COMM. MATH. SCI.
Vol. 1, No. 4, pp. 829–836

FAST COMMUNICATION

A PDE BASED TWO LEVEL MODEL OF THE MASKING PROPERTY OF THE HUMAN EAR*

JACK XIN[†] AND YINGYONG QI[‡]

Abstract. Human ear has the masking property that certain audible sound becomes inaudible in the presence of another sound. Masking is quantified by the raised thresholds from the absolute hearing thresholds in quiet. It is of scientific and practical importance to compute masking thresholds. Empirical models on masking have applications in low bit rate digital music compression. A first principle based two level model is developed with partial differential equation (PDE) at the peripheral level and a similarity transform to represent all higher levels of auditory perception. Modeled masking thresholds of banded noise by tonal signals agree well with hearing data.

1. Introduction

The scientific study of the relation between the physics of sounds and their perception is referred to as psychoacoustics where enormous data have been collected [6]. The absolute hearing threshold and masking are two fundamental phenomena in psychoacoustics. The former is the minimum intensity for the ear to detect sound at a given frequency in quiet (see circles in Fig. 1 for an illustration). The latter is described by the nonlinearly raised threshold for the ear to detect sound in the vicinity of an existing signal — the masker (see the dashed line of Fig. 1).

We developed a two level model to compute the latter given the former and the masker. The first level is a partial differential equation (PDE) model of the inner ear (cochlea), and the second level is a similarity transform, accounting for the functions of the remaining high level processes of audition. The model has a solid ground on first principles and is adaptive to nonlinearities when compared with existing data-driven, empirical models. It is expected to help us to understand human hearing and hearing-impairment.

We used the psychoacoustic database of MP3 standard to validate our model. MP3 is a music compression technology, where masking properties are used to significantly reduce the bit rate of digital sounds [1, 3]. The reduction is possible, in part, because quantization noises arising from digitization can be concealed in the presence of musical signals. No audible differences between the original and the quantized (compressed) sounds are detectable when masking properties are well exploited. MP3 is now routinely used for downloading and playing digital music. Its success lends support to the validity of the psychoacoustic database used.

It is scientifically challenging to model the masking thresholds. One of the first models goes back to Schroeder, Atal and Hall [4], where empirical formulas recapitulate key steps of human auditory information processing. The auditory pathway

---

[*] Received: July 3, 2003; accepted (in revised version): August 6, 2003. Communicated by Shi Jin.

[†] Department of Mathematics and ICES, University of Texas at Austin, Austin, TX 78712, USA. Corresponding author, (jxin@math.utexas.edu). The work was partially supported by ARO grant DAAD 19-00-1-0524, NSF grant ITR-0219004, and a fellowship from the John Simon Guggenheim Memorial Foundation.

[‡] Qualcomm Inc, 5775 Morehouse Drive, San Diego, CA 92121, USA.

from the inner ear to the auditory cortex is a coupled multi-level sound information processing system [6]. The preprocessing of sound takes the form of mechanical vibrations on cochlea's basilar membrane (BM) mediated by peripheral neural activities. At this level, nonlinear phenomena can be succinctly described by PDEs in classical mechanics modified with peripheral neural responses [5]. Our knowledge becomes more limited as the sound information goes up to higher auditory centers along nerve fibers to the brain, though a rich collection of data is available at the final perception level [6]. Interestingly, as sound information reaches higher levels, many nonlinear features remain. An example is tonal suppression, which refers to the ability of a tone of enough intensity to reduce the effect of a tone at a nearby frequency. Tonal suppression is observed on the BM ([5] and references therein), in the neural responses [2] as well as in psychoacoustics (tonal masking [6]).

Though mechanical PDEs capture well the nonlinear interactions of sounds of different frequencies [5], they have no sense of perception. The idea is to provide these PDEs with psychoacoustic information, i.e. absolute thresholds, and define a perceptive threshold on BM. The audibility of noise is measured on BM using this perceptive BM threshold. The masked threshold curves so obtained turned out to be lower and flatter than the ones in MP3, yet similar looking. The difference is not surprising, as we bypassed the intermediate processing centers, whose role is known to modify the responses [6]. Here, we devise a similarity transformation to compensate for the difference and match with the masked noise threshold curves in MP3. The rest of the paper is organized as follows. Section 2 contains the computational procedure and results, concluding remarks are in section 3.

