Distortion-product otoacoustic emissions measured using synchronized swept-sines

Václav Vencovský[1], Antonin Novak[2], Ondřej Klimeš[1], Petr Honzík[1], Aleš Vetešník[3]

[1] Department of Radioelectronics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic

[2] Laboratoire d'Acoustique de l'Université du Mans (LAUM), UMR 6613, Institut d'Acoustique - Graduate School (IA-GS), CNRS, Le Mans Université, France

[3] Department of Nuclear Chemistry, Faculty of Nuclear Science and Physical Engineering, Czech Technical University in Prague, Prague, Czech Republic

*vaclav.vencovsky@gmail.com

Abstract

Swept-sines provide a tool for fast and high-resolution measurement of evoked otoacoustic emissions. During the measurement, a response to swept-sine(s) is recorded by a probe placed in the ear canal. Otoacoustic emissions can then be extracted by various techniques, e.g., Fourier analysis, the heterodyne method, the least-square-fitting (LSF) technique. This paper employs a technique originally proposed with exponential swept-sines, which allows for direct emission extraction from the measured intermodulation impulse response. It is shown here that the technique can be used to extract distortion-product otoacoustic emissions (DPOAEs) evoked with two simultaneous swept-sines. For proper extraction of the DPOAE phase, the technique employs previously proposed adjusted formulas for exponential swept-sines generating so-called synchronized swept-sines. Here, the synchronized swept-sine (SSS) technique is verified using responses derived from a numerical solution of a cochlear model and responses measured in human subjects. Although computationally much less demanding, the technique yields comparable results to those obtained by the LSF technique, which has been shown in the literature to be the most noise-robust among the emission extraction methods.

The archived file is not the final published version of the article Vencovský, V., Novak, A., Klimeš, O., Honzík, P. & Vetešník, A. Distortion-product otoacoustic emissions measured using synchronized swept-sines, J Acoust Soc Am, 153 (5): 2586, 2023

The definitive publisher-authenticated version is available online at https://doi.org/10.1121/10.0017976,

1. Introduction

Otoacoustic emissions are acoustical signals generated from within the inner ear (cochlea) [Kemp(1978), Probst *et al.*(1991)]. If the ear is stimulated with at least two tones with near frequencies f_1 and f_2, interference between the tones generates distortion-product otoacoustic emissions (DPOAEs) due to the non-linear basilar membrane (BM) response [Goldstein(1967), Rhode(1978), Johnstone *et al.*(1986), Robles and Ruggero(2001)]. DPOAE with frequency $f_{\mathrm{DP}} = 2f_1 - f_2$ is called cubic (low-side) DPOAE or cubic difference tone (CDT) DPOAE. DPOAEs may serve as an effective and frequency-specific tool for diagnosis of hearing loss, because their level and their estimated threshold correlate with hearing sensitivity, [**?**, e.g.,]]Gaskill1990, Nelson1992,Boege2002. However, the accuracy of hearing threshold prediction based on the DPOAE level or on an estimated threshold is affected by the interference between DPOAE components [Mauermann and Kollmeier(2004), Dalhoff *et al.*(2013), Zelle *et al.*(2017), Zelle *et al.*(2020)]. DPOAE is composed of two components which differ in their generation mechanism: a nonlinear-distortion component and a coherent-reflection component [Shera and Guinan(1999), Shaffer *et al.*(2003)]. Interaction between these two components causes a fine structure in DP-gram amplitude. A DP-gram shows the DPOAE amplitude and phase as a function of frequency [Kemp and Brown(1983), Brown *et al.*(1996), He and Schmiedt(1997)].

The nonlinear-distortion and coherent-reflection components of DPOAEs can be separated by various techniques. One way is to present a third tone near the DP frequency, which would suppress the component generated by coherent reflection [Kemp and Brown(1983), Heitmann *et al.*(1998)]. Another way is to measure the onset of the DPOAE signal, which is not affected by a long-latency, reflected component [Vetešník *et al.*(2009), Zelle *et al.*(2017)]. If a DP-gram is measured with sufficient frequency resolution, the nonlinear-distortion and the coherent-reflection component can be separated by the inverse Fourier transform of the DP-gram [Stover *et al.*(1996), Dhar *et al.*(2002)] or by time-frequency filtering methods [Moleti *et al.*(2012)].

Swept-sines, also called chirps, are a tool for measuring DP-grams with sufficient frequency resolution but within a relatively short time [Choi *et al.*(2008), Long *et al.*(2008)]. A signal recorded by an OAE probe during the measurement must be *post hoc* analyzed in order to extract the measured emission. [Kalluri and Shera(2013)] compared three different methods used to extract OAEs: a digital heterodyne method [Choi *et al.*(2008)], a method employing Fourier analysis [Kalluri and Shera(2001)], and a modeling technique using least-square fitting (LSF) [Long *et al.*(2008)]. Although more computationally demanding than, e.g., Fourier analysis, the LSF technique has been shown to outperform the other techniques due to its noise robustness [Kalluri and Shera(2013)].

In this paper, we present another method allowing for DPOAE extraction. The method is based on synchronized swept-sines, introduced by [Novak *et al.*(2015)] for analysis of nonlinear systems. Synchronized swept-sines are a special type of exponential, or sometimes called logarithmic, swept-sine signals [Novak *et al.*(2015)]. The synchronized swept-sine technique can be used for analysis of nonlinear systems in terms of block-oriented models, e.g. Generalized Hammerstein models, or Diagonal Volterra Series [Novak *et al.*(2010)]. However, its main advantage consists in separating the frequency-dependent higher harmonics from each other. The technique of [Novak *et al.*(2015)] is adapted here for the estimation of intermodulation distortion products (DPs). DPOAEs are intermodulation DPs, and the current paper presents the synchronized-swept sine (SSS) technique for extracting them. The SSS technique is easy to implement,computationally inexpensive, and is not dependent on the type of nonlinearity in the system. All this, together with the ability of the technique to separate DPOAE components of different latencies, makes the SSS technique a promising tool for use during DPOAE measurements. The paper shows that in terms of noise robustness, the SSS technique is comparable with the LSF technique.

The paper is organized as follows. In Sec. 2 the theoretical background of the application of the synchronized swept-sine for estimating DPOAE is provided. The SSS technique is then combined with a windowing method for background noise reduction and DPOAE component separation and is verified on simulations (Sec. 3). Then for experiments (Sec. 4), a method for sound artifact rejection is presented which was designed to be used with the

SSS technique and the windowing method. A summary of the advantages of the proposed method is discussed next in Sec. 5. All the scripts, including the simulated and experimental responses to synchronized swept-sines and implementation of the cochlear model that is used can be downloaded from https://gitlab.fel.cvut.cz/vencovac/Prj04_OAEsweptsine_measurement_public.

2. Synchronized swept-sine for DPOAE

In recent decades, the use of swept-sine signals has proven to be highly effective for the analysis and identification of nonlinear systems [Farina(2000)]. The recently developed synchronized swept-sine signal [Novak *et al.*(2015)] has unique properties that enable quick analysis of the amplitude and the phase of frequency-dependent distortion products (DPs). While the method is widely used for analysing higher harmonics, it can be easily adapted to intermodulation products. The remainder of this section introduces the theoretical background of adapting the synchronized swept-sine for DPOAE.

2.1 Synchronized swept-sine

A synchronized swept-sine signal is a special case of an exponential swept-sine defined as

$$s(t) = \sin\Big(\varphi(t)\Big), \tag{1}$$

with the phase

$$\varphi(t) = 2\pi f_a L \exp\left(\frac{t}{L}\right). \tag{2}$$

The coefficient L related to the sweep rate is

$$L = \frac{T}{\ln\left(\frac{f_b}{f_a}\right)}, \tag{3}$$

for which the instantaneous frequencies are f_a (the start frequency) and f_b (the stop frequency) at times $t = 0$ and $t = T$, respectively.

Such a swept-sine signal has a particular property related to the higher harmonics, i.e. frequencies that are positive integer multiples of the fundamental frequency. As demonstrated in [Novak *et al.*(2015)], multiplication of the phase $\varphi(t)$ by a positive integer m corresponds to the generation of the m-th harmonic of the swept-sine, and is also equivalent to a time shift of the phase by Δt_m

$$m\,\varphi(t) = \varphi(t - \Delta t_m), \tag{4}$$

where

$$\Delta t_m = -L \ln(m). \tag{5}$$

Therefore, we can define a higher-harmonic version of the synchronized swept-sine as

$$s_m(t) = \sin\Big(m\,\varphi(t)\Big) = \sin\Big(\varphi(t - \Delta t_m)\Big). \tag{6}$$

This property is very useful for estimating frequency-dependent higher harmonics [Novak *et al.*(2015)] and also, as demonstrated below, for estimating frequency-dependent intermodulation distortion.

2.2 Frequency-dependent harmonic distortion

When a pure sine wave of frequency f_0 is used as the input to a nonlinear system, higher harmonics may appear at the output of the system as multiples of the input frequency f_0, as shown in Fig. 1A. Each of these harmonics has an amplitude A_m and a phase ϕ_m, m indicating the index of the harmonic. In addition, these harmonics can be frequency-dependent, resulting in Higher Harmonic Frequency Responses (HHFR) $H_m(f) = A_m(f)e^{\phi_m(f)}$. The

corresponding Higher Harmonic Impulse Responses (HHIR) are $h_m(t) = \mathcal{F}^{-1}H_m(f)$, with $\mathcal{F}^{-1}$ being the inverse Fourier transform.[1]

Thanks to the properties of the synchronized swept-sine summarized above [see Eq. (6)], it can be used for measuring $h_m(t)$. The higher harmonics of the swept-sine are generated at the output of the nonlinear system, as shown in the spectrogram of Fig. 1B. For the input signal being a synchronized swept-sine, this can be described as a convolution sum

$$\begin{aligned} y(t) &= \sum_m h_m(t) * s_m(t) \\ &= \sum_m h_m(t) * \sin\Big(\varphi(t-\Delta t_m)\Big) \\ &= \sum_m h_m(t-\Delta t_m) * \sin\Big(\varphi(t)\Big) \\ &= s(t) * \sum_m h_m(t-\Delta t_m) \qquad (7) \\ &= s(t) * h(t) \qquad (8) \end{aligned}$$

where the operator $*$ stands for convolution, $y(t)$ is the output signal, and $h_m(t)$ is the HHIR corresponding to the m-th harmonic. Equation (8) can be expressed as a simple convolution $y(t) = s(t) * h(t)$, which is commonly used in linear system theory to represent the system's impulse response, $h(t)$. Therefore, $h(t)$ is considered as a "virtual" impulse response (in the remainder, we drop the "virtual").[2] The latter consists of a sum of delayed HHIRs $h_m(t)$ (Fig. 1C), which can be extracted from the measured impulse response $h(t)$ by windowing. Finally, the Fourier transform of each extracted HHIR $h_m(t)$ is the HHFR $H_m(f)$.

