Tone-vs.-Noise Dichotomy in Psychoacoustics Could be Resolved by Careful Consideration of Cochlear Signal Processing

Yi-Wen LIU* and Lu-Ming YU

*Department of Electrical Engineering, National Tsing Hua University,
101 Sec. 2 Kuang-Fu Rd., Hsinchu, Taiwan 30013*

Received May 25, 2012; final version accepted August 13, 2012

Psychoacoustic research suggests that the human auditory system processes noise and tones differently. For example, narrowband noise has a stronger masking effect than a pure tone at the same sound pressure level. Can this be explained by cochlear mechanics, or is the distinction due to central neural processing? In this paper, we review a computational model that simulates wave propagation in the cochlea. The model incorporates recent findings in cochlear electrophysiology and describes wave propagation in both the forward and the backward direction. Simulation shows that the model produces compressive nonlinear responses to tones but quasilinear responses to noise. We may therefore argue that the distinction between tone- and noise-processing is an *epiphenomenon* of cochlear mechanics. If this is true, we can simplify the computational organization of audio signal-processing front-ends for applications such as hearing aids and auditory prostheses.

KEYWORDS: cochlea, signal processing, nonlinear mechanics, psychoacoustics

1. Introduction

Psychoacoustic masking refers to the fact that a target signal, often called a *probe*, becomes inaudible when a masker sound is present in its time-frequency vicinity [2]. The spectral spread of masking depends on whether the masker is noise-like or tone-like [15]; at any given masker level, the masking threshold is typically 10 dB higher for a tone to be masked by noise than for noise to be masked by a tone [4]. These facts have been incorporated in perceptual audio coding standards [3], so it became crucial on the encoder side to determine whether maskers in arbitrary audio signals are noise-like or tone-like based on spectro-temporal cues. However, spectral contents in audio signals tend to be complex and can vary rapidly in time. It is often impossible to recognize whether content in a sub-band is noise-like or tone-like before assigning the correct spread function for masking-threshold calculation. In this paper, we argue that the noise-vs.-tone dichotomy is perhaps unnecessary if cochlear mechanics are considered carefully.

In auditory physiology, distinctions have also been noted between noise- and tone-responses in mammalian cochleae. Based on measurements of basilar-membrane (BM) motion in live animals, responses to tones were found to be highly compressive [14] whereas responses to noise is *quasi-linear*; that is, the noise responses could be predicted with high accuracy by convolving the stimuli with level-dependent Wiener kernels [13]. Though quasi-linearity was expected by de Boer [1] to be a result of equivalent-nonlinear (EQ-NL) theory even before any experiment was conducted, accuracy of the theory was not evaluated until recently. Liu and Neely [6] verified the EQ-NL theory by showing that Recio-Spinoso *et al.*'s noise responses could be reproduced by a cochlear mechanics model which follows de Boer's guidelines but incorporates recent findings in the physiology of outer hair cells (OHCs) [8]. Liu and Neely's model, originally aimed to investigate traveling waves amplification, resolved the dichotomy regarding how tones and noise are processed by the cochlea. In §2, we review the Liu–Neely cochlear mechanics model and how it responds to tones and noise. Implications for processing audio and other time-domain signals are discussed in §3.

2. Methods and Results

The Liu–Neely model treated cochlear traveling waves as propagation along a bidirectional transmission line [8]. The series impedance of the transmission line depended on cochlear fluid density and the area of cochlear cross section, and calculation of the shunt admittance involved an OHC feedback mechanism [7] — the feedback was instantaneous and nonlinear due to properties of the mechano-electrical transduction channels of OHCs' hair bundles. Physical parameters such as elasticity, mass, and resistance of the cochlear partition, and the membrane capacity and electromotilities of the OHCs, were carefully chosen to satisfy the following criteria: first, the parameter values should

* Corresponding author. E-mail: ywliu@ee.nthu.edu.tw

Fig. 1. The Liu-Neely transmission-line model of cochlear mechanics. The cochlea is partitioned into 700 segments, and each segment has an outer hair cell (OHC) module inserted in between reticular lamina (RL) and basilar membrane (BM). Sounds transmit from eardrum (ED) through the ossicular chain to enter the cochlea through the oval window (OW). Cochlea mechanics follow Newton's principle (shown as $\partial_x P = -(\rho/A)\partial_t U$ in the figure) and cochlear-partition motion follows Hook's law (shown as $\partial_x U = \eta \partial_t P$), where P and U denote acoustic pressure and volume velocity, respectively, and η incorporates any nonlinear feedback that is provided by the OHCs.

