forum acusticum 2026
8-12 September · Graz, Austria

AN END-TO-END TUNABLE COCHLEAR MODEL WITH INSTANTANEOUS NONLINEARITY FOR SIMULATING OTOACOUSTIC EMISSIONS

Yong-Yue Xu[1], Yung-Chia Huang[1], Yi-Wen Liu[1]

[1] *Department of Electrical Engineering, National Tsing Hua University, Taiwan*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

Computational cochlear models are essential for auditory research. However, personalization to individual audiological data remains a challenge. While previous work achieved personalization using transient-evoked otoacoustic emissions (TEOAEs), simulating distortion product OAEs (DPOAEs) requires a more explicit bidirectional framework. We introduce TDLM-SNLv3, a fully differentiable model implemented in JAX. Unlike traditional models using time-varying parameters, TDLM-SNLv3 employs instantaneous nonlinearity to represents the rapid mechano-electrical transduction of outer hair cells. The primary innovation of this work is an extended backward path that couples nonlinear distortion products directly into a filter chain while maintaining linear reflection mechanisms for TEOAEs. Experimental results demonstrate that TDLM-SNLv3 enables simultaneous TEOAE fitting and DPOAE simulation, providing an end-to-end tunable framework for personalized hearing applications.

Keywords: *Otoacoustic Emissions, Auditory Modeling, Differentiable Signal Processing*

1. Introduction

Efficient modeling of cochlear signal processing is a challenging task because of the conflicting desires of capturing the biophysical properties while lowering the computation load. Existing models typically fall into two categories – parallel or cascade [1], [2], [3], [4]. Parallel models are robust in reproducing the place-dependent filtering effects that are the hallmark of cochlear mechanics, while cascade models have the advantage of capturing wave propagation explicitly along the basilar membrane. While standard cascade auditory models effectively simulate forward wave propagation, model personalization – the process of fitting these models to individual audiological data—remains challenging due to the complex trade-offs between biophysical fidelity and computational efficiency [5].

Recently, a systematic method was proposed for personalizing the CARFAC (Cascade of Asymmetric Resonators with Fast-Acting Compression) model [6] within the JAX environment, utilizing its automatic differentiation to tune parameters against measured transient-evoked otoacoustic emission (TEOAE) waveforms. However, conventional cascade models that achieve dynamic range compression by time-varying parameter adjusting [3], [4] often face architectural hurdles when simulating distortion product OAEs (DPOAEs) [7]. Also, capturing how nonlinear inter-modulation products are generated and propagated backward to the ear canal requires an explicit bidirectional framework that traditional models often lack.

In this work, we modify the TDLM-SNL (Time-Domain Loudness Model with Static Nonlinearity) [8] to the JAX framework. Unlike traditional models, TDLM-SNL employs instantaneous static nonlinearity, a design choice biophysically motivated by the rapid mechano-electrical transduction of outer hair cells (OHCs). By keeping filter coefficients fixed and inserting memoryless nonlinear functions within the cascade sections, the model exhibits compression and distortion product generation that are typical in animal experiments.

The primary innovation of this research is an extended backward path specifically designed for emission generation. While our previous research utilized a backward path to characterize coherent reflections for TEOAE, we have modified this architecture to accommodate DPOAEs. By directly coupling the output of the nonlinear function into a backward low-pass filter (LPF) chain, the model can now propagate inter-modulation distortion products back to the middle ear. This unified structure provides a fully differentiable framework that enables simultaneous TEOAE fitting and DPOAE simulation within a single, end-to-end tunable system.

