



Measuring Distortion-Product Otoacoustic Emission With a Single Loudspeaker in the Ear: Stimulus Design and Signal Processing Techniques

Wei-Chen Hsiao, Yung-Ching Chen and Yi-Wen Liu*

Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan

OPEN ACCESS

Edited by:

Qinglin Meng,
South China University of Technology,
China

Reviewed by:

Ivan Miguel Pires,
Universidade da Beira Interior,
Portugal
Shixiong Chen,
Shenzhen Institutes of Advanced
Technology, Chinese Academy of
Sciences (CAS), China

*Correspondence:

Yi-Wen Liu
ywliu@ee.nthu.edu.tw

Specialty section:

This article was submitted to
Connected Health,
a section of the journal
Frontiers in Digital Health

Received: 13 June 2021

Accepted: 09 August 2021

Published: 01 September 2021

Citation:

Hsiao WC, Chen YC and Liu YW
(2021) Measuring Distortion-Product
Otoacoustic Emission With a Single
Loudspeaker in the Ear: Stimulus
Design and Signal Processing
Techniques.
Front. Digit. Health 3:724539.
doi: 10.3389/fdgth.2021.724539

The distortion-product otoacoustic emission (DPOAE) is a backward propagating wave generated inside the cochlea during the wave amplification process. The DPOAE signal can be detected rapidly under relatively noisy conditions. In recent years, the earphone industry demonstrated interest in adopting DPOAE as an add-on feature to make their product “intelligent” of inner-ear status. However, a technical challenge remains to be tackled—the loudspeaker in an earphone generates its own cubic distortion at the same frequency as DPOAE. Unfortunately, the intensity of loudspeaker distortion is typically comparable to that of the DPOAE, if not higher. In this research, we propose two strategies, namely *compensation* and *cancellation*, to enable DPOAE measurement with a single loudspeaker. The compensation strategy exploits the part of the growth function of the loudspeaker distortion which is almost linear, and thus suppresses the distortion it generates while retaining a larger portion of DPOAE in the residual signal. The cancellation strategy utilizes a one-dimensional Volterra filter to remove the cubic distortion from the loudspeaker. Testing on normal-hearing ears shows that the compensation strategy improved the DPOAE-to-interference ratio by approximately 7 dB, resulting in a cross-correlation of 0.62 between the residual DPOAE level and the true DPOAE level. Meanwhile, the cancellation strategy directly recovered both the magnitude and the phase of DPOAE, reducing the magnitude estimation error from 15.5 dB to 3.9 dB in the mean-square sense. These pilot results suggest that the cancellation strategy may be suitable for further testing with more subjects.

Keywords: hearing, otoacoustic emissions, intermodulation distortion, nonlinear signal processing, Volterra filtering

1. INTRODUCTION

Otoacoustic emissions (OAEs) are sounds generated in the cochlea that propagate backward to emit from the ear (1). OAEs can be classified into two types (2)—spontaneous OAEs (SOAEs) and evoked OAEs. SOAEs occur in the absence of external stimulus, and evoked OAEs can be regarded as acoustic responses to external stimulus. Within the family of evoked OAEs, the distortion-product OAE (DPOAE) is widely used as an objective tool for detecting hearing impairment

associated with outer hair cell (OHC) dysfunctions (3, 4). To measure DPOAE, a pair of primary tones at frequencies $f_1 < f_2$ are delivered to an earphone inserted to the ear canal. With appropriately chosen intensities and frequencies of the primary tones [e.g., $f_2/f_1 = 1.22$, (5)], the most prominent distortion product would occur at $2f_1 - f_2$ and it can be recorded from a microphone in the ear canal. Because the primary tones' excitation patterns mainly overlap near the f_2 characteristic place in the cochlea (6), the sound-pressure level (SPL) of DPOAE at $f_{DP} = 2f_1 - f_2$ represents the cochlea's ability to process signals normally at frequency f_2 . Thus, DPOAE serves as a robust and non-invasive tool for assessing cochlear functions in a frequency-specific manner (4). It has been applied clinically for hearing screening (7, 8), and diagnosis of acute hearing loss (9) and other kinds of hearing impairment (10, 11).

Typically, a clinical DPOAE probe consists of two loudspeakers and one microphone; for each ear, the primary tones at f_1 and f_2 are separately delivered to the two speakers to avoid generating intermodulation distortion (IMD) electrically (12). As a rare exception, a single-speaker configuration was adopted for measuring vibration caused by DPOAE on insect tympanal organs (13); however, it was emphasized that one should avoid over-driving the speaker and thus producing IMD artifacts (14). In the field of cochlear neurophysiology, nonetheless, a combination of 5–7 tones with carefully arranged frequencies could be delivered simultaneously to a single speaker to elicit auditory-nerve responses (15); in their study, loudspeaker IMD was not a concern because the neural response by nature contains strong quadratic-distortion components which actually facilitate efficient estimation of the cochlear tuning curve at the auditory-nerve level.

Recently, the two-loudspeaker hardware design has been adopted by a commercial headphone that promotes at-home DPOAE measurement as a means of providing personalized frequency response adjustment (16). The two-speaker design seems necessary because, even with a high-quality headphone or earphone, the total harmonic distortion (THD) can reach 3% when driven to its full dynamic range (17). This THD level is acceptable for listening to music; however, when delivering two pure tones simultaneously, we found that the distortion generated by such speakers would significantly interfere with the DPOAE from the ear since the cubic distortion of the speaker also occurs at f_{DP} .

Nevertheless, human DPOAE and loudspeaker IMD have different generation mechanisms even though they may occur at the same frequency. For example, the DPOAE signal is comprised of a direct component plus a reflective component (18, 19); the direct component travels back from the f_2 characteristic place in the cochlea, while the reflective component travels further to the f_{DP} place and changes direction due to *coherent reflection* (20). The two components thus have different latency in the range of 5–20 ms, which allows them to be separated via envelope-tracking techniques (21). Also, they may superpose constructively or destructively depending on their relative phase. In comparison, the loudspeaker IMD is perhaps elicited nearly instantaneously, so we expect that its latency and rate of growth with respect to the primary tone levels L_1 and L_2

might differ from that of DPOAE. In this research, we seek to exploit these differences and develop methods for estimation of DPOAE levels using a *single speaker*, despite of interference from loudspeaker IMD.

In particular, we propose stimulus design and signal processing strategies that handle the interference issues due to loudspeaker IMD. The first strategy is called *compensation* and it involves finding a combination of L_1 and L_2 such that the IMD level grows *almost linearly* with respect to simultaneous increment in (L_1, L_2) . The second strategy is called *cancellation* and it utilizes 3rd-order one-dimensional Volterra filter (22) to subtract the loudspeaker IMD from the signal. The organization of the remaining part of this paper follows the standard order of Methods, Results, Discussion, and Conclusion.

2. METHODS

In this section, we first review the mathematics of IMD generated by two tones. Typical spectrums of DPOAE and loudspeaker IMD will be shown so we can examine the similarities and differences. Then, the compensation strategy and the cancellation strategy will be described. This section ends with brief descriptions of the recording equipment and the human subjects who participated in the testing.

2.1. IMD and Mathematical Notations

Assuming that an acoustic or electrical stimulus, called the input signal, contains two frequency components f_1 and f_2 , so the signal can be expressed as follows,

$$x(t) = A_1 \cos(2\pi f_1 t + \phi_1) + A_2 \cos(2\pi f_2 t + \phi_2), \quad (1)$$

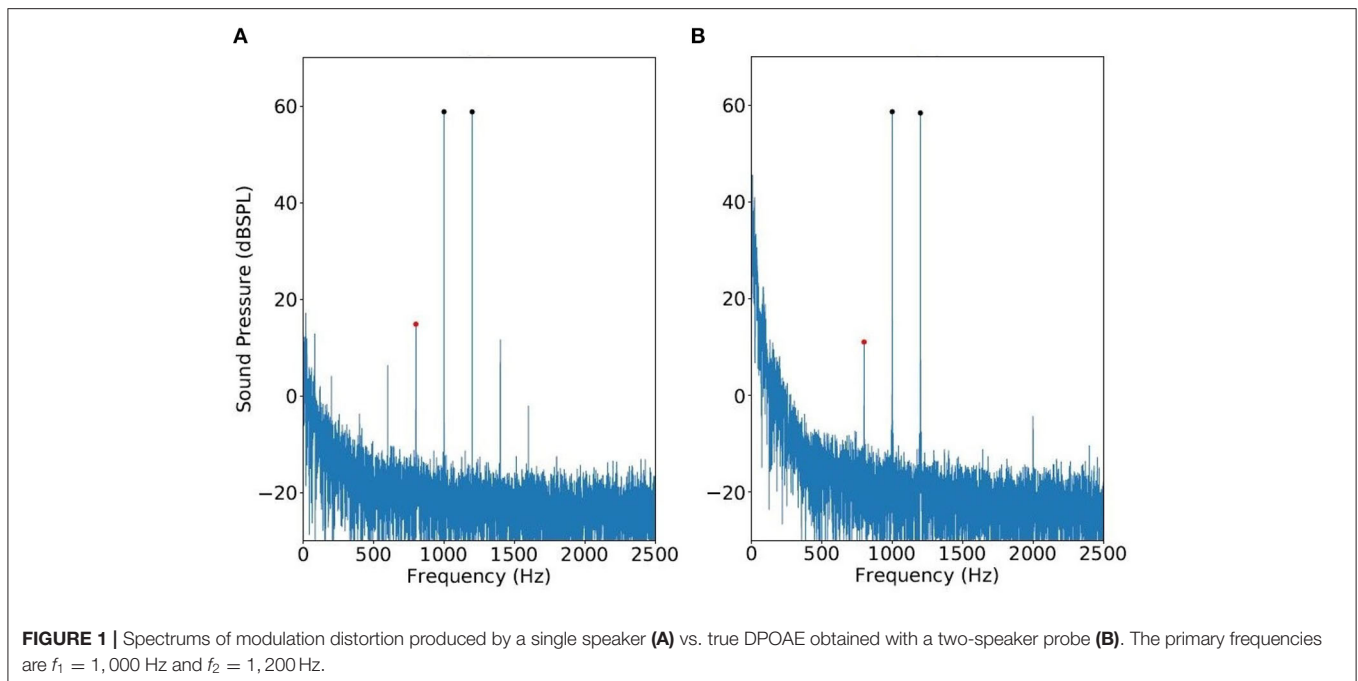
where A_1, A_2 and ϕ_1, ϕ_2 denote the amplitude and phase for two components, respectively. Assume that the stimulus is delivered to a nonlinear system G so that the response $y(t)$ can be denoted as $y(t) = G(x(t))$. When G is instantaneous, it can be expanded by Taylor's series near the origin; that is,