Fig. 2. Nonlinear compression of magnitude response on the reticular lamina (RL) at the 4-kHz characteristic place of the model. Curves show responses to pure-tone stimuli in 10 dB steps from 0 to 100 dB SPL.

be physically realistic; secondly, the model should produce dynamic range compression on the BM that is typical of lab animals [14] and distortion-product emissions that are typical of humans [5]. Structure of the model is depicted in Fig. 1.

Simulation was conducted by simultaneously solving the cochlear-partition motion, the cochlear fluid flow, and OHC membrane potentials; a finite-difference method is used. Figure 2 shows the model's magnitude responses at the 4-kHz characteristic place. Responses were calculated for fixed stimulus levels in 10-dB steps from 0 to 100 dB sound pressure level (SPL). At 0–30 dB SPL, the magnitude response remains approximately the same shape, but the shape gradually changes when stimulus level increases beyond 40 dB SPL. In particular, the best frequency shifts down by approximately 25%, or 0.4 octave, as the input level increases from 20 to 80 dB SPL. Finally, the peakiness of tuning is entirely lost as the input level increases beyond 90 dB SPL. To summarize, the input–output (I/O) function is *highly compressive* near 4-kHz.

Figure 3 shows results of Wiener-kernel analysis [13] of the model's noise responses. Liu and Neely [6] showed that, by convolving quasi-random signals with these kernels, intra-cochlear responses can be predicted with high correlation ($\rho > 0.9$), similar to what was obtained by Recio-Spinoso *et al.* [13]. Further, the high correlation persists across all stimulus levels and along the entire cochlea. Nonlinear compression is present in the sense that the kernel amplitude

Fig. 3. Wiener kernel functions for noise stimulus at 34–94 dB (SPLs). The left panels shows responses at a sites that is 3.5 mm more basal (i.e., toward the oval window) than what is shown on the right. The Wiener kernels on the left are therefore concentrated at higher frequencies and have lower group delays compared to the right.

decreases if noise power increases. Meanwhile, it is noteworthy that the duration of ringing also decreases as noise power increases. This parallels the fact that cochlear tuning is not as sharply at high intensities as it is at low intensities, as shown in Fig. 2.

To summarize, results shown in Figs. 2 and 3 resemble experimental findings reported by Ruggero *et al.* [14] and Recio-Spinoso *et al.* [13], respectively. These results demonstrate that the same cochlear model can emulate both tonal and noise responses of mammalian cochleae.

3. Discussions

Signal processing methods inspired by cochlear mechanics have been developed both in analog [10] and digital [9] domains. Nonlinearity was not desired in Mandal *et al.*'s analog circuit, while cochlear nonlinearity was mimicked by terminating every filter stage with a half-wave rectifier in Lyon's work. Recio-Spinoso *et al.* [13] argued that a linear filter in cascade with a single-stage of instantaneous nonlinearity cannot approximate noise responses in the cochlea very well, for such models are highly nonlinear and will produce noise responses that do not match well with Wiener-kernel prediction. We thus argue that Lyon *et al.*'s method does not reflect the true signal processing flow in the cochlea.

Similar signal processing flows can also be found in auditory research literature; Termed *dual-resonance nonlinear* (DRNL) filtering, BM responses were emulated by summing up two paths, one of which being passive and linear and the other path consists of a single stage of saturation nonlinearity (SNL) in cascade with gain-providing gamma-tone filters [11]. Though such construction could faithfully reproduce compressive tone responses, we argue that it might fail to provide realistic, quasi-linear noise responses based on Recio-Spinoso *et al.*'s explanation.

In contrast, Liu and Neely's model (2010, 2012) treated SNL as components (to be exact, the hair bundles of OHCs) distributed along the entire cochlea. It therefore avoids lumping up all saturation effects into a single cascade stage along the signal processing path while maintaining highly compressive tone responses. Due to this distributive nonlinear filtering structure, the method resolves the tone-vs.-noise dichotomy that arose previously due to attempts to decipher experimental results in cochlear mechanics. Figure 4 compares the difference between single-stage vs. distributed nonlinear cochlea models.

It remains unclear whether the method presented in this paper can reproduce the drastically different psychoacoustic masking effects for tonal vs. noise-maskers. Further simulation on two-tone suppression can be conducted to quantify the difference in the masking efficiency of noise-vs. tone maskers. Also, note that the computation load of full-blown cochlear mechanics simulation (e.g., 700 spatial segments in the present research) is generally much higher than Fast Fourier Transform-enabled signal-processing frontends (such as Lyon, 2010). Nevertheless, we can argue that the main idea of the present research—to distribute nonlinear components throughout the cochlear model—can also be applied to models that do not require full-blown cochlear-mechanics simulation. This distributive nonlinear design has recently been incorporated in a loudness model which simulates loudness density with a resolution of 12 filters per octave [12]; the resolution is lower than that of the present research, but still sufficient for most audio applications.