***Corresponding author**: ywliu@ee.nthu.edu.tw.*

2. Method

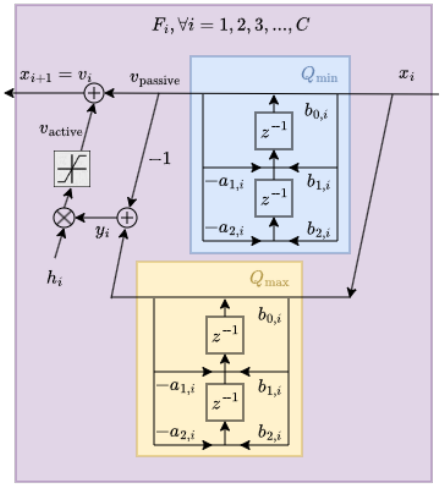


Figure 1: The structure of TDLM-SNLv3. Compared to TDLM-SNL [8], the difference is in the introduction of an OHC health coefficient h_i.

2.1. JAX-based TDLM-SNL Architecture

The TDLM-SNL structure is shown in Fig. 1, where each channel F_i receives an input x_i and produces an output v_i to be delivered to the next channel F_{i+1}. The coefficients of two 2nd-order filters are acquired from TDLM [7], and they jointly provide the place-dependent frequency responses that broaden when the stimulus level increases. In this study, we implemented TDLM-SNL within the JAX environment, creating a computational framework named TDLM-SNLv3 to leverage high-performance computing and automatic differentiation. The software architecture of TDLM-SNLv3 adopts a structure similar to CARFAC v2 [6], where model components are organized into four data classes, including *Design Parameters*, *Hyperparameters*, *Weights*, and *States* (internal variables). Only the *Weights* class is trainable. This design is specifically optimized for JAX's compilation and computation processes.

To ensure the model's differentiability required for automatic differentiation, we adopted the following asymmetric function as the static nonlinearity in TDLM-SNLv3,

$$\phi(y) = \begin{cases} \sigma_{\max} \tanh(y/\sigma_{\max}), & \text{if } y \geq 0 \\ \sigma_{\min} \tanh(y/\sigma_{\min}), & \text{if } y < 0, \end{cases} \quad (1)$$

where $+\sigma_{\max}$ and $-\sigma_{\min}$ are the saturation limits when the instantaneous input y approaches $+\infty$ and $-\infty$, respectively [8].

Additionally, we introduced an OHC-health vector $\mathbf{H} = [h_1, h_2, h_i, \ldots, h_C]$ to the *Weights* class to characterize the functional integrity of the OHCs at various characteristic frequency locations. Each h_i corresponds to an individual cochlear channel i and C is the total number of channels in TDLM-SNLv3. Also, elements in $\mathbf{H}$ scales from 0 (representing a passive response) to 1 (representing a fully healthy state). As shown in Fig. 1, before entering the static nonlinear function, the intermediate variable y_i is first multiplied by h_i; that is,

$$v_{\text{active},i} = \phi(h_i \times y_i). \quad (2)$$

This feature enables TDLM-SNLv3 to fit OAE data at different OHC conditions. Figure 2 depicts the compressive I/O functions under different $\mathbf{H}$ conditions ($h_i = \text{h}$ for all i). The model with forward middle-ear filter (MEF) coefficient $\alpha = 30$ dB, was presented with a 1-kHz pure-tone stimulus, and the output was measured at the channel corresponding to the 1 kHz characteristic frequency.

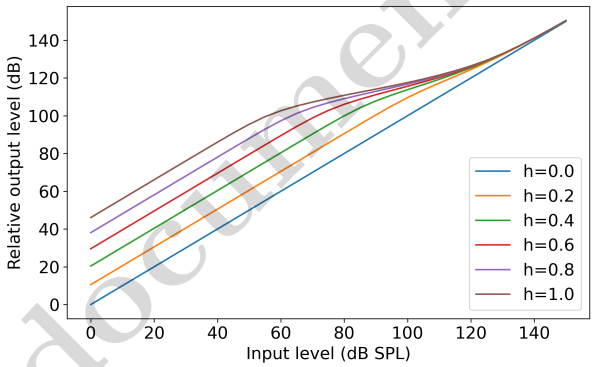


Figure 2: The I/O function under different $\mathbf{H}$ conditions.