2.3 Frequency dependent intermodulation distortion

Two-tone intermodulation distortion is measured by applying two pure sine waves of frequencies f_1 and f_2 simultaneously to the input of a nonlinear system. In addition to harmonic distortion, the output signal also consists of intermodulation DPs at frequencies equal to the sums and differences of integer multiples of frequencies f_1 and f_2, denoted as follows

$$f_{m,n} = m\,f_1 + n\,f_2, \qquad (9)$$

where $m, n \in \mathbb{Z}$ (Fig. 2A). These components are called either harmonics, when $m = 0$ or $n = 0$, or intermodulation products, when both $m \neq 0$ and $n \neq 0$. For example, the frequency components $f_{m,0}$ correspond to the harmonics of f_1, i.e. $f_{1,0} = f_1$, $f_{2,0} = 2\,f_1$, etc., while the components $f_{0,n}$ correspond to the harmonics of f_2, i.e. $f_{0,1} = f_2$, $f_{0,2} = 2\,f_2$, etc. The intermodulation frequency component $2\,f_1 - f_2$ (CDT) is therefore noted as $f_{2,-1}$. By defining the ratio of the two input frequencies as

$$\alpha = \frac{f_2}{f_1} \qquad (10)$$

the frequency components present in the output signal can also be expressed with respect to f_1 as

$$f_{m,n} = f_1\,(m + \alpha n)\,. \qquad (11)$$

This indexing system is used in the remainder of the paper not only for the frequency components, but also to index the swept-sine signals and the impulse responses related to intermodulation products.

[1] As the terms impulse response and frequency response are associated with linear system theory, we use the prefix "Higher Harmonic" or "Intermodulation" in front of the terms "impulse response" and "frequency response" to distinguish between the entire measured impulse response using linear system theory (with no prefix) and the products appearing in the measured impulse response due to nonlinear systems (with a prefix).

[2] In contrast to linear system theory, such an impulse response is level dependent. However, because linear system theory is used to obtain it, the term "impulse responses" is commonly used with the swept-sine technique also for non-linear systems.

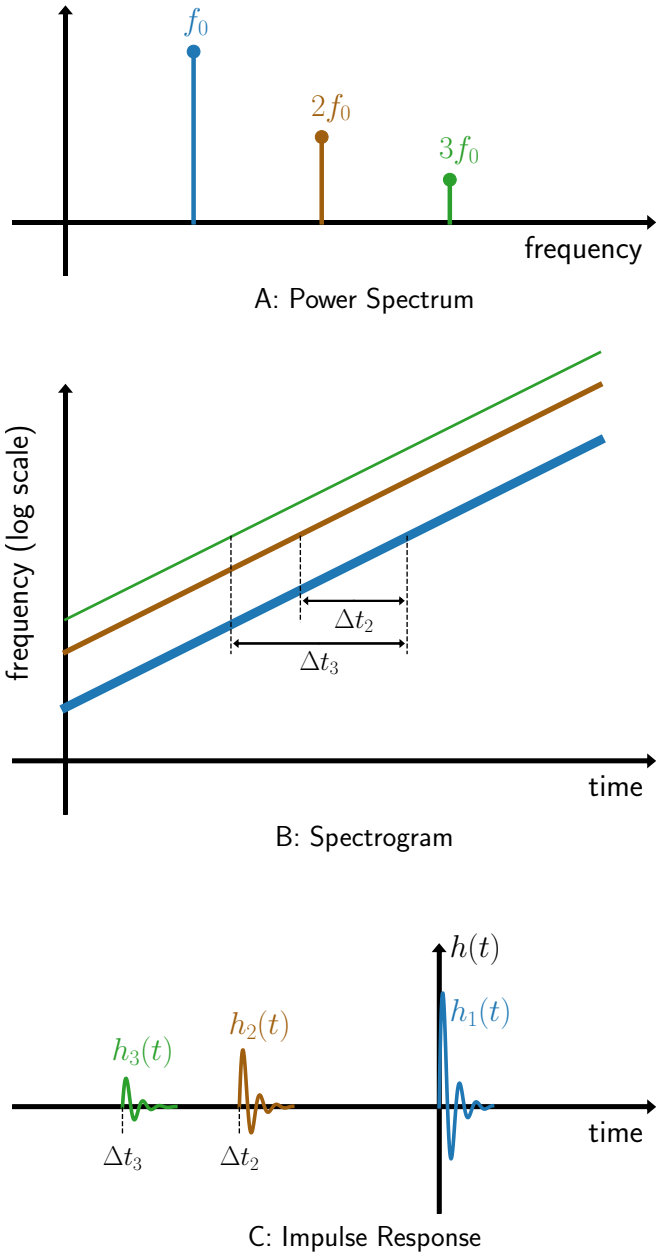

Figure 1. Demonstration of harmonic distortion caused by a nonlinear system. A: Power spectrum of a distorted harmonic signal with higher harmonics, B: spectrogram of a distorted swept-sine signal, C: impulse response after deconvolution.

Each of these frequency components (harmonic or intermodulation) has its amplitude $A_{m,n}$ and its phase $\phi_{m,n}$, and can be frequency dependent. As in the previous case, we use the notation $H_{m,n}(f) = A_{m,n}(f)e^{\phi_{m,n}(f)}$ and $h_{m,n}(t) = \mathcal{F}^{-1}H_{m,n}(f)$.

As with harmonic distortion, the synchronized swept-sine can be used to measure frequency-dependent intermodulation products using the property provided by Eq. (6). The input signal $x(t)$ is a sum of two synchronized swept-sines

$$x(t) = \sin\Big(\varphi_1(t)\Big) + \sin\Big(\varphi_2(t)\Big), \tag{12}$$

with

$$\varphi_1(t) = 2\pi f_{1a} L \exp\left(\frac{t}{L}\right), \tag{13}$$

$$\varphi_2(t) = 2\pi f_{2a} L \exp\left(\frac{t}{L}\right). \tag{14}$$

The starting frequencies f_{1a} and f_{2a} of each synchronized swept sine are related by the coefficient $\alpha = \frac{f_{2a}}{f_{1a}} = \frac{f_{2b}}{f_{1b}}$. The coefficient L [Eq. (3)] is equal for the two synchronized swept-sines, and is defined as

$$L = \frac{T}{\ln\left(\frac{f_{1b}}{f_{1a}}\right)} = \frac{T}{\ln\left(\frac{f_{2b}}{f_{2a}}\right)}, \tag{15}$$

f_{1b} and f_{2b} being respectively the stop frequencies of each swept-sine.

It is possible to generalize Eqs. (4), (5), and (6) [considering $\varphi(t) \equiv \varphi_1(t)$] as

$$(m + \alpha n)\ \varphi_1(t) = \varphi_1(t - \Delta t_{m,n}), \tag{16}$$

$$\Delta t_{m,n} = -L \ln\left(m + \alpha n\right), \tag{17}$$

and

$$s_{m,n}(t) = \sin\Big((m + \alpha n)\,\varphi_1(t)\Big) = \sin\Big(\varphi_1(t - \Delta t_{m,n})\Big). \tag{18}$$

Therefore, the output signal $y(t)$ can be expressed as a sum of all the harmonic distortion and intermodulation products, represented on a spectrogram in Fig. 2B, as

$$\begin{aligned} y(t) &= \sum_m \sum_n h_{m,n}(t) * s_{m,n}(t) \\ &= \sum_m \sum_n h_m(t) * \sin\Big(\varphi_1(t - \Delta t_{m,n})\Big) \\ &= \sum_m \sum_n h_m(t - \Delta t_{m,n}) * \sin\Big(\varphi_1(t)\Big) \\ &= \sin\Big(\varphi_1(t)\Big) * \sum_m \sum_n h_{m,n}(t - \Delta t_{m,n}) \\ &= \sin\Big(\varphi_1(t)\Big) * h(t), \end{aligned} \tag{19}$$

which results in a convolution between the first component of the excitation signal $\sin(\varphi_1(t))$ and an impulse response

$$h(t) = \sum_m \sum_n h_{m,n}(t - \Delta t_{m,n}). \tag{20}$$

The impulse response $h(t)$ consists of Intermodulation Impulse Responses (ImIR) $h_{m,n}(t)$ delayed by $\Delta t_{m,n}$ [Fig. 2C]. Each of these ImIR $h_{m,n}(t)$ can be windowed and Fourier transformed to obtain the Intermodulation Frequency

Responses (ImFR) $H_{m,n}(f)$, which provide information on the frequency-dependent amplitude and phase of each intermodulation product with indices m and n.

Note that the above derivation theory uses as reference the frequency f_1 and, therefore, the first component of the synchronized swept-sine $\sin(\varphi_1(t))$. On the other hand, the derivation could also be done for the second component f_2, or for any other frequency component.

2.4 DPOAE extraction from the swept-sine measurement

The SSS technique described above and applied to DPOAE measurement can be summarized in the following steps. First, the parameters of the two-component synchronized swept-sine, i.e. the start and stop frequencies f_{1a} and f_{2a} and the time duration T of the first component, and coefficient α for the second component, are chosen. The input signal, consisting of the sum of both swept-sines, is generated using Eqs. (12)-(14) and (15).

Once the measurement has been made and the output signal $y(t)$ has been acquired, the impulse response $h(t)$ is obtained by a deconvolution process. The convolution of Eq. (19) is written in the frequency domain as $Y(f) = S_1(f)H(f)$, where $S_1(f)$ is the Fourier transform of the synchronized swept-sine $\sin(\varphi_1(t))$. Its analytical form, derived in [Novak *et al.*(2015)], is

$$S_1(f) = \frac{1}{2}\sqrt{\frac{L}{f}}\exp\left\{j2\pi fL\left[1-\ln\left(\frac{f}{f_{1a}}\right)\right]-j\frac{\pi}{4}\right\}. \tag{21}$$

The impulse response $h(t)$ is then obtained using frequency domain deconvolution as

$$h(t) = \mathcal{F}^{-1}\left\{\frac{Y(f)}{S_1(f)}\right\}. \tag{22}$$

Finally, the ImIR $h_{2,-1}(t)$ corresponding to the DPOAE intermodulation product related to $2f_1 - f_2$ (CDT) is extracted from $h(t)$ by windowing. Its temporal position is given by Eq. (17), namely

$$\Delta t_{2,-1} = -L\ln(2-\alpha), \tag{23}$$

as depicted in Fig. 2C.

