Analysis of click-evoked otoacoustic emissions by concentration of frequency and time: Preliminary results from normal hearing and Ménière’s disease ears

Tzu-Chi Liu, Hau-Tieng Wu, Ya-Hui Chen, Ya-Han Chen, Te-Yung Fang, Pa-Chun Wang, and Yi-Wen Liu

Citation: AIP Conference Proceedings **1965**, 170005 (2018); doi: 10.1063/1.5038538
View online: https://doi.org/10.1063/1.5038538
View Table of Contents: http://aip.scitation.org/toc/apc/1965/1
Published by the American Institute of Physics

Analysis of Click-Evoked Otoacoustic Emissions by Concentration of Frequency and Time: Preliminary Results from Normal Hearing and Ménière's Disease Ears

Tzu-Chi Liu[1], Hau-Tieng Wu[2], Ya-Hui Chen[3], Ya-Han Chen[1], Te-Yung Fang[3],
Pa-Chun Wang[3] and Yi-Wen Liu[1,a)]

[1]*Dept. Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan*
[2]*Dept. Mathematics, University of Toronto, Canada*
[3]*Dept. Otolaryngology, Cathay General Hospital, Taipei, Taiwan*

a)Corresponding author: ywliu@ee.nthu.edu.tw

Abstract. The presence of click-evoked (CE) otoacoustic emissions (OAEs) has been clinically accepted as an indicator of normal cochlear processing of sounds. For treatment and diagnostic purposes, however, clinicians do not typically pay attention to the detailed spectrum and waveform of CEOAEs. A possible reason is due to the lack of noise-robust signal processing tools to estimate physiologically meaningful time-frequency properties of CEOAEs, such as the latency of spectral components. In this on-going study, we applied a modern tool called concentration of frequency and time (ConceFT, [1]) to analyze CEOAE waveforms. Randomly combined orthogonal functions are used as windowing functions for time-frequency analysis. The resulting spectrograms are subject to nonlinear time-frequency reassignment so as to enhance the concentration of time-varying sinusoidal components. The results after reassignment could be further averaged across the random choice of windows. CEOAE waveforms are acquired by a linear averaging paradigm, and longitudinal data are currently being collected from patients with Ménière's disease (MD) and a control group of normal hearing subjects. When CEOAE is present, the ConceFT plots show traces of decreasing but fluctuating instantaneous frequency against time. For comparison purposes, same processing methods are also applied to analyze CEOAE data from cochlear mechanics simulation.

INTRODUCTION

CEOAE was discovered almost 40 years ago [2]. Ever since, it has been noticed that high frequency components occur earlier than low frequency parts in CEOAE waveforms, and this is adequately explained by the fact that a low frequency component reaches further into the cochlea before getting coherently reflected around its characteristic-frequency (CF) place [8]. For clinical applications, the presence of CEOAE indicates normal cochlear functioning. However, detailed signal properties such as the phase-gradient delay function of CEOAEs or stimulus-frequency (SF) OAEs have not been utilized by clinicians for diagnosis purposes. Part of the reason would be that, though the concept of higher frequency returning first is simple and clear, the actual signal is intrinsically fluctuating both in frequency and amplitude [3]. The picture is even murkier if one considers that the reverse traveling waves from the cochlea get reflected at the stapes, so the actual OAE measured in the ear canal would be a superposition of multiple reflections. Therefore, when analyzing the temporal-spectral properties of CEOAEs, traditional short-time Fourier transform (STFT) might suffer from interference for it lacks the flexibility to "focus" on local properties.

Various ways have been proposed to visualize and extract information from CEOAEs (or its spectral equivalent) in the two dimensional time-frequency (T-F) space. For instance, Jedrzejczak et al. [4] built a dictionary of asymmetric Gabor functions to span a linear space and applied matching pursuit algorithms to identify the best fit to transient-evoked (TE) OAEs. Latency as a function of frequency could be inferred and empirical fits were reported. The continuous wavelet transform (CWT) was utilized to infer the hearing functionality of neonates through their TEOAEs [5] and to investigate the relationship of TEOAE latency and stimulus level [6]. CWT was also applied for viewing SFOAE on the T-F plane [3]; because of the multi-resolution property, CWT provides a certain extent of