2.2. TEOAE Fitting and Optimization Strategy

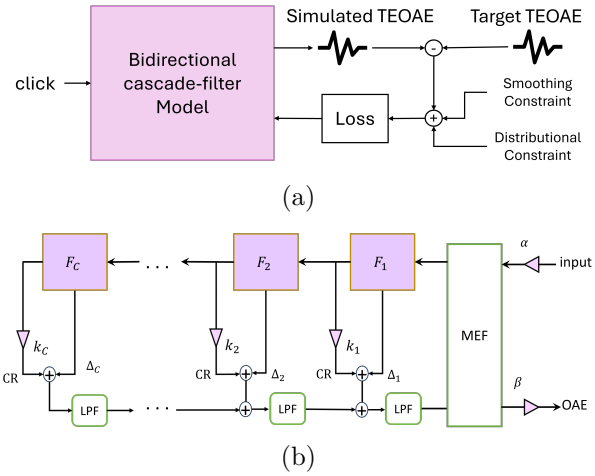


Figure 3: Simulation and model personalization using TEOAE. (a) The personalization framework. (b) The structure of wave scattering and backward propagation.

To evaluate the optimization capacity of the TDLM-SNLv3, we replicated the personalization experiment in [5]. As illustrated in Fig. 3(a) and (b), model parameters $\{\alpha, \beta, \mathbf{H}\}$ are iteratively updated via gradient

descent, with a goal of minimizing the discrepancy between the simulated and target TEOAEs while employing biophysical constraints for regularization. The TEOAEs simulation is based on the linear reflection path in Fig. 3(b), which will be described in details in Section 2.3.1.

In this study we use simulated TEOAEs from Verhulst et al.'s transmission-line (TL) model [9] as our target. To simulate TEOAE with different cochlear conditions, besides normal hearing profile, we also perturbed model poles to simulate notch patterns Notch3K/2K, which represent localized cochlear damage at 3 and 2 kHz, respectively. The parameters were calibrated empirically using ($\alpha = 15$ dB, $\beta = -54$ dB, and $h_i = 1$ for all i) as initial value to align to TL model TEOAE level at normal hearing condition. Additionally, a 2.0 ms post-stimulus blanking window was implemented to suppress transient artifacts from the stimulus.

The composite loss function is defined as a weighted sum of five components [5]:

$$\begin{aligned} L =& \lambda_{\text{stft}} L_{\text{stft}} + \lambda_{\text{time}} L_{\text{time}} + \lambda_{\text{freq}} L_{\text{freq}} \\ &+ \lambda_{\text{smooth}} L_{\text{smooth}} + \lambda_{\text{dist}} L_{\text{dist}}. \end{aligned} \tag{3}$$

The detail of each loss term is defined as follows: L_{stft} is a multi-resolution short-time Fourier transform (STFT) loss [10] with window lengths 512, 256, and 128 under the sampling rate 44.1 kHz. L_{time} is a root-mean-square (RMS) error of time domain waveform.

$$L_{\text{time}} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (p[n] - p_{\text{target}}[n])^2}, \tag{4}$$

where $p[n]$ and $p_{\text{target}}[n]$ denote the TEOAE generated by the proposed model and the target TEOAEs, respectively, and N denotes the number of samples. L_{freq} is a log-spectral magnitude error of frequency-domain mismatches.

$$\begin{aligned} L_{\text{freq}} =& \frac{1}{K} \sum_{k=0}^{K-1} (\log_{10}(|P[k]| + \epsilon) \\ &- \log_{10}(|P_{\text{target}}[k]| + \epsilon))^2, \end{aligned} \tag{5}$$

where P and P_{target} are the fast Fourier transform (FFT) of $p[n]$ and $p_{\text{target}}[n]$, respectively, calculated between 0.6 and 6 kHz, and K denotes the number of frequency bins in that range. The term ϵ is set at 10^{-6} to prevent the log-of-zero error. L_{smooth} is a biophysical constraint to penalize the first-order difference of $\mathbf{H}$.