2.5 Effect of the sweep rate

If we sweep a sine exponentially between f_{1a} and f_{1b} frequencies, the sweep rate r described in octaves per second determines the entire duration of the swept-sine; namely

$$T = \frac{\log_2(\frac{f_{1b}}{f_{1a}})}{r}, \tag{24}$$

which relates the coefficient L [Eq. (15)] with the sweep rate r; namely

$$L = \frac{\log_2(\frac{f_{1b}}{f_{1a}})}{r\ln(\frac{f_{1b}}{f_{1a}})} = \frac{1}{r\ln 2}. \tag{25}$$

We can see that only the sweep rate and the frequency ratio between two tones (f_2/f_1) determines the temporal position of $h_{2,-1}(t)$ [add Eq. (25) into Eq. (23)] and hence the time difference between $h_{1,0}(t)$ and $h_{2,-1}(t)$. For the "optimal" sweep rates of 0.5 oct/sec. suggested in [Abdala *et al.*(2015)] and the commonly used "optimal" ratio $f_2/f_1 = 1.2$ yielding the largest DPOAE amplitude, the time difference between these impulse responses is about 644 ms, meaning that they are well separated. These stimulus parameters are used in the present paper.

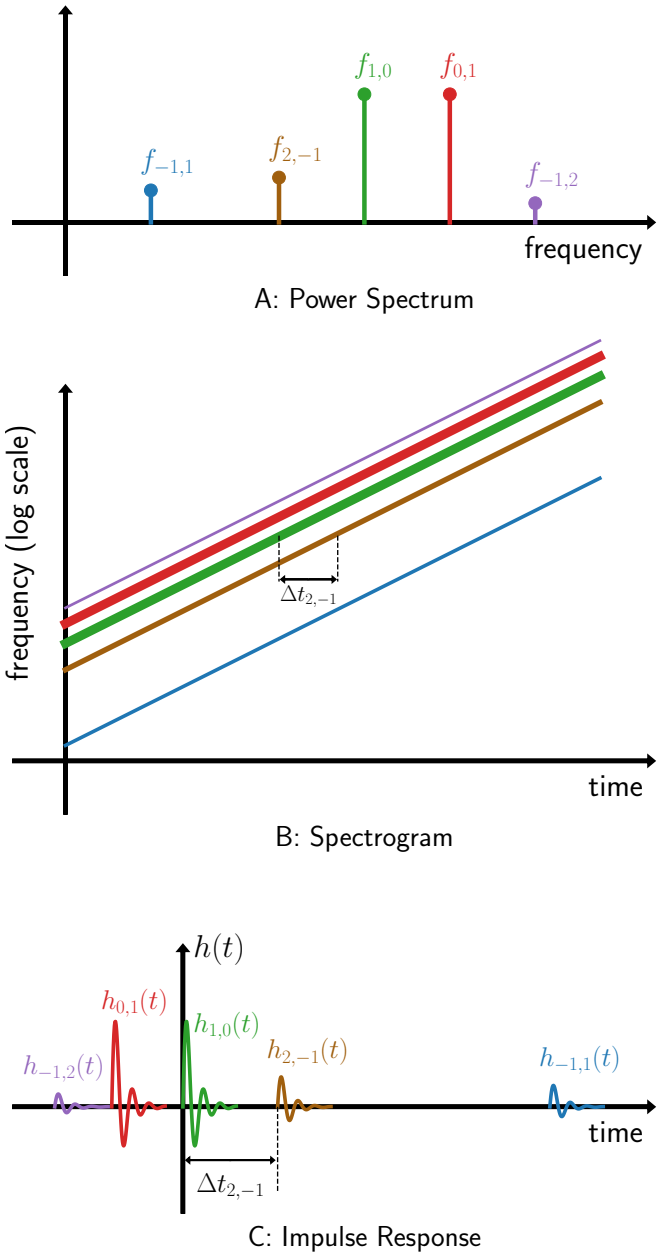

Figure 2. Demonstration of harmonic and intermodulation distortion caused by a nonlinear system. A: Power spectrum of a distorted two-tone signal with higher harmonics and intermodulation products, B: spectrogram of a distorted two-component swept-sine signal, C: impulse response after deconvolution.

3. Simulation verification

This section extends the SSS technique for DPOAEs with a windowing method allowing for extraction of DP-gram components and background noise suppression. In this section, DPOAEs derived from a cochlear model are presented. The cochlear model is used because it allows for accurate separation of nonlinear-distortion and coherent-reflection DPOAE components. In addition, the cochlear model that is used generates the coherent-reflection component of DPOAEs with roughly similar latencies to those observed experimentally, e.g. in [Moleti *et al.*(2012)] (compare their Fig. 7 with Fig. 4 presented below). Section 4 uses the same windowing method for DPOAEs measured in normally hearing human subjects.

3.1 Separation of DPOAE components

Figure 3A depicts a DP-gram (CDT at f_{DP} frequency shown in Fig. 4) in the temporal domain (ImIR). Most of the signal energy is located near the zero time delay, but we can also identify a long-latency component with most energy between 2 and 10 ms. The mechanism of nonlinear distortion generates a DPOAE component whose phase changes slowly with frequency – a short-latency component – and the mechanism of coherent reflection (due to mechanical irregularities) generates a DPOAE component whose phase changes rapidly with frequency – a long-latency component [Shera and Guinan(1999)]. Therefore, both components can be separated from the DP-gram [Stover *et al.*(1996), Konrad-Martin *et al.*(2001), Dhar *et al.*(2002), Knight and Kemp(2001), Kalluri and Shera(2001)]. Because the SSS technique first calculates the DP-gram in the time domain – the ImIR calculated with Eq. (20) – suitable windows can be applied to separate the short-latency and long-latency components of the DP-gram before it is transformed into the frequency domain.

The latency of OAEs evoked due to reflection of forward traveling waves by localized irregularities in the micromechanics of the organ of Corti is frequency dependent; it shortens as the frequency increases [Moleti *et al.*(2012), Kemp(1978), Shera and Guinan(2003), Shera and Bergevin(2012), Bergevin *et al.*(2012)]. For separation of DP-gram components, it is useful to employ frequency-dependent window duration, i.e. to shorten the window duration as the frequency increases. The advantage of shorter windows is greater suppression of the background noise which contaminates experimental data. Another advantage of the frequency dependent window is that it can remove multiple internal reflections [Shera and Zweig(1991), Dhar *et al.*(2002)].

In this paper, we use temporal windows constructed using recursive exponential windows.[3] Figure 3A depicts an example of such constructed windows. Note that both windows are asymmetrical because they are constructed by combining two halves of recursive exponential windows. A half of a window for $\tau_c = 1$ ms is used for negative times, and the second part of the window is constructed from the half window for frequency-dependent τ_c, given by Eq. (26) taken from [Moleti *et al.*(2012)]; namely,

$$\tau_c[n] = \frac{a}{2}\left[\frac{f_c[n]}{1\,\mathrm{kHz}}\right]^{-b}, \tag{26}$$

where $f_c[n]$ is frequency in Hz, which is in the equation divided by 1 kHz, $b = -0.8$ [Moleti *et al.*(2012)], and n indicates that latency is assumed for the nth window in the set. Parameter a then determines whether the window set is intended to separate the short-latency component of the DP-gram ($a = 0.01$ sec.) or to extract the entire

[3] [Shera and Zweig(1993)] presents equations for a recursive exponential window. The window of the nth order is defined by $\hat{S}(\tau, \tau_c) = 1/\Gamma_n(\lambda_n \tau/\tau_c)$, where τ is time, τ_c is the window length (cutoff time), $\Gamma_n(z)$ is a recursively defined function $\Gamma_{n+1}(z) = \exp(\Gamma_n(z) - 1)$ and $\Gamma_1(z) = \exp(z^2)$. $\lambda_n = \sqrt{\gamma_n}$, where $\gamma_{n+1} = \ln(\gamma_n + 1)$ and $\gamma_1 = 1$. With these parameters, the window $\hat{S}(\tau, \tau_c)$ has a maximum value of 1 at $\tau = 0$ and the value of $1/e$ at $\tau = \tau_c$ [Kalluri and Shera(2001)]. The windows used in this paper have the order $n = 10$ and latencies given by Eq. (26) for $t \geq 0$ second or set to 1 ms for $t < 0$; see Fig. 3A. For positive τ, $\hat{S}(\tau, \tau_c)$ represents one half of a recursive exponential window decreasing in amplitude from 1 to 0. The windows depicted in Fig. 3A are constructed by mirroring the samples for $\hat{S}(\tau, 0.001)$ and concatenating them with samples for $\hat{S}(\tau, \tau_c)$, where τ_c is given by Eq. (26). This process creates the entire (asymmetrical) windows used in this paper.

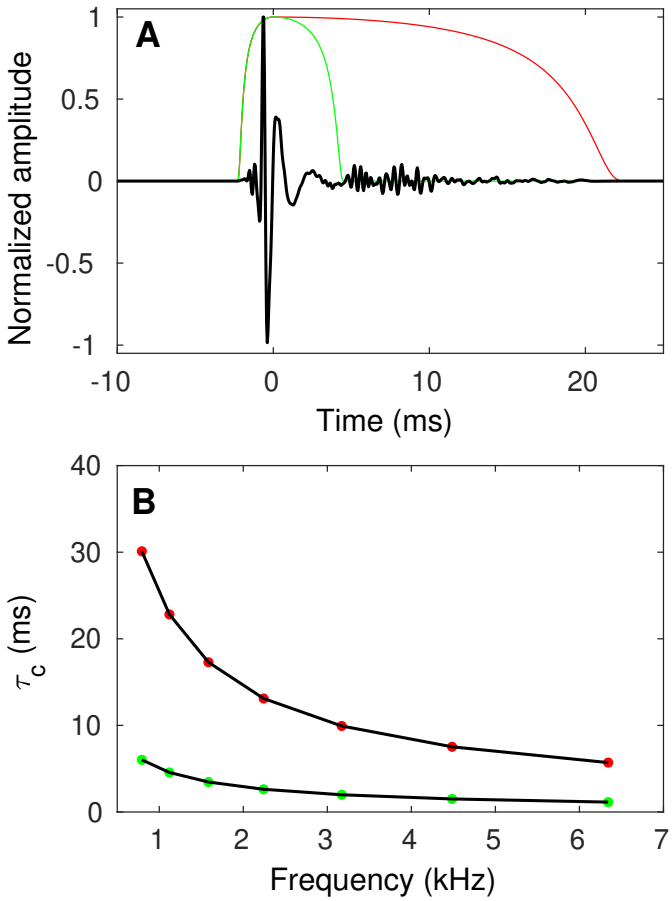

Figure 3. A: ImIR, i.e., time domain representation of the CDT DP-gram. The windows are constructed from recursive exponential windows calculated for latency τ_c given by Eq. (26) for $f_c = 3172\,\mathrm{Hz}$ – the green line depicts a window for short-latency DP-gram component separation and the red line depicts a window for the entire DP-gram. B: Cutoff times of the recursive exponential windows applied to the ImIRs in order to obtain the complete DP-gram (red dots) and in order to separate the nonlinear-distortion DPOAE component (green dots).