$$G(\eta) = G(0) + \sum_{k=1}^{\infty} \frac{G^{(k)}(0)}{k!} \eta^k. \quad (2)$$

By setting $\eta = x(t)$ and through simple trigonometry, one can show that

$$y(t) = \text{Re} \left[\sum_m \sum_n B_{m,n} e^{j2\pi(mf_1 + nf_2)t} \right], \quad (3)$$

where $B_{m,n}$ are complex-valued coefficients, and m and n sum over all integers such that $mf_1 + nf_2 > 0$. The components $B_{m,0}$ or $B_{0,n}$ are referred to as harmonics; the additional components $B_{m,n}$ when m and n are both nonzero are called the intermodulation products and they can be classified by their order $|m| + |n|$. In the context of DPOAE measurement with a single loudspeaker, $y(t)$ thus contains not only the primary frequencies f_1 and f_2 but also higher-order components at $mf_1 + nf_2$. Empirically, the 3rd-order intermodulation products $B_{2,-1}$ and $B_{-1,2}$ are most prominent from a loudspeaker (see **Figure 1A**), and a few



other components such as $B_{3,-2}$, $B_{-2,3}$ can also be identified above the noise floor. In comparison, from the DPOAE spectrum (measured by the conventional two-speaker approach), the $B_{-1,2}$ component corresponding to frequency $2f_2 - f_1$ usually cannot be detected (see **Figure 1B**). The reason is because, even though the intermodulation product at $2f_2 - f_1$ is generated due to OHC nonlinearity near the f_2 place, it is prohibited from backward propagation along the basilar membrane (23).

By inspecting **Figure 1**, note that the loudspeaker components occurring at $2f_1 - f_2$ has the same frequency as DPOAE. This 3rd-order component from the loudspeaker is referred to as IMD3 hereafter, and we shall investigate how to estimate DPOAE regardless of the presence of IMD3.

2.2. The Compensation Strategy

The DPOAE level, denoted as L_{DP} , depends systematically on parameters (L_1, L_2, f_1, f_2) . The relation $L_{DP} = L_{DP}(L_1, L_2, f_1, f_2)$ was comprehensively measured from a cohort of 20 normal-hearing human subjects (24) with a purpose to recommend the optimal choice of $L_1 = L_1^{\text{opt}}$ that maximizes L_{DP} given L_2 . When L_1 increases beyond L_1^{opt} , L_{DP} starts to decrease due to two-tone suppression (25). The same phenomenon has also been reproduced *in silico* by simulation of cochlear mechanics (23). In this section, we report on how differently the IMD3 level depends on the parameters, and hence devise a way to suppress IMD3 by considering two sets of primary-tone level (L_1, L_2) jointly.

2.2.1. Growth Function of IMD3

In contrast to cochlear mechanics, the loudspeaker nonlinearity does not demonstrate two-tone suppression; for instance, **Figure 2** shows IMD3 level as a function of (L_1, L_2) with $f_1 = 1,000$ Hz and several different ratios f_2/f_1 . The results were

obtained by delivering the primary tones to one of the two loudspeakers of a DPOAE probe (see section 2.4 for details) and measuring the response inside a syringe of approximately 2.0 cc. For any fixed L_2 , as L_1 increases, we do not find a clear L_1^{opt} beyond which L_{IMD3} starts to decrease, and this is quite unlike what was observed in human subjects with the most commonly used frequency ratio $f_2/f_1 \approx 1.2$ (24, **Figure 1**).

The contour plot of IMD3 also differs from that of human DPOAE in the rate of growth with respect to L_1 and L_2 . In particular, when L_1 and L_2 increases proportionately as they move toward the top-right corner of the plot along the straight lines $L_1 = L_2 + 10$ or $L_1 = L_2$, the rate of growth of L_{IMD3} with respect to L_2 seems to be close to 1.0 dB/dB for a wide range of L_2 and across different f_2/f_1 ratio (see the “growth functions” in **Figure 3**). The slope of these growth functions are shown in **Figure 4**, and the path $L_1 = L_2$ happens to have the slope that is closest to 1.0 dB/dB across different primary-frequency ratios. In comparison, the average human DPOAE growth rate when $L_1 = L_2$ falls in the range of 0.3 – 0.5 if $f_2/f_1 \approx 1.2$ (24, **Figure 1**). By exploiting this difference between the growth function of loudspeaker IMD3 and human DPOAE, we present a method that suppresses the IMD3 level while partially retaining the DPOAE.

2.2.2. Signal Acquisition Protocol

To leverage the part of IMD3 growth function that is almost linear (i.e., 1 dB/dB slope), we can devise the following performance metric,

$$J = \frac{|\beta G_{\text{DP}}(P_1, P_2) - G_{\text{DP}}(\beta P_1, \beta P_2)|}{|\beta G_{\text{IMD3}}(P_1, P_2) - G_{\text{IMD3}}(\beta P_1, \beta P_2)|}, \quad (4)$$

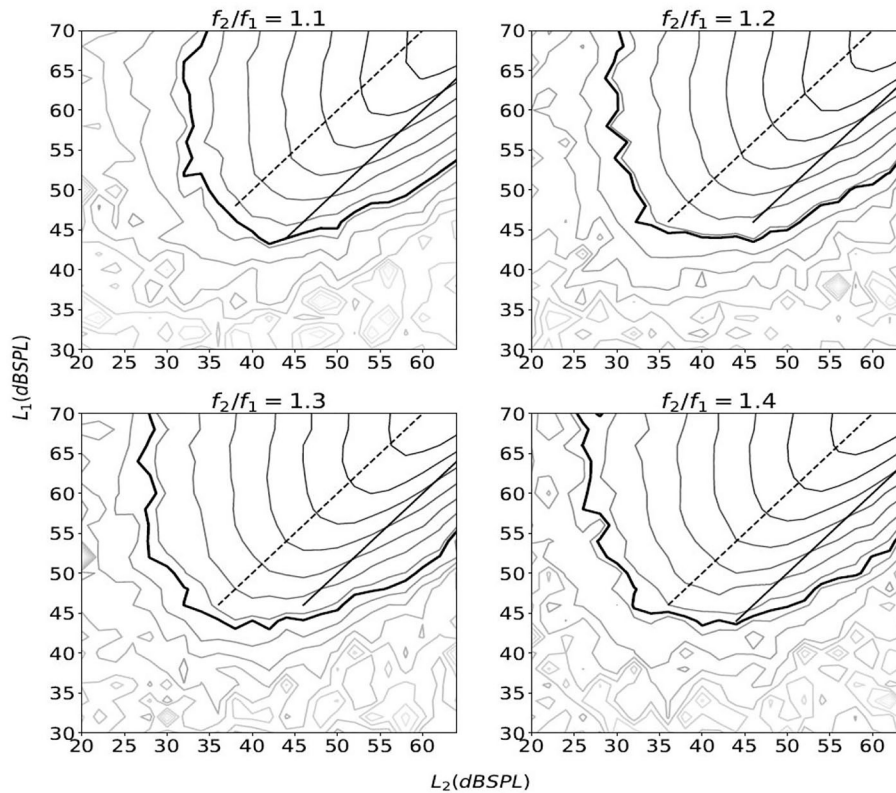


FIGURE 2 | The equal-level contour plots of IMD3 as a function of (L_1, L_2) . The thick line marks $L_{\text{IMD3}} = 0$ dB SPL, and other lines are in 4-dB steps. The IMD3 level were obtained by varying L_1 from 30 to 70 dB SPL and L_2 from 20 to 64 dB SPL in 2 dB steps. For each combination of (L_1, L_2) , the stimulus lasted for 1.0 s with a 2.5-ms raised cosine ramp for the rising and falling edges. The stimulation was repeated five times and the average IMD3 level is shown. The solid line and the dashed line represent $L_1 = L_2$ and $L_1 = L_2 + 10$ (dB), respectively.

where (P_1, P_2) denotes primary-tone sound pressure in Pa, $\beta > 1$ denotes a scaling factor, and G_{IMD3} and G_{DPOAE} denote the sound pressure of IMD3 and DPOAE in Pa, respectively. Conceptually, the goal of the compensation strategy is to choose (P_1, P_2) and β such that J is maximized. However, since the numerator (referred to as *DPOAE residual*) would vary among individuals, we seek to minimize the denominator in J , referred to as the *IMD3 residual*. Based on the results shown in **Figure 4**, we selected $L_1 = L_2$ (i.e., $P_1 = P_2$) for the remaining parts of this paper. As implied by **Figure 3**, choosing L_1 anywhere between 45 dB to 65 dB SPL should work well in reducing the IMD3 residual since $G_{\text{IMD3}}(\beta P_1, \beta P_2) \approx \beta G_{\text{IMD3}}(P_1, P_2)$. In contrast, we expect a larger proportion of DPOAE would remain in the DPOAE residual because the rate of growth against (P_1, P_2) is sub-linear (i.e., < 1 dB/dB).