$$L_{\text{smooth}} = \sum_{i=1}^{C-1} (h_{i+1} - h_i)^2. \tag{6}$$

L_{dist} is a probability-distribution constraint to ensure that the reflectance-coefficient vector $\mathbf{R} = [r_1, r_2, ..., r_C]$ approximately follows the zero-mean Gaussian distribution with unit variance [5]. Empirically, we set the weights of each loss term as $[\lambda_{\text{stft}}, \lambda_{\text{time}}, \lambda_{\text{freq}}, \lambda_{\text{smooth}}, \lambda_{\text{dist}}] = [20, 1, 200, 100, 200]$.

2.3. Extended Backward Path for DPOAE

To achieve a comprehensive simulation of human hearing, the TDLM-SNLv3 architecture is modified into a bidirectional system capable of generating both TEOAE and DPOAE.

2.3.1. Linear Reflection Path

The generation of the linear reflection component is based on coherent reflection theory [11], [12], which attributes emissions to random mechanical irregularities along the basilar membrane (BM). In this model, we adopt the linear backward architecture from our previous research [5]. Small, channel-specific scattering coefficients, enumerated as a vector $\mathbf{K} = [k_1, k_2, ..., k_C]$, are introduced along the cascade filters to represent these irregularities. $\mathbf{K}$ is defined as $\mathbf{R}$ multiplied by a small number δ:

$$\mathbf{K} = \delta \mathbf{R}, \tag{7}$$

where we use $\delta = 0.05$. These coefficients scale the forward-traveling wave at each channel, effectively creating linearly reflected backward-propagating waves. These components are extracted directly from the main signal path.

2.3.2. Nonlinear Coupling Path

The second mechanism for DPOAE generation involves the nonlinear coupling of distortion products generated by the instantaneous nonlinearity. In the TDLM-SNL architecture, the active path output $v_{\text{active},i}$ is defined as the output of the static nonlinearity. To isolate the generated distortion products, a distortion source term Δ_i is calculated as the difference between the signal after and before the nonlinear transformation:

$$\Delta_i = v_{\text{active},i} - h_i y_i. \tag{8}$$

This residual Δ_i captures the inter-modulation distortion components (such as $2f_1 - f_2$) generated at the OHC level. These nonlinear components are then directly coupled into the backward transmission path as shown in Fig. 3(b).

2.3.3. Backward Transmission path

The backward transmission path is adopted from [5] to model passive backward wave propagation. In addition, this adoption can capture the fact that any spectral component higher than the characteristic frequency would enter the evanescent mode and will not travel in the backward direction [13]. By coupling the linear reflection component and the nonlinear distortion residual Δ_i into this chain, DPOAE simulations is enabled. Also, as shown in Fig. 3(b), a middle ear module is included to ensure the final simulated OAE waveforms are in the dimensionality of acoustic pressure. The following DPOAE simulation was run at $\alpha = 15$ dB and $\beta = -35$ dB.

3. Results

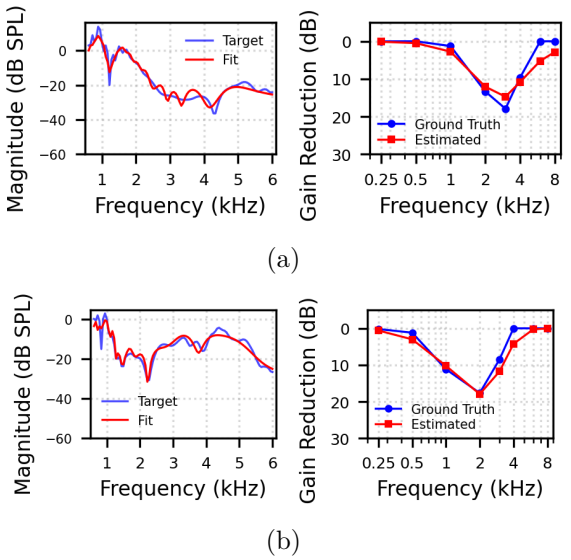


Figure 4: The TEOAE fitting results. Panels on the left are TEOAE magnitude spectrums, and panels on the right show cochlear gain reduction in dB. (a) The Notch3K pattern. (b) The Notch2K pattern.

