CANCELLING INTERMODULATION DISTORTIONS FOR OTOACOUSTIC EMISSION MEASUREMENTS WITH EARBUDS

Berken Utku Demirel, Khaldoon Al-Naimi, Fahim Kawsar, Alessandro Montanari

Nokia Bell Labs, Cambridge (UK)

ABSTRACT

This paper presents a novel cancellation method of Intermodulation Distortions (IMDs) for earbud speakers used to measure Distortion Product Otoacoustic Emissions (DPOAE). Speakers' non-linear behaviour is a significant problem for earbuds with small loudspeakers due to limitations in cone movement. Linear and non-linear speaker modelling enables us to inject exact distortion inverse to cancel what is introduced by speakers' non-linearities. Our proposed method is compared against state-of-the-art related works in terms of harmonic reduction ratio. Simulation results and evaluation on real hardware show a 77% to 95% reduction in the harmonic distortions of a focused frequency region at the output of the loudspeaker, outperforming existing works by 6% to 14%.

Index Terms— Otoacoustic Emission, Intermodulation Distortion, Nonlinear Speakers.

1. INTRODUCTION

According to the World Health Organisation about 5% of the world population suffer from hearing impairments and it is estimated that this number is set to increase within the next 30 years, reaching the worrisome ratio of 1 hearing-impaired person in every 10 people. Additionally, noise-induced hearing loss (NIHL) is a growing health concern nationally and globally[1]. Most hearing loss occurs as a result of damage to the cochlea, specifically the outer hair cells.

Otoacoustic emissions (OAEs) are sounds generated from the outer hair cells (OHCs) in the cochlea, that propagate outward from the ear [1]. Evoked OAEs (EOAEs) [2], which are acoustical responses to an external stimulus, enable the study of the cochlea's ability to process signals normally at different frequencies by using different excitation signals. EOAEs have been widely used as a non-invasive tool for measuring the cochlear mechanics in a frequency-specific manner for newborns [3], as well as a possible objective tool for evaluating hearing sensitivity level [4]. Measuring OAE does not require active cooperation from subjects (no user feedback is needed) as compared to audiometry where subjects need to provide feedback (i.e. clicking a button) to evaluate how well they can hear. OAE testing equipment is expensive and requires specialised operators, which limits hearing screening in developing countries [5]. Using off-the-shelf and affordable earbuds to evaluate OAEs could usher in a new era of mobile sensing, making hearing screening more accessible and continuous. This approach aligns with the growing trend of using off-the-shelf earbuds for vital signs sensing [6, 7, 8].

Different OAE measurements have been proposed in the literature. The most widely used method for detecting hearing loss at specific frequencies is the Distortion Product Otoacoustic Emission (DPOAE), which is a response generated when the cochlea is stimulated simultaneously by two pure tone frequencies and can be recorded from the ear canal using a microphone [2, 9]. Current medical devices that measure DPOAEs use a single microphone with two loudspeakers [10, 11]. This configuration allows the generation of two pure tones without any other distortion (Figure 1, top). However, to measure DPOAEs from off-the-shelf earbuds only a single loudspeaker can be used, due to space constraints inside the devices. This introduces the challenge of removing Intermodulation Distortions (IMDs) which result from two or more signals being present at the input of a non-linear device (i.e. the loudspeaker inside earbuds). This issue is a bigger problem for earbuds compared to the classic loudspeakers which are large enough to allow for a large movement of the cone, enabling the production of sound at various frequencies. The small loudspeakers used in earbuds show higher non-linear behaviours due to limitations in the movement of the cone [12]. Crucially, the frequency location of the cubic distortion $(2f_1 - f_2)$ due to the loudspeaker non-linearities is at the same location where the cochlea response is measured for DPOAEs, as shown in Figure 1. The detection of OAEs requires a stimulus without any distortion. To avoid masking the emission from the cochlea, distortion cancellation is crucial to eliminate speaker non-linearities appearing at the same frequency as DPOAEs.