To the Ear and Back Again - Advances in Auditory Biophysics
AIP Conf. Proc. 1965, 170005-1–170005-7; https://doi.org/10.1063/1.5038538
Published by AIP Publishing. 978-0-7354-1670-3/$30.00

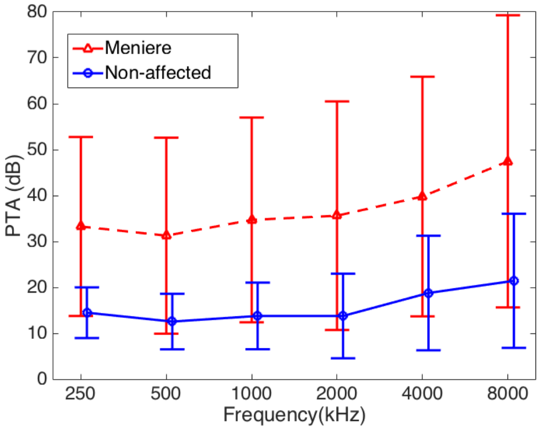

FIGURE 1. Summary of pure tone audiogram (PTA) data from MD subjects. All were affected in only one ear, and data were also obtained from the other ear (non-affected) for comparison purposes. The bars indicate ±1 standard deviation.

flexibility to focus on instantaneous signal properties of the OAEs. Further, by filtering on the T-F plane and then applying the inverse CWT, it is possible to extract the single-reflection component from its mixture with higher-order reflections even though they overlap in time.

In this paper, we explore the possibility of analyzing CEOAEs by a relatively new T-F processing technique called *concentration of frequency and time (ConceFT)*, which is a nonlinear-type T-F analysis technique. It has been established that, if the signal of interest can be modeled as a sum of sinusoids that satisfy a *well-separated* condition and certain *slow-varying* assumptions (see appendix), then ConceFT helps produce sharpened traces on the T-F plane that represent the signal [1, 9]. In the next section, mathematical definition of the technique will be described. Subsequently, examples of CEOAEs obtained from normal hearing (NH) and Ménière's disease (MD) ears, and results of ConceFT T-F plots will be shown. Finally, comparison to simulation will be briefly discussed before conclusions are given.

METHODS

In this section, we first describe how CEOAE waveforms were recorded and estimated. Then, ways to transform the estimated CEOAE signals into T-F plots are described in mathematical details.

Signal Acquisition

CEOAE waveforms were recorded using an ER-10C microphone (Etymotic Research Inc., Elkgrove Village, IL, USA) and an UltraLite-mk3 Hybrid soundcard (MOTU, Cambridge, MA, USA). The click stimulus was generated by a python script and delivered through the soundcard to the ER-10C earphone. The peak sound pressure level (peSPL) was about 74 dB. The click rate was maintained at 10 per second. The CEOAEs have been recorded in a quiet office at Cathay General Hospital, Taipei. The noise floor has been about 23-27 dB SPL, varying between measurements. For each recording session, 3000 clicks were presented. For MD subjects, the non-affected ears were also measured. A linear-phase bandpass filter was applied off line to reject the frequencies below 1 kHz or above 8 kHz. Then, artifact rejection was performed; a frame (0.1 sec long) was discarded if the instantaneous SPL was greater than 60 dB at anywhere more than 3 ms after the onset of the click. Afterwards, the mean acoustic pressure $\bar{p}(t)$ was calculated by averaging across all the artifact-free frames.

Short-Time Fourier Transform, Synchrosqueezing, and ConceFT

Given the CEOAE signal $\bar{p}(t)$, its STFT associated with a window function $h(t)$ can be defined as:

$$V_{\bar{p}}^{(h)}(t,\eta) := \int \bar{p}(s)h(s-t)e^{-i2\pi\eta(s-t)}\,\mathrm{d}s, \tag{1}$$