2. Numerical Algorithm and Results

We employ the nonlinear nonlocal cochlear PDE of transmission line type [5] to describe the vibration pattern on BM. The BM damping function contains a convolution integral of BM amplitude, and has been shown [5] to generate suppressions among two and three tones. For input of tones and noise, the output is a BM displacement pattern over the BM, denoted by $u(x,t)$ (on the scale of nanometer), where the longitudinal space variable x in $[0,3.5]$ (cm), time t in $[0,20]$ ms (millisecond). Such a duration is chosen as in Schroeder et al [4]. We refer to our previous work [5] for details of the PDE, the numerical method, and related parameters. Let us then perform the following steps:

(A) Input banded noise at absolute threshold levels with center frequencies f_n ($n=1,\cdots,N$) ranging from 300 Hz (Hertz) to 10 kHz (kilo-Hertz). The bandwidth is 100 Hz below and 0.2 times the center frequency above 500 Hz, so called critical band widths (1). Calculate the BM root mean square (rms) excitation function for each n:

$$E_n(x) = (\frac{1}{J}\sum_{j=1}^{J} |u_n(x,t_j)|^2)^{1/2}, \;\; n=1,2,\cdots,N,$$

t_j's range over $[10,20]$ ms in 0.01 ms increment. The interval $[10,20]$ ms is to remove initial transients. Define:

$$E_{th} = \text{envelope}(E_1, E_2, \cdots, E_N),$$

as the BM perception threshold. Here envelope is the operation of taking the envelope of all N excitation functions. Each excitation function has a pulse shape, and the envelope is numerically constructed in Matlab by a cubic spline interpolation of all local maxima of excitation functions.

(B) Given a signal S (tones), consider a noise Ns. Compute two BM responses u_S and u_{S+Ns}, for input S and $S+Ns$; then compute the rms excitation pattern E_{SNs} from the difference $u_S - u_{S+Ns}$ as in Step A. Find threshold intensity I_{Ns} of Ns so that E_{SNs} is just below E_{th}. Output all I_{Ns} as a function of center frequencies of Ns, this is the masked noise threshold curve by signal S.

Step B is to measure the amount of noise after interacting with the signal S, and calibrate it with the BM perception threshold in Step A. The rms norm is used to smooth out irregularities in the output due to noise. Fig. 2 shows a comparison of model masked noise thresholds (lines) by 86 dB (decibel) tones at (1,2,3,4.3) kHz. These four frequencies are a few integral number of critical band widths apart, and represent the region of sensitive hearing. The model curves are seen to be lower and flatter than the MP3 counterparts, yet qualitatively similar. There is a slight decrease in peak altitudes with increasing signal frequencies, and the masked threshold curves tend to broaden towards higher frequencies. Also the left and right sides of the curves are asymmetric with left side being steeper. The masked threshold curves at other signal frequencies (up to 6 kHz) and intensities (76, 66) dB show similar properties.

Finally, step C, we map the model output to MP3 with the similarity transform:

$$F_c = F_c(f) = \lambda_2(F_b - F_a)(f_0 + \lambda_1(f - f_0)) + F_a(f),$$

where λ_1, λ_2 are positive constants, f_0 the signal (masker) frequency, F_a the absolute threshold, F_b (F_c) the model threshold output at Step B (C). The λ_1 takes two different values: $\lambda_{1,-}$ on the left ($f < f_0$), and $\lambda_{1,+}$ on the right side ($f > f_0$) to enhance the asymmetry. We also let λ_1 and λ_2 depend on f_0 in that $\lambda_{1,-}$ makes one adjustment above 2 kHz while keeping the ratio $\lambda_{1,-}/\lambda_{1,+}$ fixed; and λ_2 makes an adjustment for f in $[2,3]$ kHz, namely 5 parameters total. The four transformed model threshold curves (lines) are plotted in Fig. 3 with MP3 curves (dashed lines). Fig. 4 shows that the transformed masked noise thresholds by 76 dB signals are also near the MP3 ones.

3. Concluding Remarks

The two level model offers a promising framework for first principle based modeling of psychoacoustic data on normal and hearing-impaired individuals. In future work, it is interesting to investigate more accurate partial differential equations at the peripheral level, and enhance the precision of model masking curves prior to similarity transform.