Volterra filters (VF) have been used to model loudspeakers' non-linearities in order to reduce distortions [13, 14]. These approaches are adapted from an established technique for linearising conventional loudspeakers [15]. A significant drawback of Volterra kernels is that their weights need to be continually adapted as the input level changes [16]. This is especially true for the amplitude of low-order harmonic components (IMDs) resulting from high-order non-linearity which considerably varies over time [17], leading to accuracy degradation of the designed model. Therefore, if Volterra filters are used to model loudspeaker non-linearities in a OAE detection system (to predominately reduce 2nd and 3rd order harmonics [18]), its weights need to be continuously calibrated prior to performing the DPOAE test, due to person to person ear-canal variation and the need to have certain dB SPL intensities for the tone pair at the ear-drum [19, 9]. Additionally, Volterra filter implementation in practical systems is challenging due to its high computational complexity. The number of coefficients grows exponentially with the order of the VF. This has led researchers to develop simplified versions of Volterra filters, such as the One-Dimension Volterra Filter (ODVF) [20] at a cost of reduced performance, even though still requiring around 200 taps [13].

Previous methods achieve IMD cancellation by injecting an anti-distortion signal into the speaker's original signal [21, 22]. While this method works well for loudspeakers, earbuds present a more challenging working environment, as shown in Figure 5a. Speakers and microphones in an earbud are in such close proximity to one another that the microphone not only captures the non-linear speaker distortions but also ear-canal reflections and OAE emissions. Addi-

[1] https://www.who.int/publications-detail-redirect/world-report-on-hearing

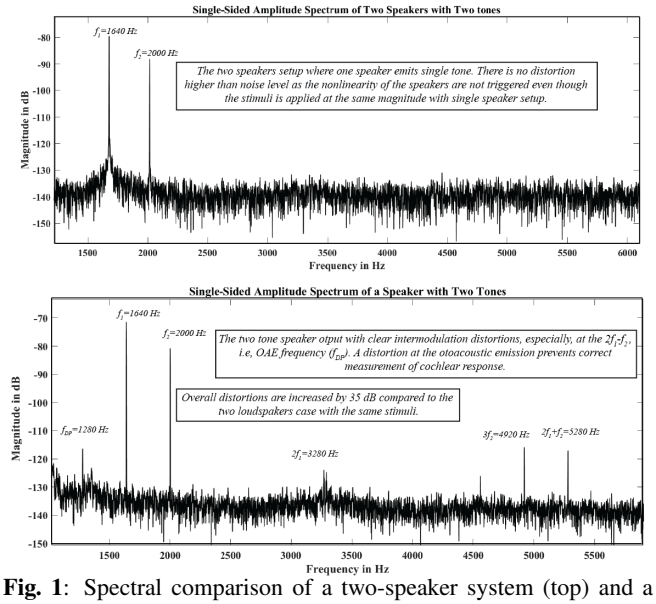


Fig. 1: Spectral comparison of a two-speaker system (top) and a single speaker (bottom) using two tones stimuli.

tional hardware has been suggested as a possible solution [23, 24], which cannot be generalised and poses manufacturing problems. A significant drawback of previous solutions is that none of them focused on decreasing specific frequency distortion components, which has crucial importance for OAE measurement. There is, therefore, a strong need to design an accurate and frequency region-focused distortion cancellation method that can adapt to different ear canal geometries to enable OAE detection with earbuds.

In this work, we propose a frequency region-focused modelling method to cancel IMDs for time-varying input signals as an initial step toward measuring OAEs with earbuds. Since earbuds are ubiquitous, using them to measure OAEs opens a new era for continuous hearing assessment and enhancement. Results show that our proposed method achieves a high percentage reduction in IMD. The contributions of our work are summarised as follows:

- We propose a new IMD cancellation method for DPOAE measurement using earbuds with a single speaker.
- We propose a novel and efficient pipeline by utilising multi-rate signal processing to inject an optimised second sinusoidal for cancelling the IMDs.
- We compare simulations of the proposed solution against widely accepted speaker models as well as data collected from our prototype, both of which show that our method achieves a substantial reduction between 77% and 95% in the harmonic distortions at the output of the loudspeaker while outperforming existing works by approximately 6% to 14%.