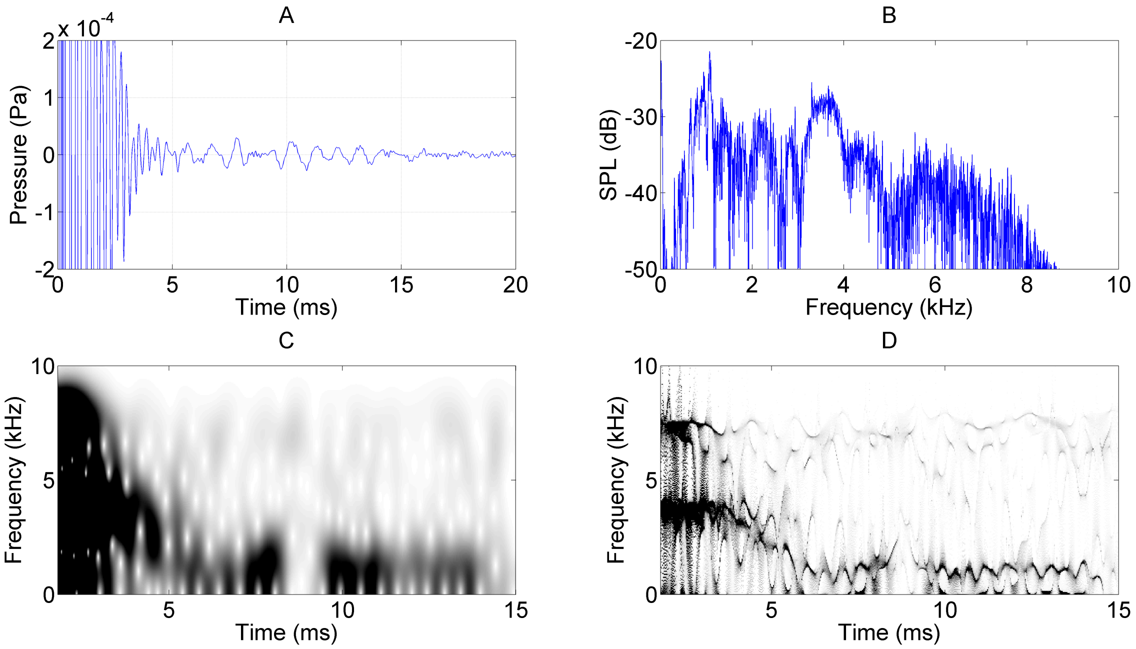

FIGURE 2. Typical results of CEOAE analyzed by the present methods. Data were obtained from a NH ear. (A) Time domain after averaging, (B) magnitude spectrum, (C) STFT, and (D) ConceFT, obtained with $N = 50$ random windows in Eq. (3).

where $\eta \in \mathbb{R}^+$ is the frequency, and h is the window function chosen by the user. The ConceFT is composed of the nonlinear manipulations of the STFT – the synchrosqueezing transform (SST) to sharpen the spectrum and the nonlinear multitapering to stabilize the spectrum. The STFT-based SST is defined as:

$$S_{\bar{p}}^{(h,\Gamma,\alpha)}(t,\xi) := \int_{\mathfrak{R}_t} V_{\bar{p}}^{(h)}(t,\eta) g_\alpha\left(|\xi - \Omega_{\bar{p}}^{(h,\Gamma)}(t,\eta)|\right) \mathrm{d}\eta, \tag{2}$$

where $\mathfrak{R}_t := \{\eta : |V_{\bar{p}}^{(h)}(t,\eta)| > \Gamma\}$, $0 < \alpha \ll 1$ are chosen by the user, $g_\alpha(\cdot) := \frac{1}{\alpha} g(\frac{\cdot}{\alpha})$, g is a smooth function so that $g_\alpha \to \delta$ in the weak sense as $\alpha \to 0$, and $\Omega_{\bar{p}}^{(h,\Gamma)}(t,\omega)$ is the frequency reassignment rule, which is given in every points (t,ω) by:

$$\Omega_{\bar{p}}^{(h,\Gamma)}(t,\omega) = -\Im \frac{V_{\bar{p}}^{(\mathcal{D}h)}(t,\eta)}{V_{\bar{p}}^{(h)}(t,\eta)} \text{ when } |V_{\bar{p}}^{(h)}(t,\eta)| > \Gamma \quad \text{and} \quad \Omega_{\bar{p}}^{(h,\Gamma)}(t,\omega) = -\infty \text{ when } |V_{\bar{p}}^{(h)}(t,\eta)| \le \Gamma,$$

where $\Im$ means taking the imaginary part, $\Gamma \ge 0$ is the chosen hard threshold to reduce the numerical error and noise influence and $\mathcal{D}h = h'$, the first derivative of h.

