \cdot x[i-1] + C \cdot x[i-2] - D \cdot y[i-1] - E \cdot y[i-2], \]

\[ (A4) \]

where the coefficients are

\[ C = \frac{1}{1 + \sqrt{2} \cot \theta + \cot^2 \theta}, \]

\[ D = 2C(1 - \cot^2 \theta), \]

\[ E = C(1 - \sqrt{2} \cot \theta + \cot^2 \theta), \]

\[ (A5) \]

and

\[ \theta = \pi f_c dt, \]

\[ (A5d) \]

where \( f_c \) is the 3-dB-down cut-off frequency of the low-pass filter, and \( dt \) is the sampling period of the digital signal.

3. The linear gain

The gain in the linear path of the DRNL filter was implemented digitally in the time domain as follows:

\[ y[i] = g \cdot x[i], \]

\[ (A6) \]

where \( [i] \) refers to the \( i \)th sample of the digital signal, and \( x \) and \( y \) are the input and output signals to/from the linear gain stage, respectively.

4. The nonlinearity

The time domain digital implementation of the “broken-stick” nonlinearity was as follows:

\[ y[i] = \text{sign}(x[i]) \cdot \min(a|x[i]|, b|x[i]|^c), \]

\[ (A7) \]

where \( [i] \) refers to the \( i \)th sample of the digital signal, \( x \) and \( y \) are the input and output signals to/from the nonlinearity, and \( a, b, \) and \( c \) are parameters.