2. TOWARDS IMD CANCELLATION FOR EARBUDS

In this work, we focus on DPOAEs which are part of evoked OAEs and can be performed under moderately noisy conditions [2]. To measure DPOAEs, stimuli, composed of two primary tones f_1 and f_2 generally chosen according to $f_2/f_1 = 1.22$ formula, with intensities L_1 and L_2 are sent to the ear canal. The most prominent distortion product would occur at $2f_1 - f_2$ and it can be recorded from the ear canal using a microphone. The excitation region of the frequency pair mainly overlaps near the f_2 characteristic region in the cochlea, and the sound-pressure level (SPL) of DPOAE at $f_{DP} = 2f_1 - f_2$

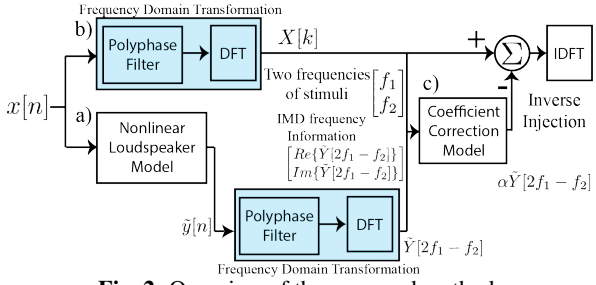


Fig. 2: Overview of the proposed method.

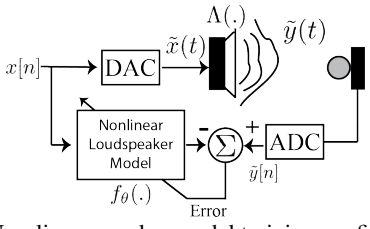


Fig. 3: Non-linear speaker model training configuration.

shows the cochlea's ability to process signals at frequency f_2. This can be used to assess a person hearing health. As shown in Figure 1, when using a single speaker to emit f_1 and f_2, distortions emerge at the same f_{DP} frequency, making it impossible to measure DPOAEs.

In our proposed solution, we present an IMD cancellation method that is frequency region-focused for use in DPOAE. The solution is divided into three parts as shown in Figure 2, namely:

- modelling the non-linear speaker characteristics to predict IMDs;
- frequency region-focused transformation (such as Discrete Fourier Transformation) which results in magnitude and phase information for both the $x[n]$ and $\tilde{y}[n]$;
- determining the correction coefficient for multiplying the secondary waveform before mixing it back into $X[k]$.

Finally, the signal that is composed of stimuli and cancellation sinusoid is transformed back to the time domain through Inverse DFT (IDFT) and sent to the speaker.

2.1. Non-linear Loudspeaker Model

We modified the widely used Wave-U-Net architecture [25] with a single source and 4 layers to approximate the non-linear speaker function, $\Lambda(.)$. During the training phase, the desired signal $(x[n])$ (a composite signal of two tones f_1 and f_2 at intensities L_1 and L_2 as defined by the DPOAE method) is simultaneously sent to the loudspeaker and the model. Model weights are updated to decrease the Mean Square Error (MSE) between the model output, $f_\theta(x[n])$, and the speaker output, $\tilde{y}[n]$, measured from the in-ear microphone, as shown in Figure 3. For the simulations, $\tilde{y}[n]$ is obtained by feeding the desired signal to corresponding speaker functions.