The ConceFT could then be defined by the following multitapering step:

$$\mathtt{CSST}_f(t,\xi) := \frac{1}{N} \sum_{n=1}^{N} S_f^{(h_{[n]},\Gamma,\alpha)}(t,\xi), \tag{3}$$

where each window function $h_{[n]}$ is a random linear combination of the first J orthonormalized Hermite functions, denoted as $\{h_1, \ldots, h_J\}$. In particular, h_1 is the Gaussian function. In practice, J could be chosen as small as 2, while N could be chosen as large as the user wishes, but a number of $N = 30$ is in general good enough. Theoretical properties of ConceFT based on SST has been explored in [1].

We could see the motivation for defining Ω_f and $S_f^{(h,\Gamma,\alpha)}(t,\xi)$ by considering a harmonic function $f(t) = A\exp(i2\pi\xi_0 t)$, where $A, \xi_0 > 0$. By a direct calculation, $V_f^{(h)}(t,\eta) = A\hat{h}(\xi_0 - \eta)e^{i2\pi\xi_0 t}$ and $V_f^{(\mathcal{D}h)} = A\xi_0\hat{h}(\xi_0 - \eta)e^{i2\pi\xi_0 t}$

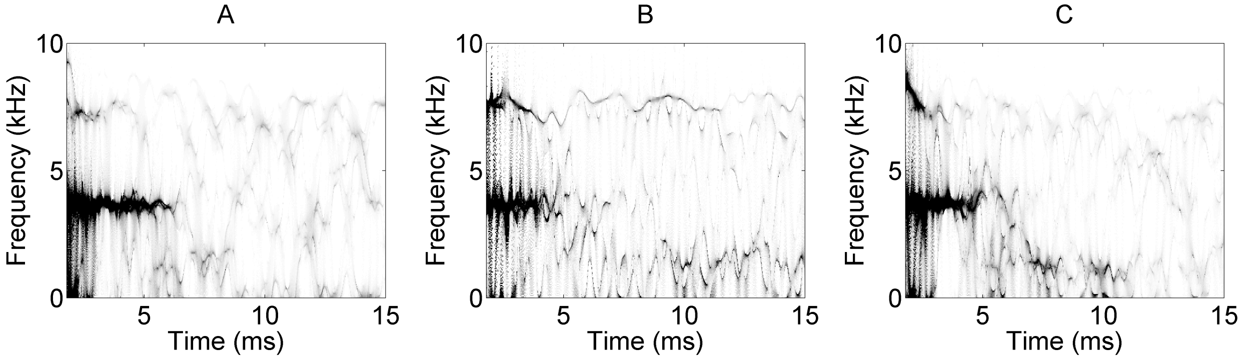

FIGURE 3. Longitudinal ConceFT plots of CEOAE obtained from a MD ear over a span of three months.(A) First visit to the clinic. (B) Three-week follow-up after continuous treatment. (C) Three months follow-up.

(here, we use the "hat" symbol $\hat{h}$ to denote the Fourier transform of the corresponding time-domain function $h(t)$). Thus, by definition, $\Omega_f(t,\omega) = \xi_0$ when $A|\hat{h}(\xi_0 - \omega)| > \Gamma$, which results in a sharper time-frequency representation, $S_f^{(h,\Gamma,\alpha)}(t,\xi)$.

Subjects

We have recruited 9 NH subjects and 26 MD subjects to participate in this research under the IRB approval at Cathay General Hospital, Taipei, Taiwan. MD subjects were recruited with a consecutive manner. CEOAE, distortion-product (DP) OAE, and pure-tone audiogram (PTA) were collected from both ears of each subject. For the MD subjects, data were first collected during their first visit (after diagnosed with MD). After they received treatment, the same data were collected longitudinally during the follow-up appointments. The age distribution of MD patients is 46.1 ± 10.3 yrs and the age of the normal group is 23.6 ± 1.4 yrs. Figure 1 shows the distribution of PTA of the MD patients during their first visits.