Acknowledgments. While the work was in progress, Y-Y Qi was a visiting fellow at the Institute of Computational Engineering and Sciences (ICES) at the University of Texas at Austin.

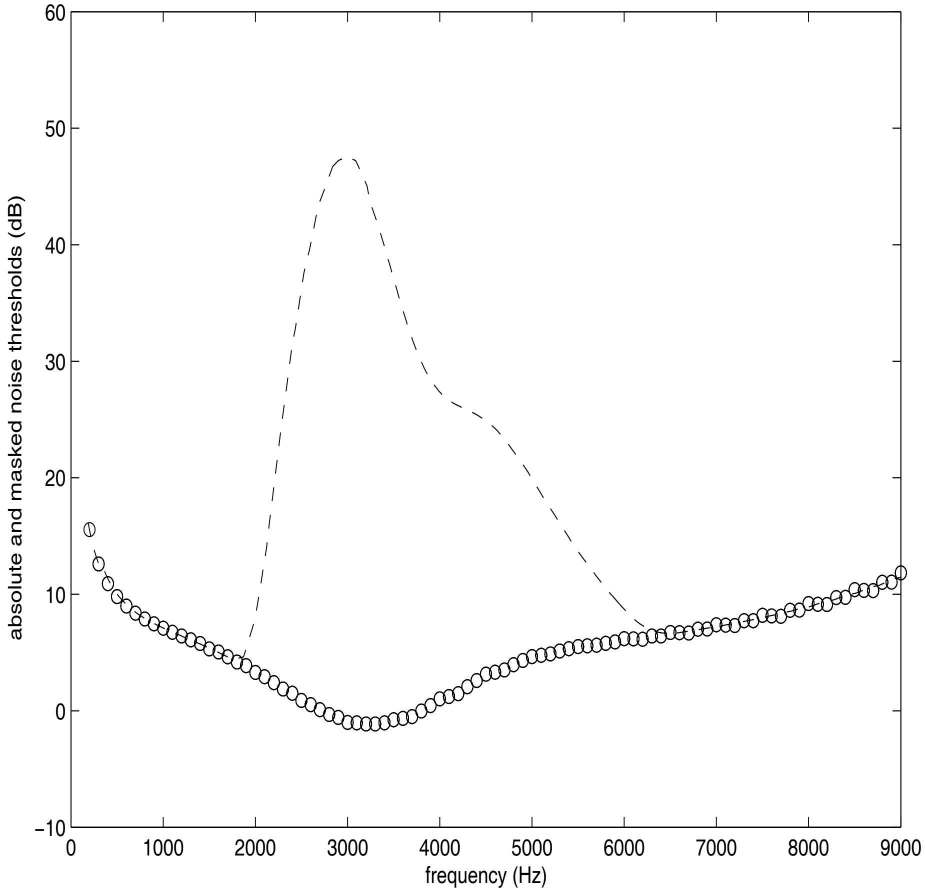

Fig. 3.1. *Absolute thresholds (circles) and masked noise thresholds (dashed line) by a 86 dB tone at 3 kHz generated from MP3 standard.*