2.2. Frequency Domain Transformation

For frequency domain transformation, we combine polyphase filter banks (channeliser) with Discrete Fourier Transform (DFT). This increases efficiency compared to the direct DFT, and provides good resolution and improved accuracy when compared to Welch-based estimators. In the transformation block, we aim to focus on a specific frequency region, rather than the whole spectra. The frequency bands of interest are known, i.e. $(f_1, f_2,$ and $2f_1 - f_2)$ for both stimuli and distorted signals. Therefore, the resolution of spectral

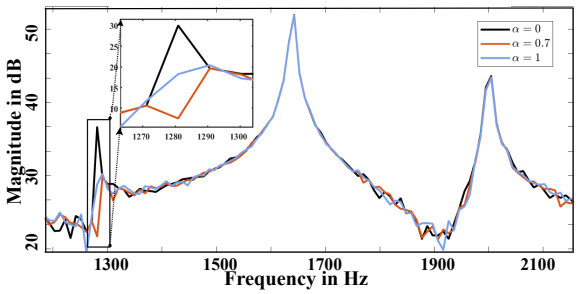


Fig. 4: Effect of coefficient α to cancel distortion at f_{DP} = 1.28 kHz.

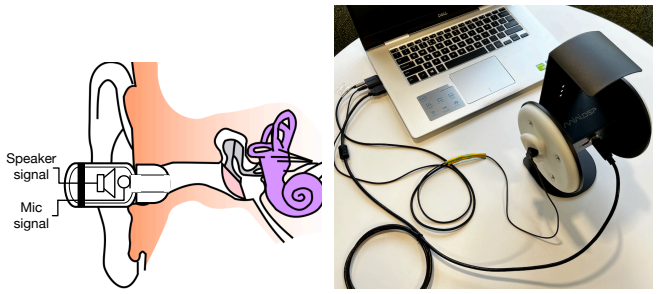
analysis is increased at those bands rather than the complete spectrum. In our system, we have used a 4-phase polyphase FIR filter and an 8-point FFT to divide the broadband signal into 4 sub-bands corresponding to each DPOAE measurement frequency. The desired frequency band is fed to a 32-band filter bank estimator that contains a 32-phase polyphase FIR filter and a 256-point FFT is used with the expected OAEs band in order to compute the spectral estimate with higher resolution.

2.3. Coefficient Correction Model

We assume that the distortion portion added (i.e., for IMD cancellation) to the original stimuli will not generate further distortions because of its small amplitude and therefore within the linear region of operation for the speaker. Variability among users (due to outer-ear and ear-canal size and shape differences) is captured through the correction coefficient leading to the best IMD cancellation performance. We have used a data-driven method to find the best correction coefficient, (α). The coefficient correction model is trained after the non-linear loudspeaker model has been trained.

The designed model is a multi-layer perceptron that has 3 layers, including input. The model input is a 4 by 1 vector while the output is the coefficient, α. The first 2 values of the input vector correspond to the sinusoid frequencies (f_1, f_2) (i.e. DPOAE stimuli frequencies). The frequencies are given in kHz to prevent gradient explosion. The last 2 values are the real $(\Re\tilde{Y}[2f_1 - f_2])$ and imaginary $(\Im\tilde{Y}[2f_1 - f_2])$ parts of the distortion at the specific frequency, i.e., $2f_1 - f_2$, to give information about the phase and magnitude of the distortion. As the coefficient correction model is a data-driven solution, it needs the best coefficient values corresponding to the input vectors for training. Empirically, we have found that the best coefficient values are between 0.6 and 1.4. Therefore we have searched over values starting from 0.4 to 1.7 with a 0.02 increment. The best value is chosen by comparing the distortion reduction ratio. Each input vector to the model and its best correction coefficient is used to train the model to reduce the error between the best-measured correction coefficient (labelled) and the system output (model prediction). The effect of the coefficient correction model can easily be seen in Figure 4. If the magnitude of the injected sinusoidal is set to the same as the distortion signal $\alpha = 1$, the distortion reduction ratio is less than when $\alpha = 0.7$.

Figure 2 shows how the two models fit within the entire system. Once both models converged, the desired signal with f_1 and f_2 is sent to the non-linear speaker model to predict the expected IMDs at the DPOAE frequency (i.e., $\tilde{Y}[2f_1 - f_2]$). The input signal is also sent to the frequency domain transformation to obtain $X[k]$. $X[k]$ and the real and imaginary parts of the predicted distortion are fed to the Coefficient Correction Model for predicting the optimal coefficient. The obtained correction coefficient is used to generate an additional sinusoidal signal at the same frequency as the distortion in an inverse phase. The generated signal is added to the desired signal



(a) Prototype earbud. (b) Test setup.