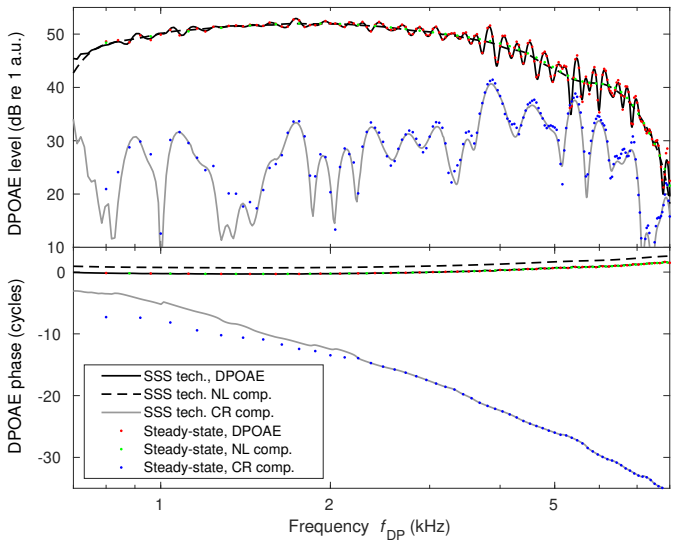

Figure 4. Amplitude and phase of a DP-gram derived from the cochlear model. The solid and dashed lines depict DP-grams derived by the SSS technique. The red, green and blue dots depict DP-grams derived from the same model with steady-state tones. The entire DP-gram derived from the model with irregularities generating the secondary component of DPOAE due to coherent reflection is depicted with a solid black line and red dots. The short-latency component yielded by the SSS technique using a specific window set is depicted with a black dashed line. The gray line depicts the DP-gram calculated as the difference between the entire DP-gram (black solid line) and the short-latency DP-gram (black dashed line). For the steady-state tones, the short-latency component of the DP-gram was derived from the model without irregularities. The coherent-reflection component depicted with blue dots is calculated as a vector difference between DP-grams for the model with irregularities and for the model without irregularities.

DP-gram, including the short-latency and long-latency components ($a = 0.05$ sec.); suppressing only higher-order reflections and background noise. Both values of a were set empirically to achieve the best visual agreement between the steady-state simulated DP-gram and the DP-gram derived by the SSS technique for $L_1 = 60$ dB, $L_2 = 50$ dB, $f_2/f_1 = 1.2$, and f_2 swept[4] between 1 kHz and 12 kHz with a speed of 0.5 oct/sec. (see Fig. 4). Figure 3B depicts the calculated latencies used in the two sets of recursive exponential windows. For DPOAE filtering with wavelets, [Moleti *et al.*(2012)] set the parameter $a = 0.013$ sec. for the short-latency component and multiplied the latency by 2 for the entire DPOAE without higher order reflections.

Because the SSS technique is computationally inexpensive, we decided to process the entire response to the swept-sines repeatedly for a set of windows constructed of the recursive exponential windows for various f_c values. Each window setting gives the entire DP-gram, i.e. the DP-gram for the entire frequency range used during the measurement. However, the final DP-gram is constructed post-hoc by assuming only those portions of DP-grams which were calculated with windows for that frequency region. We tuned the presented approach using swept-sines for 0.5 oct/sec. For this setting, we empirically chose 1.5-sec. long time frames and calculated the center frequencies f_c used in Eq. (26) for the instantaneous frequency of the swept-sines given by

$$f_{\text{inst}}(t) = f_a \exp\left(\frac{t}{L}\right), \tag{27}$$

derived from Eq. (2). The set of N center frequencies f_c then determines the number of windows and hence the number of ImIRs that need to be calculated for the entire DP-gram or for the short-latency component of the DP-gram. We denote each ImIR for the DP component $\hat{h}^{\text{DP}}$ and assume that in Matlab, this ImIR is composed of samples $\hat{h}^{\text{DP}} = \left(h^{\text{DP}}[1], h^{\text{DP}}[2], ..., h^{\text{DP}}[M]\right)$, where M is the entire number of samples. We can then expand these ImIRs into a matrix where each row is $\hat{h}^{\text{DP}}$; namely

$$\mathbf{h}^{\text{DP}} = \begin{pmatrix} \hat{h}_1^{\text{DP}} \\ \hat{h}_2^{\text{DP}} \\ \vdots \\ \hat{h}_N^{\text{DP}} \end{pmatrix}, \tag{28}$$

where N is the number of center frequencies $f_c[n]$. We construct a new matrix $\mathbf{R}$ composed of recursive exponential windows $\hat{r}_i$ whose shape is depicted in Fig. 3A. Each row contains window samples with latency given by Eq. (26) for $f_c[n]$ for $n = 1, 2, ..., N$; namely

$$\mathbf{R} = \begin{pmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \vdots \\ \hat{r}_N \end{pmatrix}, \tag{29}$$

DP-grams for each specific windows can then be saved in the matrix, which is calculated as a dot-product of matrices $\mathbf{h}^{\text{DP}}$ and $\mathbf{R}$; namely

$$\mathbf{H}^{\text{DP}} = \mathcal{F}\left[\mathbf{h}^{\text{DP}} \cdot \mathbf{R}\right]. \tag{30}$$

Each row in matrix $\mathbf{H}^{\text{DP}}$ represents a DP-gram calculated for a specific window $\hat{r}_n$. The samples of the DP-gram are distributed along the frequency axis $f_x = (i-1)f_s/M$, where the sampling frequency $f_s = 44.1$ kHz. The final DP-gram is constructed from the matrix $\mathbf{H}^{\text{DP}}$, assuming only those samples which are in the neighborhood of the given $f_c[n]$. A simple algorithm assumed in this paper consists of the following steps:

[4]Equations for the SSS technique are presented in Sec. 2 for the f_1 tone at 0 time, which is useful because $f_{\text{DP}} = 2f_1 - f_2$ is adjacent to the f_1 component. This setting could be changed without any effect on the accuracy of the technique. On the other hand, DP-grams are often depicted with f_2 frequency on the x-axis, because the assumed generation region for DPOAEs is near the f_2 best frequency place.

1. For each row in the matrix $\mathbf{H}^{\mathrm{DP}}$, set the samples which lie out of the specific interval to $\emptyset$ – empty set symbol denoting outliers or, as [Fang and Liu(2022)] suggested, "a don't care condition", i.e., do not take these samples for the final averaging.
   - If $n == 1$ then $H_n^{\mathrm{DP}}[i] = \emptyset$, $\forall i$ for which $f_x[i] \geq f_c[n+1]$
   - Else if $n == N$ then $H_n^{\mathrm{DP}}[i] = \emptyset$, $\forall i$ for which $f_x[i] \leq f_c[n-1]$
   - Else $H_n^{\mathrm{DP}}[i] = \emptyset$, $\forall i$ for which $f_x[i] \leq f_c[n-1]$ and $f_x[i] \geq f_c[n+1]$
2. Calculate mean across the rows (column-wise mean) of matrix $\mathbf{H}^{\mathrm{DP}}$

$$H_{\mathrm{DP}}[i] = \frac{1}{N}\sum_{n=1}^{N} H_n^{\mathrm{DP}}[i] \tag{31}$$

Based on the chosen window set given by the parameter a in Eq. (26), we can calculate a DP-gram H_{DP} which contains short and long latency components (for $a = 0.05$) and a DP-gram $H_{\mathrm{DP}}^{\mathrm{NL}}$ which contains only the nonlinear-distortion (short latency) component (for $a = 0.01$). Then, the DP-gram containing the long-latency component can be calculated as a difference; namely

$$H_{\mathrm{DP}}^{\mathrm{CR}} = H_{\mathrm{DP}} - H_{\mathrm{DP}}^{\mathrm{NL}}. \tag{32}$$

The entire DP-gram – the nonlinear-distortion (short latency) component of the DP-gram, and the coherent-reflection (long-latency component) of the DP-gram – is depicted in Fig. 4. Figure 4 shows very good agreement between the DP-gram components derived by the presented SSS technique and the DP-gram components derived from the steady-state responses using a cochlear model with roughness (impedance irregularities) and a smooth cochlear model which allows for a perfect decomposition of the nonlinear-distortion DPOAE component and the coherent-reflection DPOAE component.

4. Experimental verification

Section 3 expanded the SSS technique with a windowing method allowing for extraction of DP-gram components and larger suppression of background noise. This adapted SSS technique is used in this section to extract DP-grams from the swept-sine responses recorded in normally hearing human subjects. Because the OAE recordings may be affected by various forms of excessive noises, e.g. due to subject swallowing, this section extends the SSS technique by an artifact rejection method adapted from the method presented in [Fang and Liu(2022)].

4.1 Methods

4.1.1 Subjects

DPOAEs were measured in four normally hearing subjects. Their pure tone hearing thresholds were within the range of 20 dB re hearing level (HL) for frequencies between 0.125 and 8 kHz. The age of the subjects ranged between 22 and 24, with a median of 23.

4.1.2 Stimuli and data acquisition

DPOAEs were measured with synchronized swept-sines of various stimulus levels ($L_1 = 60$ dB SPL and $L_2 = 50$ dB SPL, or $L_1 = 50$ dB SPL and $L_2 = 45$ dB SPL), the frequency ratio between the stimuli $f_2/f_1 = 1.2$, and f_2 tone swept at a rate of 0.5 oct/sec. between 1 kHz and 12 kHz (the same frequency range as in the simulations in Sec. 3). The onset and offset of the swept-sines was shaped with 20-ms long raised-cosine ramps.

All measurements were made in an audiological booth using custom software written in Matlab. Sound signals were generated in a computer and were presented by an RME Fireface UCX sound card connected to an Etymotic ER10C probe. The probe was calibrated inside the ear canal of the subjects before each measurement. To reduce

the noise floor, the swept-sines were presented repeatedly. The measurement was stopped when the DP-gram and the estimated background noise were almost unchanging. This criterion required 23 repetitions for subject s013, 23 repetitions for subject s014, 20 repetitions for subject s015, and 20 repetitions for subject s17. This means that, on average, the measurement required about 2 minutes and 40 seconds, including a few hundred ms long pauses between signal presentations. The experiment was conducted under the permission of the Ethics Committee of the Czech Technical University in Prague.

4.1.3 DPOAE extraction

DPOAEs were extracted by the SSS technique and by the LSF technique [Long *et al.*(2008)],[5] which served here as a reference. The LSF technique was used with "optimal" parameters for the given frequency range, as suggested in [Abdala *et al.*(2015)] to be: 125 ms (5512 samples for 44.1 kHz sampling frequency) time window for a 0.5 oct/sec. sweep for the 1-4 kHz range. In addition, because the LSF technique can smooth the DPOAE fine-structure – separate the nonlinear-distortion component – if a longer analysis window is used, we also extracted DP-grams with 500 ms (22050 samples for 44.1 kHz sampling frequency), as recommended in [Abdala *et al.*(2015)]. For both window lengths (125 ms and 500 ms), the windows were shifted with a step of 200 samples (4.5 ms for 44.1 kHz sampling frequency).

