A computational model of loudness density

Stephen T. Neely, Joyce Rodriguez, Yi-Wen Liu, Walt Jesteadt, and Michael P. Gorga

Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

Submission Date: 2009-10-20

Running Title: Loudness-density model

Abstract

Loudness density is the distribution of loudness across dimensions of time and frequency. A computational model of loudness density is described that reproduces equal-loudness level contours from time-domain representations of tones. The model also demonstrates increase in loudness as the bandwidth of a multi-tone complex increases, an observation that is consistent with psychophysical measurements. The model provides an efficient method for computing the loudness of arbitrary acoustic waveforms. When loudness is assumed to be the decision variable for psychophysical tasks such as intensity discrimination and masked thresholds, model-predicted performance can be compared with psychophysical measurements. The similarity between model predictions and psychophysical estimates of suppression is noteworthy because the model provides no separate mechanism for suppression. Suppression in the model is only possible through the same mechanism that provides compression for single tones. Its ability to predict suppressive effects suggests that it also may be a useful tool for exploring more general mechanisms of auditory signal processing.

PACS numbers: 43.66.Ba, 43.66.Cb, 43.66.Dc

I. Introduction

Auditory models focused on psychophysics have traditionally represented cochlear frequency analysis by a linear filter bank followed by a nonlinear mechanism intended to represent cochlear compression (*e.g.*, Patterson and Holdsworth, 1996). Some recent nonlinear auditory models that are more physiologically based still use a separate frequency analysis to control level dependence of the primary frequency analysis (e.g., Zhang *et al*., 2001; Heinz *et al.*, 2001; Jepsen *et al*., 2008). However, the separation of frequency analysis from compression is inconsistent with physiological observations of cochlear signal processing. Physiologically, frequency analysis and compression in the cochlea are simultaneous and inseparable (*e.g.*, Rhode and Robles, 1974). The time-domain loudness model (TDLM) described here differs from most auditory models by combining compression and frequency analysis into the same nonlinear filter sections. In this way, the TDLM implements cochlear compression and frequency analysis in a manner that is consistent with physiological observations.

The output of the TDLM represents loudness distributed across two dimensions: time and characteristic frequency (CF). In this paper, the two-dimensional representation of loudness is called *loudness density*. CF is the frequency represented by a particular location in the cochlea or by a particular auditory filter. Integration of loudness density across time produces *band-limited loudness*, which spans a range of CFs. Integration of loudness density across the CF produces *instantaneous loudness*, which varies with time. In general, loudness density is a random variable due to the stochastic nature of spike generation in the auditory nerve. The integration of loudness density by the central auditory system across limited ranges of time and CF may provide the decision variable that determines listeners' performance on basic psychophysical detection and discrimination tasks (Moore *et al*., 1997; Glasberg and Moore, 2005; Neely and Jesteadt, 2005).

In previous work with computer simulations of cochlear mechanics (Neely *et al.*, 2000), an estimate of loudness was based on the summed output of the cochlear model, including both temporal and spatial integration, and shown to resemble the loudness data of Fletcher and Munson (1933) in its rate of growth with increasing stimulus level. Distortion-product otoacoustic emissions (DPOAE) exhibit similar compressive growth (*e.g.*, Neely *et al.*, 2003). Cochlear models offer a means to link compressive growth of loudness and DPOAEs to a common underlying mechanism, which is the compressive growth of basilar membrane (BM) responses within the cochlea (Neely *et al.*, 2000). A consensus is emerging in the literature on loudness that the function describing the growth of loudness with level is a reflection of peripheral compression (*e.g.*, Allen and Neely, 1997; Buus *et al.*, 1997; Moore *et al.*, 1997).

Suppression has been observed in auditory-nerve (AN) responses (*e.g.*, Galambos and Davis, 1944; Sachs and Kiang, 1968; Sachs and Abbas, 1974; Abbas and Sachs, 1976) and in BM responses (*e.g.,* Rhode, 1977). In single-fiber AN recordings, suppression is often quantified as a horizontal shift of a rate-level function, where *rate* can be either total spike discharge rate (*e.g.*, Sachs and Kiang, 1968) or the component of the spike rate that is synchronized to the stimulus (*e.g.*, Javel *et al.*, 1983). BM suppression largely accounts for AN suppression when the suppressor frequency is above the probe frequency, but suppressor frequencies below the probe frequency produce less BM suppression than AN suppression (*e.g.*, Geisler and Nuttall, 1997). The discrepancy between BM and AN low-frequency suppression may indicate the presence of a “second filter” interposed between BM and AN (Hall, 1980).

It is generally accepted that suppression is a potential contributor to psychophysical masking (*e.g.*, Delgutte, 1990; Moore and Vickers, 1997); however, quantifying the contribution of suppression to masking has been difficult. Recently, Rodriguez *et al.* (2010) found similarity between psychophysical and physiological estimates of suppression. Their psychophysical estimate was the dB difference in the amount of masking (DAM) between simultaneous masking and forward masking at a given masker level. Their physiological estimate was obtained by measuring DPOAE input/output (I/O) functions and observing the rightward shift in these I/O functions in the presence of a suppressor tone. This definition of suppression parallels the one based on rate-level curves for AN measurements. The results of their study comparing DAM to DPOAE data in the same group of subjects were consistent with the view that most of the difference between simultaneous and forward masking is due to suppression, provided that the signal delay in the forward masking is minimal.

In the present study, the amount of suppression produced by the TDLM is compared to our recent psychophysical estimates of suppression (Rodriguez *et al.*, 2010). Although the TDLM is implemented in the time domain and can potentially compute loudness density for any acoustic waveform, the present study focuses on steady-state responses because the current implementation of the TDLM lacks any representation of the transient effects of neural adaptation. TDLM results are presented that depend on comparing the loudness of two tones (*viz.*, loudness equality, intensity discrimination) and the reduction of loudness of one tone in the presence of a second tone (*i.e.*, two-tone suppression).

II. Time-domain loudness model

The cochlea performs frequency analysis and dynamic-range compression mechanically, prior to neural transduction. Cochlear mechanics are often modeled by a *transmission line* with properties that vary along its length (e.g., Zweig *et al.*, 1976). Cochlear filtering is accomplished primarily by incremental removal of the highest frequencies contained in the fluid wave that travels from the base to the apex of the cochlea. Outer hair cells (OHCs) amplify the cochlear traveling wave in a level-dependent manner that compresses its dynamic range. Neural transduction is performed by inner hair cells (IHCs) that detect BM displacements. The additional filtering that may occur between the BM and the IHCs is sometimes called the *second filter*. See Allen and Neely (1992) for a review of relevant features of cochlear mechanics.

The TDLM is implemented as a cascade arrangement of filter sections that resembles a cochlear transmission line. An additional "second filter" is attached to the output of each cascade section to simulate additional filtering in the cochlea between the BM and IHCs. This arrangement of filter sections is illustrated in Fig. 1. In this figure, the cascade-filter sections are labeled f_1, $f_2 \ldots f_N$ and the second-filter sections are labeled g_1, $g_2 \ldots g_N$. Each cascade-filter and second-filter section is implemented as a second-order discrete-time filter with time-varying filter coefficients. The cascade arrangement of filter sections requires that frequency analysis be accomplished incrementally, so that the transfer function between the input to the first section and the output of any given section is the accumulation of filter gains provided by each preceding cascade-filter section up to that particular section. The TDLM filter sections were designed to provide transfer functions that are similar in shape to gamma-tone filters, which are often used in psychophysical auditory models (*e.g.*, Patterson and Holdsworth, 1996; Meddis *et al.*, 2001; Jepsen *et al*., 2008) because of their similarity to physiological measures of BM vibrations (*e.g.*, Rhode and Robles, 1974).

Each cascade-filter section has a low-pass frequency response, as shown in the upper panel of Fig. 2. Each second-filter section has a high-pass frequency response, as shown in the lower panel of Fig. 2. In addition to CF, each filter section has a quality-factor (Q). Frequency responses are shown in Fig. 2 for two examples of quality factors. The first example, Q=1 (dashed lines), is the minimum value for all CFs and represents a hearing-impaired (HI) condition with no cochlear amplifier. The second example, Q=5 (solid lines), is the maximum value for CF=5 kHz and represents a normal-hearing (NH) condition with a fully-functioning cochlear amplifier. Equations for calculating discrete-time filter coefficients from CF and Q parameters are described in Appendix A.

The Q associated with each TDLM filter section varies with time and is dependent on the instantaneous input signal to that section. The Q has its maximum value when the signal amplitude is small and its minimum value when the signal amplitude is large. At intermediate signal amplitudes, the Q is assigned an intermediate value. The maximum Q value (Q_{max}) decreases as the CF of the filter section decreases. The variation of Q_{max} with frequency is shown in Fig. 3 (thick solid line). Although the decrease of Q_{max} with CF is consistent with both physiological (*e.g.*, Shera *et al.*, 2002) and psychophysical (*e.g.*, Glasberg and Moore, 1990) measurements, the values selected for the TDLM are based primarily on producing agreement with equal-loudness contours (see section III). Q_{min}, is shown as thin line (at a quality factor = 1) that is independent of frequency. This is the expected quality factor when cochlear mechanics lack any contribution from OHC motility.

The peak gain of each cascade-filter section is proportional to its Q. Besides causing a decrease in gain, lower Q values also cause transfer functions to be broader and have fewer cycles of delay. The sharpness of tuning of the transfer functions can be quantified by the ratio of their peak frequency to their equivalent rectangular bandwidth (ERB). This ratio is called Q_{erb}**.** Although, the TDLM was not designed specifically to reproduce measured filter bandwidths, it is of interest to compare TDLM-predicted Q_{erb} with measured values. In Fig. 3, the maximum sharpness of tuning of the TDLM (circles), representing the TDLM transfer functions for extremely low signal levels is compared with tuning estimated from tone-threshold measurements with simultaneous notched-noise maskers (dashed line; Glasberg and Moore, 1990). Note that Q_{erb} in the TDLM is greater than the psychophysical estimate at low frequencies and less at high frequencies. Although it may be possible to adjust model parameters to produce Q values in closer agreement with measured values, it is not known how such changes would affect TDLM-predicted intensity discrimination or suppression.

Another filter metric of interest is the ratio of Q_{erb} to the maximum number of cycles of delay. This ratio is between 0.97 and 1.29 for low-level TDLM transfer functions, which is smaller than the values of 2.52 estimated for humans or 3.0 for cats (Shera *et al.*, 2002), but similar to the value of 1.25 estimated for chinchilla (Shera *et al.*, 2010). Because the Q-delay ratio is a characteristic feature of a filter, the discrepancy between the TDLM and measured values for humans may indicate that the filter shapes used in the TDLM are not realistic. Whether it is possible to reduce this discrepancy in a future version of the TDLM or whether this is a fundamental limitation of this modeling approach is not known.

The level-dependent Q of each filter section causes it to compress signal levels in a manner similar to the way OHCs compress basilar-membrane vibrations in the cochlea. The similarity is twofold: (1) gain is distributed and provided to each frequency component *en route* to its characteristic place and (2) level dependence is likewise distributed and controlled by local input signals.

To complete the simulation of peripheral compression, an additional stage of compression is appended to the output of the second-filter sections to represent IHC transduction and neural-rate saturation. IHC compression is implemented as an instantaneous nonlinearity that is expansive at low levels and compressive at high levels. The expansive compression at low levels plays a role in fitting TDLM-predicted results in both the NH and HI conditions and is consistent with other auditory models (*e.g.*, Jepsen *et al.*, 2008). The current implementation of the TDLM lacks any representation of the transient effects of neural adaptation, which emphasize the leading edge of auditory signals. The net TDLM compression at 1 kHz is shown in Fig. 4 for simulations of both the normal-hearing condition (thick solid line) and the hearing-impaired condition (medium solid line).

Compression in Fig. 4 is defined as the reciprocal of the slope of loudness level[1] versus signal level when both quantities are expressed in decibels. Normal-hearing compression in the TDLM combines compression due to the cascade-filter and second-filter sections with IHC compression. TDLM-predicted compression at 1 kHz for the NH condition (thick line) is compared in Fig. 4 with compression derived from the classic Fletcher-Munson (1933) loudness function (thin line) and with estimates of compression (circles and squares) from some recent studies (Neely *et al.*, 2005; Neely and Jesteadt, 2005; Schairer *et al.*, 2003). The agreement between model predictions and previous data suggests that the TDLM does a reasonable job of describing cochlear processing, at least in the case of normal hearing. If desired, the TDLM could be adjusted to replicate other loudness functions (*e.g.*, Hellman and Zwislocki, 1961; Hellman, 1976) by adjusting model parameters associated with (1) the level-dependent Q and (2) the instantaneous nonlinearity that represents IHC transduction. These parameters are described in Appendix A. The selection of model parameters for the TDLM that produce a compression function similar at low levels to the quadratic compression function suggested by Neely and Jesteadt (2005) facilitates a replication in the TDLM of their demonstration of intensity discrimination based on loudness differences (see section VI).