Based on the above-mentioned concept, we propose the following signal acquisition protocol.

- Step 1: Calibrate the stimulus levels A_1 and A_2 in Equation (1) such that $P_1 = P_2$ in the ear canal.
- Step 2: Transform the recorded signal to the frequency domain, and calculate the magnitude at $2f_1 - f_2$, which is the vector sum of IMD3 and DPOAE. Denote the result as $Y(P_1, P_2)$.

- Step 3: Repeat Step 2 with increased primary-tone levels βP_1 and βP_2 . Denote the result as $Y(\beta P_1, \beta P_2)$.
- Step 4: Calculate the magnitude of residual at $2f_1 - f_2$, defined as $|\beta Y(P_1, P_2) - Y(\beta P_1, \beta P_2)|$.

To summarize, the goal of the signal acquisition protocol is to keep a large portion of DPOAE in the residual while maximally suppressing IMD3 at the same time.

2.3. The Cancellation Strategy

To describe the cancellation strategy, since digital adaptive filtering techniques are involved, we change the time variable from t to the integer index n (not to be confused with the index n in Equation 3). We follow a standard digital signal processing notation in defining $y[n] = y(nT)$ (26), where $y(t)$ is a continuous-time signal, T denotes the sampling period, and $y[n]$ denotes the result after sampling in time.

The idea behind this strategy is to cancel IMD3 instantaneously. To achieve this goal, we utilize two techniques, namely a phase-controlled exponential swept-sine chirp and the one-dimensional Volterra filters (ODVFs), to adjust the input signal before sending it to the speaker. The workflow is shown in **Figure 5** where $H(\omega)$ denotes a frequency

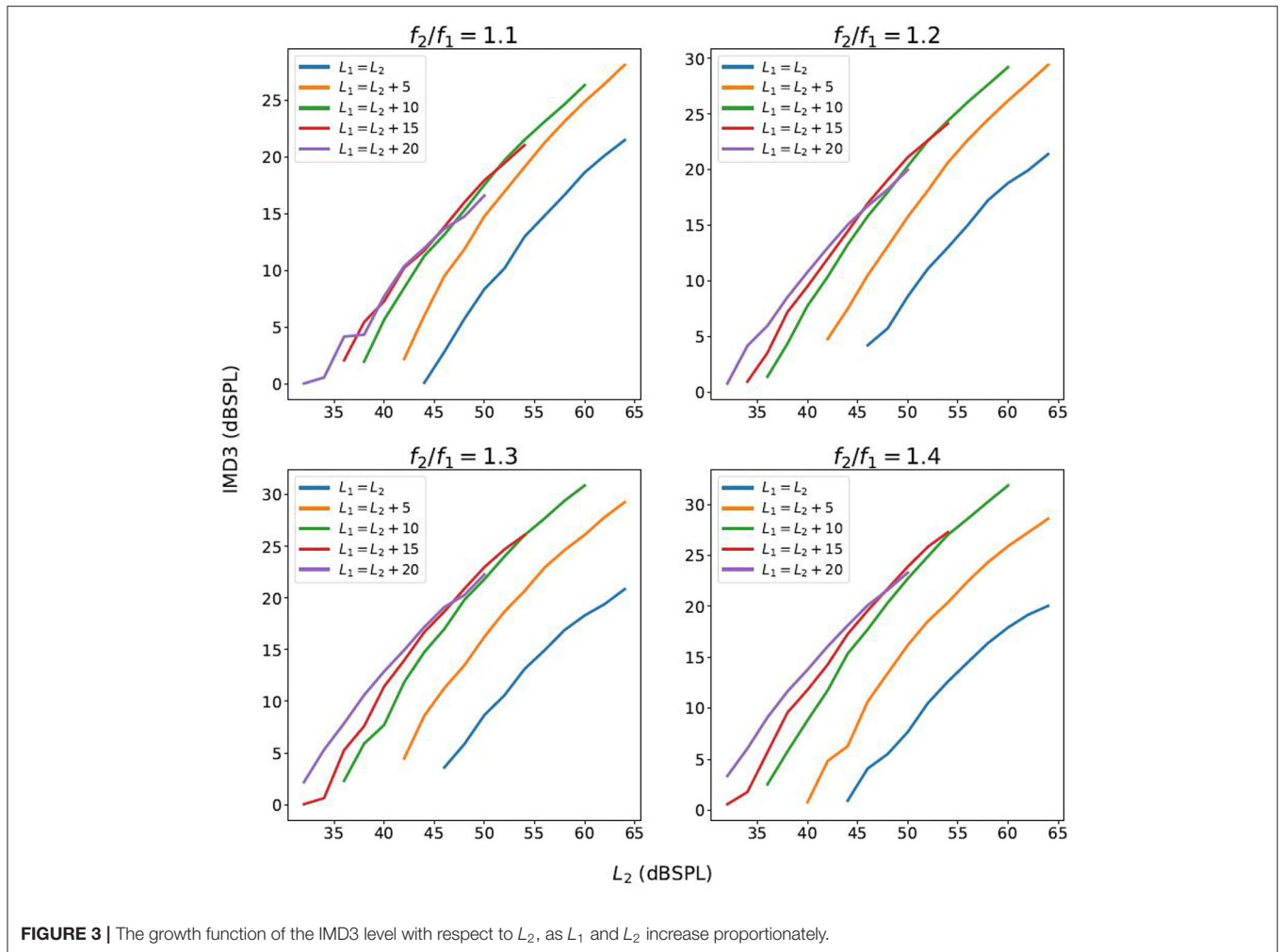


FIGURE 3 | The growth function of the IMD3 level with respect to L_2 , as L_1 and L_2 increase proportionately.

response measured by the phase-controlled exponential swept-sine chirp (to be described in section 2.3.1), and the normalization factor ensures that the input to ODVF is limited to the range $[-1, 1]$. The input signal which contains two pure tones for DPOAE measurement is first transformed from time domain to frequency domain through DFT, then transformed back to time domain after multiplying with the frequency response of the single speaker $H(\omega)$. The filter coefficients of ODVF are meant to be obtained *offline* by adaptive LMS algorithm (to be described in section 2.3.2); when applied online, the ODVF filter coefficients are fixed. Subsequently, the multiplication by $H^{-1}(\omega)$ is to compensate the gain and phase change due to multiplication by $H(\omega)$. This workflow produces an adjusted input signal to be delivered to the single speaker for the purpose of measuring DPOAE.

The rationale of this design is to artificially generate intermodulation distortions by the ODVF filter but with an inverted polarity, so they can cancel the real intermodulation distortion generated by the loudspeaker. We shall see in section 2.3.2 that the ODVF is trained off-line to emulate the situation of using two separate loudspeakers.

2.3.1. Linear System Estimation by Phase Controlled Exponential Swept-Sine Chirp

A phase-controlled exponential swept-sine chirp (27) is used to obtain the linear coupling response from the loudspeaker to the microphone. This chirp exhibits an instantaneous frequency that increases exponentially with time as follows,

$$s[n] = A \sin \left(\frac{\pi L}{2^Q \ln(2^Q)} \left(2^{\frac{Q}{N}} \right)^n \right), \quad (5)$$

where A is the amplitude of the sine wave, Q is an integer number of octaves, L is the ideal chirp length, and N is the real chirp length which is L rounded to the nearest integer. We also need an inverse chirp to convolve with, which can be expressed as

$$s^{-1}[n] = \frac{Q \ln 2}{A^2(1 - 2^{-Q})} \cdot s[N - n] \left(2^{\frac{Q}{N}} \right)^{-n}. \quad (6)$$

The result of convolving $s[n]$ and $s^{-1}[n]$ is approximately a Dirac delta impulse, and the gain is 0 dB for all frequencies.