The SSS technique was used together with the windowing method presented in Sec. 3. The SSS technique extracted the entire DP-gram and the nonlinear-distortion component of the DP-gram.

4.1.4 Artifact rejection

To reduce the background noise in the experimental data, several (approximately 20) presentations of the swept-sine stimuli were needed for averaging. Some sound artifacts can be relatively pronounced but affect only a small number of samples, e.g., artifacts due to swallowing. Because the SSS technique processes the entire response, we decided not to reject the entire response contaminated with artifacts. Instead, we adapted the "point-wise artifact rejection method" presented by [Fang and Liu(2022)] for transient-evoked OAEs. This method rejects only those samples in the response that are affected by a pronounced sound artifact. We had to adapt the method slightly, because in addition to the OAE signal and the noise, our records also contain the evoking stimulus. The adapted technique is described below. The same artifact rejection method is also used for DP-grams extracted by the LSF technique and presented in this paper.

The recorded responses are collected in a matrix $\mathbf{Y}$; namely

$$\mathbf{Y} = \begin{pmatrix} y_1[1] & y_1[2] & \cdots & y_1[M] \\ y_2[1] & y_2[2] & \cdots & y_2[M] \\ \vdots & \vdots & \ddots & \vdots \\ y_N[1] & y_N[2] & \cdots & y_N[M] \end{pmatrix}, \tag{33}$$

where N is the total number of recorded responses (stimulus repetitions) and M is the total number of samples in a single recorded response. This means that each row in $\mathbf{Y}$ contains samples from a single recorded response.

[Fang and Liu(2022)] rejected responses which were heavily affected by artifacts (detected by assuming a fixed threshold value for the response) and then calculated the mean across the rows of the matrix $\mathbf{Y}$ (across the responses). Because our responses contain the evoking stimulus, our adapted method does not reject heavily affected samples but calculates the median $y^{\mathrm{med}}[i]$ across the responses. The median signal is then subtracted from each response to obtain a noise matrix $\mathbf{U}$; namely

$$u_n[i] = y_n[i] - y^{\mathrm{med}}[i], \tag{34}$$

[5] Matlab implementation of the LSF technique available in the OAETOOLBOX [OAE()] is used in this paper.

which creates the matrix

$$\mathbf{U} = \begin{pmatrix} u_1[1] & u_1[2] & \cdots & u_1[M] \\ u_2[1] & u_2[2] & \cdots & u_2[M] \\ \vdots & \vdots & \ddots & \vdots \\ u_N[1] & u_N[2] & \cdots & u_N[M] \end{pmatrix}. \tag{35}$$

The noise matrix is used to calculate the threshold, which is then used to detect samples affected by sound artifacts, i.e., the samples which were then abandoned. For the threshold calculation, a standard deviation is calculated across all the samples in the noise matrix $\mathbf{U}$; namely,

$$\sigma' = \sqrt{\frac{1}{NM}\sum_{n=1}^{N}\sum_{i=1}^{M} u_n^2(i)}. \tag{36}$$

The threshold $\theta = 2\sigma'$ was set to avoid the strongest artifacts in accordance with [Fang and Liu(2022)].

The adapted point-wise rejection method can then be described in several steps.

1. Initialize $\mathbf{Y}'' = \mathbf{Y}$, and $\mathbf{U}'' = \mathbf{U}$.

2. For all $n = 1, 2, ..., N$ and $i = 1, 2, ..., M$ if $|u_n(i)| > \theta$ then $y''_n[i] = \emptyset$ and $u''_n[i] = \emptyset$, where the empty set symbol $\emptyset$ indicates that this sample is not taken into account in the final averaging of the X'' and U'' matrices; [Fang and Liu(2022)] states that the symbol $\emptyset$ "denotes a don't-care condition".

3. The final one-dimensional set of samples with an averaged signal response and an averaged noise response (estimated background noise) is calculated by averaging $\mathbf{Y}''$ and $\mathbf{U}''$ across the rows (columns-wise mean) of the matrices; namely

$$\bar{y}[i] = \frac{1}{N}\sum_{n=1}^{N} y_n[i], \tag{37}$$

$$\bar{u}[i] = \frac{1}{N}\sum_{n=1}^{N} u_n[i], \tag{38}$$

where $i = 1, 2, ..., M$. $\bar{y}[i]$ is then used to calculate the DP-gram of the experimental data presented in this paper, and $\bar{u}[i]$ is used to estimate the background noise. To estimate the background noise, the SSS or LSF techniques are applied on the averaged noise signal $\bar{u}[i]$.

4.2 Results

Figures 5, 6, 7, and 8 depict DP-grams measured in four subjects using swept-sines (0.5 oct/sec.) of various levels indicated in the figure captions, $f_2/f_1 = 1.2$ and f_2 swept between 1 kHz and 12 kHz. The figures compare DPOAEs extracted by the SSS technique and DPOAEs extracted by the LSF technique. The data were chosen to cover various conditions: DP-grams with a less pronounced fine structure due to interaction between nonlinear-distortion and coherent-reflection components (Fig. 5 and 6), a DP-gram with a pronounced fine structure (Fig. 7), and a DP-gram with a large background noise level, leading to a small DPOAE to background noise ratio (Fig. 8). Panels A-1,2 depict entire DP-grams, including the nonlinear-distortion (short latency, SL) component and the coherent-reflection (long-latency) component. These DP-grams were extracted by the SSS technique with frequency-dependent windows (see Sec. 3), and by the LSF technique with a 125-ms long time window (recommended by [Abdala *et al.*(2015)] for 0.5 oct/sec. swept-sines). To demonstrate the ability of the SSS technique to extract the nonlinear-distortion component of a DP-gram (the short-latency component), panels B-1,2 in the figures compare DP-grams derived by the SSS technique with short, frequency-dependent windows, and by the LSF technique for 500-ms frames. In

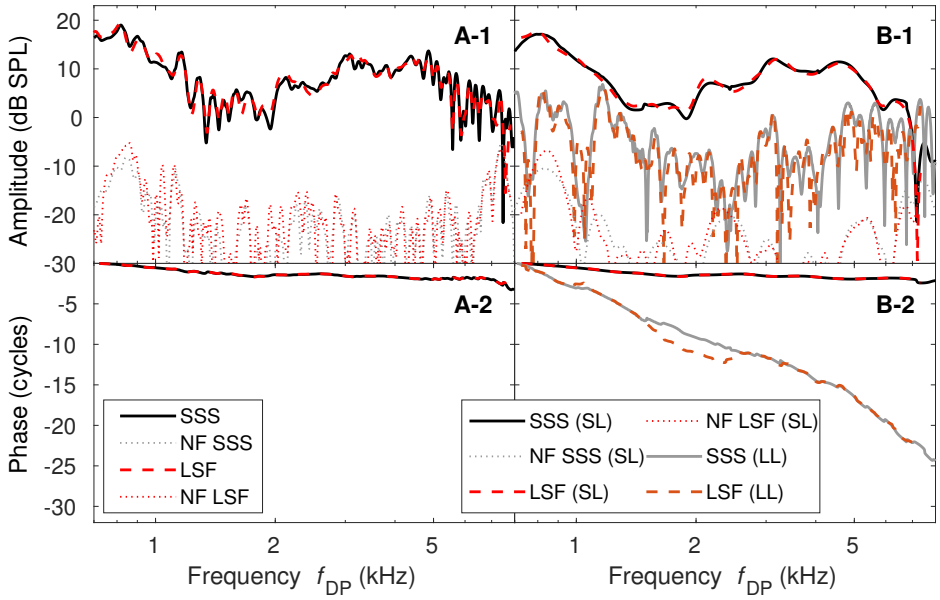

Figure 5. Amplitude and phase of CDT DP-grams extracted from the ear canal responses evoked with synchronized swept-sines in normally hearing subject s013. The stimulus parameters were $L_1 = 60$ dB SPL, $L_2 = 50$ dB SPL, $f_2/f_1 = 1.2$, and f_2 swept from 1 kHz to 12 kHz. Panels A-1,2 respectively depict the amplitude and the phase for the SSS technique with variable window size and the LSF technique for 125 ms window. Black solid lines and dotted lines, respectively, depict DP-grams and noise floors (NF) extracted by the SSS technique. Red dashed and dotted lines, respectively, depict DP-grams and noise floors extracted by the LSF technique. Panels B-1,2 depict the amplitude and the phase of the short-latency (SL) and long-latency (LL) components of the DP-gram. Using the SSS technique, the SL component is depicted with a black solid line and the LL component is depicted with a gray solid line; the noise floor was estimated for the SL component and is depicted with the black dotted line. Using the LSF technique, the SL component is depicted with a red dashed line and the LL component is depicted with an orange dashed line; the noise floor was estimated for the SL component and is depicted with a red dotted line.

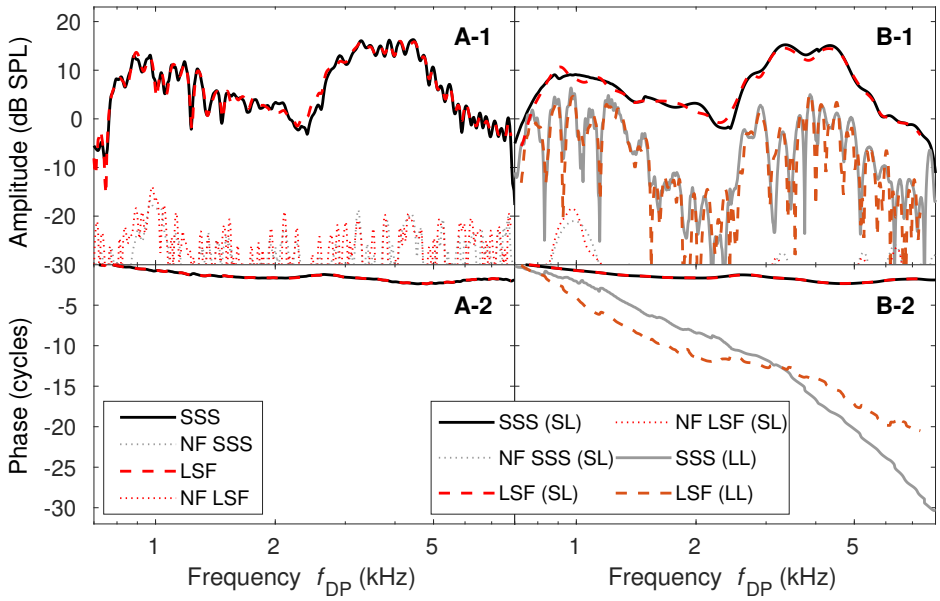

Figure 6. Amplitude and phase of CDT DP-grams extracted from the ear canal responses evoked with synchronized swept-sines in normally hearing subject s014. The stimulus parameters were $L_1 = 60\,\text{dB SPL}$, $L_2 = 50\,\text{dB SPL}$, $f_2/f_1 = 1.2$, and f_2 swept from 1 kHz to 12 kHz. The description of the figure is otherwise the same as for Fig. 5.

addition, panels B-1,2 also depict the long-latency component of the DP-grams, which should be generated due to coherent reflection. This component is obtained as a vector subtraction of the short-latency component depicted in panels B from the entire estimated DP-grams depicted in panels A. The agreement between the long-latency components estimated by both techniques is very good. The largest discrepancies are visible in Fig. 8, where the estimated long-latency component is very close to the estimated noise floor.