TDLM-predicted compression in the HI condition (medium-weight line at the bottom of Fig. 4) is entirely due to the IHC stage of the TDLM, so it simulates hearing impairment due to loss of OHC function. IHC compression is less than one (*i.e.*, expansive) at low levels and greater than one (*i.e.*, compressive) at high levels. An equation for the instantaneous nonlinearity that implements this compression is described in Appendix A. The form of this equation is partly based on previous cochlear modeling work that transformed BM displacement into an estimate of loudness (Neely *et al.*, 2000). The IHC nonlinearity was adjusted in the model primarily to produce results for the HI condition that were in agreement with measurements, such as the rate of growth of the loudness function (see section IV) and intensity discrimination (see section VI). Although the design of the IHC stage of the model was motivated to some extent by physiological properties of IHCs, the TDLM provides no output that is appropriate for direct comparison with physiological measurements such as IHC receptor potentials or neural firing rates. The TDLM outputs are an abstraction of neural firing rates that could include influences of signal processing in the central auditory system.

Some of the frequency dependence observed in equal-loudness-level contours (ELLCs) may be due to the transfer characteristics of the middle-ear. To allow middle-ear influence to be represented, a middle-ear filter section is prepended to the TDLM. The middle-ear filter characteristics were designed to minimize differences between the TDLM-predicted and ISO-226 (2003) ELLCs, which are compared in the next section. The gain of the middle-ear transfer function is shown in Fig. 5. The middle-ear filter is implemented as a level-independent, sixth-order discrete-time filter, which is described in Appendix A. The middle-ear transfer function in Fig. 5 is similar to the one suggested by Glasberg and Moore (2006) in having a mid-frequency 6-dB notch, but the notch in Fig. 5 is at 1.4 kHz and theirs is at 2.5 kHz. Their middle-ear stage provides a few dB less gain at low frequencies (e.g., 0.03 to 0.05 kHz) and high frequencies (e.g., 3 to 5 kHz) than the one in Fig. 5. The TDLM reduces gain at low frequencies beyond the middle-ear stage by reducing the maximum Q of the cascade and second filters.

To demonstrate the level dependence of TDLM tuning, Fig. 6 shows the transfer gain between input to the middle-ear stage and output of the IHC stage for a frequency-swept tone at two different stimulus levels. When the stimulus is 0 dB SPL at the input to the middle-ear stage, the TDLM gain is similar to the low-level linear gain shown in Fig. 3, although slightly reduced due to temporal variation in the Q of the filter sections. At 80 dB SPL, the TDLM gain is further reduced by 30 to 40 dB, depending on frequency.

III. Equal loudness

For any acoustic waveform applied to its input, the TDLM provides a prediction of loudness density. Loudness density may be reduced to a single, composite loudness value by integrating across some range of CFs and some interval of time. The integration across CF is accomplished by summing outputs from the IHC stage of the TDLM across any desired range of CFs. The TDLM has 12 outputs per octave.

In this paper, *tone loudness* is defined by integrating loudness density across a two-octave band of CFs centered at the tone frequency and by taking the maximum value of the resulting instantaneous loudness after the initial transient response. This two-octave integration range is reduced for frequencies below 31 Hz or above 16 kHz because the lowest CF output is 15.6 Hz and the highest CF output is 32 kHz. The limited bandwidth used to define the loudness of tones is justified because the contribution to loudness outside this two-octave range is negligible. The advantage of the restricted range is that it reduces computation time by allowing lower sampling rates when high frequencies are absent and eliminating low-CF filter sections when low frequencies are absent. In general, for arbitrary stimuli, the entire range of audible frequencies (0.0156 to 32 kHz) is included in the calculation of loudness.

In the TDLM, loudness density is multiplied by an arbitrary scale factor to make tone loudness equal to one for a 1-kHz tone at 0 dB SPL for consistency with the definition of *perceptual intensity* suggested by Neely and Jesteadt (2005). TDLM-predicted loudness is equivalent to *sones* when divided by 92, which is the TDLM-predicted loudness for a 1-kHz tone at 40 dB SPL.

TDLM-predicted ELLCs, such as those shown in Fig. 7 (solid lines), are obtained by (1) determining loudness level as a function of input level for all frequencies of interest, (2) taking horizontal slices across these loudness-level functions, and (3) plotting the intersecting input levels as functions of the corresponding tone frequencies. For comparison, standard ELLCs (ISO-226, 2003) are also shown in Fig. 7 (dashed lines). By definition, each ELLC has a loudness level in *phons* equal to the input level (SPL) at 1 kHz. The strategy for obtaining agreement between the TDLM-predicted and standard ELLCs involved two different types of adjustments: (1) the middle-ear transfer function was designed to minimize the average TDLM-ISO difference in a set of higher-level contours from 50 to 90 phons and (2) Q_{max} was adjusted (at CF=0.05 and CF=5) to match a set of lower-level contours from 10 to 50 phons. In Fig. 7 near 0.1 kHz, for example, the better agreement between the dashed and solid lines along the 70 phon contour is due to selection of middle-ear parameters, while the better agreement along the 10 and 20 phon contours is due to selection of Q_{max}. In the calculation of the contours produced by the TDLM, Q_{max} was the *only* filter parameter that was allowed to vary with CF.

Overall, the agreement between the TDLM-predicted and standard ELLCs is good, especially considering that the ISO-226 ELLCs are defined as being level-dependent in the frequency domain and the TDLM is defined and implemented in the time domain. When the set of TDLM and ISO curves from 10 to 90 phons is compared, the rms difference is 2.4 and 0.4 dB at 0.05 and 5 kHz, respectively. The largest difference between the TDLM-predicted and ISO-226 contours is 12.5 dB, which occurs in the 0-phon contour at 40 Hz, a frequency of little importance in human perception. Although the TDLM-predicted contours do not match the ISO-226 contours as closely at frequencies below 1 kHz as those predicted by the loudness model of Glasberg and Moore (2006), the TDLM-predicted contours have better agreement with the ISO-226 contours in the 1 to 8 kHz range. This agreement is remarkable considering that Q_{max} was the only model parameter that varied with CF (see Fig. 3).

IV. Cochlear hearing loss

Hearing impairment due to total loss of OHC function is simulated in the TDLM by setting Q_{max}=1 for all CFs, so that cascade-filter sections provide no gain. Comparison of HI and NH loudness-level functions at 1 kHz (see Fig. 8) indicates that this simulation of cochlear hearing loss elevates threshold by 43 dB. HI loudness grows more rapidly than NH loudness, so that HI loudness becomes nearly equal to NH loudness at 100 dB SPL. If the horizontal separation between the HI and NH loudness functions is taken as a measure of the gain required to restore normal hearing, then Fig. 8 illustrates that the normative gain is a decreasing function of stimulus level. Note that this simulation of HI loudness growth was implemented by removing the gain and compression associated with OHC function, contrary to recent suggestions that loudness recruitment is not observed in peripheral auditory responses (Heinz *et al.*, 2005; Cai *et al.*, 2009).

Assuming that hearing threshold coincides with a loudness level of zero in Fig. 8, the slope of the HI loudness function at threshold is about twice the slope of the NH loudness function. Although this model result is inconsistent with recent psychophysical estimates (e.g., Buus and Florentine, 2001; Moore, 2004) that indicate the NH and HI slopes should be about the same near threshold, the factor of two difference between NI and HI is about the same as predicted by Hellman and Meiselman (1990) for a hearing loss of 40 dB. In Fig. 8, a loudness level of 19.5 is 40 dB above threshold on the NH curve and only 18.3 dB above threshold on the HI curve. The TDLM loudness match of 40 dB above threshold for a NH ear with 18.3 dB above threshold for an ear with 40 dB loss is about the same as observed in loudness-matching experiments reported by Miskolczy-Fodor (1960, Fig. 3). The TDLM predicts that 2 dB above threshold on the NH loudness function matches 1.3 dB above threshold on the HI loudness function, which differs by only 0.7 dB from Moore's (2004) observation that a tone 2 dB above NH threshold matches the loudness of a tone 2 dB above HI threshold.

V. Loudness dependence on bandwidth

The loudness of a multi-tone complex presented at a constant rms level increases as the frequency span of the complex increases beyond a critical band. This feature of normal hearing can be observed by matching the loudness of a 1-kHz tone to the loudness of multi-tone complexes of varying bandwidth centered at 1-kHz. In Fig. 9, TDLM-predicted loudness matches in the NH condition (squares) are compared with measurements obtained by Leibold at al. (2007) and show good agreement in the range of bandwidths between 0.2 and 2 kHz. The TDLM predictions are closer than predictions obtained by Leibold *et al*. using the Moore *et al*. (1997) frequency-domain loudness model.

Increase in loudness with increasing bandwidth was first observed with band-pass noise (*e.g.*, Zwicker *et al*. 1957) and is known to be reduced or absent in ears with hearing impairment (*e.g.*, Scharf and Hellman, 1966). Note in Fig. 9 that TDLM-matches in the HI condition exhibit less loudness increase for bandwidths above 0.2 kHz. This result is consistent with band-pass noise loudness-match measurements (*e.g.*, Scharf and Hellman, 1966); however, loudness match measurements with the multi-tone complex in HI ears are not yet available for comparison.

VI. Intensity discrimination

Loudness is presumed to be a random variable and a just-noticeable difference (JND) in the intensity of a tone is proportional to the standard deviation of this variable. In general, the variance associated with basic psychophysical tasks may be assumed to have quadratic dependence on loudness (Neely and Jesteadt, 2005):

$$\sigma_N^2 = \eta_0 + \eta_1 N + \eta_2 N^2, \tag{1}$$

where N is a loudness associated with the task and the coefficients (η_0, η_1, η_2) may depend on the task and on signal parameters such as duration. The coefficient η_1 may be associated with Poisson internal noise (Allen and Neely, 1997). In case of TDLM-predicted intensity discrimination, N is the tone loudness and the variance coefficients are $\eta_0 = 0$, $\eta_1 = 0.435$, and $\eta_2 = 0.00128$. In general, the coefficients of the variance polynomial must be determined empirically and will depend on the reliability of the decision cues available for the task.

Suppose that the loudness JND for intensity discrimination is $\Delta N \approx \sigma_N$ and let α be defined as the reciprocal of the slope of a loudness-level function, which is the same as the definition of *compression* in Fig. 4. Under these assumptions, Allen and Neely (1997) suggest that the Weber fraction $\Delta I / I$ is approximated by the following equation:

$$\frac{\Delta I}{I} \approx \alpha \frac{\Delta N}{N}, \qquad (2)$$

where ΔI is a JND in stimulus intensity. This approximation is plotted in Fig. 10 for both the NH and HI states of the TDLM. For comparison, previously published NH intensity-discrimination measurements are also shown (squares; Jesteadt *et al.*, 1977) in this figure.

The agreement between the measurements and the model suggests that Eq. (1) may be an appropriate description of the variance that determines the JND for NH listeners (Neely and Jesteadt, 2005). However, the TDLM-predicted intensity JND for wide-band noise (not shown) also decreases at high levels, which is contrary to expectations. The reason for this discrepancy is not known, but should be resolved in future implementations of the DLM by adjustment of model parameters and/or the manner in which noise variance is calculated.

The TDLM-predicted JND for HI listeners in Fig. 10 is nearly the same as for NH listeners when the stimulus level is above 40 dB, despite the steeper slope of the loudness function. The fact that steeper loudness functions do not always have smaller intensity JNDs has been noted previously (*e.g.*, Zwislocki and Jordan, 1986). Examination of Eq. (2) reveals that the reason for the JND similarity must be that, at a given SPL above threshold, $\Delta N / N$ is increased for the HI listener by about the same amount that α is decreased. The similarity between NH and HI JNDs at high input levels is consistent with previous measurements obtained from NH and HI listeners (e.g., Florentine *et al.*, 1993; Schroder *et al.*, 1994).

VII. Suppression

TDLM-predicted two-tone suppression may be quantified as the increase in signal level required to keep the signal at threshold in the presence of a masker. Because the TDLM has no mechanism to provide time-varying adaptation, TDLM-predicted two-tone suppression is comparable to the psychophysical measurement of the difference between simultaneous and forward masking at the same masker level.

Listener performance on masked-threshold tasks is quantified by d', which is the ratio of the mean value of a random decision variable to its standard deviation (e.g., Green and Swets, 1966). As in the case of intensity discrimination (see section VI), d' is equal to the difference between two loudness values divided by the square root of the loudness variance:

$$d' = \frac{N_2 - N_1}{\sigma_N} . \quad (3)$$

In this equation, N_1 is the loudness of the masker alone and N_2 is the loudness of the signal plus masker. It is assumed that $N_2 > N_1$. As mentioned earlier, loudness integrated across a two-octave range centered at the signal-tone frequency. However, because the psychophysical task in a masked threshold experiment is not identical to *intensity discrimination*, it is appropriate to represent the variance associated with this task differently. For simplicity, loudness variance for the masked threshold task is represented as a constant, which makes d' proportional to the increment in loudness due to the presence of the signal. The variance was selected to fit the psychophysical suppression data of Rodriguez *et al*. (2010): $\sigma_N^2 = 5000$. In terms of Eq. (1), the loudness variance for TDLM-predicted increment detection has $\eta_1 = \eta_2 = 0$.

Because the loudness-decision theory is incomplete with respect to the representation of loudness variance for different psychophysical tasks (Neely and Jesteadt, 2005), its functional form must be inferred from comparisons with measurements. The assumption of constant variance for masked detection suggests a decision mechanism that is different than the one used for intensity discrimination. For example, the masked-detection decision could be based on the output of a modulation filter-bank (Dau *et al.*, 1997; Gallun and Hafter, 2006). However, the exact representation of variance for masked detection is uncertain and subject to change in implementations of the TDLM.