We set the chirp's frequency to glide from 47 to 24,000 Hz and the chirp length was 10.5 s. Then, we delivered the chirp to

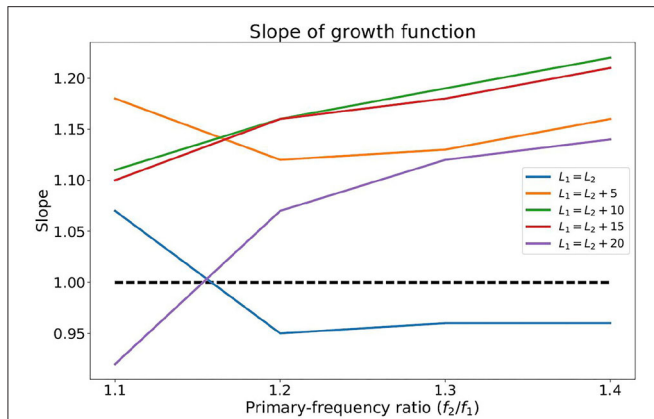


FIGURE 4 | The slope of the growth functions of the IMD3 shown in **Figure 3** with various primary-frequency ratios and primary-level differences.

drive the single speaker and recorded the sound simultaneously. Subsequently, the recorded signal was convolved with the inverse chirp (Equation 6) to obtain the impulse response $h[n]$ that characterizes the linear coupling from the speaker to the microphone. The Fourier transform of $h[n]$ is denoted as $H(\omega)$ in **Figure 5**.

2.3.2. IMD3 Cancellation by One-Dimensional Volterra Filtering

Since the intermodulation is due to loudspeaker nonlinearity, the behavior of its inverse system can ideally be characterized by Volterra series expansion (28). However, the full-scale Volterra series expansion requires estimate of multi-variate kernel functions which may be computationally impractical to implement and its estimation might be slow in convergence. In this research, we adopted a simplified version called one-dimensional Volterra filters (ODVF) (22) — assume that the inverse system can be modeled as follows,

$$y[n] = \sum_{i=0}^{M_1-1} h_1[i]x[n-i] + \sum_{i=0}^{M_2-1} h_2[i](x[n-i])^2 + \dots + \sum_{i=0}^{M_p-1} h_p[i](x[n-i])^p + \dots \quad (7)$$

where $x[n]$ and $y[n]$ denote the input and output of the inverse system, respectively, $h_p[i]$ denotes the p th-order kernel of ODVF, and M_p is the length of the p th-order kernel. Since this research focuses on canceling a cubic distortion, a partial ODVF retaining only the 1st-order and 3rd-order kernels was used; that is,

$$y[n] \approx \sum_{i=0}^{M_1-1} h_1[i]x[n-i] + \sum_{i=0}^{M_3-1} h_3[i](x[n-i])^3. \quad (8)$$

The filter coefficients could be obtained by the adaptive least mean square (LMS) method (29). The LMS method involves updating the filter coefficients constantly as the time index n

proceeds. First, denote the filter coefficients in the following vector form,

$$\mathbf{h}_{1,n} = (h_1[0], h_1[1], \dots, h_1[M_1 - 1])^T \in \mathbb{R}^{M_1},$$

and similarly,

$$\mathbf{h}_{3,n} = (h_3[0], h_3[1], \dots, h_3[M_3 - 1])^T \in \mathbb{R}^{M_3}.$$

Note that the subscript n indicates that both vectors are updated as n increases. Then, an approximation error signal can be defined as follows,

$$e[n] = d[n] - y[n] \approx d[n] - \mathbf{h}_{1,n}^T \mathbf{x}[n] - \mathbf{h}_{3,n}^T \mathbf{x}_3[n], \quad (9)$$

where $d[n]$ is a desired signal to be defined shortly, and vectors $\mathbf{x}[n]$ and $\mathbf{x}_3[n]$ are defined as follows,

$$\mathbf{x}[n] = (x[n], x[n-1], \dots, x[n-M_1+1])^T \in \mathbb{R}^{M_1}$$

and

$$\mathbf{x}_3[n] = ((x[n])^3, (x[n-1])^3, \dots, (x[n-M_3+1])^3)^T \in \mathbb{R}^{M_3}.$$

Finally, the update equations are given as follows,

$$\mathbf{h}_{1,n+1} = \mathbf{h}_{1,n} + \alpha e[n] \mathbf{x}[n], \quad (10)$$

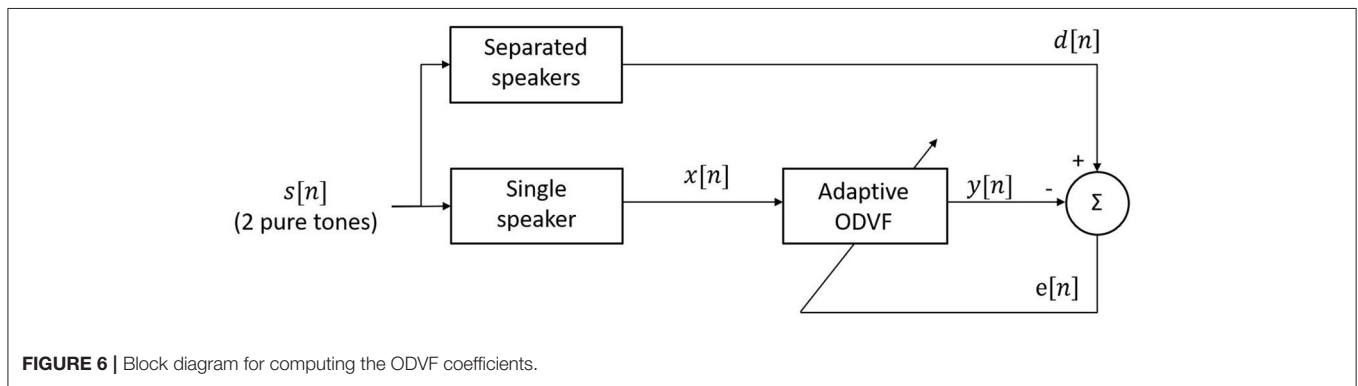
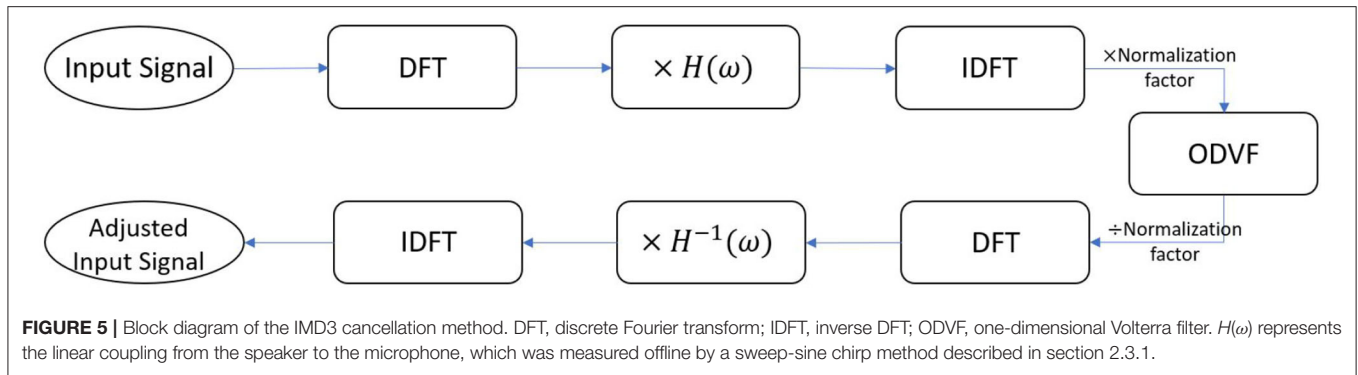
and similarly,

$$\mathbf{h}_{3,n+1} = \mathbf{h}_{3,n} + \alpha e[n] \mathbf{x}_3[n], \quad (11)$$

where $\alpha = 6 \times 10^{-3}$ denotes a stepsize that was chosen empirically.

Figure 6 shows how the ODVF coefficients were obtained in this research. The stimulus $s[n]$ contained two primary tones, and we set $d[n]$, the desired signal, to be the signal recorded by sending the two tones to separate speakers of a reliable reference probe, so $d[n]$ was free of IMD. Meanwhile, the input $x[n]$ to the ODVF was the signal recorded by using one single speaker while $y[n]$ denotes the output of the ODVF. Here, we emphasize that the adaptive procedure was performed offline just for one time, and it was not necessary to repeat the procedure when measuring DPOAE from individual ears. In practice, we first recorded $x[n]$ and $d[n]$ separately in the same 2-cc syringe with the same stimulus $s[n]$. Then, the filter coefficients $\mathbf{h}_{1,0}$ were initialized at 1 and $\mathbf{h}_{3,0}$ were initialized at 0, and we computed the updates iteratively according to Equations (9–11). After the filter coefficients converged, we could expect that the variance of $e[n]$ would be minimized and $y[n]$ should approximate $d[n]$, which is a cubic distortion-free signal, in a stochastic sense.

In practice, we found that the 1st-order kernel $h_1[n]$ tended to converge to a band-pass filter around the primary-tone frequencies which simply rejected all the intermodulation components linearly. This caused the “training” of $\mathbf{h}_{3,n}$ to fail. Therefore, we set the 1st-order kernel length M_1 to 1 to ensure that $h_3[n]$ learns to cancel the cubic distortion.