The agreement between the DP-grams derived by the SSS technique and by the LSF technique is very good under the currently presented conditions. There seems to be slightly better agreement between the DP-gram phases, except in Fig. 7, A-2, where the pronounced notches in the DP-gram amplitude for the SSS technique cause the unwrapped phases to depart by 1 cycle from the DPOAE phase derived by the LSF technique. However, the general trend in the DP-gram phase is the same for both extraction techniques.

Figures 5–8 show that the LSF technique yields a shallower fine structure of the DP-gram amplitude, in comparison with the SSS technique. Figure 7 in [Abdala *et al.*(2015)] shows that the chosen analysis window (frame) duration affects the fine structure of the DP-gram. If we were to choose shorter frames than 125-ms suggested in [Abdala *et al.*(2015)], we would get a deeper fine structure in the DP-grams derived by the LSF technique. Therefore, the cause of the discrepancy between the LSF and SSS techniques is the chosen analysis window duration for the LSF technique and parameter $a = 0.05\,\text{sec.}$ for the SSS technique.

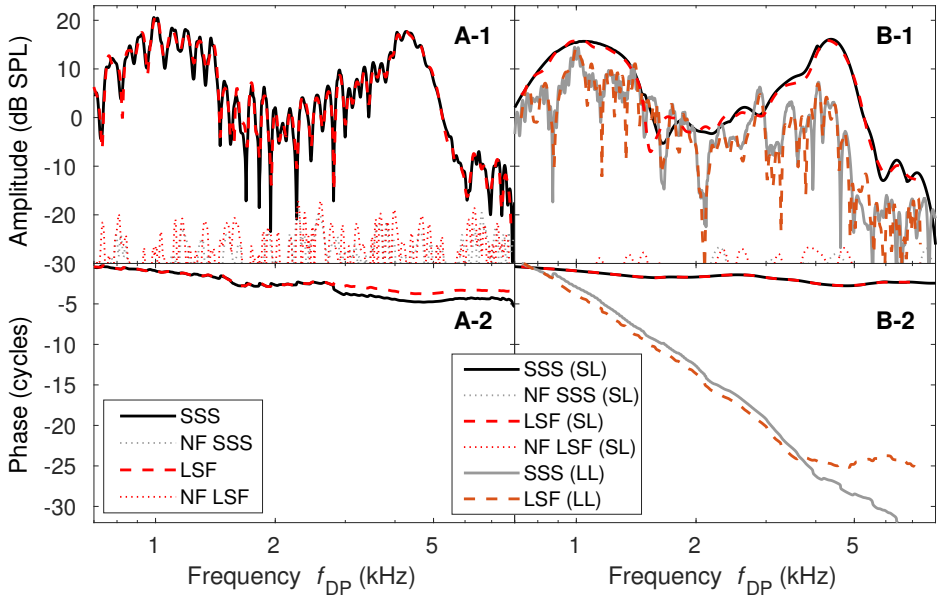

Figure 7. Amplitude and phase of CDT DP-grams extracted from the ear canal responses evoked with synchronized swept-sines in normally hearing subject s015. The stimulus parameters were $L_1 = 50\,\text{dB SPL}$, $L_2 = 45\,\text{dB SPL}$, $f_2/f_1 = 1.2$, and f_2 swept from 1 kHz to 12 kHz. The figure description is otherwise the same as for Fig. 5.

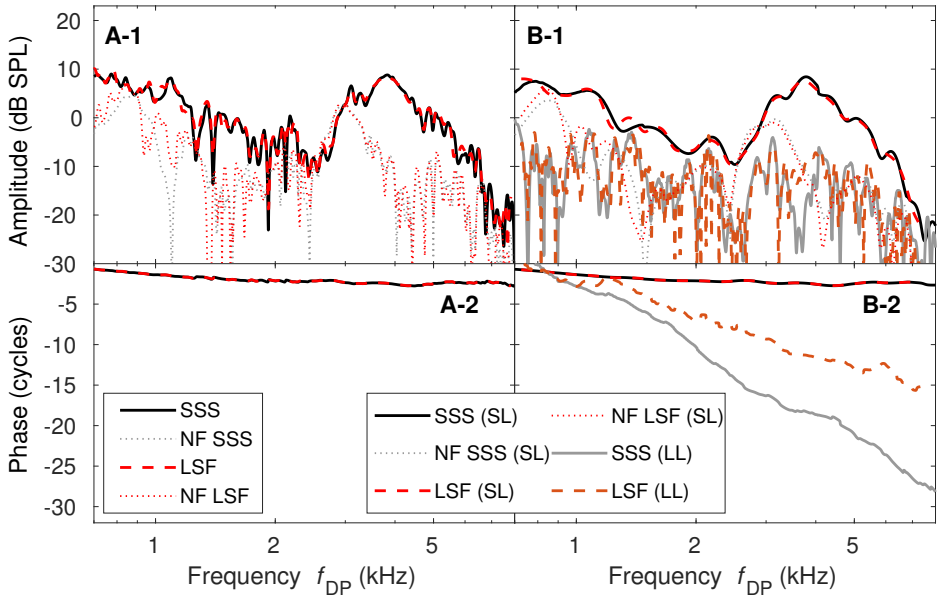

Figure 8. Amplitude and phase of CDT DP-grams extracted from the ear canal responses evoked with synchronized swept-sines in normally hearing subject s017. The stimulus parameters were $L_1 = 50\,\text{dB SPL}$, $L_2 = 45\,\text{dB SPL}$, $f_2/f_1 = 1.2$, and f_2 swept from 1 kHz to 12 kHz. The figure description is otherwise the same as for Fig. 5.

5. Discussion and conclusion

This paper has presented an approach for extracting DPOAEs from the ear canal responses evoked with synchronized swept-sines (SSS) [Novak *et al.*(2015)]. The paper is composed of three parts. The first part (Sec. 2) presents the theory describing how the SSS technique extracts intermodulation DPs from the responses to two simultaneous swept-sines. This technique could also be useful for other possible applications focused on intermodulation DPs, not only for CDT DPOAEs at $2f_1 - f_2$ as shown in this paper. The technique can be easily adapted for any intermodulation DPs generated by a system. The remaining two parts of the paper (Sec. 3 and Sec. 4) then extend the SSS technique with methods which we designed in order to allow for separation of the DP-gram component and for suppressing the background noise (Sec. 3) and for rejecting sound artifacts during a measurement (Sec. 4). The reader can come up with different methods and adapt the SSS technique based on his/her needs.

As designed in this paper, the SSS technique estimates DP-grams with similar accuracy as the least-square-fitting (LSF) technique, which has been suggested to be the most noise robust [Kalluri and Shera(2013)]. The SSS technique is computationally inexpensive. The measured DP-gram can therefore be calculated during the measurement and presented to the experimenter almost instantaneously. In addition, because the DP-gram is also available in the time domain during the calculation, nonlinear-distortion and coherent-reflection components of DPOAE can be extracted by temporal windows and provided to the experimenter during the measurement. The experimenter can thus decide which number of stimulus repetitions could be adequate for DP-gram measurement. To conclude, the presented SSS technique is not suggested to be superior to other DPOAE extraction techniques (e.g., the LSF technique, the heterodyne technique, a technique based on the Fourier transform), which can still be used post-hoc for verification and data analysis. The time efficiency of the SSS technique makes it suitable for data presentation during a measurement and, therefore, for example, in clinical equipment used for hearing loss assessment based on a DP-gram.

5.1 Application of synchronized swept-sines for DPOAE measurement

A DPOAE signal recorded in the ear canal is a weak acoustical signal with a level close to the background noise floor in quiet. To increase the noise robustness of the SSS technique, it is useful to multiply the calculated ImIR for the DP component with a suitable window which suppresses the noise in the time samples outside of the required time interval. In addition to suppressing background noise, multiplication of the ImIR with a temporal window allows for decomposition of DPOAE components while they are being measured (see Fig. 4). DPOAE at $2f_1 - f_2$ (and possibly also other low-side DPs) is generated by at least two sources: a source due to inter-modulation distortion generating DP wavelets which travel backward into the cochlea-middle-ear boundary, and DP wavelets which travel forward toward the DP tonotopic place. At the DP tonotopic place, these wavelets are partly reflected by impedance irregularities, which are the second source of DPOAEs [Shera and Guinan(1999)].[6] The DPOAE component evoked with the first source (also called the primary or nonlinear-distortion component) has a short latency, whereas the source due to coherent reflection (also called the secondary source of DPOAEs) has a long latency, which decreases as a function of frequency [Shera and Guinan(1999), Kalluri and Shera(2001), Moleti *et al.*(2012)].

We have solved the issue of background noise suppression and component separation by using temporal windows with frequency-dependent parameters (see Sec. 3). For the application of a frequency-dependent window, we can either separate the response into shorter parts or process the entire response for a set of windows and select only a subset of frequency samples from each set based on the frequency region for which the specific window was constructed. As described in Sec. 3, we chose the latter approach using a set of windows. We saw an advantage in

[6] A recent paper by [Vetešník *et al.*(2022)] presented an additional source of DPOAEs due to perturbation of the nonlinear force in the generation region of DPOAEs. However, this source was shown to have a long latency, comparable to the latency of the coherent-reflection source of DPOAEs. Therefore, the currently presented windowing technique would combine the DP-gram component evoked due to the coherent-reflection source and the DP-gram component evoked due to perturbation of the nonlinear force.

this approach because the separation of the swept-sine response into shorter time frames caused pronounced ringing in the DP-gram amplitude if the short temporal windows needed for extraction of the short-latency DP-gram component were used (data not shown).

The accuracy of the extraction method is verified using a nonlinear cochlear model (Sec 3) and for experimental data by comparison with DP-grams extracted by the LSF technique [Long *et al.*(2008)]. The LSF technique also allows for DP-gram component extraction and background noise suppression [Abdala *et al.*(2015)]. The technique fits an assumed DPOAE signal into the swept-sine response in the time-domain. The fitting is done within a temporal frame of fixed duration. For sines swept exponentially at a rate of 0.5 oct/sec., the 125-ms window was suggested for the entire DP-gram including short and long latency components, and a 500-ms window was suggested for the short-latency component only [Abdala *et al.*(2015)]. The use of a fixed-duration window for responses obtained with exponentially swept-sines is equivalent to the use of frequency-dependent windows with the SSS technique. As the temporal window shifts across the swept-sine response, a larger frequency range falls into the windowed response because the frequency of the swept-sine increases exponentially with time. However, we should mention that the use of the fixed window duration in the LSF technique is more elegant than the method suggested in this paper in Sec. 3 , which requires post-hoc construction of the final DP-gram.