For a two-alternative, forced-choice task, d' is related to proportion correct by the cumulative-normal distribution function:

$$P_c = \Phi\left(\frac{d'}{\sqrt{2}}\right). \tag{4}$$

TDLM-predicted psychometric functions (PFs) are shown in Fig. 11 for various masking conditions. The signal frequency was 4 kHz. The masker frequency was 4.281 kHz for the PFs in the upper panel and 2.141 kHz for the PFs in the lower panel. These frequencies were selected for comparison with the psychophysical suppression data of Rodriguez *et al.* (2010). The higher frequency has been observed to be the most effective frequency for suppressing distortion product otoacoustic emissions when the f_2 primary frequency is 4 kHz (Gorga *et al*, 2008). The lower frequency is an octave below the higher frequency and is thought to be sufficiently low to elicit essentially uncompressed response growth at the 4 kHz location in the cochlea. In both panels of Fig. 11, the leftmost PF represents the condition with no masker and the other PFs represent masker levels increasing from 0 to 80 dB SPL in 10 dB steps as one moves towards the right side of each panel. For the suppression results, loudness was defined by integrating over a half-octave range of CFs (centered at 4 kHz) instead of the two-octave range used to define single-tone loudness. This reduced integration range included the 4.281 kHz masker frequency, but not the 2.141 kHz masker frequency, and may be justified by the ability of the auditory system to assign a separate loudness to the signal and masker when they are widely separated in frequency. However, such modeling details are uncertain and subject to change.

For comparison purposes, suppression was quantified as the rightward shift of the PF at the point where $d'=1$. Estimates of the amount of suppression derived from TDLM predictions of PFs (provided in Fig. 11) are shown as filled symbols in Fig. 12, along with our psychophysical, empirically-based estimates of suppression for the same masker and signal frequencies.

The psychophysical estimate is the difference in amount of masking (DAM) between simultaneous masking and forward masking at fixed masker levels. DAM is thought to provide an indirect estimate of suppression when signal delay is kept short in order to minimize the reduction in forward masking due to recovery from adaptation. Support for this view was provided by Rodriguez *et al.* (2010) by showing agreement between DAM and a physiological estimate of suppression based on distortion-product otoacoustic emission measurements.

The agreement between TDLM-predicted suppression and DAM suppression in Fig. 12 supports the view that compression and suppression are due to the same mechanism in the cochlea. Not only is the amount of TDLM-predicted suppression similar to DAM, there is also agreement in the signal levels over which this suppression is observed. For the model predictions shown in Fig. 11, when the masker level was 80 dB SPL, the signal level at threshold was 65 and 64 dB SPL for the masker frequencies 2.141 and 4.281 kHz, respectively. In our psychophysical study (Rodriguez *et al.*, 2010), when the masker level was 80 dB SPL, the signal level at threshold was 60 and 65 dB SPL for masker frequencies of 2.141 and 4.281 kHz, respectively. The amount of TDLM-predicted suppression depends strongly on the amount of TDLM compression (see Fig. 4), so the combined agreement in amount of TDLM suppression and range of signal levels with psychophysical measurements suggests that the amount of TDLM compression is realistic.

VIII. Discussion

The TDLM described here is not the first model to use a cascade arrangement of filter sections to model cochlear signal processing. Lyon (1982) described a cascade-parallel pole-zero filter design that was similar to the TDLM described here, but with constant coefficients. A simplified version of this model, with an all-pole cascade, was implemented as an integrated circuit by Watts *et al.*, (1992). Kates (1991) suggested a model of cochlear signal processing with both cascade-filter and second-filter sections, with adjustable Q values in all sections, but the control signal for the Q adjustment was derived from the filter outputs. The present TDLM differs from other auditory models by adjusting filter coefficients based on the input to each filter section. Other models (e.g., Zhang et al., 2001) typically use a separate frequency analysis or "control path" to derive a control for the primary frequency analysis. In contrast, the TDLM uses the instantaneous input to each filter section as the control for that filter section without any further processing of this input signal.

Although certain aspects of the TDLM design were motivated by cochlear physiology, the TDLM resembles the psychophysical model of Moore *et al*. (1997) more than the physiology-based auditory models of based on the work of Carney (1994) (see Zhang *et al*., 2001; Zilany and Bruce, 2006). The TDLM shares with physiology-based models the idea of implementing compression and suppression with a single mechanism. However, the TDLM provides no outputs for comparison with any physiological measurement. In particular, although the final stage of the TDLM is the IHC and is intended to represent mechanical-to-electrical transduction only in an abstract way, the TDLM is not designed to model any measureable response of the auditory nerve. By relying on discrete-time second-order sections to implement frequency analysis, the TDLM gains computational efficiency, but loses direct correspondence with cochlear physiology. For this reason, further development of the TDLM will be motivated by comparisons with psychophysical measurements. A model that directly represents the physics of cochlear structures and the physiology of OHCs (*e.g.*, Liu and Neely, 2010) is more appropriate for comparisons with physiological measurements. The TDLM is primarily an auditory signal-processing model.

The term *dynamic loudness model* (DLM) was used by Chalupper and Fastl (2002) to describe their approach to auditory modeling, which is similar to the present approach in its focus on loudness density. Although the motivation and suggested application of their DLM is similar to our TDLM, many implementation details differ. For example, spread of masking in their DLM is represented in a separate stage that follows a linear filter bank. Our TDLM represents spread of masking by means of suppression, which is inherent in the level dependence of the same filters that separate frequencies.

Allen and Sen (2003) have modeled the role of suppression in the upward spread of masking (USM) and argued that suppression and the USM are both due to the same mechanism located in cochlear micromechanics (*e.g.*, Allen, 2007). The view that USM is primarily due to suppression contrasts with the alternative view that USM is primarily due to spread of excitation (*e.g.*, Delgutte, 1990; Moore and Vickers, 1997). Rodriguez *et al.* (2010) have recently obtained experimental evidence suggesting that suppression is a significant contributor to simultaneous masking. Their data show that much of the difference in the amount of masking between simultaneous and forward masking can be attributed to suppression. Their data also highlight the view that suppression and masking are linked, and not the consequence of independent processes.

Although the second-order discrete-time filters used in the TDLM can be implemented efficiently, especially in a dedicated processor designed for digital signal processing, the requirement for 12 filters per octave is computationally expensive relative to some alternative methods of frequency analysis. This computational demand may limit practical application of the TDLM to devices for which power consumption is not a primary consideration. For example, the TDLM could be incorporated into a sound-level meter to provide real-time quantification of loudness in *sones*. The output of the TDLM could potentially provide appropriate signals for stimulating electrodes in a cochlear implant because it provides adequate CF resolution and wide-dynamic range compression in a computationally efficient manner. However, the limited availability of power in such a device may preclude its use in this application. The TDLM should be useful in basic research studies focused on understanding signal processing mechanisms in the auditory system, where it may have value in helping to explain several different physiological and psychophysical phenomena in a single time-domain model.

The cascade arrangement of filters was selected to mimic cochlear signal processing; however, the TDLM was not intended to reproduce any specific BM or AN measurements. Still, this model is physiological in its distribution of level-dependent gain and may provide a computational advantage for implementation of level-dependent filtering when compared to the multi-stage filtering mechanisms used in some auditory models (e.g., Goldstein, 1995; Meddis *et al.*, 2001). Because the TDLM does not require a separate frequency analysis to implement compression and assuming equivalent frequency resolution of 12 outputs per octave, the computation advantage of the TDLM over the other models is estimated to be about a factor of two.

The expansive nonlinearity at low levels in the OHC stage of the TDLM was partly motivated by observations of AN spike rates at low levels (Yates *et al.*, 1990). The TDLM is not unique in this respect, as similar expansive nonlinearities have previously been incorporated into other auditory models (e.g., Goldstein, 1995; Dau *et al.*, 1997; Plack and Oxenham, 1998; Neely *et al.*, 2000 Jepsen *et al*., 2008).

In this study, the loudness of a tone was quantified as the maximum instantaneous loudness integrated across a two-octave span centered at the tone frequency. Other definitions of tone loudness based on the specific-loudness output of the TDLM are possible, but have not been investigated. Because the TDLM, as currently implemented, has 12 logarithmically-spaced outputs per octave, a two-octave span included 25 adjacent outputs, 12 above and 12 below the output with CF equal to the tone frequency. The restricted frequency span inherent in our definition of tone loudness could potentially have limited the extent to which *spread of excitation* was represented in the growth of our loudness functions; however, little quantitative differences were observed in the results when the CF bandwidth used to calculate loudness was decreased to 1/3 octave or increased to four octaves. The main reason that the CF bandwidth was limited to less than the entire range of hearing was to reduce computation time.

Weber’s Law suggests that the intensity JND should be independent of level. The observation that the intensity JND for tones decreases as level increases, a deviation from Weber’s Law known as the “near miss,” has been attributed to spread of excitation (e.g., Zwicker, 1956; Florentine and Buus, 1981) and is supported by data showing the effect of masking noise on the near miss (Moore and Raab, 1974). Although loudness growth in the TDLM is influenced to a limited extent by spread of excitation, it is primarily the decreasing value of the denominator of Eq. (3), which represents internal noise, relative to the numerator that causes the “near miss” in the TDLM (e.g., McGill and Goldberg, 1968). In other words, the “near miss” is due to the internal equivalent of the Weber fraction decreasing more rapidly with level than the corresponding increase in compression. This interpretation is quantified by Eq. (2). The traditional explanation for the “near miss” in terms of “spread of excitation” (e.g., McGill and Goldberg, 1968) is not necessarily inconsistent with the view expressed here, but requires consideration of multiple independent neural pathways, whereas, TDLM-predicted loudness, being integrated over essentially all cochlear outputs, provides only a single pathway to the central auditory system.

Even though transient responses of the TDLM are of interest, only steady-state responses of the TDLM are described here because the transit effects of neural adaptation have not yet been implemented. A widely held view is that persistence of masker excitation (*i.e.*, temporal smearing) is the dominant mechanism of forward masking (*e.g.*, Oxenham, 2001). An alternative view is that adaptation at the IHC synapse provides a more consistent and physiologically plausible interpretation of forward-masking data (*e.g.*, Meddis and O’Mard, 2005; Neely and Jesteadt, 2005). In the current TDLM, level-dependent aspects of steady-state adaptation are included in the instantaneous nonlinearity of the IHC stage. The addition of a *temporal contrast enhancement* stage to the TDLM could simulate forward masking. Similarly, the addition of a *modulation filter bank* following the IHC stage would provide a more complete representation of increment detection (Dau *et al.*, 1997; Gallun and Hafter, 2006; Jepsen *et al*., 2008).

IX. Conclusions

Loudness density is a quantity hypothesized to be available to the central auditory system upon which detection and discrimination decisions may be derived. Because the TDLM described in this paper is calibrated against standard ELLCs and produces the expected intensity-discrimination performance, its output helps us understand other psychophysical tasks, such as masking. The fact that the TDLM exhibits suppression in the presence of a masking tone that is similar to empirical estimates of suppression, without a separate suppressive mechanism, supports the view that suppression in the cochlea is due to the same mechanism that provides dynamic range compression.

Acknowledgements

Work supported by the NIH/NIDCD (R01 8318, R01 6648, R01 2251, and P30 4662).

Appendix A: Implementation of the time-domain loudness model

Technical details of the time-domain loudness model (TDLM) are described here. A MATLAB implementation of the TDLM is available from the first author upon request.

Middle-ear section

The middle-ear section of the model is a sixth-order discrete-time filter (*e.g.*, Oppenheim and Shafer, 1975) described by the following equation:

$$H_{me}(z) = s_{me} \cdot B\left(z \mid f_{0z}, Q_{0z}, f_{0p}, Q_{0p}\right) \cdot B\left(z \mid f_{1z}, Q_{1z}, f_{1p}, Q_{1p}\right) \cdot B\left(z \mid f_{2z}, Q_{2z}, f_{2p}, Q_{2p}\right), \quad \text{(A1)}$$

which incorporates three *biquad* filter sections:

$$B\left(z \mid f_z, Q_z, f_p, Q_p\right) = \frac{\left[1 + \dfrac{\sin(\varphi_z)}{2Q_z}\right] - 2\cos(\varphi_z)\, z^{-1} + \left[1 - \dfrac{\sin(\varphi_z)}{2Q_z}\right] z^{-2}}{\left[1 + \dfrac{\sin(\varphi_p)}{2Q_p}\right] - 2\cos(\varphi_p)\, z^{-1} + \left[1 - \dfrac{\sin(\varphi_p)}{2Q_p}\right] z^{-2}}. \quad \text{(A2)}$$

In Eq. (A2), $\varphi_z = 2\pi\, f_z / f_s$ and $\varphi_p = 2\pi\, f_p / f_s$, where f_s is the sampling rate. The function B represents a biquad discrete-time filter section and is the ratio of two quadratic functions of the sample-delay variable z^{-1}. Parameter values for the middle-ear section were selected to minimize differences between TDLM-predicted and ISO-226 ELLCs and are listed in Table A-I. The resulting middle-ear transfer function is shown in Fig. 5.

The input to the middle-ear section is the only signal in the TDLM with a physical interpretation and should be scaled in sound pressure units of pascal.