2.4. Equipment

All recordings were collected using a Python script that controls the ER-10C DPOAE probe-microphone (Etymotic Research) system via a 24-bit soundcard (Fireface UFX II, RME). The sampling frequency was set to 48 kHz. All recordings were done in a sound-proof room with the noise floor of approximately 19-21 dB SPL (30).

2.5. Human Subjects

Twelve subjects between age 22 and 32 participated in the research, including 8 males and 4 females. The compensation strategy was tested on the data from 5 of the subjects, while the cancellation strategy was tested on the data from 7 subjects. All the subjects did not have ear infection or report any hearing problems at the time of experiment. The recruitment of human subjects was approved by the IRB of National Tsing Hua University (No. 10912HE101).

3. RESULTS

Here we report the efficacy of applying the compensation and cancellation strategies.

3.1. The Compensation Strategy

We first tested the signal acquisition protocol in a 2-cc syringe. The parameters being tested were $\beta = 1.5, 2.0$ or 3.0 , and $P_1 = P_2 = 6.3, 11.3$, and 20.0 mPa, which correspond to 50, 55, and 60 dB SPL, respectively. The primary frequencies were $f_1 = 1,000$

TABLE 1 | The suppression index K for different combination of parameters.

	$L_1 = L_2$ (dB SPL)		
	50	55	60
$\beta = 1.5$	21.7	21.9	22.2
$\beta = 2.0$	16.2	15.9	16.5
$\beta = 3.0$	12.0	13.1	12.4

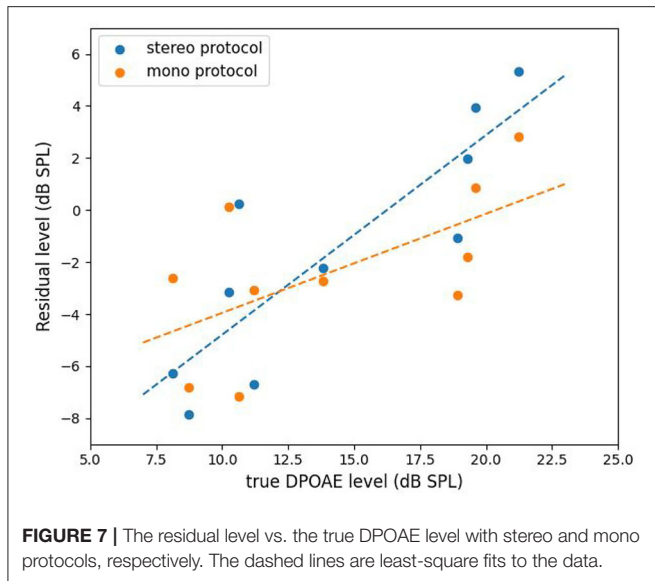
and $f_2 = 1,200$ Hz. We quantify the performance of IMD3 suppression by the following index,

$$K = 20 \log_{10} \left| \frac{G_{\text{IMD3}}(P_1, P_2)}{\beta G_{\text{IMD3}}(P_1, P_2) - G_{\text{IMD3}}(\beta P_1, \beta P_2)} \right|. \quad (12)$$

Here, K is just the IMD3 to IMD3 residual ratio in dB scale.

The values K for different combinations of β and L_1 are listed in **Table 1**. Note that K is quite insensitive to change in L_1 , while $\beta = 1.5$ gives the highest K . Therefore, $\beta = 1.5$ was chosen for testing in the ear. Also, $L_1 = L_2 = 60$ dB SPL was selected in order to maximize DPOAE and its residual.

DPOAEs residuals were recorded from five subjects for their left and right ears using the compensation strategy. Two protocols were considered. The first one is called the “stereo” protocol which uses separate speakers to obtain the ground truth of DPOAE and the residual thereof after applying the compensation strategy. The second is called “mono” protocol and it uses a single speaker mentioned in section 2.2.2 to obtain



the DPOAE residual subject to the IMD3 interference. The residual obtained with the stereo protocol should be regarded as a performance upper-bound for the compensation strategy since the IMD3 component is negligible when both speakers are used. We then study the correlation between the residual level obtained by both protocols and the true DPOAE level to evaluate the effectiveness of the compensation strategy.

The results of mono vs. stereo protocols are plotted in **Figure 7**. The x-axis is the true DPOAE level obtained with two speakers, and the y-axis represents the DPOAE residual levels obtained with the two protocols. The dashed lines show the results of linear regression. For the stereo protocol, the regression line is $y = 0.77x - 12.46$, and for the mono protocol, the regression line is $y = 0.38x - 7.76$. By using the stereo protocol, the correlation between DPOAE residual and true DPOAE level, both in dB SPL, equals to 0.86 with a high significance ($p < 0.001$). However, the DPOAE residual levels were 15 dB lower than the real DPOAE levels in average. This indicates that the DPOAE was suppressed by about 80% after applying the compensation strategy. Note that, nevertheless, the IMD3 component was suppressed by 22 dB under the same settings (see **Table 1**). Thus, we can say that the compensation strategy improved the DPOAE to IMD3 ratio by 7 dB in average. By using the mono protocol, the correlation between the residual and the true DPOAE level is lowered to 0.62. Nevertheless, with a $p < 0.01$, the correlation is deemed significant for this particular set of data, in the sense that the null hypothesis (no correlation) is rejected.

3.2. The Cancellation Strategy

We obtained different sets of ODVF coefficients at $f_1 = 1,000, 1,200, 1,500, \text{ and } 2,000$ Hz, respectively, while $f_2/f_1 = 1.2$ was fixed. The amplitude of both tones was set to 20 mPa (60 dB SPL), and the recordings ran for 10.5 s with 2.5 ms raised cosine ramp for the rising and falling edges of the stimulation. Empirically, the recording time was sufficiently long to ensure convergence

of the filter coefficients. **Figure 8** shows the resulting 3rd-order filter coefficients $h_3[n]$ for $f_1 = 1,000$ and 2,000 Hz, respectively. These $h_3[n]$ coefficients were subsequently used for validating the proposed cancellation strategy depicted in **Figure 5**.

Following the workflow described in **Figure 5**, we first tested the adjusted input signal in the 2-cc syringe. The results are shown in **Figure 9**. **Figure 9A** is the signal recorded by using a single speaker to play two primary tones. Therefore, the spectrum contains intermodulation distortion. **Figure 9B** is the signal recorded by using separated speakers and it therefore does not contain intermodulation distortion, and **Figure 9C** is the signal by using the adjusted input signal produced by the workflow of **Figure 5**. The black dots represent the magnitude of primary tones f_1 and f_2 , the red dot represents the magnitude of the IMD3. All the three recordings ran for 2.5 s with 2.5 ms raised cosine ramp for the rising as well as falling edges of the stimulation. The result shows that IMD3 is largely reduced to submerge below the noise floor. Note that the $2f_2 - f_1$ component at 1400 Hz is also suppressed, though not as perfectly as at $2f_1 - f_2$. The fifth-order distortions at 600 and 1600 Hz remain unchanged.