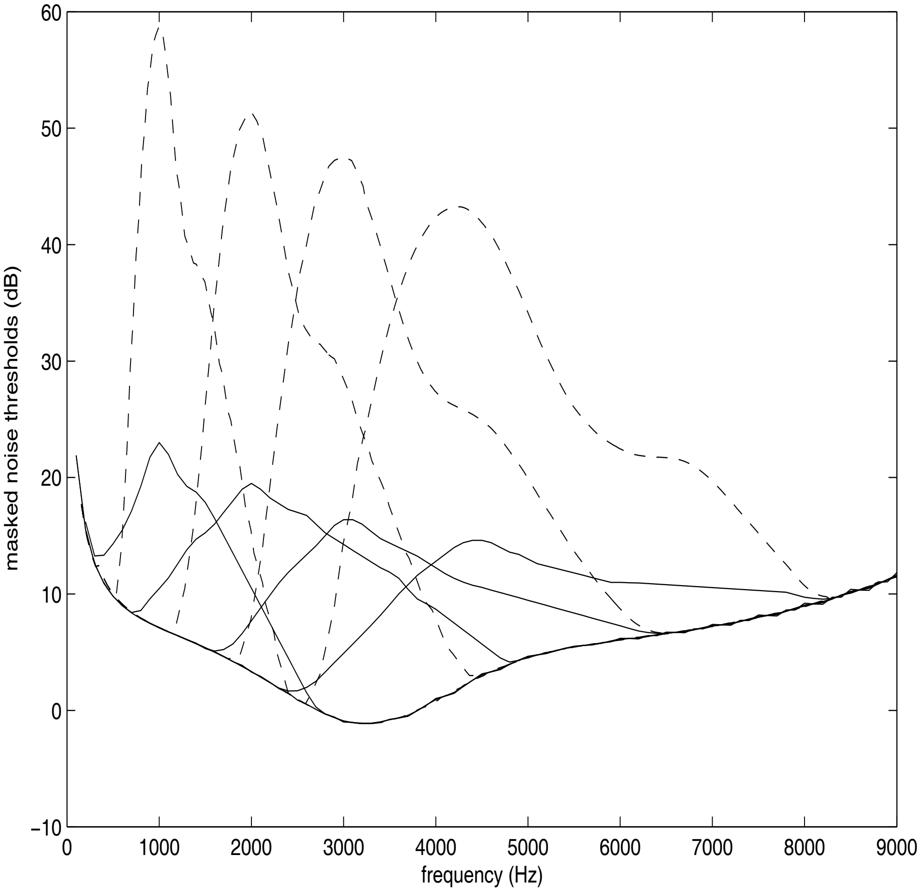

FIG. 3.2. *Comparison of masking thresholds of the PDE model (lines, steps A-B) with those of MP3 (dashed lines) by 86 dB tones of (1,2,3,4.3) kHz. Qualitative features agree, such as left-right asymmetry, lowering and broadening towards higher frequencies.*

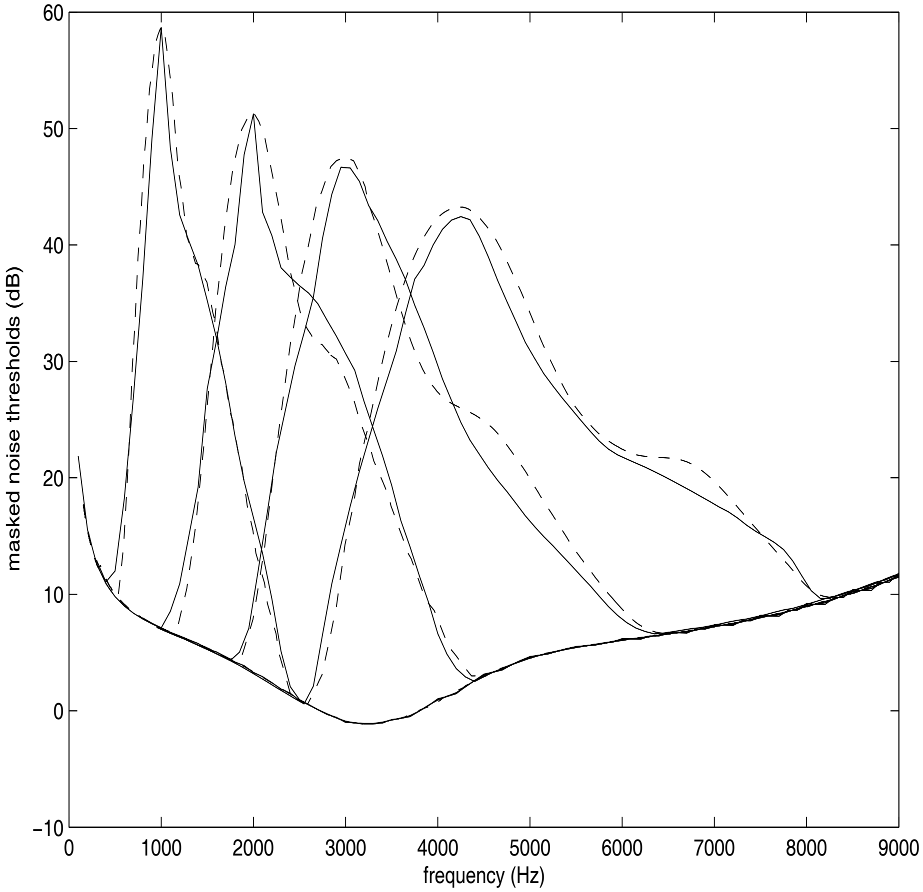

Fig. 3.3. *Comparison of similarity transformed masking thresholds of the PDE model (lines, steps A-C) with those of MP3 (dashed lines) by 86 dB tones of (1,2,3,4.3) kHz.*