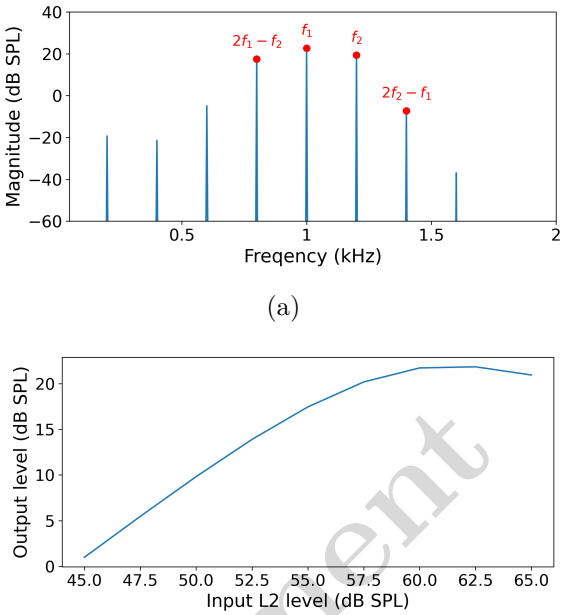


Figure 5: Simulated two-tone DPOAE characteristics (a) Spectral response for primary frequencies $f_1 = 1$ kHz and $f_2 = 1.2$ kHz (b) DPOAE I/O function at the $2f_1 - f_2$ component for input levels L_2 ranging from 45 to 65 dB SPL

3.1. TEOAE Fitting and Validation

The optimization results demonstrate that the TDLM-SNLv3 effectively captures the essential characteristics of TEOAEs. As shown in Fig. 4, the model successfully reconstructs the Notch3K/Notch2K feature in the frequency domain, with the estimated OHC gain reduction profile showing high accuracy compared to the TL model's ground truth; here, gain reduction was calculated as the dB difference in RMS energy between the impaired model's response and the response with the baseline (healthy) parameters. Thus, it serves as a proxy for quantifying hearing loss.

3.2. DPOAE Simulation

3.2.1. Two-tone DPOAE Simulation

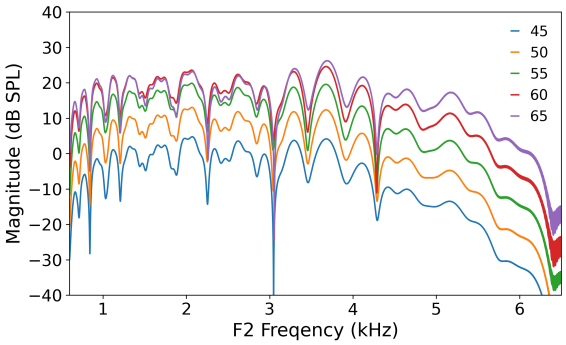
To simulate two-tone DPOAE under standard conditions, a stimulus paradigm of divergent primary tone levels [14] was employed to determine the input levels ($L_2 = 55$ dB SPL, $L_1 = 0.4L_2 + 39 = 61$ dB SPL), with a fixed frequency ratio of $f_2/f_1 = 1.2$. At these input levels, the spectral output Fig. 5(a) clearly demonstrates the characteristic nonlinear distortion inherent in the TDLM-SNLv3 architecture. The cubic distortion product ($2f_1 - f_2$) was observed at a magnitude of 17 dB SPL. In contrast, the upper sideband product ($2f_2 - f_1$) exhibited a significantly lower magnitude of -7 dB SPL. This pronounced suppression of the $2f_2 - f_1$ component validates that our backward low-pass filter chain effectively captures the physiological "low-pass" characteristic of DPOAE propagation, where high-frequency distortion products are attenuated as they travel toward the basal end of the cochlea. In Fig. 5(a), note that the peaks at f_1 and f_2 do not correspond to the primary-tone levels L_1 and L_2, since in Fig. 3(b) the input and the OAE signals are not summed in the ear canal. Instead, the spectral peaks at f_1 and f_2 should be more correctly interpreted as the *stimulus frequency OAE* [11] due to coherent reflection.