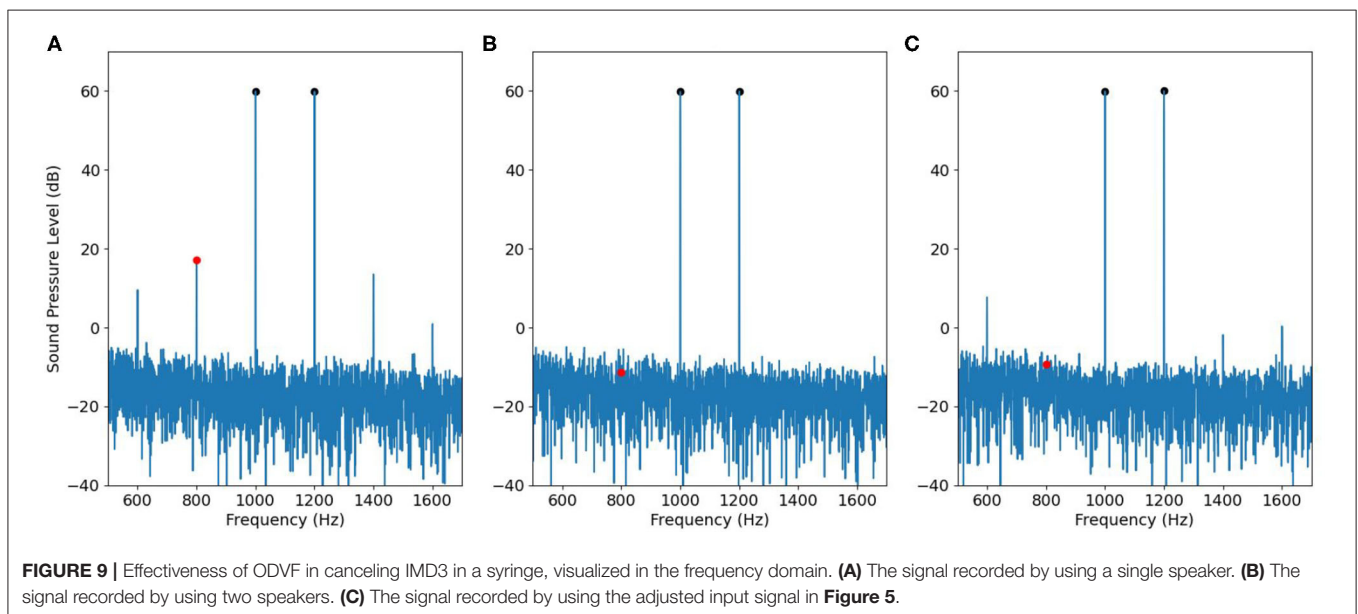
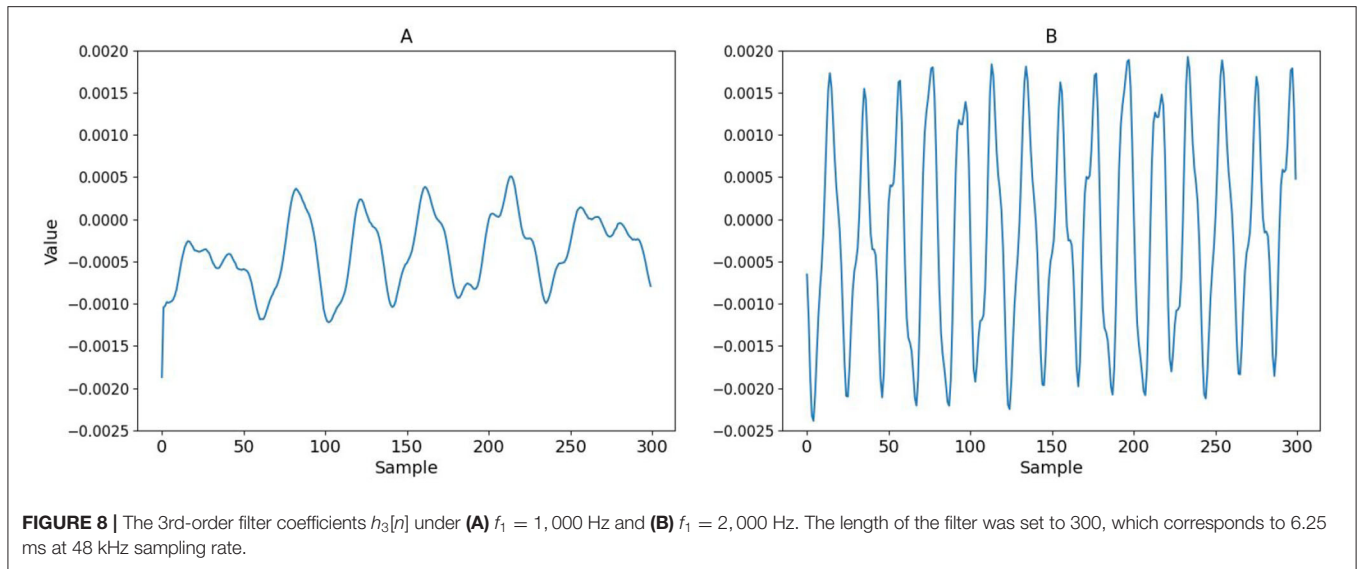
Then we applied the same recording procedure in human ears. **Figures 10, 11** show some typical results. **Figure 10A** contains DPOAE interfered with the original IMD3, **Figure 10B** shows the ideal signal recorded by using separate speakers for the primary tones, and the results are regarded as the true DPOAE signal to compare against. **Figure 10C** is the signal recorded by using the adjusted input signal described in **Figure 5**; the signal contains DPOAE interfered with the remaining IMD3.

We also extended the experiments to $f_1 = 1,200, 1,500, \text{ and } 2,000$ Hz while $f_2/f_1 = 1.2$. The results without (Figures 12A,C) and with (Figures 12B,D) the IMD3 cancellation strategy are shown in **Figure 12**. In **Figures 12A,B**, the horizontal axis is the true DPOAE level, and the vertical axis is the magnitude at f_{DP} . In **Figures 12C,D**, the horizontal axis is the true DPOAE phase, and the vertical axis is the phase measured at f_{DP} . The linear regression lines are also shown for visualization.

Without applying IMD3 cancellation strategy, the regression lines of the magnitude and the phase are $y = 0.12x + 19.87$ and $y = 1.03x - 0.25$, respectively, and the root mean square errors (RMSE) are 15.52 dB and 1.50 rad, respectively. After applying IMD3 cancellation strategy, the regression lines of the magnitude and the phase become $y = 0.76x + 2.64$ and $y = 1.02x - 0.06$, respectively, and the RMSE are reduced to 3.88 dB and 0.76 rad, respectively.

After applying the IMD3 cancellation strategy, we calculated the correlation between the estimated and the true DPOAE sound pressure levels, and the correlation between the estimated and the true DPOAE phase, respectively. Subsequently, we ran the Wald test with t-distribution to evaluate the significance of correlation. The correlation between the estimated and true DPOAE levels equals to 0.81 with a high significance ($p < 0.001$); also, the correlation between the estimated and the true DPOAE phase equals to 0.93 with a high significance ($p < 0.001$).

In contrast, without applying the IMD3 cancellation strategy, the correlation between the measured DPOAE and the true DPOAE sound pressure levels equals to 0.25 with a low



significance ($p = 0.057$). Although the phase estimation error is higher without applying the cancellation strategy, the cross-correlation between the measured and the true DPOAE phase was still high (0.80) and significant ($p < 0.001$).

The number of data points turns out to be 55. We recruited 7 subjects, test both ears at 4 frequencies, resulting in 56 DPOAE magnitudes and phases. However, the DPOAE level of one of the ears was below the noise floor when $f_1 = 1,000$ Hz. So that single point of data was abandoned.

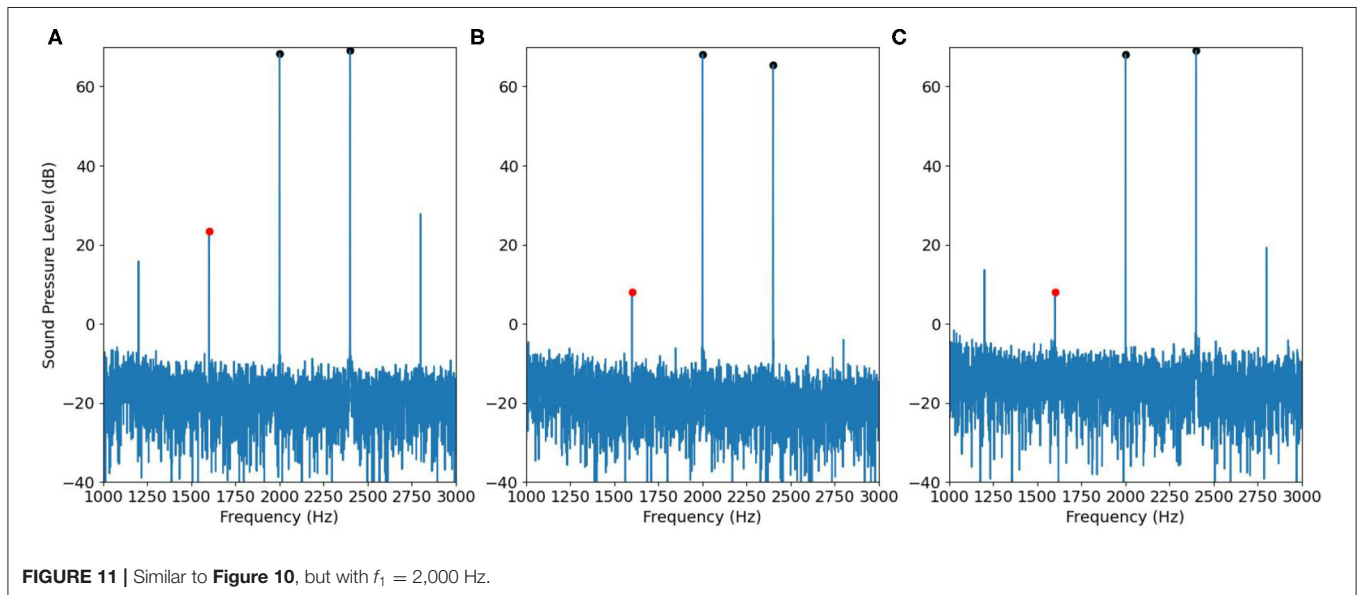
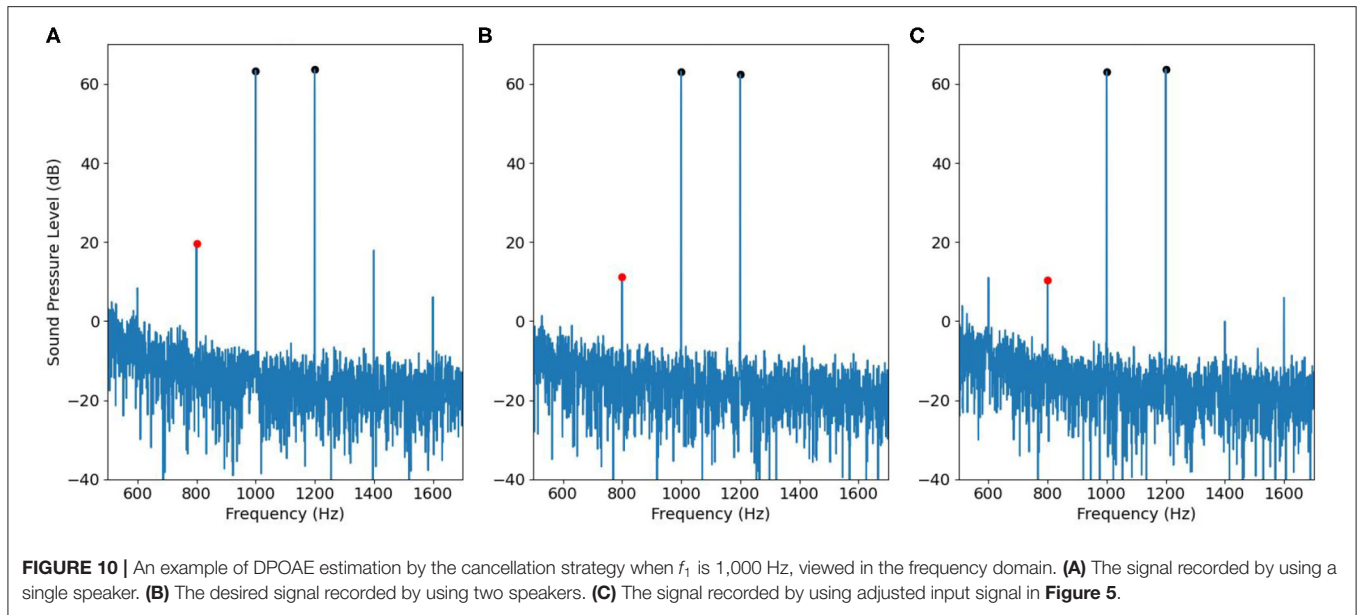
4. DISCUSSION