Cascade sections

Cascade-filter sections and associated second-filter sections are each biquad filters as described by Eq. (A2). The CFs of adjacent cascade-filter sections are separated by one semitone, which is one-twelfth of an octave. Every second-filter section has the same CF as the cascade-filter section to which it is attached. The transfer functions of the cascade-filter and second-filter sections have the following form:

$$H_c(z \mid f_c, Q_c) = \left(\frac{f_{cp}}{f_{cz}}\right)^2 B\left(z \mid f_{cz}, Q_{cz}, f_{cp}, Q_{cp}\right) \cdot C\left(f_{cz}, f_{cp}\right) \tag{A3}$$

$$H_s(z \mid f_c, Q_c) = B\left(z \mid f_{sz}, Q_{sz}, f_{sp}, Q_{sp}\right) \cdot C\left(f_{sz}, f_{sp}\right), \tag{A4}$$

where

$$C\left(f_z, f_p\right) = \left(\frac{f_z}{f_p}\right)^2 \frac{1-\cos\left(2\pi\, f_p / f_s\right)}{1-\cos\left(2\pi\, f_z / f_s\right)}. \tag{A5}$$

In these equations, Q_c is time varying and depends on the instantaneous value of the signal at the input to the filter. For the cascade-filter section, the pole and zero frequencies are $f_{cp} = f_c$ and $f_{cz} = 1.3 f_c$, respectively. For the second-filter section, the pole and zero frequencies are $f_{sp} = f_c$ and $f_{sz} = 0.5 f_c$, respectively. The Q value of the poles and zeros are related to Q_c by the following equations:

$$Q_{cz} = Q_c \cdot C\left(f_{cz}, f_{cp}\right)^{e_{cz}} \tag{A6}$$

$$Q_{cp} = Q_c \cdot C\left(f_{cz}, f_{cp}\right)^{e_{cp}} \tag{A7}$$

$$Q_{sz} = Q_c \cdot C\left(f_{sz}, f_{sp}\right)^{e_{sz}} \tag{A8}$$

$$Q_{sp} = Q_c \cdot C\left(f_{sz}, f_{sp}\right)^{e_{sp}} . \tag{A9}$$

Equations (A6)-(A9) all incorporate the function C from Eq. (A5), but with differing exponents e_{cz}, e_{cp}, e_{sz}, and e_{sp}, which have values of -4.6, -2.3, 2, and 5, respectively. These exponents were selected to maintain similarity between the shapes of TDLM transfer functions and gamma-tone filters over a wide range of CF and Q values. Note that $C \approx 1$ when $f_c << f_s$.

The value of Q_c varies with the instantaneous amplitude of the signal $x(t)$ at the input to the filter:

$$Q_c = Q_{\min} + \frac{Q_{\max} - Q_{\min}}{1 + \dfrac{\left|s_q x\right|}{c_q + \left|s_q x\right|^{e_q}}} . \tag{A10}$$

The parameters in Eq. (A10) are $s_q = 400$, $c_q = 0.1$, and $e_q = 0.5$, respectively. When x is large, the minimum value of Q_c is $Q_{\min} = 1$. When x is small, the maximum value of Q_c is $Q_{\max}$. The value of $Q_{\max}$ depends on the CF of filter. This dependence is a quadratic function of log(CF), except that $Q_{\max}$ is never less than $Q_{\min}$. Quadratic coefficients were selected to give $Q_{\max}$ values of 2.0, 4.3, and 5 when CF is 0.05, 1, and 5 kHz, respectively. The dependence of $Q_{\max}$ on CF is shown in Fig. 3.

Although each cascade-filter and second-filter section is a single biquad, which can be implemented with just a few multiply and add operations, complexity in the TDLM design is introduced in the equations that relate CF and Q to the corresponding discrete-time filter coefficients. These equations are important because they ensure well-behaved filter responses over a wide range of CF and Q values. For efficiency, discrete-time filter coefficients may be precomputed for the known range of Q values, so that only *table lookups* are required when processing a signal. A summary of parameter values for cascade-filter and second-filter sections is provided in Table A-II.

IHC sections

The output of each second-filter section is input to an IHC section. The IHC section consists of three parts: (1) a first-order high-pass filter, (2) an instantaneous nonlinearity, and (3) a first order low-pass filter. The high-pass filter represents mechano-electric transduction at the hair bundle of the IHC (*e.g.*, Howard *et al.*, 1988):

$$H_{hb}(z) = \frac{1 - z^{-1}}{1 - \exp\left(\frac{-1}{\tau_{hb}\, f_s}\right) z^{-1}} . \tag{A11}$$

In this equation, $\tau_{hb} = 1$ ms and f_s is the sampling rate. The low-pass filter represents IHC membrane:

$$H_{cm}(z) = \frac{N_0}{1 - \exp\left(\frac{-1}{\tau_{cm}\, f_s}\right) z^{-1}} . \tag{A12}$$

In this equation, $\tau_{cm} = 1$ ms. The scale factor $N_0 = 2.4$ makes the loudness equal to one for a 1-kHz tone at 0 dB SPL.

The instantaneous nonlinearity transforms the output of the high-pass filter y_{hb} into the input to the low-pass filter x_{cm}:

$$x_{cm} = \frac{\left|s_v \, y_{hb}\right|^4}{1 + \left|s_v \, y_{hb}\right|^{e_v}} . \tag{A13}$$

The scale factor and exponent in this equation are $s_v = 60$ and $e_v = 3$, respectively. The effect of this nonlinearity is illustrated by the lowest solid line in Fig. 7, which does not include any contribution from cascade-filter and second-filter sections. At low levels, the exponent 4 in the numerator of Eq. (A13) makes the compression equal to 1/2 (see the medium solid line in Fig. 4), because loudness is considered to be a perceptual *intensity*. At high levels, the exponent 4 in the numerator is reduced by 3 in the denominator for a net exponent of 1, which makes compression equal to 2. Parameter values for the IHC section are listed in Table A-III. Note that the absolute value in Eq. (A13) provides full-wave rectification, unlike the half-wave rectification observed in physiological measurements. As with many of features of the model, this detail is uncertain and subject to change upon further investigation. The model output is scaled to provide a loudness value equal to 1 when the input to the middle-ear stage is a 1-kHz tone at 0 dB SPL.

Although the TDLM is not expected to account for membrane and synaptic processes directly, the current implementation of the IHC stage is admittedly oversimplified and must incorporate additional elements in future versions in order to provide a mechanism for forward masking.

References

Abbas, P. J. and Sachs, M. B. (**1976**). “Two-tone suppression in auditory-nerve fibers: Extension of a stimulus-response relationship,” J. Acoust. Soc. Am. **59**, 112-122.

Allen, J. B. (**2007**). “Nonlinear cochlear signal processing and masking in speech perception,” in *Handbook on Speech Processing and Speech Communication* , edited by J. Benesty, M. M. Sondhi, and Y. Huang (Springer, Berlin), pp. 27-60.

Allen, J. B., and Neely, S. T. (**1992**). ‘‘Micromechanical models of the cochlea,’’ Phys. Today **45**, 40–47.

Allen, J. B. and Neely, S. T. (**1997**). “The relation between the intensity JND and loudness for pure tones and wide-band noise,” J. Acoust. Soc. Am. **102**, 3628-3646.

Allen, J. B. and Sen, D. (**2003**). “The role of micromechanics in explaining two tone suppression and the upward spread of masking,” in *Biophysics of the Cochlea: From Molecules to Models,* edited by A. W. Gummer, E. Dalhoff, M. Nowotny, M. P. Scherer (World Scientific, Singapore), pp. 383-392.

Buus, S., and Florentine, M. (**2002**). “Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered,” J. Assoc. Res. Otolaryngol. **3**, 120–139.

Buus, S., Florentine, M., and Poulsen (**1997**). “Temporal integration of loudness, loudness discrimination, and the form of the loudness function,” J. Acoust. Soc. Am. **101**, 669-680.

Cai, S., Ma, W.-L., Young, E. D. (2009). “Encoding intensity in ventral cochlear nucleus following acoustic trauma: implications for loudness recruitment,” J. Assoc. Res. Otolaryngol., **10**, 5-22.

Carney, L. H. (**1994**). “Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness,” Hear. Res. **76**, 31–44.

Chalupper, J. and Fastl, H. (**2002**). “Dynamic loudness model (DLM) for normal and hearing-impaired listeners,” Acustica, **88**, 378-386.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). “Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers,” J. Acoust. Soc. Am. **102**, 2892-2905.

Delgutte, B. (1990). “Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers,” J. Acoust. Soc. Am. **87**, 791-809.

Florentine, M. and Buus, S. (**1981**). “An excitation-pattern model for intensity discrimination,” J. Acoust. Soc. Am. **70**, 1646-1654.

Florentine, M., Reed, C. M., Rabinowitz, W. M., Braida, L. D., Durlach, N. I., and Buus, S. (**1993**). “Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss,” J. Acoust. Soc. Am. **94**, 2575-2586.

Galambos, R. and Davis H. (**1944**). “Inhibition of activity in single auditory nerve fibers by acoustic stimulation,” J. Neurophysiol. **7**, 287-303.

Gallun, F. J. and Hafter, E. R. (**2006**). “Amplitude modulation sensitivity as a mechanism for increment detection,” J. Acoust. Soc. Am. **119**, 3919-3930.

Geisler, C. D. and Nuttall, A. L. (1997). “Two-tone suppression of basilar membrane vibrations in the base of the guinea pig cochlea using ‘low-side’ suppressors,” J. Acoust. Soc. Am. **102**, 430-440.

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

Glasberg, B. R., and Moore, B. C. J. (**2005**). “Development and evaluation of a model for predicting the audibility of time-varying sounds in the presence of background sounds,” J. Audio Eng. Soc. **53**, 906-918.

Glasberg, B. R., and Moore, B. C. J. (**2006**). “Prediction of absolute thresholds and equal-loudness contours using a modified loudness model,” J. Acoust. Soc. Am. **120**, 585-588.

Goldstein, J. L. (**1995**). “Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: data and MBPNL model,” Hear. Res. **89**, 52–68.

Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Kristin Groenenboom, K., Tan, H., and Stiegemann, B. (**2008**). “Low-frequency and high-frequency distortion product otoacoustic emission suppression in humans," J. Acoust. Soc. Am. **123**, 2172-2190.

Green D., and Swets J. (**1966**). *Signal Detection Theory and Psychophysics*. Wiley, (New York), pp. 64-69.

Hall, J. L. (**1980**). "Cochlear models: two-tone suppression and the second filter," J. Acoust. Soc. Am. **67**, 1722-1728.

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neur. Comput. **13**, 2273-2316.

Heinz, M. G., Issa, J. B. and Young, E. D. (**2005**). "Auditory-nerve rate responses are inconsistent with common hypotheses for the neural correlates of loudness recruitment," J. Assoc. Res. Otolaryngol. **6**, 1525-3961.

Hellman, R. P. (**1976**). "Growth of loudness at 1000 and 3000 Hz," J. Acoust. Soc. Am. **60**, 672-679.

Hellman, R. P., and Meiselman, C. H. (**1990**). "Loudness relations for individuals and groups in normal and impaired hearing," J. Acoust. Soc. Am. **88**, 2596-2606.

Hellman, R. P., and Zwislocki, J. J. (**1961**). "Some factors affecting the estimation of loudness," J. Acoust. Soc. Am. **35**, 687-694.

Howard, J., Roberts, W. M, Hudspeth, A. J. (**1988**). "Mechanoelectrical transduction by hair cells," Ann. Rev. Biophys. Biophys. Chem. **17**, 99-124.

ISO-226 (2003). “Acoustics—Normal equal-loudness-level contours,” International Organization for Standardization, (Geneva, Switzerland).

Javel, E., McGee, J., Walsh, E. J., Farley, G. R., and Gorga, M. P. (1983). “Suppression of auditory nerve responses. II. Suppression threshold and growth, iso-suppression contours,” J. Acoust. Soc. Am. **74**, 801-813.

Jepsen, M. L., Ewert, S. D., and Dau, T. (**2008**). “A computational model of human auditory signal processing and perception,” J. Acoust. Soc. Am. **124**, 422-438.

Jesteadt, W. and Leibold, L. J. (**2010)**. “Loudness in the laboratory: steady-state sounds,” in *Loudness*, edited by M. Florentine (Springer, New York), pp. 42-48.

Jesteadt, W., Wier, C., and Green, D. (**1977**). “Intensity discrimination as a function of frequency and sensation level,” J. Acoust. Soc. Am. **61**, 169-177.

Kates, J. M. (**1991**). “A time-domain cochlear model,” IEEE Trans. Sig. Proc. **39**, 2573-2592.

Leibold, L. J., Tan, H. Khaddam, S., and Jesteadt, W. (2007). “Contributions of individual components to the overall loudness of a multitone complex,” J. Acoust. Soc. Am. 121 2822-2831.

Liu, Y.-W. and Neely, S.T. (**2010**) “Distortion product emissions from a cochlear model with nonlinear mechanoelectrical transduction in outer hair cells,” J. Acoust. Soc. Am. **127**, 2420-2432.

Lyon, R. F. (**1982**). "A computational model of filtering, detection, and compression in the cochlea", in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp 1148-1151.

McGill, W. J. and Goldberg, J. P. (**1968**). “A study of the near-miss involving Weber’s Law and pure-tone intensity discrimination,” Perception & Psychophys. **4**, 105-109.

Meddis, R., and O'Mard, L. P. (**2005**). “A computer model of the auditory-nerve response to forward-masking stimuli,” J. Acoust. Soc. Am. **117**, 3787-3798.

Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). “A computational algorithm for computing nonlinear auditory frequency selectivity,” J. Acoust. Soc. Am. **109**, 2852-2861.