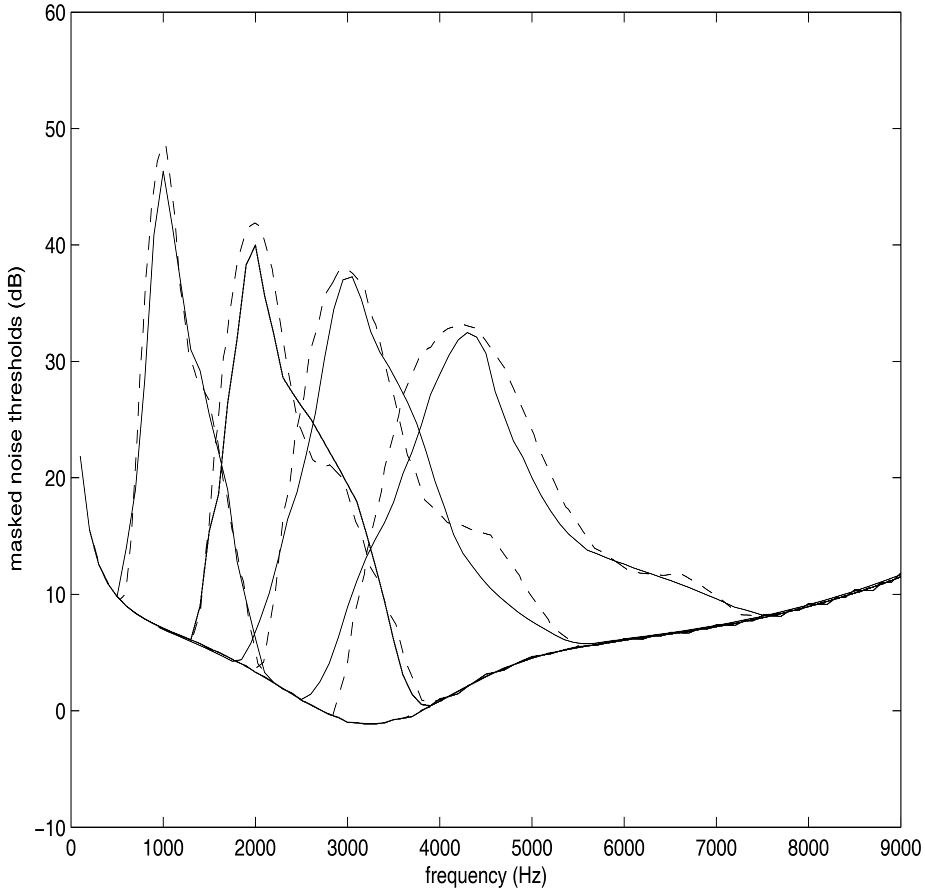

FIG. 3.4. *Comparison of similarity transformed masking thresholds of the PDE model (lines, steps A-C) with those of MP3 (dashed lines) by 76 dB tones of (1,2,3,4.3) kHz.*

REFERENCES

[1] K. Brandenburg and M. Bosi, *Overview of MPEG Audio: Current and Future Standards for Low-Bit-Rate Audio Coding*, J. Audio Eng. Soc., 45(1/2):4–21, 1997.

[2] C.D. Geisler, *From Sound to Synapse*, Oxford University Press, Oxford, 1998.

[3] K. Pohlmann, *Principles of Digital Audio*, 4th edition, McGraw-Hill Video/Audio Professional, 2000.

[4] M.R. Schroeder, B.S. Atal, and J.L. Hall, *Optimizing digital speech coders by exploiting properties of the human ear*, Journal Acoust. Soc. America, 66(6):1647–1652, 1979.

[5] J. Xin, Y.Y Qi, and L. Deng, *Time Domain Computation of a Nonlinear Nonlocal Cochlear Model with Applications to Multitone Interaction in Hearing,* Communications in Math Sciences, 1(2):211–227, 2003. (www.intlpress.com/CMS)

[6] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models*, Springer Series in Information Sciences, 22, 2nd edition, New York, 1999.