As true wireless, noise-cancellation earphones are gaining popularity in recent years, the ear canal also becomes an

over-booked space for various body sensors to enter and make the earphone intelligent (31, 32). Since active noise-cancellation earphones are indeed equipped an internal microphone¹, there is no reason why the microphone cannot measure DPOAE. The main factor that hinders such application might be the interference due to loudspeaker IMD. As much as we are aware of, consumer earphones are not typically designed to have two speakers in one ear², so sending the primary tones to separate speakers would not be a choice. In this research, we demonstrated

¹For example *AirPods Pro* and *AirPods Max Active Noise Cancellation and Transparency mode*, <https://support.apple.com/en-us/HT210643>.

²For example “Qualcomm QCC3056 is an ultra-low power, single-chip solution, optimized for use in truly wireless earbuds and hearables.” The chip features mono audio playback. <https://www.qualcomm.com/products/qcc3056>.



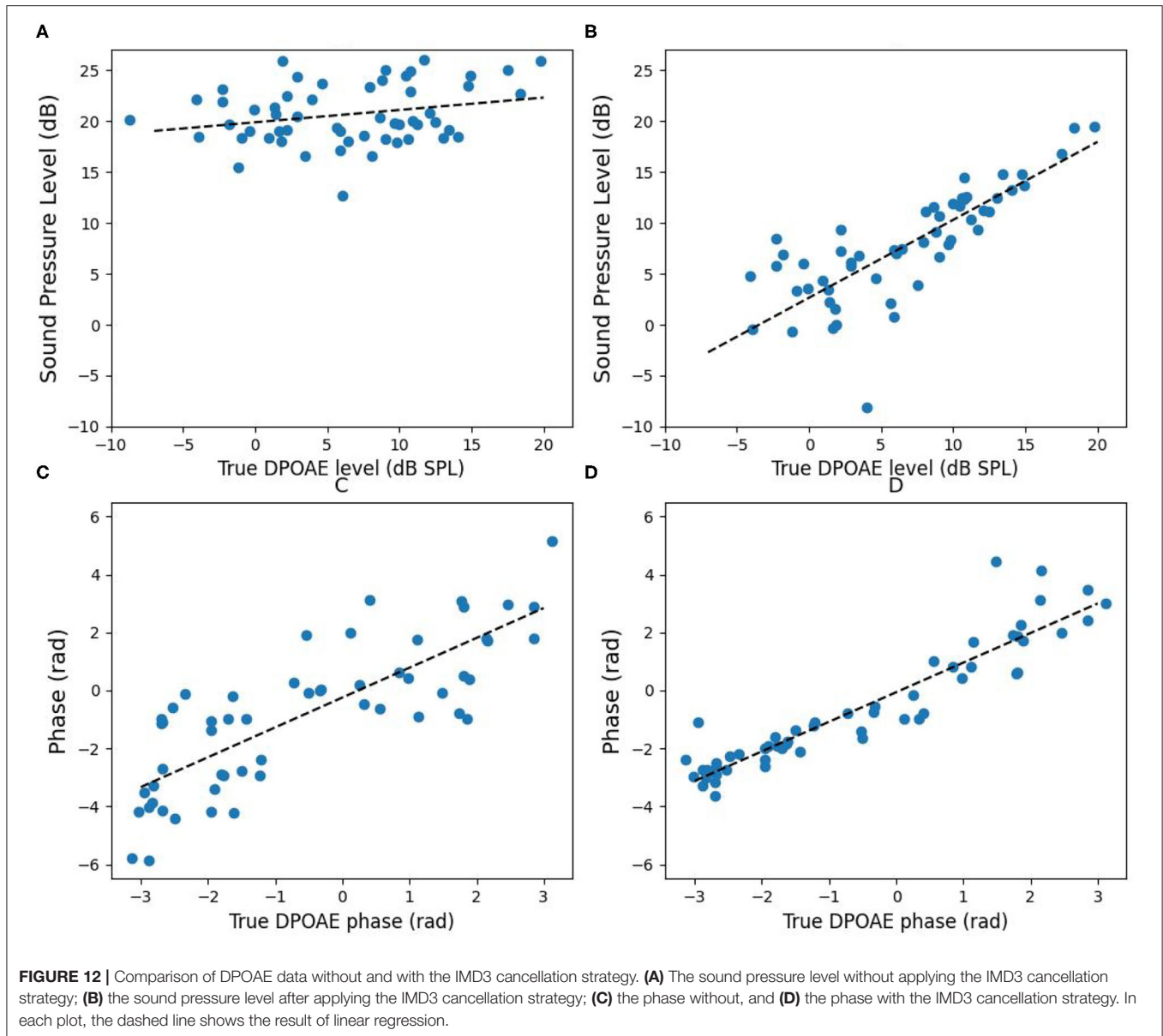
that “wrongly” using one single loudspeaker to play the f_1 and f_2 tones may still work as long as we can cancel the IMD3 it generates. Hence, we hope that this research can serve as a feasibility study for the earphone industry to promote DPOAE as a service to consumers of active noise cancellation earphones. We envision that making DPOAE available at home could also enrich any remote hearing care program in the future.

Among the two proposed strategies, cancellation outperforms compensation in producing a more accurate prediction of the true DPOAE level. On one hand, the cancellation strategy achieves a higher cross-correlation to the true DPOAE level (0.81) than the compensation strategy (0.62); on the other hand, it also provides a direct estimate of the DPOAE magnitude and phase, instead of just a residual. It remains to be seen in the

future whether similar results would be obtained with a larger sample size.

The usage of Volterra filters also brings up many research questions. For instance, **Figure 8** shows that the 3rd-order function $h_3[n]$ are different for different choices of (f_1, f_2) . Presently, we are uncertain whether (a) this is a limitation due to omitting all the off-diagonal elements of the Volterra filter to make it one-dimensional, or (b) it is possible to apply certain transformation akin to pre-whitening so the adaptive system eventually “learns” a universal ODVF for canceling all the cubic distortions given any input signal. Apparently, there is still much room to explore on this topic.

Meanwhile, as much as the compensation strategy is concerned, it is surprising that we found a large region on the



(L_1, L_2) plane where the IMD3 level grows *quasi-linearly* when L_1 and L_2 increase proportionately. In practice, it may be interesting to see if a similar property can be observed in other DPOAE probes or consumer earphones. We speculate that the quasi-linear growth is an epiphenomenon because cubic distortion is by nature a third-order component. Under the light of Taylor expansion in Equation (2), we can expect IMD3 to demonstrate a 3 dB/dB growth when the stimulus is at low intensity—and so does DPOAE (23). As the intensity of η increases, higher order terms $G^{(k)}(0)\eta^k/k!$ begin affecting the growth function. In particular, all the odd-order terms should jointly reduce the slope of growth of $B_{2,-1}$ and account for its saturation (it is straightforward to show that the even-order terms do not

contribute to $B_{2,-1}$). So the fact that we observe nearly 1 dB/dB growth at the working range of (L_1, L_2) may just be a coincidence.

Some other techniques might be worth consideration for estimating the DPOAE level under IMD3 interference. As mentioned in section 1, DPOAE itself has two components — the direct one and the reflection. Based on the difference in latency, Vetešník et al. (21) designed short pips to elicit DPOAE and applied envelope tracking techniques to *separate* the two components. If the loudspeaker distortion is generated within a shorter time before DPOAE emerges, one might be able to identify an early peak by tracking the instantaneous amplitude at the $2f_1 - f_2$ frequency. Thus, the loudspeaker IMD3 and DPOAE might be separated in the time domain. This may

require careful re-thinking of the stimulus design and is left for future exploration.