Miskolczy-Fodor, F. (**1960**). “Relation between loudness and duration of tonal pulses. III. Response in cases of abnormal loudness function,” J. Acoust. Soc. Am. **32**, 486-492.

Moore, B. C. J. (2004). “Testing the concept of softness imperception: Loudness near threshold for hearing-impaired ears,” J. Acoust. Soc. Am. **115**, 3103-3111.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). “A model for the prediction of thresholds, loudness and partial loudness,” J. Audio Eng. Soc. **45**, 224-240.

Moore, B. C. J., and Raab, D. H. (**1974**). "Pure-tone intensity discrimination: some experiments relating to the 'near-miss' to Weber's Law," J. Acoust. Soc. Am. **55**, 1049-1054.

Moore, B. C. J. and Vickers, D. A. (**1997**). "The role of spread excitation and suppression in simultaneous masking," J. Acoust. Soc. Am. **102**, 2284-2290.

Neely, S. T. and Jesteadt, W. (**2005**). "Quadratic compression model of auditory discrimination and detection," Acta Acustica united with Acustica **91,** 980-991.

Neely, S. T., Gorga, M. P. and Dorn, P. A. (**2000**). "Distortion product and loudness growth in an active, nonlinear model of cochlear mechanics," in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama and T. Koike (World-Scientific Publ. Ltd., Singapore), pp. 237-243.

Neely, S. T., Gorga, M. P. and Dorn, P. A. (**2003**). "Cochlear compression estimates from measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **114**, 1499-1507.

Neely, S. T., Schairer, K. and Jesteadt, W. (**2005**). "Estimates of cochlear compression from measurements of loudness growth," in *Auditory Signal Processing: Physiology, Psychoacoustics and Models*, edited by D. Pressnitzer, A. deCheveigne, S. McAdams and L. Collet (Springer, New York), pp. 42-48.

Oppenheim, A. V. and Schafer, R. W. (**1975**). *Digital Signal Processing*. Prentice-Hall, (Englewood Cliffs, New Jersey), pp. 195-283.

Oxenham, A. J. (**2001**). “Forward masking: Adaptation or integration?” J. Acoust. Soc. Am. **109**, 732-741.

Patterson, R. D and Holdsworth, J. (**1996**). “A functional model of neural activity patterns and auditory images,” in *Advances in Speech, Hearing and Language Processing*, edited by A. W. Ainsworth (JAI Press, London), pp. 547-563.

Plack, C. J., and Oxenham, A. J. (**1998**). “Basilar-membrane nonlinearity and the growth of forward masking” J. Acoust. Soc. Am. **103**, 1598–1608.

Rhode, W. S. (**1977**). “Some observations of two-tone interactions measured using the Mössbauer effect,” in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 27–38.

Rhode, W. S. and Robles, L. (**1974**). “Evidence from Mössbauer experiments for nonlinear vibration in the cochlea,” J. Acoust. Soc. Am. **55**, 588-596.

Rodriguez, J., Neely, S. T., Patra, H., Kopun, J., Jesteadt, W., Tan, H., and Gorga, M. P. (**2010**). “The role of suppression in psychophysical tone-on-tone masking,” J. Acoust. Soc. Am. **127**, 361-369.

Sachs, M. B. and Abbas, P. J. (**1974**). “Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli,” J. Acoust. Soc. Am. **56**, 1835-1847.

Sachs, M. B. and Kiang, N. Y.-S. (**1968**). “Two-tone inhibition in auditory nerve fibers,” J. Acoust. Soc. Am. **43**, 1120-1128.

Schairer, K. S., Nizami, L., Reimer, J. F., and Jesteadt, W. (**2003**). "Effects of peripheral nonlinearity on psychometric functions for forward-masked tones," J. Acoust. Soc. Am. **113**, 560-1573.

Scharf, B., and Hellman, R. P. (**1966**). "Model of loudness summation applied to impaired ears," J. Acoust. Soc. Am. **40**, 71–78.

Schroder, A. C., Viemeister, N. F., Nelson, D. A. (1994). "Intensity discrimination in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **96**, 2683-2693.

Shera, C. A., Guinan, J. J, and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," PNAS **99**, 3318-3323.

Shera, C. A., Guinan, J. J, and Oxenham, A. J. (**2010**). "Otoacoustic Estimation of Cochlear Tuning: Validation in the Chinchilla," J. Assoc. Res. Otolaryngol. **(in press)**.

Watts, L., Kerns, D., Lyon, R., and Mead, C. (**1992**). "Improved implementation of the silicon cochlea", IEEE Journal of Solid-state Circuits, **27**(5), 692-700.

Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," Hear. Res. **45**, 203–220.

Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). “A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression,” J. Acoust. Soc. Am. **109**, 648-670.

Zilany, M. S. A., and Bruce, I. C. (**2006**). “Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery,” J. Acoust. Soc. Am. **120**, 1446–1466.

Zweig, G., Lipes, R., and Pierce, J. R. (**1976**). “The cochlear compromise,” J. Acoust. Soc. Am. **59**, 975-982.

Zwicker, E. (**1956**). “Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs (The basic principles for determining the information capacity of hearing)," Acustica **6**, 356-381.

Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). “Critical bandwidth in loudness summation,” J. Acoust. Soc. Am. **29**, 548–557.

Zwislocki, J. J., and Jordan, H. N. (**1986**). “On the relations of intensity jnds to loudness and neural noise,” J. Acoust. Soc. Am. **79**, 772-780.

Endnote

[1]The loudness units (LU) of Fletcher and Munson (1933, Table III) were converted to decibels by taking $10 \cdot \log_{10}(LU)$. This formula converts an intensity ratio to decibels and assumes that loudness represents a perceptual intensity. The mean psychometric-function slope from Schairer *et al*. (2003, Fig. 7) was converted to compression by taking its multiplicative inverse. The other compression data in Fig. 4 (Neely *et al*. 2005; Neely and Jesteadt, 2005) was previously published in compatible units.

Tables

Table A-I. Middle-ear parameters used in Eqs. (A1)-(A2). Frequencies are in kHz.

f_{0z}	0	Q_{0z}	0.73
f_{0p}	0.0267	Q_{0p}	0.16
f_{1z}	1.394	Q_{1z}	4.00
f_{1p}	1.297	Q_{1p}	2.36
f_{2z}	8.688	Q_{2z}	0.94
f_{2p}	3.776	Q_{2p}	1.55
		s_{me}	0.493

Table A-II. Cascade-filter and second-filter parameters used in Eqs. (A3)-(A10). Arguments of $Q_{\max}$ are CFs in kHz.

f_{cz}/f_{cp}	1.3	e_{cz}	-4.6
f_{sz}/f_{sp}	0.5	e_{cp}	-2.3
$Q_{\min}$	1.0	e_{sz}	2.0
$Q_{\max}(0.05)$	2.0	e_{sp}	5.0
$Q_{\max}(1)$	4.3	s_q	400
$Q_{\max}(5)$	5.0	c_q	0.1
		e_q	0.5

Table A-III. IHC parameters used in Eqs. (A11)-(A13). Time constants τ_{hb} and τ_{cm} are in ms.

τ_{hb}	1
s_v	60
e_v	3.0
τ_{cm}	1
N_0	2.4

Figure Captions

FIG. 1. Cascade arrangement of second-order filter sections. Second-order second-filter sections (g_1, g_2…g_N) are attached to each cascade filter section (f_1, f_2…f_N). Each cascade-filter and second-filter section is implemented as a second-order discrete-time filter with time-varying filter coefficients. The input signal labeled x represents cochlear vestibule pressure, which differs from the acoustic waveform by a high-pass filter representing the middle-ear transfer function. The output signals labeled y_1, y_2 …y_N represent IHC inputs. An instantaneous nonlinearity and low-pass filter are applied to the IHC inputs to produce loudness density.

FIG. 2. Frequency-responses of cascade and second filter sections at 1 kHz. The cascade filter is low-pass because the pole frequency precedes the zero frequency. The second filter is high-pass because the zero frequency precedes the pole frequency. The ratio of the zero frequency to the pole frequency is 1.3 for the cascade filter and 0.5 for the second filter. Poles and zeros in this example have either Q=5 (solid lines), which is the largest value of Q_{max} at any CF, or Q=1 (dashed lines), which is the value of Q_{min} at all CFs and is also the value of Q_{max} when simulating a hearing impairment condition.

FIG. 3. (color online) Comparison of filter quality factors as functions of their CF. The thick solid line shows how Q_{max} in the cascade-filter and second-filter sections varies with CF. The thin solid line shows Q_{min}, which does not vary with CF. The circles indicate the Q_{erb} of TDLM IHC transfer functions at octave CFs from 0.062 to 16 kHz. The dashed line is the psychophysical Q_{erb} calculated by the ERB formula described by

Glasberg and Moore (1990), which is based on measurements of tone thresholds with simultaneous notched-noise maskers.

FIG. 4. (color online) TDLM compression at CF=1 kHz with Q_{max}=4 (thick solid line) and Q_{max}=1 (medium solid line). Compression is defined as the reciprocal of the slope of loudness level versus physical signal level. The larger value of Q_{max} simulates normal hearing, while the smaller value of Q_{max} simulates hearing impairment due to loss of OHC function. Other compression estimates based on psychophysical measurements are shown for comparison: the classic Fletcher-Munson loudness data (thin solid line; Fletcher and Munson, 1933), loudness-quadrupling measurements (squares; Neely *et al.*, 2005), and forward-masking psychometric-function slopes (circles; Schairer *et al.*, 2003). The quadratic-compression function described by Neely and Jesteadt (2005) is also shown for comparison (dashed line).

FIG. 5. (color online) Gain of the middle-ear transfer function between ear-canal pressure and vestibule pressure in the TDLM. The middle-ear section of the TDLM is implemented as a sixth-order, IIR digital filter and helps to reproduce some of the frequency dependence observed in standard ELLCs.

FIG. 6. TDLM gain at two levels in response to a frequency-swept tone. The instantaneous frequency increased at a logarithmic rate from 0.03 to 30 kHz over a time span of 8 seconds. The thin and thick lines represent the TDLM gain when pressure at the input to the middle-ear stage was 0 and 80 dB SPL, respectively.

FIG. 7. (color online) Comparison of TDLM-predicted and ISO-226 ELLCs. The TDLM contours (solid lines) represent horizontal slices through loudness-level functions at various frequencies. The ISO-226 (2003) contours (dashed lines) represent 0 to 90 phones in 10 dB steps.

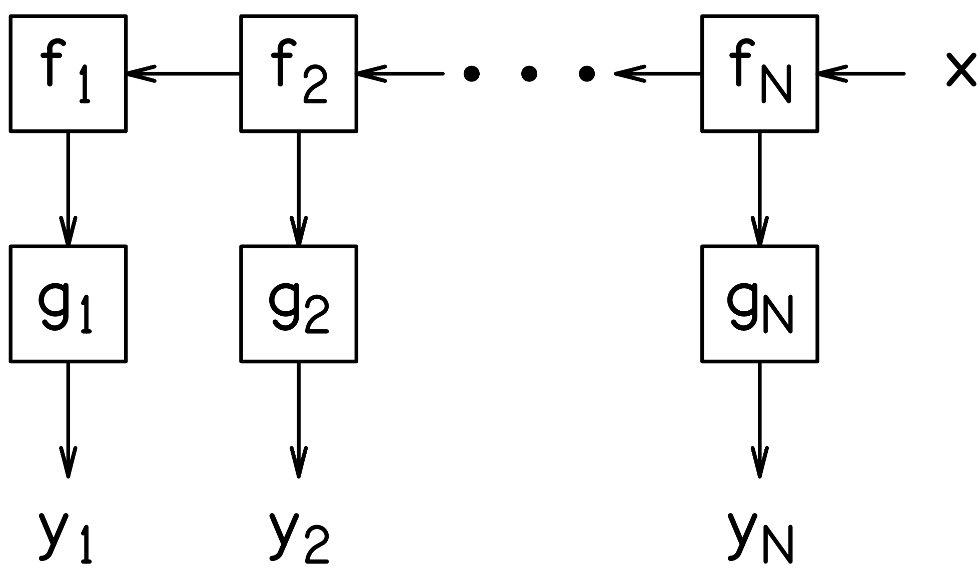
FIG. 8. (color online) Comparison of normal-hearing (thick line) and hearing-impaired (medium line) loudness functions for a 1-kHz tone. The plus symbol indicates threshold for the HI condition, which is elevated 43 dB from the NH threshold. The NH loudness function (thick line) is more compressed than the Fletcher-Munson (1933) loudness function (thin line). Compression for these three loudness functions was presented in Fig. 4.