Our long-term goal is to study if details of CEOAE signal properties help predict the chance of recovery from MD. Data collection started in Nov. 2015 and will continue till the end of 2017.

RESULTS

Figure 2 shows the CEOAE obtained from a typical NH ear and results of subsequent T-F processing. The raw signal is subject to band-pass filtering from 1 to 8 kHz and then is averaged per 100 ms. Figure 2(B) shows the spectrum of the averaged signal from 3 to 100 ms. Comparing STFT and ConceFT, it is clear that frequency reassignment produces narrow sharpened traces out of the blurry STFT representation.

Figure 3 shows how the CEOAE ConceFT plot has changed for a MD ear throughout the treatment process. The PTA of this ear indicates 35-55 dB hearing loss at 0.5 to 8 kHz during the initial visit. The patient was treated, and three weeks later the PTA improved at every frequency, dropping to 30-35 dB. Three months later, the PTA further improved by 5-10 dB at a few frequencies. In the CEOAE ConceFT plots, the dense trace near 4 kHz before $t = 5$ ms should be neglected because we believe it corresponds to vibrational coupling in the ER10C probe. We are currently working on cancelling or reducing this artifact. Other than this, changes in the ConceFT plots seem to correlate with the hearing recovery progress the patient has made based on the following observations: first, 1-kHz component seem to have emerged in (B) after $t = 10$ ms. Secondly, a descending glide from 4 kHz to 2 kHz seem to be present from $t = 5$ to 7 ms in (C).

At this moment we should avoid generalizing these particular findings. A variety of factors can lead to MD. We have seen examples of near perfect PTA (≤ 15 dB) and DPOAE (> 0 dB SPL) across the entire frequency range in an MD-affected ear, and the CEOAE conceFT also resembles that of NH ears. Further analysis of the overall data is warranted.

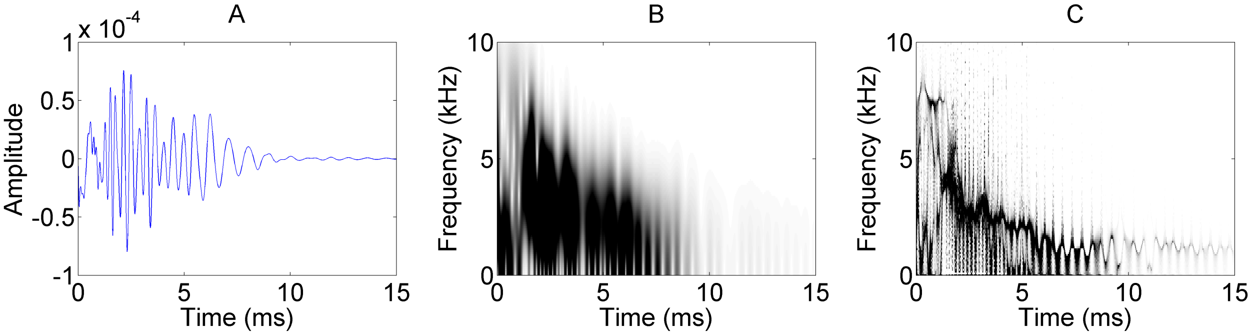

FIGURE 4. A typical result of simulated OAE and its processing by the present methods. (A) Time-domain waveform $h_{\mathrm{OAE}}(t)$ obtained by introducing 2% of irregularity to the basilar-membrane impedance, then using a symmetric Hann window of width 1.0 mm for spatial parameter smoothing. (B) and (C): STFT and ConceFT, respectively, under the same setting as in Fig. 2.