5. CONCLUSIONS

We proposed two strategies to estimate the DPOAE level subject to interference from the loudspeaker IMD3. The compensation strategy was designed to suppress IMD3 based on its quasi-linear growth with respect to primary-tone levels, in contrast to DPOAE's sub-linear growth. Results show that, although the DPOAE level was also suppressed by 80%, the residual level correlates to the true DPOAE level. The cancellation strategy adjusted the input signal nonlinearly to emulate distortion-free stimulation. It thus recovered both the DPOAE magnitude and phase directly. Overall, this research suggests that it might be feasible to use a single-loudspeaker probe to measure DPOAE. Testing with a larger sample of human subjects as well as various types of earphones shall ensue to evaluate whether commercial noise-cancellation earphones could be utilized to allow sufficiently accurate DPOAE measurement at home.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Kemp DT. Stimulated acoustic emissions from within the human auditory system. *J Acoust Soc Am.* (1978) 64:1386–91. doi: 10.1121/1.382104
- Shera CA, Guinan JJ Jr. Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs. *J Acoust Soc Am.* (1999). 105:782–98. doi: 10.1121/1.426948
- Zurek PM, Clark WW, Kim DO. The behavior of acoustic distortion products in the ear canals of chinchillas with normal or damaged ears. *J Acoust Soc Am.* (1982) 72:774–80. doi: 10.1121/1.388258
- Lonsbury-Martin B, Martin G. The clinical utility of distortion-product otoacoustic emissions. *Ear Hear.* (1990) 11:144–54. doi: 10.1097/00003446-199004000-00009
- Harris F, Lonsbury-Martin B, Stagner B, Coats A, Martin G. Acoustic distortion products in humans: systematic changes in amplitude as a function of f2/f1 ratio. *J Acoust Soc Am.* (1989) 85:220–9. doi: 10.1121/1.397728
- Kanis LJ, de Boer E. Two-tone suppression in a locally active nonlinear model of the cochlea. *J Acoust Soc Am.* (1994) 96:2156–65. doi: 10.1121/1.410157
- Barker SE, Lesperance MM, Kileny PR. Outcome of newborn hearing screening by ABR compared with four different DPOAE pass criteria. *Amer J Audiol.* (2000) 9:142–8. doi: 10.1044/1059-0889(2000/017)
- Sanford CA, Keefe DH, Liu YW, Fitzpatrick D, McCreery RW, Lewis DE, et al. Sound-conduction effects on DPOAE screening outcomes in newborn infants: test performance of wideband acoustic transfer functions and 1-kHz tympanometry. *Ear Hear.* (2009) 30:635–52. doi: 10.1097/AUD.0b013e3181b61cdc
- Brown DJ, Gibson WP. On the differential diagnosis of Ménière's disease using low-frequency acoustic biasing of the 2f1-f2 DPOAE. *Hear Res.* (2011) 282:119–27. doi: 10.1016/j.heares.2011.09.002
- Shiomi Y, Tsuji J, Naito Y, Fujiki N, Yamamoto N. Characteristics of DPOAE audiogram in tinnitus patients. *Hear Res.* (1997) 108:83–8. doi: 10.1016/S0378-5955(97)00043-9

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethic Committee, National Tsing Hua University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YCC discovered the almost linear growth in IMD3, and came up with the compensation strategy. WCH tested the compensation strategy, and conceived the cancellation strategy using ODVF and then implemented and tested the cancellation strategy. YWL supervised the research and is responsible for the manuscript preparation process. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Ministry of Science and Technology of Taiwan (grant no. 107-2221-E-007-093-MY2).

ACKNOWLEDGMENTS

We thank Drs. Yu-Ting Lin and Tang-Chuan Wang for regular exchange of research ideas and offering critiques.

- Job A, Nottet JB. DPOAEs in young normal-hearing subjects with histories of otitis media: evidence of sub-clinical impairments. *Hear Res.* (2002) 167:28–32. doi: 10.1016/S0378-5955(02)00330-1
- Whitehead ML, Stagner BB, Lonsbury-Martin BL, Martin GK. Measurement of otoacoustic emissions for hearing assessment. *IEEE Eng Med Biol Mag.* (1994) 13:210–26. doi: 10.1109/51.281681
- Moir HM, Jackson JC, Windmill JFC. No evidence for DPOAEs in the mechanical motion of the locust tympanum. *J Exp Biol.* (2011) 214:3165–72. doi: 10.1242/jeb.056465
- Moir HM, Jackson JC, Windmill JFC. Response to “Measurement of sensitive distortion-product otoacoustic emissions in insect tympanal organs.” *J Exp Biol.* (2012) 215:567. doi: 10.1242/jeb.068098
- van der Heijden M, Joris PX. Cochlear phase and amplitude retrieved from the auditory nerve at arbitrary frequencies. *J Neurosci.* (2003) 23:9194–8. doi: 10.1523/JNEUROSCI.23-27-09194.2003
- Campbell LJ, Slater KD. *Personalization of Auditory Stimulus.* United States: 9497530B1 (2016).
- der Aerschoot MV, Swanepoel DW, Mahomed-Asmail F, Myburgh HC, Eikelboom RH. Affordable headphones for accessible screening audiometry: an evaluation of the sennheiser HD202 II supra-aural headphone. *Int J Audiol.* (2016) 55:616–22. doi: 10.1080/14992027.2016.1214756
- Kalluri R, Shera CA. Distortion-product source unmixing: a test of the two-mechanism model for DPOAE generation. *J Acoust Soc Am.* (2001) 109:622–37. doi: 10.1121/1.1334597
- Shaffer LA, Withnell RH, Dhar S, Lilly DJ, Goodman SS, Harmon KM. Sources and mechanisms of DPOAE generation: implications for the prediction of auditory sensitivity. *Ear Hear.* (2003) 24:367–79. doi: 10.1097/01.AUD.0000090439.16438.9F
- Zweig G, Shera CA. The origin of periodicity in the spectrum of evoked otoacoustic emissions. *J Acoust Soc Am.* (1995) 98:2018–47. doi: 10.1121/1.413320

21. Vetešník A, Turcanu D, Dalhoff E, Gummer AW. Extraction of sources of distortion product otoacoustic emissions by onset-decomposition. *Hear Res.* (2009) 256:21–38. doi: 10.1016/j.heares.2009.06.002
22. Mu Y, Ji P, Ji W, Wu M, Yang J. Modeling and compensation for the distortion of parametric loudspeakers using a one-dimension volterra filter. *IEEE/ACM Trans Audio Speech Lang Proc.* (2014) 22:2169–81. doi: 10.1109/TASLP.2014.2363414
23. Liu YW, Neely ST. distortion-product emissions from a cochlear model with nonlinear mechano-electrical transduction in outer hair cells. *J Acoust Soc Am.* (2010) 127:2420–32. doi: 10.1121/1.3337233
24. Johnson TA, Neely ST, Garner CA, Gorga MP. Influence of primary-level and primary-frequency ratios on human distortion product otoacoustic emissions. *J Acoust Soc Am.* (2006) 119:418–28. doi: 10.1121/1.2133714
25. Hall JL. Two-tone distortion products in a nonlinear model of the basilar membrane. *J Acoust Soc Am.* (1974) 56:1818–28. doi: 10.1121/1.1903519
26. Oppenheim AV, Schaffer RW. *Discrete-time Signal Processing, 3rd Edn.* Upper Saddle River, NJ: Pearson (2010).
27. Vetter K, di Rosario S. ExpoChirpToolbox: a pure data implementation of ESS impulse response measurement. In: *Proceeding Pure Data Convention*. Weimar (2011). Available online at: https://www.uni-weimar.de/kunst-und-gestaltung/wiki/images/ExpoChirpToolbox-_a_Pd_implementation_of_ESS_impulse_response_measurement.pdf
28. Lesiak C, Krener A. The existence and uniqueness of Volterra series for nonlinear systems. *IEEE Trans Autom Control.* (1978) 23:1090–5. doi: 10.1109/TAC.1978.1101898
29. Widrow B, Stearns S. *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall (1985).
30. Liu TC, Liu YW, Wu HT. Denoising click-evoked otoacoustic emission signals by optimal shrinkage. *J Acoust Soc Am.* (2021) 149:2659–70. doi: 10.1121/10.0004264
31. Venema B, Blanik N, Blazek V, Gehring H, Opp A, Leonhardt S. Advances in reflective oxygen saturation monitoring with a novel in-ear sensor system: results of a human hypoxia study. *IEEE Trans Biomed Eng.* (2012) 59:2003–10. doi: 10.1109/TBME.2012.2196276
32. Goverdovsky V, Looney D, Kidmose P, Mandic DP. In-Ear EEG From viscoelastic generic earpieces: robust and unobtrusive 24/7 monitoring. *IEEE Sens J.* (2016) 16:271–7. doi: 10.1109/JSEN.2015.2471183

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Hsiao, Chen and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.