3.2.2. DPOAE Input/Output Characteristics

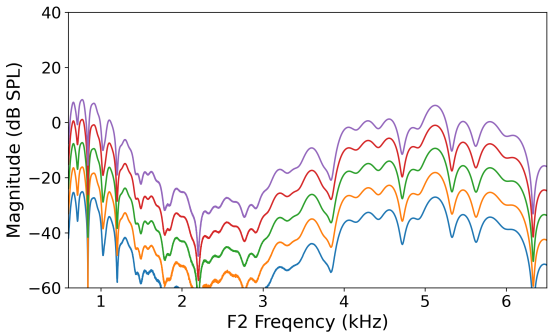
The model's nonlinear compression was further characterized by analyzing the DPOAE I/O function, with the input level L_2 ranging from 45 to 65 dB SPL in 2.5 dB increments. As illustrated in Fig. 5(b), the distortion product exhibits an expansive growth phase at lower primary levels (45 to 55 dB SPL). As the stimulus intensity increases, a distinct compressive trend emerges, with the DPOAE magnitude reaching a peak of approximately 22 dB SPL at an input of 60 dB SPL. However, a prominent "rollover" effect [15] is observed as the input exceeds this threshold; specifically, the DPOAE amplitude recedes to 21 dB SPL at $L_2 = 65$ dB SPL. This non-monotonic behavior indicates that the simulated cochlear transducer has entered a state of deep saturation. In this regime, the hard-limiting nature of the instantaneous nonlinear function, potentially coupled with phase cancellation between distributed sources, suppresses further growth of the distortion product. This characteristic aligns with physiological observations in human clinical data, where DPOAE amplitudes often plateau or decline at high stimulation levels due to the inherent saturation

of the cochlear amplifier.

3.2.3. DPOAE Frequency-Sweep and Fine Structure



(a)



(b)

Figure 6: The swept-sine DPOAE results (a) normal hearing, (b) Notch2K. The progressive compression of the I/O function is visualized by the narrowing vertical distance between the sweep curves as intensity increases.

The DPOAE frequency-sweep characteristics, extracted via the swept-sine method [16], are presented in Fig. 6(a). By aggregating responses across stimuli with four different phase-shifts ($0, \pi/2, \pi$, and $3\pi/2$), the $2f_1 - f_2$ distortion product remains while both primary tones are cancelled. The results reveal a prominent fine structure characterized by quasi-periodic ripples. This fine structure exhibits a distinct frequency dependence: the notch density is higher at lower frequencies and becomes increasingly sparse as f_2 increases, reflecting the logarithmic mapping of the cochlear tonotopic organization [17]. Regarding the growth behavior, the DPOAE magnitude shows a non-linear progression as L_2 increases from 45 to 65 dB SPL. While the curves are uniformly spaced at lower primary levels, this interval narrows at higher intensities, signaling the onset of nonlinear compression. Notably, a non-monotonic rollover effect is observed in several frequency regions, where the magnitude at $L_2 = 65$ dB SPL falls below that at $L_2 = 60$ dB SPL. The DPOAE frequency-sweep patterns with parameters $\mathbf{H}$ estimated by Notch2K TEOAE are shown in Fig. 6(b). A distinct notch is observed near 2 kHz, reflecting the impact of OHC damage. Nevertheless, the intensity of DPOAE grows rapidly as the input level increases, since the damaged OHC health keeps the $\phi(\cdot)$ from saturation in Eq. (2) even at high input levels.