DISCUSSION

Confounded by the mechanical vibration artifacts and possibly the measurement noise, we resort to simulation to ensure the validity of the present signal processing techniques. A frequency-domain model of middle-ear and cochlear mechanics is adopted [7]. Cochlear parameters in the model used to have smooth variation along the wave-traveling direction. By introducing Gaussian irregularity $\epsilon(x) \sim \mathcal{N}(0, \sigma_\epsilon^2)$ to the basilar membrane (BM) impedance, the model produces significant amount of reverse waves. Empirically, at any fixed σ_ϵ^2, we found that applying spatial smoothing to $\epsilon(x)$ helped reducing the chance of spontaneous emissions in the model. Similar to what is found in [3], the OAE magnitude can be controlled by adjusting σ_ϵ^2. Multiple reflection is also considered, so an OAE transfer function $H_{\mathrm{OAE}}(f) = P_{\mathrm{OAE}}(f)/P_{\mathrm{EC}}(f)$ can be calculated, where $P_{\mathrm{EC}}(f)$ denotes the stimulus spectrum delivered to the ear canal, and $P_{\mathrm{OAE}}(f)$ denotes the linearly generated OAE in response to the stimulus. An impulse response $h_{\mathrm{OAE}}(t)$ can be obtained by applying the inverse Fourier transform to $H_{\mathrm{OAE}}(f)$. Detailed description of the model and the way to calculate the OAE transfer function can be found in [7].

Figure 4 shows a typical result of simulated CEOAE, and its processing by re-assignment techniques. The parameters of the cochlear model has not been fine-tuned to fit human data; nevertheless, reassignment result in (C) clearly shows a decreasing trend of instantaneous frequency as it varies against time. Also, the descending trace wobbles and is not monotonic. The simulation result thus captures the complex dynamics of CEOAEs.

CONCLUSIONS

CEOAE data have been collected from NH and MD ears. A frequency reassignment technique is applied to process the data in the T-F space. Preliminary results for real data and for simulated data seem to promise that, when the OAE signal is present, its concentration on the T-F plane can be sharpened by the present techniques. The sharpening effect should be at least beneficial for data visualization purposes.

ACKNOWLEDGMENTS

This research is supported by the Ministry of Science and Technology of Taiwan under grant No. 105-2628-E-007-005-MY2.

REFERENCES

[1] Daubechies I, Wang Y, Wu HT (2016) ConceFT: Concentration of frequency and time via a multitapered synchrosqueezed transform. The Royal Society Publishing: Phil Trans A, 374:20150193.

[2] Kemp DT (1978) Stimulated acoustic emissions from within the human auditory system. J Acoust Soc Am 64:1386–1391.

[3] Shera CA, Bergevin C (2012) Obtaining reliable phase-gradient delays from otoacoustic emission data. J Acoust Soc Am 132(2):927–943.

[4] Jedrzejczak WW, Kwaskiewicz K, Blinowska KJ, Kochanek K, Skarzynski H (2009) Use of the matching pursuit algorithm with a dictionary of asymmetric waveforms in the analysis of transient evoked otoacoustic emissions. J Acoust Soc Am 126(6):3137–3146.

[5] Moleti A, Sisto R, Tognola G, Parazzini M, Ravazzani P, Grandori F (2005) Otoacoustic emission latency, cochlear tuning, and hearing functionality in neonates. J Acoust Soc Am 118(3):1576–1584.

[6] Sisto R, Moleti A (2007) Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels. J Acoust Soc Am 122(4):2183–2190.

[7] Liu TC, Liu YW (2016) Quasilinear reflection as a possible mechanism for suppressor-induced otoacoustic emission. J Acoust Soc Am 140(6):4193–4203.

[8] Zweig G, Shera CA (1995) The origin of periodicity in the spectrum of evoked otoacoustic emissions. J Acoust Soc Am 98:2018–2047.

[9] Auger F, Flandrin P, Lin YT, McLaughlin S, Meignen S, Oberlin T, Wu HT (2013) Time-frequency reassignment and synchrosqueezing: An overview. IEEE Sig Process Mag 30(6):32–41.

[10] Nuttall AH, Bedrosian E (1966) On the quadrature approximation to the Hilbert transform of modulated signals. Proceedings of the IEEE, 54(10), 1458-1459.

Appendix: Approximation Error Regarding the Use of Hilbert Transform

An *intrinsic-mode-type (IMT) function* $f : \mathbb{R} \to \mathbb{C}$ has the form

$$f(t) = A(t)e^{i2\pi\phi(t)}, \tag{4}$$

where $A(t) > 0$ and $\phi'(t) > 0$ satisfy the *slowly varying* condition: $|A'(t)| \leq \epsilon|\phi'(t)|$, and $|\phi''(t)| \leq \epsilon|\phi'(t)|$ for all $t \in \mathbb{R}$. Superpositions of IMTs are referred to as *adaptive harmonic models* (AHMs), having the form

$$f(t) = \sum_{k=1}^{K} f_k(t) = \sum_{k=1}^{K} A_k(t)e^{i2\pi\phi_k(t)}$$

for some finite $K > 0$, and the constituting IMFs satisfy the following *frequency separation* condition: $\phi'_k(t) - \phi'_{k-1}(t) \geq d$, for all $t \in \mathbb{R}$.