Fig. 5: Experimental system setup.

in the frequency domain and sent to the IFFT unit to be converted for loudspeaker input.

3. EVALUATION

3.1. Experimental Setup

We evaluate our method both in simulations and with an earbud prototype that embeds a microphone and a speaker as shown in Figure 5a. Below are the details of our experimental setup.

Non-linear Speaker models: We use three widely known non-linear speaker models in our simulations. The first model is an asymmetric loudspeaker [26] where the input-output relation is defined as:

$$r(n) = \gamma \left[\frac{1}{1 + e^{-\theta r_1(n)}} - \frac{1}{2} \right] \quad (1)$$

where, $r_1(n) = 1.5x(n) - 0.3x^2(n)$ and $\gamma = 2$ and θ is 4 if $r_1(n)$ is positive, and 0.5 if $r_1(n)$ is less or equal than zero.

The second non-linear model is given by Equation 2 which is a symmetrical soft-clipping non-linearity model [27].

$$r(n) = \begin{cases} \frac{2}{3\tau}x(n) & 0 \le |x(n)| < \tau \\ [x(n)] \ \frac{3-[2-x(n)\tau]^2}{3} & \tau \le |x(n)| < 2\tau \\ [x(n)] & 2\tau \le |x(n)| < 1 \end{cases} \quad (2)$$

with $\tau \in (0, 0.5]$. Here $\tau = 0.3$ is considered. The third model used is, sinusoidal non-linearity [28], as defined by Equation 3.

$$r(n) = e^{\frac{x(n)}{2}} \left[sin[\pi x(n)] + 0.3sin[3\pi x(n)] + 0.1sin[5\pi x(n)] \right] \quad (3)$$

These are non-linear models commonly used in practical loudspeakers modelling [29] and hence are used in the simulations in this paper. In the simulations, the input signal $x[n]$ is considered to be a DPOAE stimulus where f_2 changes from 1 kHz to 8 kHz with integer multiplies. The corresponding tone magnitude ranges between $0.1 - 0.5$ to simulate different ear canal geometries in humans. The ratio of the first to second tone was kept the same since it was proven that the DPOAEs are observed at the maximum level for $L_1 = 65$ and $L_2 = 55$ dB SPL where L_1 and L_2 are the stimulus magnitudes [2]. The output $r(n)$ of the non-linear system is corrupted with a white Gaussian noise $v(n)$ with an SNR of 30 dB.

Ear drum simulator: For the evaluation using the earbud prototype, we simulate a realistic ear canal environment using the miniDSP EARS[2] as shown in Figure 5b. Since the sound pressures at the eardrum should only be composed of two sinusoidal for the DPOAE measurements, the microphone output of the miniDSP EARS is used as an eardrum for calculating the overall distortion reduction.

Table 1: Effectiveness of reducing the distortion in percentage. Each cell displays the mean ± standard deviation computed over the testing dataset which includes different frequencies and magnitudes.

Methods	Speaker Models			
	Asymmetric	Symmetrical	Sinusoidal	Real
Ours	97.80 ± 4.7	90 ± 3.5	91 ± 3.1	77 ± 5.2
Neural Network [30]	51.5 ± 7.2	39 ± 8.3	40.4 ± 6.5	30.8 ± 9.2
ODVF [14]	77 ± 5.4	65.3 ± 5.4	73 ± 4.2	46.2 ± 7.3
SOVF [31]	88.7 ± 5.2	80.5 ± 4.5	85.3 ± 3.9	63.2 ± 6.2

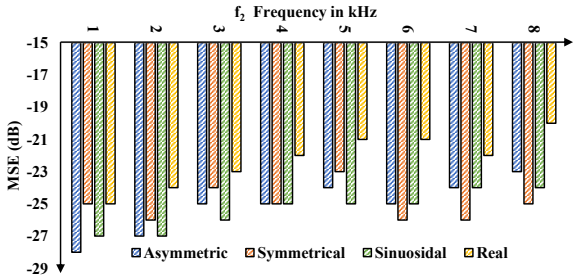


Fig. 6: Mean Squared Error between the output of our non-linear speaker model and actual non-linear speaker output.