4. Discussion

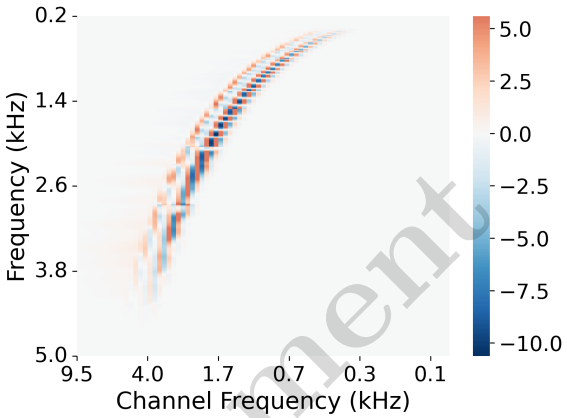


Figure 7: Jacobian matrix of the DPOAE simulation

To further elucidate the relationship between local cochlear health and OAE generation, we leverage JAX's automatic differentiation and performed a sensitivity analysis by calculating the Jacobian matrix of the DPOAE simulation. The Jacobian matrix is defined as:

$$\mathbf{J} = \frac{\partial \mathbf{P}}{\partial \mathbf{H}} = \begin{bmatrix} \frac{\partial P[0]}{\partial h_0} & \cdots & \frac{\partial P[0]}{\partial h_C} \\ \vdots & \ddots & \vdots \\ \frac{\partial P[K-1]}{\partial h_0} & \cdots & \frac{\partial P[K-1]}{\partial h_C} \end{bmatrix}, \quad (9)$$

where $\mathbf{P}$ is the FFT of the simulated four-phase DPOAE, calculated between 0.2 and 5 kHz (and K denotes the number of frequency bins in that range). The operation point of $\mathbf{H}$ is at the healthy state ($h_i = 1.0$ for all i).

As illustrated in Fig. 7, the resulting sensitivity map explicitly reveals the tonotopic organization of the human ear, where the arc-shaped distribution across channels reflects the intrinsic logarithmic frequency mapping of the cochlear partition. The alternating red and blue oscillatory patterns in the Jacobian matrix reveal the complex mechanisms of phase interference and coherent wave summation from distributed sources along the basilar membrane. These fluctuations represent regions where local parameter variations contribute either constructively (red) or destructively (blue) to the total ear canal pressure. Also, the sensitivity between frequency response and local parameters provides the theoretical basis for estimating cochlear mechanical parameters and locating OHC dysfunctions using recorded DPOAE signals.

We also acknowledge that the present personalization study was only validated on a few patterns of TL model-simulated TEOAEs. Future work should in-

volve a more extensive dataset encompassing real measured data and different types of hearing impairment. Moreover, the nonlinearity in our system currently lacks efferent control. Future development should consider incorporating a feedback loop to describe the dynamics of ascending and descending auditory neural pathways, thus allowing the efferent control to characterize the dynamic range compression more accurately.

5. Conclusion

This study successfully developed TDLM-SNLv3, a fully differentiable cochlear model within the JAX environment, providing a unified, bidirectional computational framework for TEOAE fitting and DPOAE simulation. By incorporating instantaneous static nonlinearity and extending the reverse transmission path, the model not only can reproduce individual TEOAE waveforms and spectral characteristics with high fidelity, but also accurately simulates complex physiological phenomena of DPOAEs.

Experimental results demonstrate that the model successfully captures inter-modulation distortion under two-tone stimulation, the non-monotonic rollover effect caused by cochlear transducer saturation at high intensities, and the quasi-periodic fine structure associated with the cochlear tonotopic organization. Furthermore, sensitivity analysis reveals the potential to estimate model parameters using DPOAEs. In conclusion, TDLM-SNLv3 integrates biophysical mechanisms with automatic differentiation algorithms, laying a solid foundation for future personalized hearing application based not only on TEOAEs but also on DPOAEs.