Based on the argument provided in [8], the main component of a CEOAE waveform is the superposition of coherently reflected wavelets that return from respective CF places. We could model the CEOAE signal by the real form of an AHM; that is, a CEOAE signal satisfies

$$f_{\mathbb{R}}(t) = A_k(t)\cos(2\pi\phi_k(t) + ct) = \Re f(t), \tag{5}$$

where $f(t) = A_k(t)e^{i2\pi\phi_k(t)+ct}$, $c > 0$ is a base frequency, k can be regarded as the number of intra-cochlear reflection in the context of OAEs, and all conditions of an AHM are satisfied except that the function is real. In practice, we only have an access to $f_{\mathbb{R}}$ instead of f. To recover f from $f_{\mathbb{R}}$, a common approach is applying the Hilbert transform. However, it has been well known that the Hilbert transform of $f_{\mathbb{R}}$ may not guarantee to be its complex conjugate, $\sum_{k=1}^{K} A_k(t)\sin(2\pi\phi_k(t))$. We now quantify how much error incurs from this procedure. Denote

$$f_H(t) = [\mathcal{H} f_{\mathbb{R}}](t), \quad f_Q(t) := \sum_{k=1}^{K} A_k(t)\sin(2\pi\phi_k(t) + \xi_0 t),$$

where $\mathcal{H}$ is the Hilbert transform. Clearly, $f(t) = f_{\mathbb{R}}(t) + i f_Q(t)$ but it is not clear if $f_H = f_Q$. We now follow the argument in Nuttall [10] and consider the following error $E(f) := \int_{-\infty}^{\infty} |f_Q(t) - f_H(t)|^2 dt$. Based on Parseval's theorem, we have

$$\begin{aligned} E(f) &= \int_{-\infty}^{\infty} |\hat{f}_Q(\xi) - \hat{f}_H(\xi)|^2 d\xi \\ &= \int_{-\infty}^{\infty} | - i\mathrm{sgn}(\xi)\hat{f}_{\mathbb{R}}(\xi) - \hat{f}_Q(\xi)|^2 d\xi = 2\int_{-\infty}^{-\xi_0} |\hat{f}_{\mathbb{R}}(\xi)|^2 d\xi, \end{aligned}$$

where the last equality holds due to the facts that $\hat{f}_{\mathbb{R}}(\xi) = \frac{1}{2}[\hat{f}(\xi-\xi_0)+\overline{\hat{f}(-\xi-\xi_0)}]$ and $\hat{f}_Q(\xi) = \frac{1}{2i}[\hat{f}(\xi-\xi_0)-\overline{\hat{f}(-\xi-\xi_0)}]$. Thus, by the conditions obtained from simulation or model, we know K, A_k, ϕ_k and ξ_0. Hence we could evaluate $E(f)$ precisely. In conclusion, we could count on the Hilbert transform to recover the complex form from the collected data with a controllable error.

COMMENTS & QUESTIONS

[*Online Forum*]

Bastian Epp: You also mention that "the simulation result captures the complex dynamics of CEOAEs." What do you mean by *complex dynamics*? And how do you know that the stripes/oscillations you see are not just an artifact of the method?

Authors (H.T. Wu/Y.W. Liu): To analyze a signal, you always need a model and hence an algorithm designed for it. Suitability of the current model is based on the background knowledge; we know that CEOAE originates not from a single point but from a distributed area, and multiple reflections happen inside the cochlea for the traveling waves. The algorithm has been mathematically analyzed and its viability is confirmed. The question raised here could actually be the final stage of scientific argument — does the model fit the data or capture the system behaviors? In general it is a difficult problem and we do not have a good answer now. However, we do have a partial confirmation — the simulated dataset confirms it, and the finding is consistent across subjects we have. So, we could conclude that the finding is not an algorithm-based artifact, unless the simulation is too far away from the true physiology.