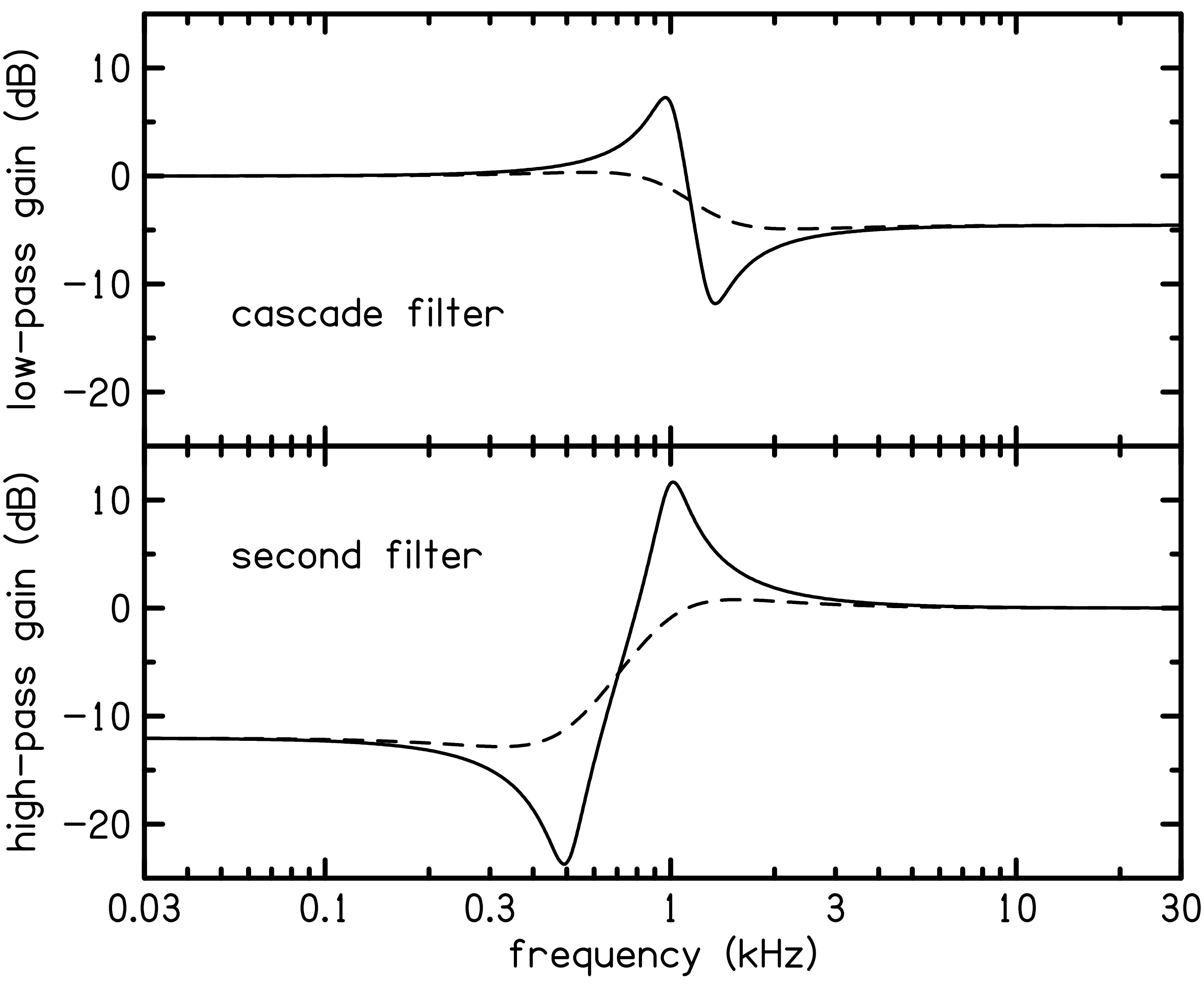
FIG. 9. (color online) Loudness of multi-tone complex as a function of bandwidth. The closed circles represent mean level, averaged across seven subjects, of a 1-kHz tone judged equal in loudness to the multi-tone complex at 60 dB SPL from Leibold et al. (2007). The open circles are estimates of the loudness-matched tone level from the Moore *et al.* (1997) model. The squares and triangles are estimates of the loudness-matched tone level from TDLM in the NH and HI conditions, respectively. The vertical line indicates the equivalent rectangular bandwidth (ERB) at 1000 Hz (Glasberg and Moore, 1990). [Figure adapted from Leibold et al. (2007, Fig. 1).]

FIG. 10. (color online) Comparison of normal-hearing and hearing-impaired intensity JNDs at 1 kHz. The NH and HI Weber factions ($\Delta I/I$) at 1 kHz are computed from Eq. (1). Previously published NH measurements (Jesteadt *et al.*, 1977) are also shown for comparison (squares).

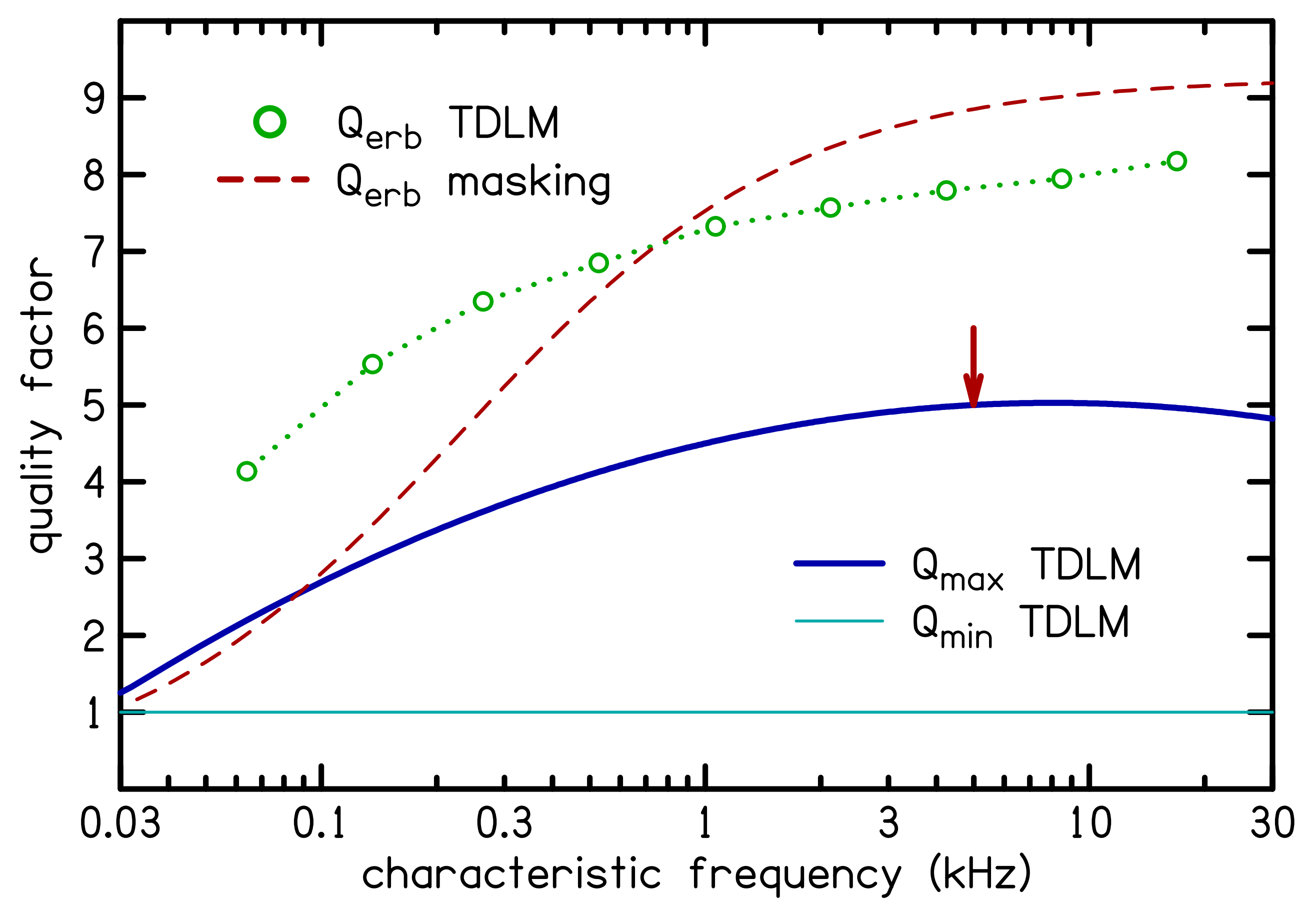
FIG. 11. (color online) Detection performance (d') as a function of signal level for different masker levels. The signal frequency was 4 kHz. The masker frequencies were 4.282 (upper panel) and 2.141 (lower panel) kHz. The leftmost curve in each panel is the same and represents the condition with no masker. The other curves represent masker levels from 0 to 80 dB SPL in 10 dB steps. The arrows illustrate the definition of suppression as the horizontal shift in the signal level for which $d' = 1$.

FIG. 12. (color online) TDLM suppression compared with psychophysical suppression estimates based on the difference in the amount of masking (DAM) between simultaneous masking and forward masking (Rodriguez *et al.*, 2010).

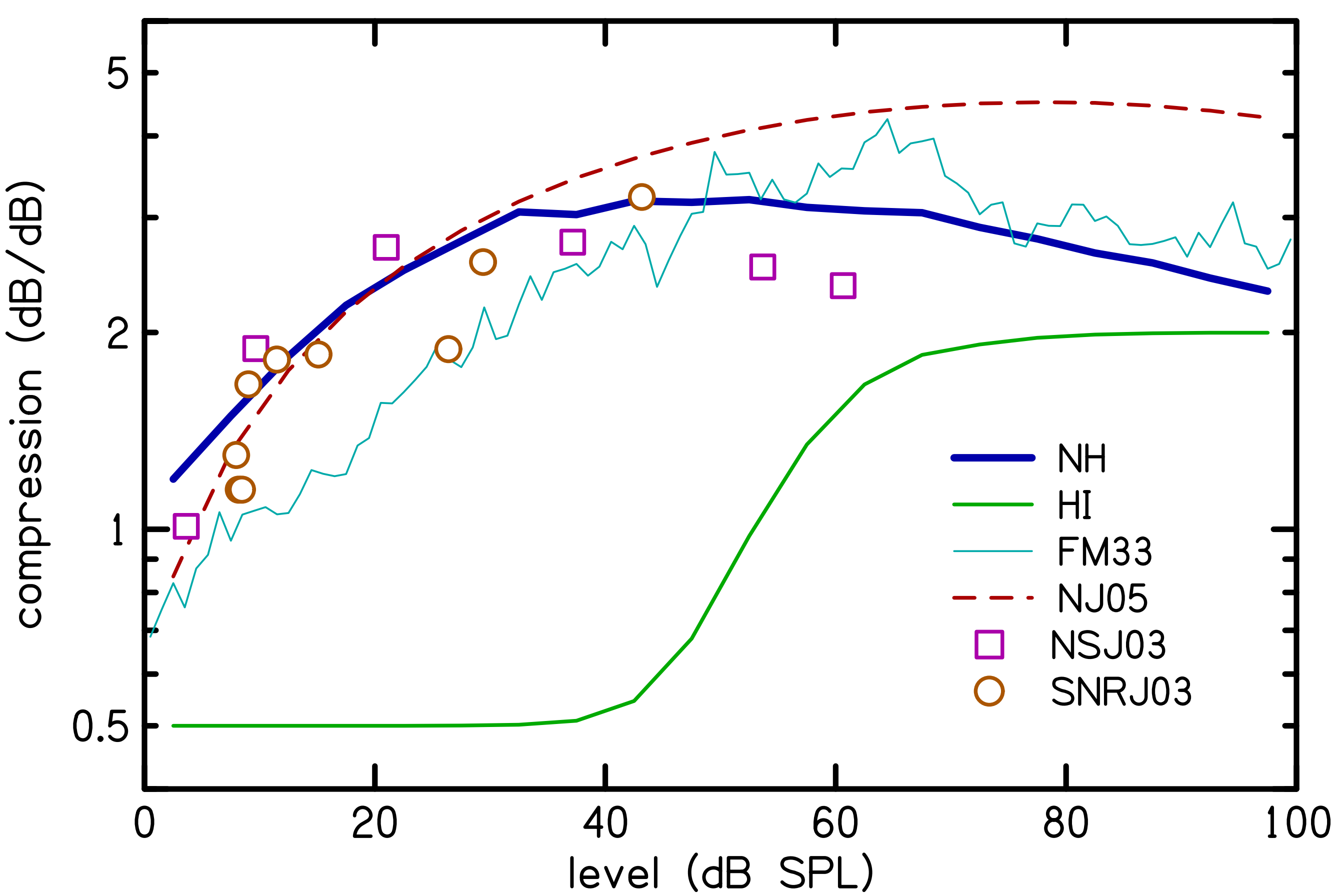

f1
f2
fN
x
g1
g2
gN
y1
y2
yN



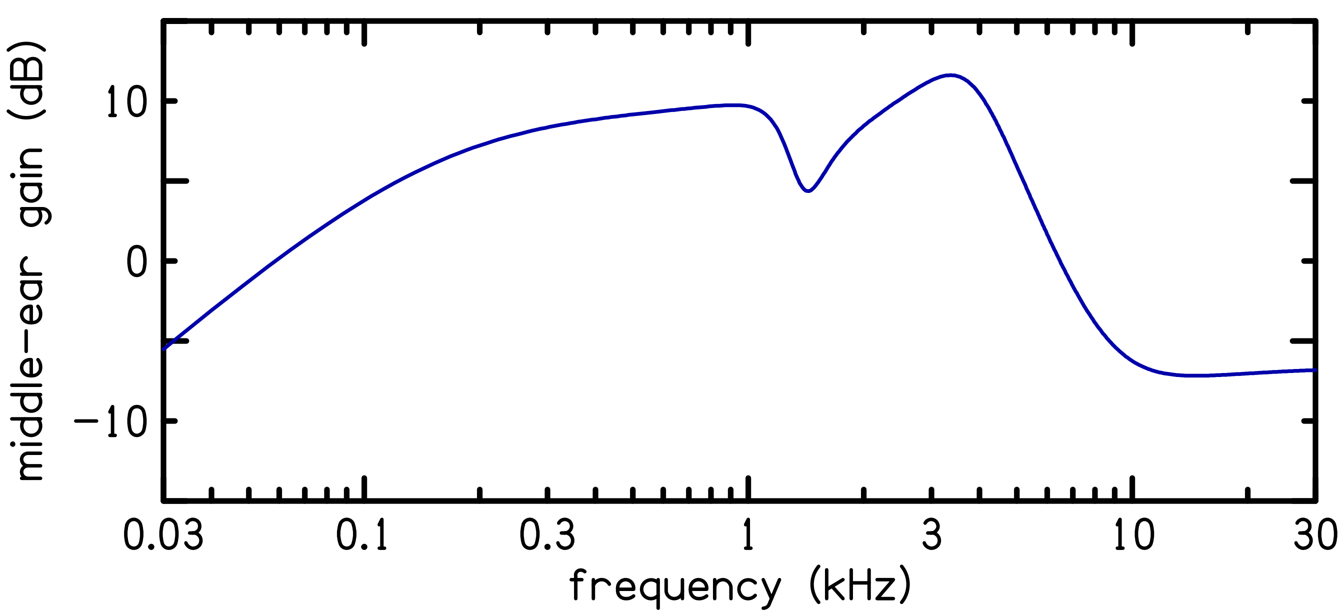
10
0
−10
−20
low-pass gain (dB)
cascade filter
10
0
−10
−20
high-pass gain (dB)
second filter
0.03
0.1
0.3
1
3
10
30
frequency (kHz)