The coherent-reflection (long-latency) component of DP-grams estimated by the SSS or LSF techniques can be extracted as a vector subtraction of the nonlinear-distortion component from the entire DP-gram, as we did in Figs. 4–8. For the coherent-reflection component, Figs. 5–8 show larger discrepancies between the LSF and SSS techniques than for the nonlinear-distortion component, which may, especially in Fig. 8, be due to the large noise floor relative to the level of the coherent-reflection component. For practical use, we think the SSS technique is suitable for extracting the nonlinear-distortion (short latency) component of a DP-gram, for which the use of a short temporal window decreases the noise floor. However, for an analysis of long-latency components, we would advise the use of time-frequency filtering techniques based on wavelets [Bergevin *et al.*(2012), Moleti *et al.*(2012)].

5.2 Artifact rejection

As designed in Secs. 2 and 3, the SSS technique processes the entire response to a swept-sine, which may be up to several seconds long based on the sweep rate (0.5 oct/sec. used in the present paper) and the frequency range (f_2 swept from 1 kHz to 12 kHz in the present paper). Relatively common measurement artifacts, caused, e.g., by subject movement or by swallowing, usually contaminate only a small number of adjacent samples of the response. An advantage of the LSF technique is that it processes the response in temporal frames, which allows for the detection of measurement artifacts in these frames. The affected frames can then be abandoned, but the rest of the response can be kept for processing. The method suggested in Sec. 3 for the SSS technique processes the entire response to swept-sine stimuli. However, Sec. 4 adapted the method of [Fang and Liu(2022)] designed for transient-evoked OAEs. This adapted method detects sound artifacts in the swept-sine response and abandons them. The method can be used with any DP-gram extraction technique.

5.3 Sweep rate limit

The SSS technique calculates impulse responses at the frequencies of the input tones and their intermodulation products (see Fig. 2C). As the sweep rate increases, the time difference between the adjacent impulse responses decreases, and for fast sweep rates the impulse responses can overlap. Equations (24) and (23) show that the time difference depends only on the sweep rate and the frequency ratio between the tones (f_2/f_1). For upward swept-sines, the impulse response for the f_1 tone approaches the impulse response of the CDT tone. If we neglect the effect of the probe transducer on the duration of the f_1 impulse response and focus only on the delay of OAEs evoked with a single tone, we can for $f_2/f_1 = 1.2$ suggest that the upper limit for the swept-sine rate is about 10 oct/sec., which gives a time difference between the impulse responses of about 32 ms. We assume that most of the long-latency

components in OAEs are within 32 ms [Moleti *et al.*(2012)]. However, this value is frequency-dependent and can increase if there are significant higher order reflections. On the other hand, preserving the measurement noise floor for higher sweep rates "requires a compensating increase in the number of sweeps presented and averaged" [Abdala *et al.*(2015)]. Therefore, it is questionable whether the speed near 10 oct/sec. is useful for DPOAE measurement. Lower sweep rates up to about 5 oct/sec. are for sure applicable with the presented SSS technique.

5.4 Efficiency

The SSS technique presented in Sec. 2 applies the fast Fourier transform on the entire swept-sine response, and performs frequency domain deconvolution and the inverse fast Fourier transform [Eq. (22)]. Then the technique extracts the ImIR and applies a window and performs the fast Fourier transform. This is in fact the total required computation. The windowing method added into the technique in Sec. 3 increases the computational time by the needed repetition of the last step for a set of windows (7 windows times 2 for the frequency range used in the paper from 1 to 12 kHz). Even this complication does not complicate real-time use of the technique implemented in Matlab or in another interpreted programming language.

In comparison, the LSF technique processes the response in time frames. For the stimulus parameters that are used: 0.5 oct/sec. sweep rate and f_2 ranged between 1 kHz and 12 kHz and a 125-ms window for the entire DP-gram extraction, 56 fittings have to be performed if the adjacent time frames were not overlapped. In reality, overlapping is often needed to achieve good frequency resolution. The LSF technique is, therefore, much more computationally demanding than the SSS technique.

To summarize, the SSS technique provides an easy to implement, fast, accurate and noise robust method for DPOAE estimation. Because the method is not computationally expensive, it can be used during a measurement to provide feedback to the experimenter. The method could be implemented in the OAE measurement systems used in clinics.

Acknowledgments

Supported by the project 23-07621J of the Czech Science Foundation (GAČR), internal grant of the Czech Technical University in Prague SGS23/185/OHK3/3T/13, and by European Regional Development Fund-Project "Center for Advanced Applied Science" (Grant No. CZ.02.1.01/0.0/0.0/16_019/0000778). Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the Projects of Large Research, Development, and Innovations Infrastructures programme (CESNET LM2015042) is greatly appreciated.

Appendix

Cochlear model

DPOAEs were derived from the numerical solution of a two-dimensional hydrodynamical cochlear model. The same variant of the model with the same parameters was used as in [Vencovský *et al.*(2019)]. For the model description and parameter values, the reader is referred to that paper. The model is based on previous work of [Sondhi(1978), Allen and Sondhi(1979), Mammano and Nobili(1993), Vetešník and Nobili(2006)]. It is a box-model of the cochlea which takes into account the height of the cochlear duct. The basilar membrane (BM) is approximated by an array of fluid coupled oscillators whose displacement $\xi(x',t)$ at the longitudinal position x'[7] and time t is given by

$$m_{\mathrm{oc}}(x')\partial_t{}^2\xi(x',t) + h_{\mathrm{oc}}(x)\partial_t\xi(x',t) - [\partial_{x'}s_{\mathrm{oc}}(x)\partial_{x'}]\partial_t\xi(x,t) + k_{\mathrm{oc}}(x)\xi(x,t) + \int_0^{L_{\mathrm{BM}}} G(x',\bar{x}')\partial_t{}^2\xi(\bar{x}',t)d\bar{x}' = -G_S(x')\partial_t^2\zeta(t) - U_{\mathrm{OHC}}(x',t), \tag{39}$$

where ∂_t and $\partial_{x'}$ denote partial derivatives with respect to t and x', respectively, $m_{\mathrm{oc}}(x')$ is the mass, $h_{\mathrm{oc}}(x')$ is the damping, and $k_{\mathrm{oc}}(x')$ is the stiffness per unit BM length. $\partial_{x'}s_{\mathrm{oc}}(x')\partial_{x'}$ accounts for the shearing viscosity between adjacent BM segments and Green's functions $G(x',\bar{x}')$, and $G_S(x')$ accounts for the BM-BM hydrodynamic coupling and the stapes-BM hydrodynamic coupling, respectively; $\zeta(t)$ is the stapes displacement. As in [Nobili and Mammano(1996)], $U(x',t)$ simulates the OHC electromechanical feedback force given by

$$U_{\mathrm{OHC}}(x',t) = u_{\mathrm{OHC}}(x')S[a\eta(x',t)], \tag{40}$$

where $u_{\mathrm{OHC}}(x)$ is a suitable spatial function controlling the degree of amplification along the BM and $S[\cdot]$ is a sigmoidal function proportional to the 2nd-order Boltzmann function, and $\eta(x,t)$ is the OHC stereocilia radial deflection calculated as the displacement of a damped harmonic oscillator

$$\partial_t^2\eta(x',t) + \gamma_{\mathrm{TM}}(x')\partial_t\eta(x',t) + \omega_{\mathrm{TM}}^2(x')\eta(x',t) = -\partial_t^2\xi(x',t), \tag{41}$$

forced by the negative BM acceleration $\partial_t^2\xi(x',t)$; $\gamma_{\mathrm{TM}}(x')$ is the damping resulting from the viscosity of the subtectorial space and the TM viscoelasticity and $\omega_{\mathrm{TM}}(x')$ is the TM resonance frequency at x'.

Due to the sigmoidal function $S[.]$, the model is nonlinear. The model parameters were set to work in the range of levels that are common for mammalian cochlea; nonlinearity in the input/output function of the simulated BM displacement is reached for levels above about 30 dB SPL [see Fig. 1 in [Vencovský *et al.*(2019)]]. The gain of the model is about 50 dB at frequencies between 1 and 5.5 kHz, which might be assumed to simulate normal-hearing cochlea at least in that frequency region. The cochlear model is coupled with a middle ear model. Therefore, the OAEs can be derived from the model as pressure changes at the ear drum. However, the model was not calibrated to predict BM displacement in physically correct units. Hence, the OAEs are expressed in arbitrary units (a.u.).

In order to simulate the interference between the nonlinear-distortion and coherent-reflection components of DPOAEs [Shera and Guinan(1999)], mechanical irregularities (perturbations) were introduced by Gaussian randomization of the undamping force. Such created roughness was introduced into the undamping term $u_{\mathrm{OHC}}(x)$ [see Eq. (3) in [Vencovský *et al.*(2019)]] by

$$\tilde{u}_{\mathrm{OHC}}(x) = u_{\mathrm{OHC}}(x)[1 + \epsilon\cdot\mathcal{N}(0,1)], \tag{42}$$

where the parameter $\epsilon = 0.05$ scales the roughness which is Gaussian distributed with zero mean and variance of unity.

[7] The symbol x' was chosen because we wanted to keep the notation x for the longitudinal position along the BM, but we wanted to distinguish between x' and x used as a symbol for the input signal in Sec. 2

The model is implemented in Matlab. All simulations in the presented paper were done numerically with an explicit Runge-Kutta (4,5) integration algorithm for 600 kHz sampling frequency. The model was composed of 800 segments.

References

[OAE()] "OAETOOLBOX" https://gitlab.com/simonhenin/oaetoolbox/, accessed: 2019-09-30.

[Abdala *et al.*(2015)] Abdala, C., Luo, P., and Shera, C. A. (**2015**). "Optimizing swept-tone protocols for recording distortion-product otoacoustic emissions in adults and newborns," J. Acoust. Soc. Am. **138**(6), 3785–3799, https://doi.org/10.1121/1.4937611.

[Allen and Sondhi(1979)] Allen, J. B., and Sondhi, M. M. (**1979**). "Cochlear macromechanics: Time domain solutions," J. Acoust. Soc. Am. **66**(1), 123–132, https://doi.org/10.1121/1.383064.

[Bergevin *et al.*(2012)] Bergevin, C., Walsh, E. J., McGee, J., and Shera, C. A. (**2012**). "Probing cochlear tuning and tonotopy in the tiger using otoacoustic emissions," Journal of Comparative Physiology A **198**(8), 617–624, https://doi.org/10.1007/s00359-012-0734-1.

[Boege and Janssen(2002)] Boege, P., and Janssen, T. (**2002**). "Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission i/o-functions in normal and cochlear hearing loss ears," J. Acoust. Soc. Am. **111**(4), 1810–1818, https://doi.org/10.1121/1.1460923.