References

[1] T. Irino and R. D. Patterson, "A dynamic compressive gammachirp auditory filterbank," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2222–2232, 2006.

[2] A. Robert and J. L. Eriksson, "A composite model of the auditory periphery for simulating responses to complex sounds," *J. Acoust. Soc. Amer.*, vol. 106, no. 4, pp. 1852–1864, 1999.

[3] S. T. Neely, J. Rodriguez, Y.-W. Liu, and M. Gorga, "A computational model of loudness density," *unpublished article*, 2012, (available at http://audres.org/cel/loud/dlmsupp-revised.pdf).

[4] R. F. Lyon, "Cascades of two-pole-two-zero asymmetric resonators are good models of peripheral auditory function," en, *J. Acoust. Soc. Amer.*, vol. 130, no. 6, pp. 3893–3904, Dec. 2011.

[5] Y.-Y. Xu and Y.-W. Liu, "Auditory model personalization based on transient-evoked otoacoustic emissions," Proc. EUSIPCO, 2026.

[6] R. F. Lyon, R. Schonberger, M. Slaney, M. Velimirović, and H. Yu, "The CARFAC v2 cochlear model in Matlab, NumPy, and JAX," *ArXiv*, vol. abs/2404.17490, 2024.

[7] Y.-W. Liu and S. Neely, "Distortion product emissions from a cochlear model with nonlinear mechanoelectrical transduction in outer hair cells," *J. Acoust. Soc. Amer.*, vol. 127, no. 4, pp. 2420–2432, 2010.

[8] Y.-C. Huang, V. Vencovský, and Y.-W. Liu, "Cascade cochlear modeling using instantaneous nonlinearity and fixed filter parameters," *JASA Express Letters*, vol. 6, no. 6, p. 064 402, 2026.

[9] S. Verhulst, A. Altoè, and V. Vasilkov, "Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss," en, *Hearing Research*, vol. 360, pp. 55–75, Mar. 2018.

[10] R. Yamamoto, E. Song, and J.-M. Kim, "Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, 2019, pp. 6199–6203.

[11] C. A. Shera and J. J. Guinan Jr, "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian oaes," *J. Acoust. Soc. Amer.*, vol. 105, no. 2, pp. 782–798, 1999.

[12] G. Zweig and C. A. Shera, "The origin of periodicity in the spectrum of evoked otoacoustic emissions," en, *J. Acoust. Soc. Amer.*, vol. 98, no. 4, pp. 2018–2047, Oct. 1995.

[13] P. Dallos, *The Auditory Periphery Biophysics and Physiology*. Elsevier, 2012.

[14] T. Janssen, P. Kummer, and W. Arnold, "Growth behavior of the 2 f1- f2 distortion product otoacoustic emission in tinnitus," *J. Acoust. Soc. Amer.*, vol. 103, no. 6, pp. 3418–3430, 1998.

[15] M. L. Mills, Y. Shen, and R. H. Withnell, "Examining the factors that contribute to non-monotonic growth of the 2 f 1-f 2 otoacoustic emission in humans," *J. Assoc. Res. Otolaryngol*, vol. 22, no. 3, pp. 275–288, 2021.

[16] V. Vencovský, A. Novak, O. Klimeš, P. Honzík, and A. Vetešník, "Distortion-product otoacoustic emissions measured using synchronized swept-sines," *J. Acoust. Soc. Amer.*, vol. 153, no. 5, pp. 2586–2599, 2023.

[17] K. Reuter and D. Hammershøi, "Distortion product otoacoustic emission fine structure analysis of 50 normal-hearing humans," *J. Acoust. Soc. Amer.*, vol. 120, no. 1, pp. 270–279, 2006.