Baselines: We compare our proposed solution with three different methods that have been proposed to cancel IMDs. The first method [30], is purely based on neural networks to control a non-linear loudspeaker. We modified the pre-filter part of the architecture according to the DPOAE frequencies. The second method [14], is based on One Dimensional Volterra filters (ODVF) to cancel distortions for measuring OAEs. The last method is the Second-Order conventional Volterra Filter (SOVF) [31] without any modifications. We generated 40000 trials, where 70% of data is used for training and the rest is used for testing, with a duration of 100 ms and 22 kHz sampling rate. For a fair comparison, we use the same number of training and testing data for all approaches and the splitting for training and testing is performed only once.

Performance Metrics: We use the mean of the distortion reduction in percentage on the testing dataset as our main performance metric.

3.2. Results

Non-linear Speaker Model Performance. We evaluated the performance of our speaker model (§ 2.1) for the real earbud and the simulated non-linear loudspeaker models at the different DPOEs we tested. The overall results, reported in Figure 6, show that the MSE between our model predicted output and the actual non-linear speaker output is ≈ -25 dB. This demonstrates that our method models non-linear speaker behaviour accurately.

Complete System Performance. Table 1 compares the performance of our proposed method with state-of-the-art works in distortion reduction for the DPOAE frequencies. We observe that our proposed method outperforms all other works in distortion reduction ratio. Our approach is consistently better at reducing IMDs when evaluated both with simulations, using three non-linear speaker models, and with our earbud prototype (*Real* column in Table 1). We observe that in general, the different approaches (including ours) achieve a higher IMD reduction in simulations rather than when tested on an actual earbud. However, our approach shows the largest reduction (i.e., 14%) when evaluated on the real earbud compared to the best-performing baseline (SOVF).

In addition to lower IMD reduction, other methods have several drawbacks. For example, the solution purely based on Neural Networks (NN) [30], applies an additional filter to the input source data prior to being received into the trained neural network to prevent input distribution shifts for the neural network, which greatly limits the range of possible inputs. This is particularly problematic for DPOAEs which require multiple frequencies to be tested in order to create a comprehensive picture of the user's hearing health. Although Volterra filters show better performance compared to the NN approach, they cannot reach the performance of our method while still requiring a high-computational power as mentioned in § 1. The usage of Volterra filters also brings additional limitations. For instance, it was shown that the optimal filter coefficients $h_3[n]$ are different for different choices of DPOAE frequency (f_2) and magnitudes L_1 and L_2 due to omitting all the off-diagonal elements of the Volterra filter to make it simpler [14].

[2]https://www.minidsp.com/products/acoustic-measurement

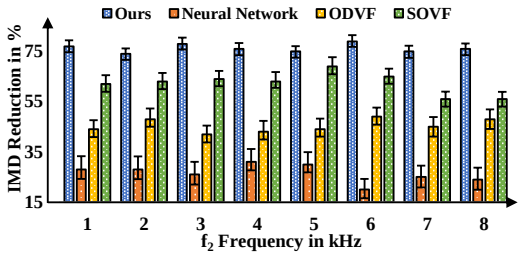


Fig. 7: IMD percentage reduction for each DPOAE frequency.

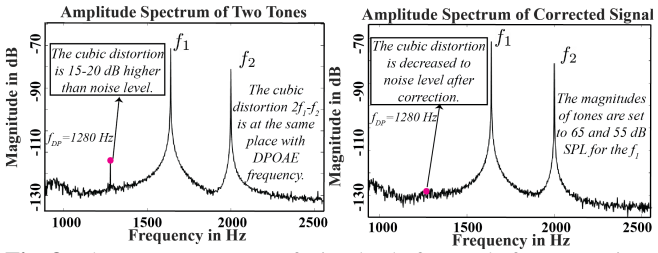


Fig. 8: The power spectrum of stimulus before and after correction.