Fig. 4. Comparison between single-stage and distributed saturation nonlinearities (SNLs) in cochlear models. **A**: the dual-resonance nonlinear filtering (DRNL) approach. For the input signal to reach any point along the BM, the nonlinear path consists of a single stage of SNL. **B**: the present method. To reach any point along the BM, the signal would propagate through multiple segments. Each segment is provided with nonlinear positive feedback that saturates at high intensity.

4. Conclusions

This paper reviews a cochlear mechanic model that can reproduce tone- and noise responses in mammalian cochleae. We believe the present research may lead to new design of signal processing front-ends for applications such as speech recognition, audio data compression, hearing aids, and auditory prostheses. For the latter two applications, better understandings of cochlear mechanics can lead to more power-efficient encoding of audio signals. Because a unified model for noise and tone stimuli makes it unnecessary to conduct sound source classification, it may reduce the computation load for advanced hearing-aid functions and therefore extend the battery life. Research along these directions is left as future work.

Acknowledgment

This research is sponsored by National Science Council of Taiwan through Grant No. 100-2221-E-007-011. The author also thanks National Tsing Hua University for the start-up package that has enabled this research to continue.

REFERENCES

- [1] de Boer, E., "Connecting frequency selectivity and nonlinearity for models of the cochlea," *Aud. Neurosci.*, **3**: 377–388 (1997).
- [2] Fastl, H., and Zwicker, E., *Psychoacoustics: Facts and Models*, 3rd Ed., Springer-Verlag, Berlin (2007).
- [3] ISO MPEG IS11172-3 ("MPEG-1"), "Information technology—Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s, Part 3: Audio" (1992).
- [4] Jayant, N., *et al.*, "Signal compression based on models of human perception," *Proc. IEEE*, 1385–1422 (1993).
- [5] Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P., "Influence of primary-level and primary-frequency ratios on

- human distortion product otoacoustic emissions,” *J. Acoust. Soc. Am.*, **119**: 418–428 (2006).
- [6] Liu, Y.-W., and Neely, S. T., “Quasilinear cochlear responses to noise can result from instantaneous nonlinearities,” in *What Fire is in Mine Ears: Progress in Auditory Biomechanics*, Shera, C. A., and Olson, E. (Eds.), *American Institutes of Physics*, pp. 417–422, New York (2012).
 - [7] Liu, Y.-W., and Neely, S. T., “Outer hair cell electromechanical properties in a nonlinear piezoelectric model,” *J. Acoust. Soc. Am.*, **127**: 751–761 (2009).
 - [8] Liu, Y.-W., and Neely, S. T., “Distortion product emissions from a cochlear model with nonlinear mechano-electrical transduction in outer hair cells,” *J. Acoust. Soc. Am.*, **127**: 2420–2432 (2010).
 - [9] Lyon, R. F., *et al.*, “Sound retrieval and ranking using sparse auditory representations,” *Neural Computation*, **22**: 2390–2416 (2010).
 - [10] Mandal, S., Zhak, S. M., and Sarpeshkar, R., “A bio-inspired active radio-frequency silicon cochlea,” *IEEE J. Solid-State Circuits*, **44**: 1814–1827 (2009).
 - [11] Meddis, R., O’Mard, L. P., and Lopez-Poveda, E. A., “A computational algorithm for computing nonlinear auditory frequency selectivity,” *J. Acoust. Soc. Am.*, **109**: 2852–2861 (2001).
 - [12] Neely, S. T., Rodríguez, J., Liu, Y.-W., Jesteadt, W., and Gorga, M. P., “A computational model for loudness density,” submitted to *Acta Acustica* in (2012).
 - [13] Recio-Spinoso, A., Narayan, S. S., and Ruggero, M. A., “Basilar membrane responses to noise at a basal site of the chinchilla cochlea: Quasi-linear filtering,” *J. Assoc. Res. Otolaryngol.*, **10**: 471–484 (2009).
 - [14] Ruggero, M. A., Rich, N. C., and Robles, L., “Basilar-membrane responses to tones at the base of the chinchilla cochlea,” *J. Acoust. Soc. Am.*, **101**: 2151–2163 (1997).
 - [15] Spanias, A., Painter, T., and Atti, V., *Audio Signal Processing and Coding*, Wiley Interscience (2007).