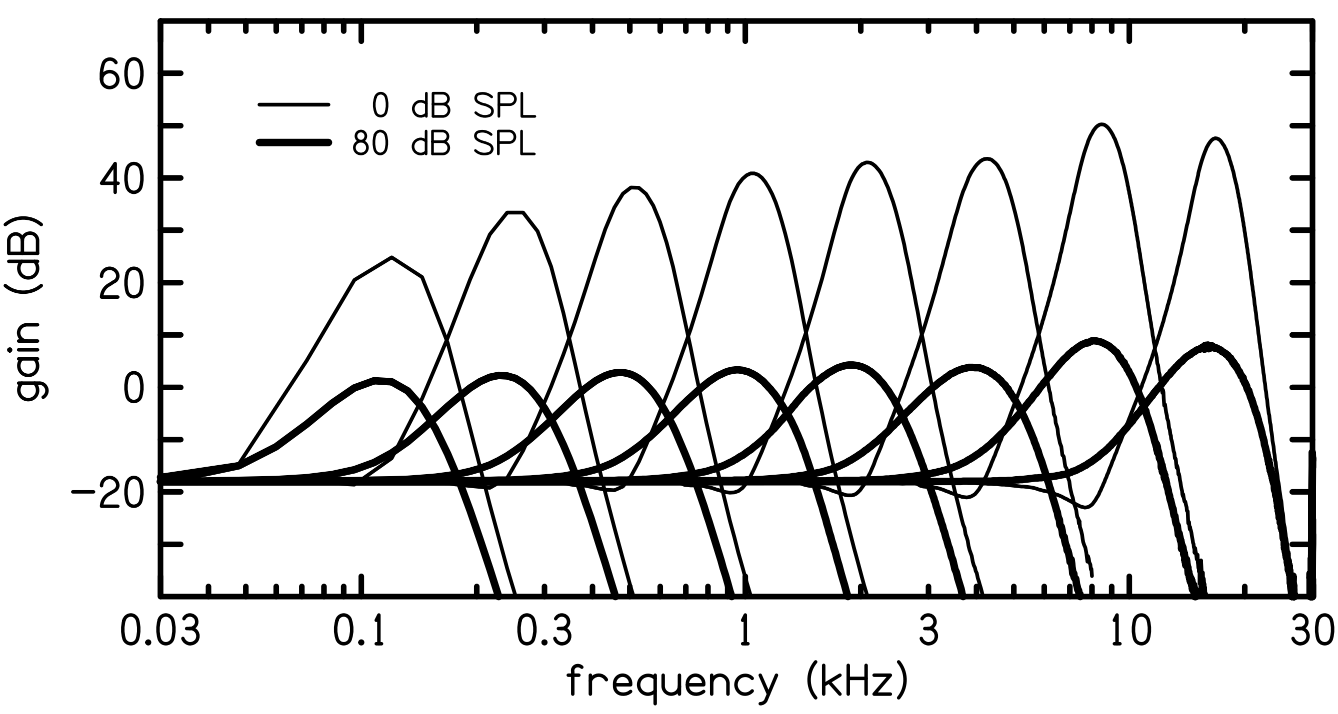


Q_erb TDLM
Q_erb masking
Q_max TDLM
Q_min TDLM
quality factor
characteristic frequency (kHz)
0.03
0.1
0.3
1
3
10
30
1
2
3
4
5
6
7
8
9

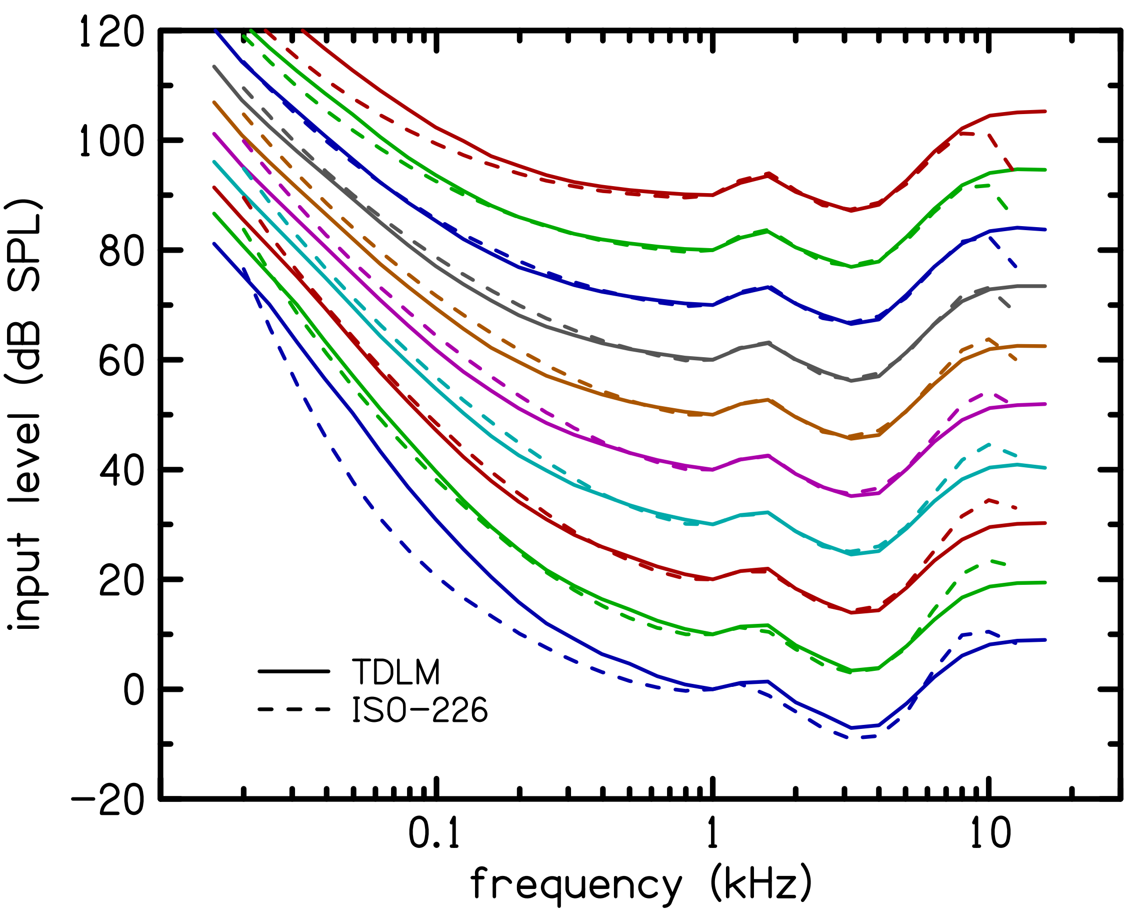


compression (dB/dB)
level (dB SPL)
5
2
1
0.5
0
20
40
60
80
100
NH
HI
FM33
NJ05
NSJ03
SNRJ03

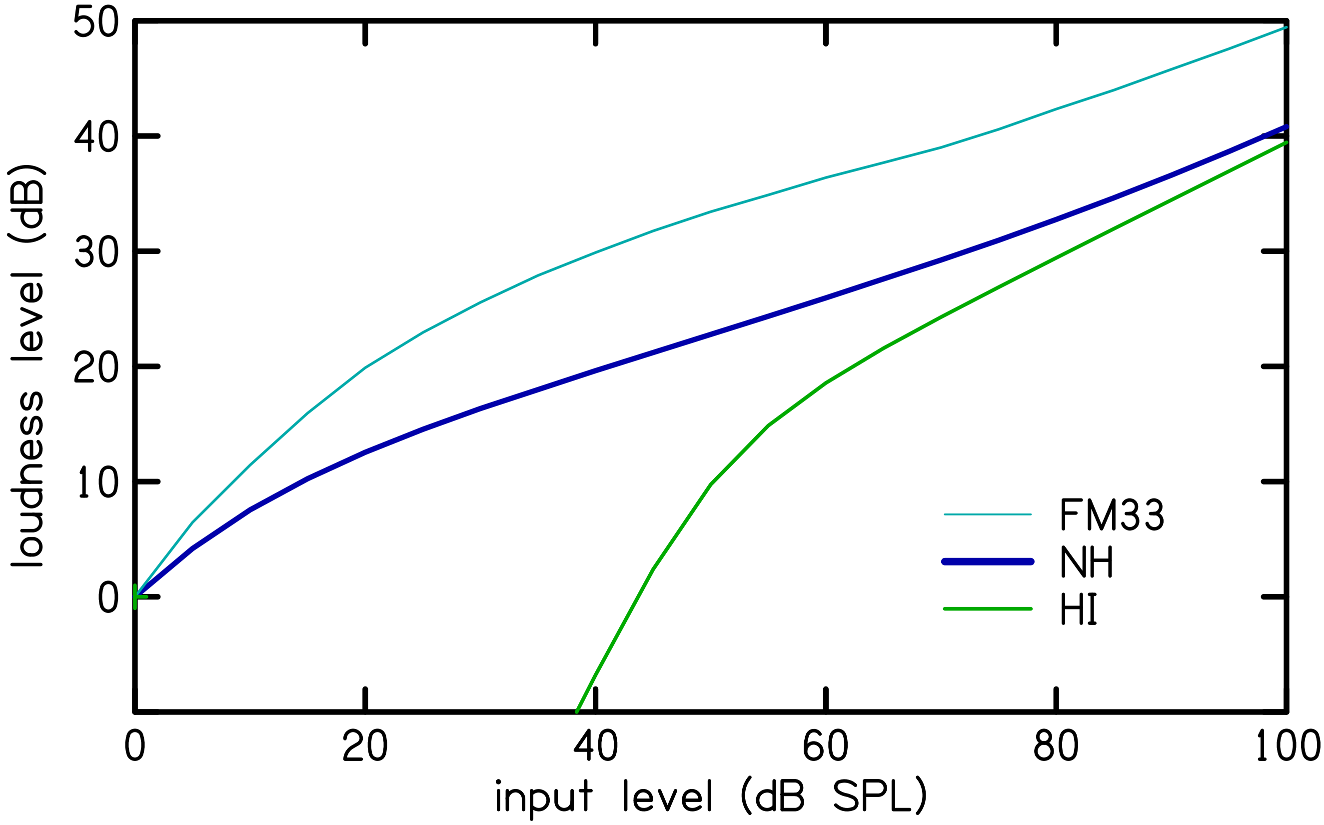

middle-ear gain (dB)
10
0
−10
0.03
0.1
0.3
1
3
10
30
frequency (kHz)