To further evaluate the performance of our approach in reducing IMDs for DPOAE frequencies, we tested different values of the (f_2) frequency (and consequently (f_1)). The results are reported in Figure 7 where the standard deviations for each method are also indicated. Our proposed method not only achieves the best performance but also has the lowest standard deviation at each frequency. Figure 8 instead, shows the signal with 2 kHz f_2 before and after our method is applied to the stimuli. As shown, the proposed method decreases the IMD at the DPOAE frequency to the noise level, which corresponds to a 20 dB decrease in the distortion frequency. These results show the feasibility of using our approach to reduce IMDs for the purpose of assessing DPOAEs using off-the-shelf earbuds. In future, more factors can be considered, such as hardware diversity, human diversity, and other relevant factors that impact signals.

4. CONCLUSION

We presented a methodology for cancelling IMDs towards enabling OAE measurements with earbuds having a single speaker. Evaluation on real and simulated speakers demonstrates that our proposed method achieves 77–95% reduction in the harmonic distortions at the speaker output while outperforming related works by ≈ 6–14%. We acknowledge that testing on humans is still necessary to fully evaluate the performance of our approach and we aim to address this in future work. Nevertheless, we believe this work is a step towards OAE measurements with off-the-shelf earbuds.

5. REFERENCES

[1] David T Kemp, "Otoacoustic emissions, their origin in cochlear function, and use," *British Medical Bulletin*, vol. 63, no. 1, pp. 223–241, 10 2002.

[2] Caroline Abdala and Leslie Visser-Dumont, "Distortion Product Otoacoustic Emissions: A Tool for Hearing Assessment and Scientific Study," *The Volta review*, vol. 103, 2001.

[3] M. Suckfüll, S. Schneewei, A. Dreher, and K. Schorn, "Evaluation of TEOAE and DPOAE Measurements for the Assessment of Auditory Thresholds in Sensorineural Hearing Loss," *Acta Oto-Laryngologica*, vol. 116, no. 4, pp. 528–533, 1996.

[4] D. Zelle, E. Dalhoff, and A. W. Gummer, "Objective audiometry with DPOAEs," *HNO*, vol. 65, no. 2, pp. 122–129, 2017.

[5] Amina Seguya, Francis Bajunirwe, Elijah Kakande, and Doreen Nakku, "Feasibility of establishing an infant hearing screening program and measuring hearing loss among infants at a regional referral hospital in south western uganda," *PloS one*, vol. 16, no. 6, pp. e0253305, 2021.

[6] Andrea Ferlini, Alessandro Montanari, Chulhong Min, Hongwei Li, Ugo Sassi, and Fahim Kawsar, "In-ear ppg for vital signs," *IEEE Pervasive Computing*, , no. 01, dec 2021.

[7] Ananta Narayanan Balaji, Andrea Ferlini, Fahim Kawsar, and Alessandro Montanari, "Stereo-bp: Non-invasive blood pressure sensing with earables," in *Proceedings of the 24th International Workshop on Mobile Computing Systems and Applications*, 2023, pp. 96–102.

[8] Hoang Truong, Alessandro Montanari, and Fahim Kawsar, "Non-invasive blood pressure monitoring with multi-modal in-ear sensing," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 6–10.

[9] Stephen T. Neely, Tiffany A. Johnson, and Michael P. Gorga, "Distortion-product otoacoustic emission measured with continuously varying stimulus level," *The Journal of the Acoustical Society of America*, vol. 117, no. 3, pp. 1248–1259, 2005.

[10] Jr Kenneth I. Vaden, Lois J. Matthews, and Judy R. Dubno, "Transient-evoked otoacoustic emissions reflect audiometric patterns of age-related hearing loss," *Trends in Hearing*, vol. 22, pp. 2331216518797848, 2018, PMID: 30198420.

[11] Luke John Campbell and Kyle Damon Slater, "Personalization of auditory stimulus," United States 9794672B2. (2017).

[12] Ylva Björk and Ebba Wilhelmsson, "Linearisation of micro loudspeakers using adaptive control," 2014.