[Brown *et al.*(1996)] Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). "Two sources of acoustic distortion products from the human cochlea," J. Acoust. Soc. Am. **100**(5), 3260–3267, https://doi.org/10.1121/1.417209.

[Choi *et al.*(2008)] Choi, Y.-S., Lee, S.-Y., Parham, K., Neely, S. T., and Kim, D. O. (**2008**). "Stimulus-frequency otoacoustic emission: Measurements in humans and simulations with an active cochlear model," J. Acoust. Soc. Am. **123**(5), 2651–2669, https://doi.org/10.1121/1.2902184.

[Dalhoff *et al.*(2013)] Dalhoff, E., Turcanu, D., Vetešník, A., and Gummer, A. W. (**2013**). "Two-source interference as the major reason for auditory-threshold estimation error based on dpoae input–output functions in normal-hearing subjects," Hear. Res. **296**, 67–82, https://www.sciencedirect.com/science/article/pii/S0378595512002961.

[Dhar *et al.*(2002)] Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**). "Multiple internal reflections in the cochlea and their effect on dpoae fine structure," J. Acoust. Soc. Am. **112**(6), 2882–2897, https://doi.org/10.1121/1.1516757.

[Fang and Liu(2022)] Fang, C.-H., and Liu, Y.-W. (**2022**). "A point-wise artifact rejection method for estimating transient-evoked otoacoustic emissions and their group delay," JASA Express Letters **2**(2), 024401, https://doi.org/10.1121/10.0009393.

[Farina(2000)] Farina, A. (**2000**). "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, http://www.aes.org/e-lib/browse.cfm?elib=10211.

[Gaskill and Brown(1990)] Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, 2f1-f2, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**(2), 821–839, https://doi.org/10.1121/1.399732.

[Goldstein(1967)] Goldstein, J. L. (**1967**). "Auditory nonlinearity," J. Acoust. Soc. Am. **41**(3), 676–699, https://doi.org/10.1121/1.1910396.

[He and Schmiedt(1997)] He, N.-j., and Schmiedt, R. A. (**1997**). "Fine structure of the 2f1-f2 acoustic distortion product: Effects of primary level and frequency ratios," J. Acoust. Soc. Am. **101**(6), 3554–3565, https://doi.org/10.1121/1.418316.

[Heitmann *et al.*(1998)] Heitmann, J., Waldmann, B., Schnitzler, H.-U., Plinkert, P. K., and Zenner, H.-P. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1 - f_2$ removes DP-gram fine structure-Evidence for a secondary generator," J. Acoust. Soc. Am. **103**(3), 1527–1531, https://doi.org/10.1121/1.421290.

[Johnstone *et al.*(1986)] Johnstone, B., Patuzzi, R., and Yates, G. K. (**1986**). "Basilar membrane measurements and the travelling wave," Hear. Res. **22**(1), 147 – 153, http://www.sciencedirect.com/science/article/pii/0378595586900900.

[Kalluri and Shera(2001)] Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for dpoae generation," J. Acoust. Soc. Am. **109**(2), 622–637, https://doi.org/10.1121/1.1334597.

[Kalluri and Shera(2013)] Kalluri, R., and Shera, C. A. (**2013**). "Measuring stimulus-frequency otoacoustic emissions using swept tones," J. Acoust. Soc. Am. **134**(1), 356–368, https://doi.org/10.1121/1.4807505.

[Kemp(1978)] Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**(5), 1386–1391, https://doi.org/10.1121/1.382104.

[Kemp and Brown(1983)] Kemp, D. T., and Brown, A. M. (**1983**). "An integrated view of cochlear mechanical nonlinearities observable from the ear canal," in *Mechanics of Hearing*, edited by E. de Boer and M. A. Viergever, Springer Netherlands, Dordrecht, pp. 75–82.

[Knight and Kemp(2001)] Knight, R. D., and Kemp, D. T. (**2001**). "Wave and place fixed dpoae maps of the human ear," J. Acoust. Soc. Am. **109**(4), 1513–1525, https://asa.scitation.org/doi/abs/10.1121/1.1354197.

[Konrad-Martin *et al.*(2001)] Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast fourier transforms in normal ears," The Journal of the Acoustical Society of America **109**(6), 2862–2879, https://doi.org/10.1121/1.1370356.

[Long *et al.*(2008)] Long, G. R., Talmadge, C. L., and Lee, J. (**2008**). "Measuring distortion product otoacoustic emissions using continuously sweeping primaries," J. Acoust. Soc. Am. **124**(3), 1613–1626, https://doi.org/10.1121/1.2949505.

[Mammano and Nobili(1993)] Mammano, F., and Nobili, R. (**1993**). "Biophysics of the cochlea: Linear approximation," J. Acoust. Soc. Am. **93**(6), 3320–3332, https://doi.org/10.1121/1.405716.

[Mauermann and Kollmeier(2004)] Mauermann, M., and Kollmeier, B. (**2004**). "Distortion product otoacoustic emission (dpoae) input/output functions and the influence of the second dpoae source," J. Acoust. Soc. Am. **116**(4), 2199–2212, https://doi.org/10.1121/1.1791719.

[Moleti *et al.*(2012)] Moleti, A., Longo, F., and Sisto, R. (**2012**). "Time-frequency domain filtering of evoked otoacoustic emissions," J. Acoust. Soc. Am. **132**(4), 2455–2467, https://doi.org/10.1121/1.4751537.

[Nelson and Kimberley(1992)] Nelson, D. A., and Kimberley, B. P. (**1992**). "Distortion-product emissions and auditory sensitivity in human ears with normal hearing and cochlear hearing loss," Journal of Speech, Language, and Hearing Research **35**(5), 1142–1159, https://pubs.asha.org/doi/abs/10.1044/jshr.3505.1142.

[Nobili and Mammano(1996)] Nobili, R., and Mammano, F. (**1996**). "Biophysics of the cochlea ii: Stationary nonlinear phenomenology," J. Acoust. Soc. Am. **99**(4), 2244–2255, https://doi.org/10.1121/1.415412.

[Novak *et al.*(2015)] Novak, A., Lotton, P., and Simon, L. (**2015**). "Synchronized swept-sine: Theory, application, and implementation," J. Audio Eng. Soc **63**(10), 786–798, http://www.aes.org/e-lib/browse.cfm?elib=18042.

[Novak *et al.*(2010)] Novak, A., Simon, L., Kadlec, F. s., and Lotton, P. (**2010**). "Nonlinear system identification using exponential swept-sine signal," IEEE Trans. Instrum. Meas. **59**(8), 2220–2229.

[Probst *et al.*(1991)] Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**(5), 2027–2067, https://doi.org/10.1121/1.400897.

[Rhode(1978)] Rhode, W. S. (**1978**). "Some observations on cochlear mechanics," J. Acoust. Soc. Am. **64**(1), 158–176, https://doi.org/10.1121/1.381981.

[Robles and Ruggero(2001)] Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Phys. Rev. **81**(3), 1305–1352, https://doi.org/10.1152/physrev.2001.81.3.1305, pMID: 11427697.

[Shaffer *et al.*(2003)] Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. M. (**2003**). "Sources and mechanisms of dpoae generation: Implications for the prediction of auditory sensitivity," Ear and Hearing **24**(5), https://journals.lww.com/ear-hearing/Fulltext/2003/10000/Sources_and_Mechanisms_of_DPOAE_Generation_.3.aspx.

[Shera and Bergevin(2012)] Shera, C. A., and Bergevin, C. (**2012**). "Obtaining reliable phase-gradient delays from otoacoustic emission data," J. Acoust. Soc. Am. **132**(2), 927–943, https://doi.org/10.1121/1.4730916.

[Shera and Guinan(1999)] Shera, C. A., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**(2), 782–798, https://doi.org/10.1121/1.426948.

[Shera and Guinan(2003)] Shera, C. A., and Guinan, J. J. (**2003**). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," The Journal of the Acoustical Society of America **113**(5), 2762–2772, https://doi.org/10.1121/1.1557211.

[Shera and Zweig(1991)] Shera, C. A., and Zweig, G. (**1991**). "Reflection of retrograde waves within the cochlea and at the stapes," J. Acoust. Soc. Am. **89**(3), 1290–1305, https://doi.org/10.1121/1.400654.

[Shera and Zweig(1993)] Shera, C. A., and Zweig, G. (**1993**). "Noninvasive measurement of the cochlear traveling-wave ratio," J. Acoust. Soc. Am. **93**(6), 3333–3352, https://doi.org/10.1121/1.405717.

[Sondhi(1978)] Sondhi, M. M. (**1978**). "Method for computing motion in a two-dimensional cochlear model," J. Acoust. Soc. Am. **63**(5), 1468–1477, https://doi.org/10.1121/1.381893.

[Stover *et al.*(1996)] Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). "Latency and multiple sources of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **99**(2), 1016–1024, https://doi.org/10.1121/1.414630.

[Vencovský *et al.*(2019)] Vencovský, V., Zelle, D., Dalhoff, E., Gummer, A. W., and Vetešník, A. (**2019**). "The influence of distributed source regions in the formation of the nonlinear distortion component of cubic distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **145**(5), 2909–2931, https://doi.org/10.1121/1.5100611.

[Vetešník *et al.*(2022)] Vetešník, A., Vencovský, V., and Gummer, A. W. (**2022**). "An additional source of distortion-product otoacoustic emissions from perturbation of nonlinear force by reflection from inhomogeneities," The Journal of the Acoustical Society of America **152**(3), 1660–1682, https://doi.org/10.1121/10.0013992.

[Vetešník and Nobili(2006)] Vetešník, A., and Nobili, R. (**2006**). "The approximate scaling law of the cochlea box model," Hear. Res. **222**(1), 43–53, https://www.sciencedirect.com/science/article/pii/S0378595506002280.

[Vetešník *et al.*(2009)] Vetešník, A., Turcanu, D., Dalhoff, E., and Gummer, A. W. (**2009**). "Extraction of sources of distortion product otoacoustic emissions by onset-decomposition," Hear. Res. **256**(1), 21 – 38, http://www.sciencedirect.com/science/article/pii/S0378595509001415.

[Zelle *et al.*(2020)] Zelle, D., Bader, K., Dierkes, L., Gummer, A. W., and Dalhoff, E. (**2020**). "Derivation of input-output functions from distortion-product otoacoustic emission level maps," J. Acoust. Soc. Am. **147**(5), 3169–3187, https://doi.org/10.1121/10.0001142.

[Zelle *et al.*(2017)] Zelle, D., Lorenz, L., Thiericke, J. P., Gummer, A. W., and Dalhoff, E. (**2017**). "Input-output functions of the nonlinear-distortion component of distortion-product otoacoustic emissions in normal and hearing-impaired human ears," J. Acoust. Soc. Am. **141**(5), 3203–3219, https://doi.org/10.1121/1.4982923.