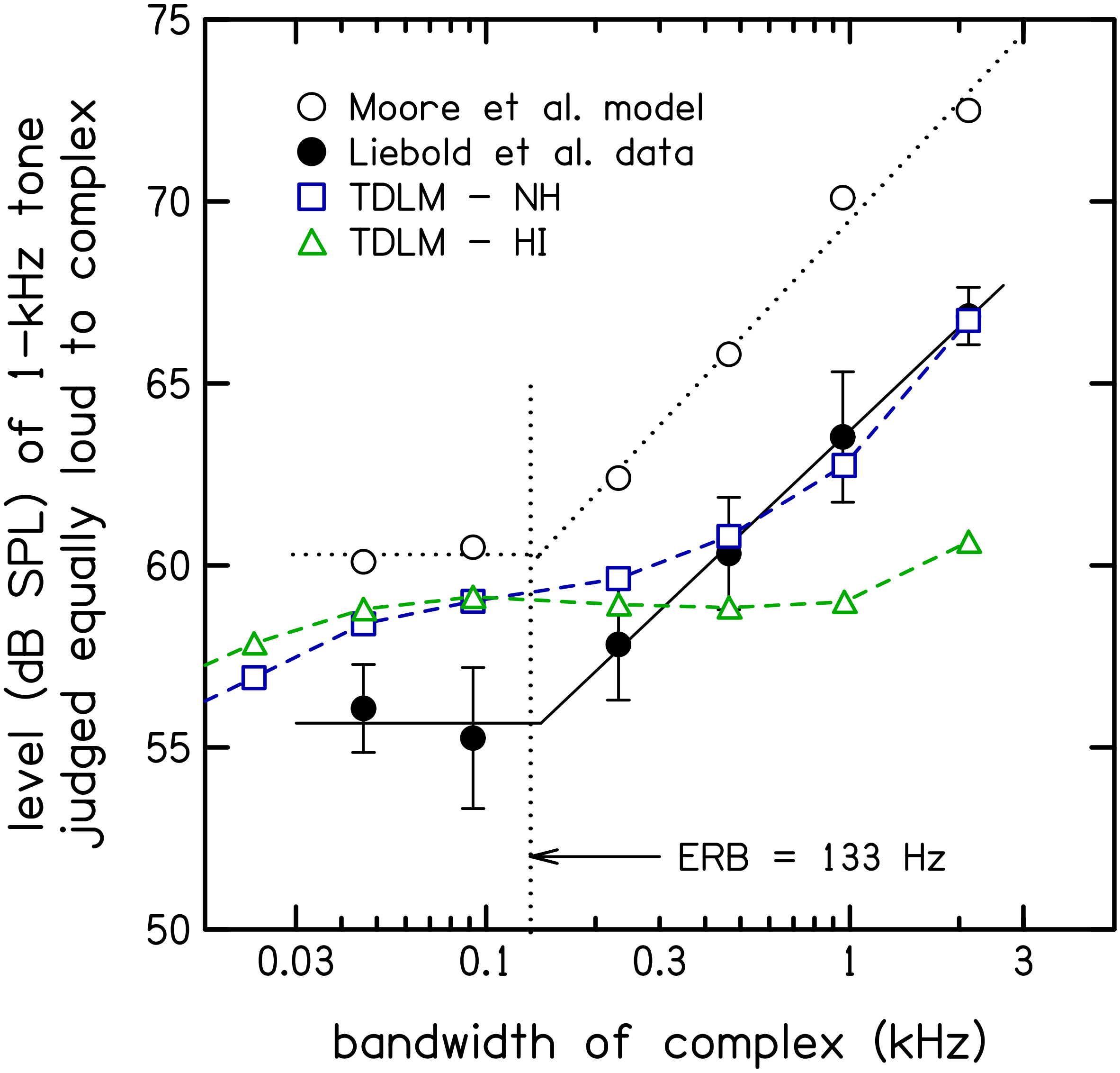
0 dB SPL
80 dB SPL
60
40
20
0
−20
gain (dB)
0.03
0.1
0.3
1
3
10
30
frequency (kHz)



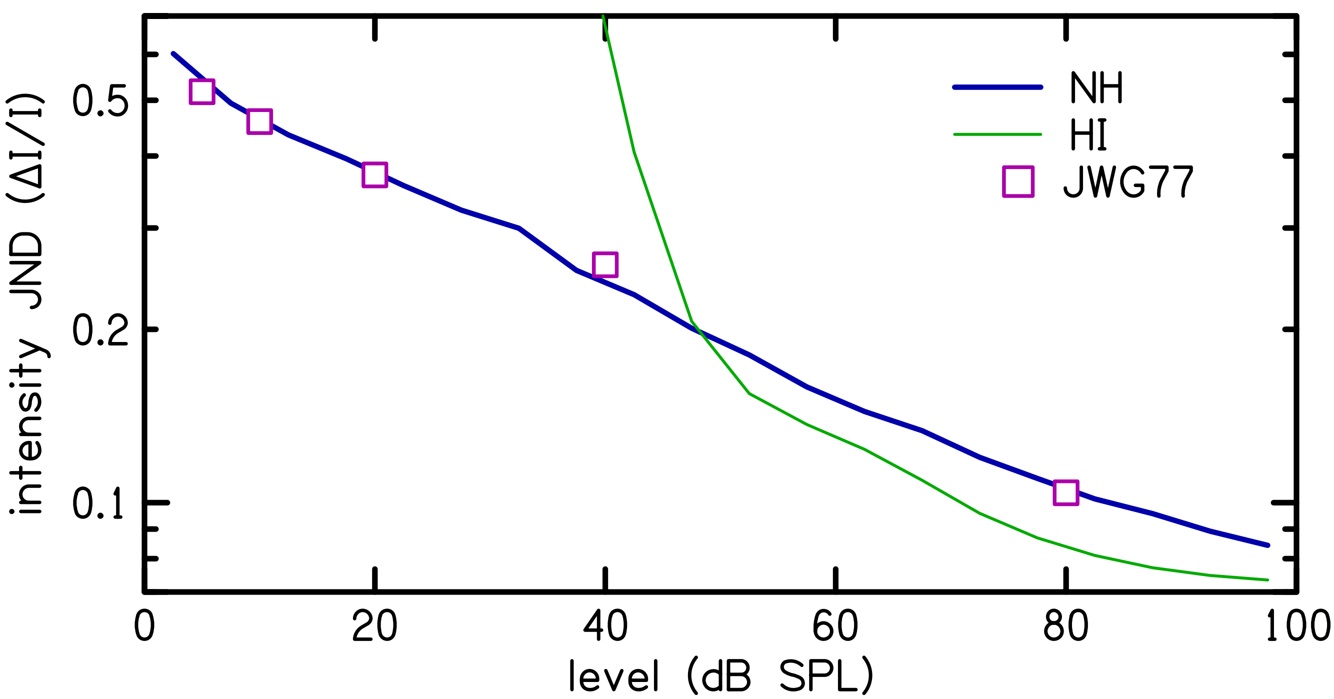
input level (dB SPL)
120
100
80
60
40
20
0
−20
0.1
1
10
frequency (kHz)
TDLM
ISO−226



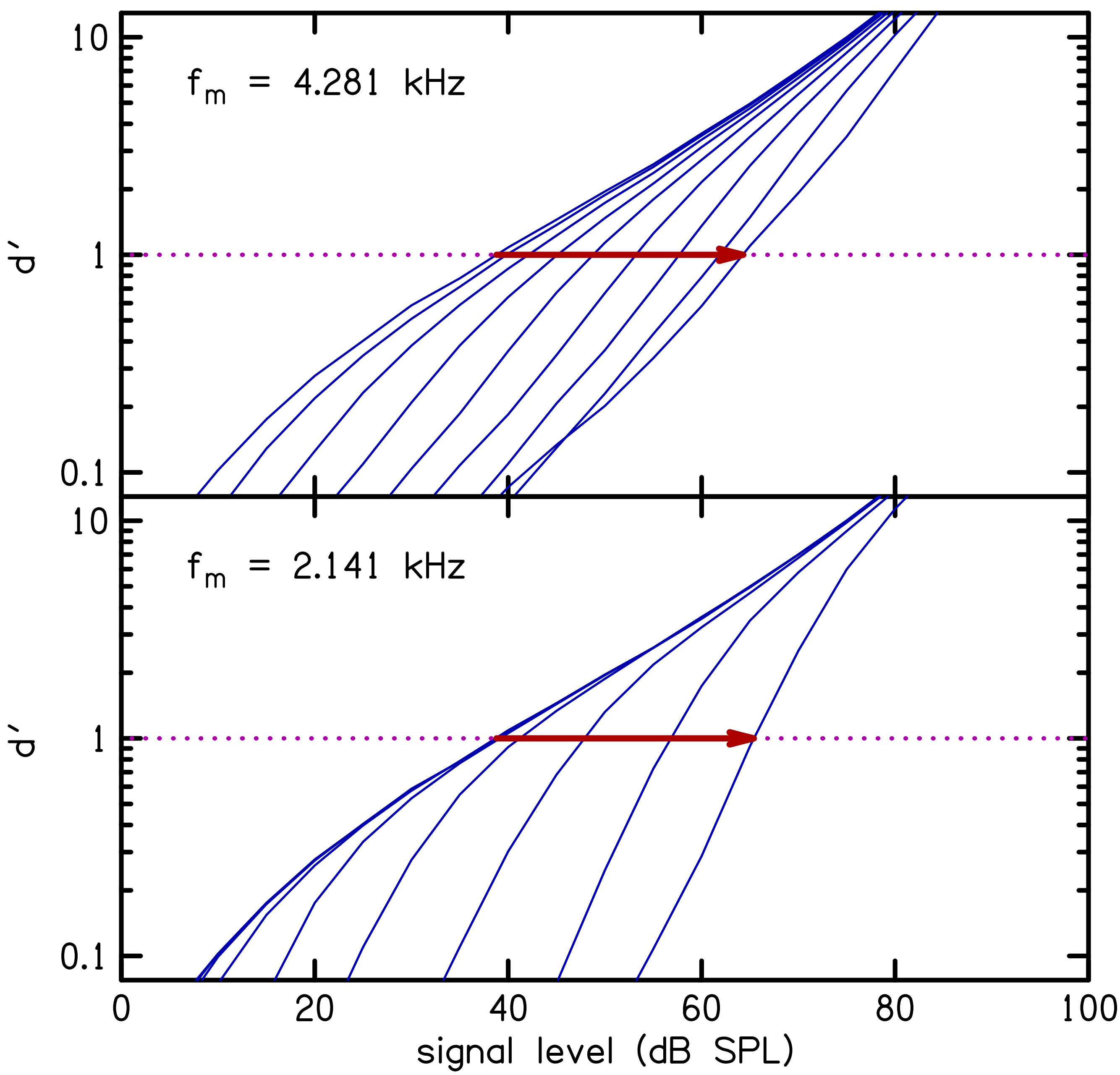
50
40
30
20
10
0
0
20
40
60
80
100
loudness level (dB)
input level (dB SPL)
FM33
NH
HI



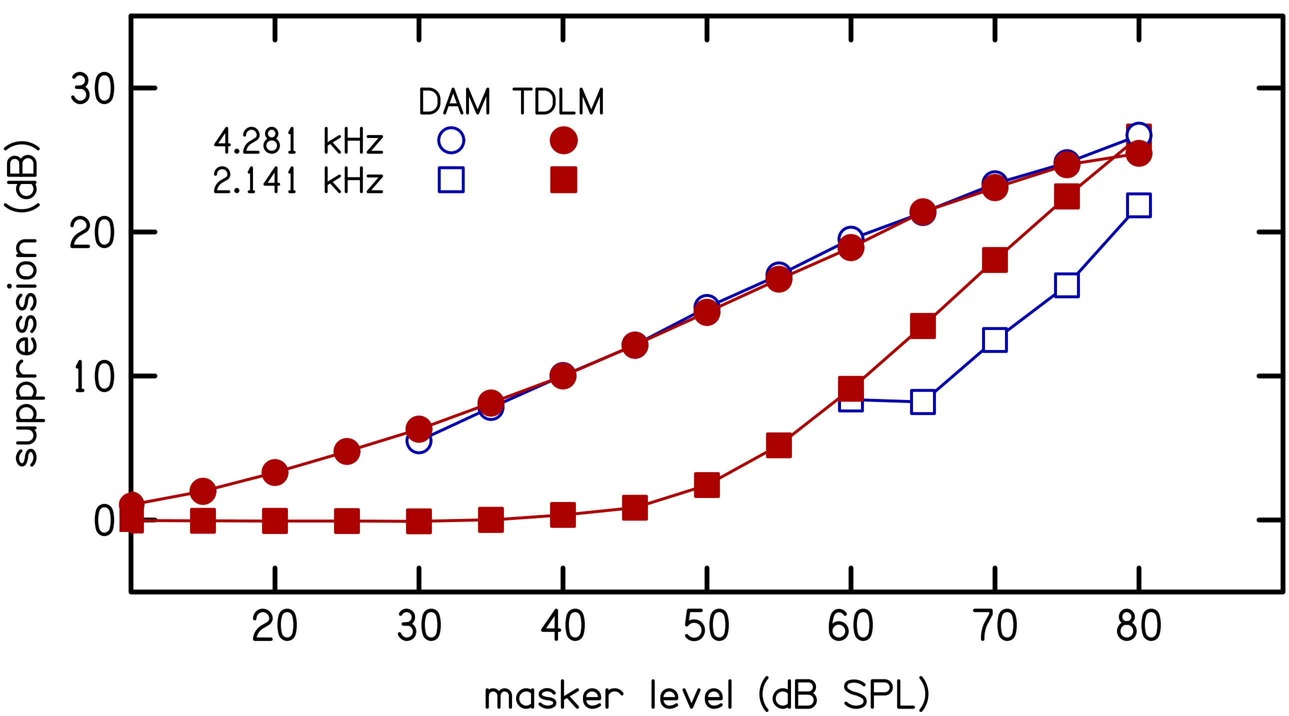
Moore et al. model
Liebold et al. data
TDLM – NH
TDLM – HI
ERB = 133 Hz
75
70
65
60
55
50
0.03
0.1
0.3
1
3
level (dB SPL) of 1-kHz tone judged equally loud to complex
bandwidth of complex (kHz)



NH
HI
JWG77
0.5
0.2
0.1
0
20
40
60
80
100
intensity JND (ΔI/I)
level (dB SPL)



f_m = 4.281 kHz
f_m = 2.141 kHz
d′
d′
10
1
0.1
10
1
0.1
0
20
40
60
80
100
signal level (dB SPL)



DAM
TDLM
4.281 kHz
2.141 kHz
suppression (dB)
masker level (dB SPL)
0
10
20
30
20
30
40
50
60
70
80