[13] Yongsheng Mu, Peifeng Ji, Wei Ji, Ming Wu, and Jun Yang, "Modeling and compensation for the distortion of parametric loudspeakers using a one-dimension volterra filter," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2169–2181, 2014.

[14] Wei-Chen Hsiao, Yung-Ching Chen, and Yi-Wen Liu, "Measuring distortion-product otoacoustic emission with a single loudspeaker in the ear: Stimulus design and signal processing techniques," *Frontiers in Digital Health*, vol. 3, 2021.

[15] G.L. Sicuranza and A. Carini, "Filtered-x affine projection algorithm for multichannel active noise control using second-order volterra filters," *IEEE Signal Processing Letters*, vol. 11, no. 11, pp. 853–857, 2004.

[16] Y Hatano, C Shi, S Kinoshita, and Y Kajikawa, "A study on compensating for the distortion of the parametric array loudspeaker with changing nonlinearity," in *Proc. Western Pacific Acoust. Conf. 2015*, 2015.

[17] Yuta Hatano, Chuang Shi, and Yoshinobu Kajikawa, "Compensation for nonlinear distortion of the frequency modulation-based parametric array loudspeaker," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1709–1717, 2017.

[18] Yongsheng Mu, Peifeng Ji, Wei Ji, Ming Wu, and Jun Yang, "On the linearization of a parametric loudspeaker system by using third-order inverse volterra filters," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*, 2013, pp. 570–574.

[19] Michael R. Stinson and B. W. Lawton, "Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution," *The Journal of the Acoustical Society of America*, vol. 85, no. 6, pp. 2492–2503, 1989.

[20] Angelo Farina, Alberto Bellini, and Enrico Armelloni, "Nonlinear convolution: A new approach for the auralization of distorting systems," *Preprints - Audio Engineering Society*, 2001.

[21] Wolfgang Klippel, "Active attenuation of nonlinear sound," Dec. 1999, US6005952A.

[22] Wolfgang Klippel, "Method and arrangement for controlling an electro-acoustical transducer," Dec. 2019, EP2910032B1.

[23] Wolfgang Klippel, "Arrangement and method for converting an input signal into an output signal and for generating a predefined transfer behavior between said input signal and said output signal," Dec. 2016, US9326066B2.

[24] Wolfgang Klippel, "Methods and apparatus for reduced distortion balanced armature devices," Dec. 2010, US8385583B2.

[25] Daniel Stoller, Sebastian Ewert, and Simon Dixon, "Wave-u-net: A multi-scale neural network for end-to-end audio source separation," *arXiv preprint arXiv:1806.03185*, 2018.

[26] Danilo Comminiello, Michele Scarpiniti, Luis A. Azpicueta-Ruiz, Jerónimo Arenas-García, and Aurelio Uncini, "Functional link adaptive filters for nonlinear acoustic echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1502–1512, 2013.

[27] Danilo Comminiello, Michele Scarpiniti, Luis A. Azpicueta-Ruiz, Jerónimo Arenas-García, and Aurelio Uncini, "Nonlinear acoustic echo cancellation based on sparse functional link representations," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 7, pp. 1172–1183, 2014.

[28] Vinal Patel, Vaibhav Gandhi, Shashank Heda, and Nithin V. George, "Design of adaptive exponential functional link network-based nonlinear filters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2016.

[29] Udo Zölzer, Xavier Amatriain, Daniel Arfib, Jordi Bonada, Giovanni De Poli, Pierre Dutilleux, Gianpaolo Evangelista, Florian Keiler, Alex Loscos, Davide Rocchesso, et al., *DAFX-Digital audio effects*, John Wiley & Sons, 2002.

[30] Pascal M. Brunet and Yuan Li, "Nonlinear control of a loudspeaker with a neural network.," Dec. 2022, US11356773B2.

[31] S. Kinoshita, Y. Kajikawa, and Y. Nomura, "Volterra filters using multirate signal processing and their application to loudspeaker systems," in *IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, 2001.