

# Derivation Technique for Headphone Transfer Functions Based on Sine Sweeps and Least Squares Minimization

Piero Rivera Benois, Purbaditya Bhattacharya and Udo Zölzer

Department of Signal Processing and Communications, Helmut Schmidt University, Germany

Email: piero.rivera.benois@hsu-hh.de

**Abstract**—System identification for ANC headphones considers two transfer functions: one between the outer and inner microphones, called primary path; and one between the control speaker and the inner microphone, known as secondary path. The accuracy of their measurements can be directly translated into the maximum reachable attenuation and also into stability, in case of feedback approaches. Using an exponential sine sweep technique, we compute two auxiliary impulse responses from an external external loudspeaker to the outer and inner microphones attached to the headphone shell. Based on them, the equivalent primary path's finite length impulse response is indirectly derived. For computing the derivation we compare two derivation techniques, one in frequency-domain, based on cross-power and auto-power spectra, and the other one in time-domain, based on an ad hoc adapted Least Squares approach proposed by us. Comparisons between both techniques on a real system with different lengths for the primary path impulse response show a mean squared error reduction of at least 10 dB and up to 24 dB using the proposed derivation.

**Index Terms**—Measurement techniques I-INCE Classification of Subjects Number(s): 72.9.

## I. INTRODUCTION

Active Noise Control (ANC) headphones provide the user with an attenuation of the unwanted environmental noise that surrounds him. This protection is a combined effect of the passive attenuation characteristic of the materials used in the construction of the ear-cups and the active noise control applied over the remaining noise. In practice, the construction materials produce a low-pass filtering effect over the incoming noise, that is capable of attenuating only the medium to high frequency range. The low frequency range, has to be addressed actively by generating destructive superposition of the noise with a sound pressure control signal radiated from the headphone's speaker. To achieve this, the ANC headphones are equipped with an outer microphone in order to have a measured reference of the incoming noise from the outside, and an inner microphone, to sense the remaining noise inside the ear-cup.

The aforementioned microphones and the control speaker define two main transfer functions: the primary path, between the outer and inner microphone; and the secondary path, between the headphone's speaker and the inner microphone (see Fig. 1). Alternatively, a third transfer function, the feed-

back path, can be used in the overall analysis of the system, which is the acoustic path between headphone's speaker and outer microphone. The exactness in magnitude and phase of the measurements of these transfer functions can be directly translated into the maximum achievable noise attenuation that the system may be able to produce [1], and into stability, in case of feedback approaches. In the following a discrete time notation will be used. The impulse response  $h(n)$  of a system will have a transfer function  $H(f)$ , which is the discrete-time Fourier transform of  $h(n)$ .

Transfer function measurements on acoustic systems have improved over the years in several aspects [2] [3] [4], specially since the introduction of the technique presented by Griesinger in [5] and further developed by Farina in [6] and [7]. The usage of long sine sweeps can minimize the influence of harmonic distortions on the impulse response, and allows to almost completely separate them from the linear behavior of the system under test [8]. In particular, the usage of exponential sine sweeps provides higher energy in the low frequency range, where the environmental noise is frequently present, thus improving the SNR of the measurement. These advantages also perfectly suit the requirements of high accuracy in the low-frequency spectrum that ANC headphones transfer functions require.

The measurement of an impulse response between a speaker and a microphone, as the one required for the secondary path, is quite a standard practice for the characterization of an acoustic path. In the case of the primary path, however, there is no speaker involved, only two microphones, and therefore no impulse response per definition. To indirectly derive a transfer function and generate an equivalent impulse response,  $p(n)$ , a physical setup as the one depicted in Fig. 1 is required. Here, the external loudspeaker and both microphones define two auxiliary impulse responses:  $q(n)$ , going from loudspeaker to inner microphone; and  $r(n)$ , going from loudspeaker to outer microphone. Subsequently, these two impulse responses define the primary path,  $p(n)$ , which, similar to a regular impulse response, is valid for a very well defined spatial position of the source and sinks. This means that, although the transfer function is intuitively defined between the two fixed positioned microphones, its validity depends also on the relative position of the source generating the excitation.

This work is an ad hoc adaption of the Least Squares

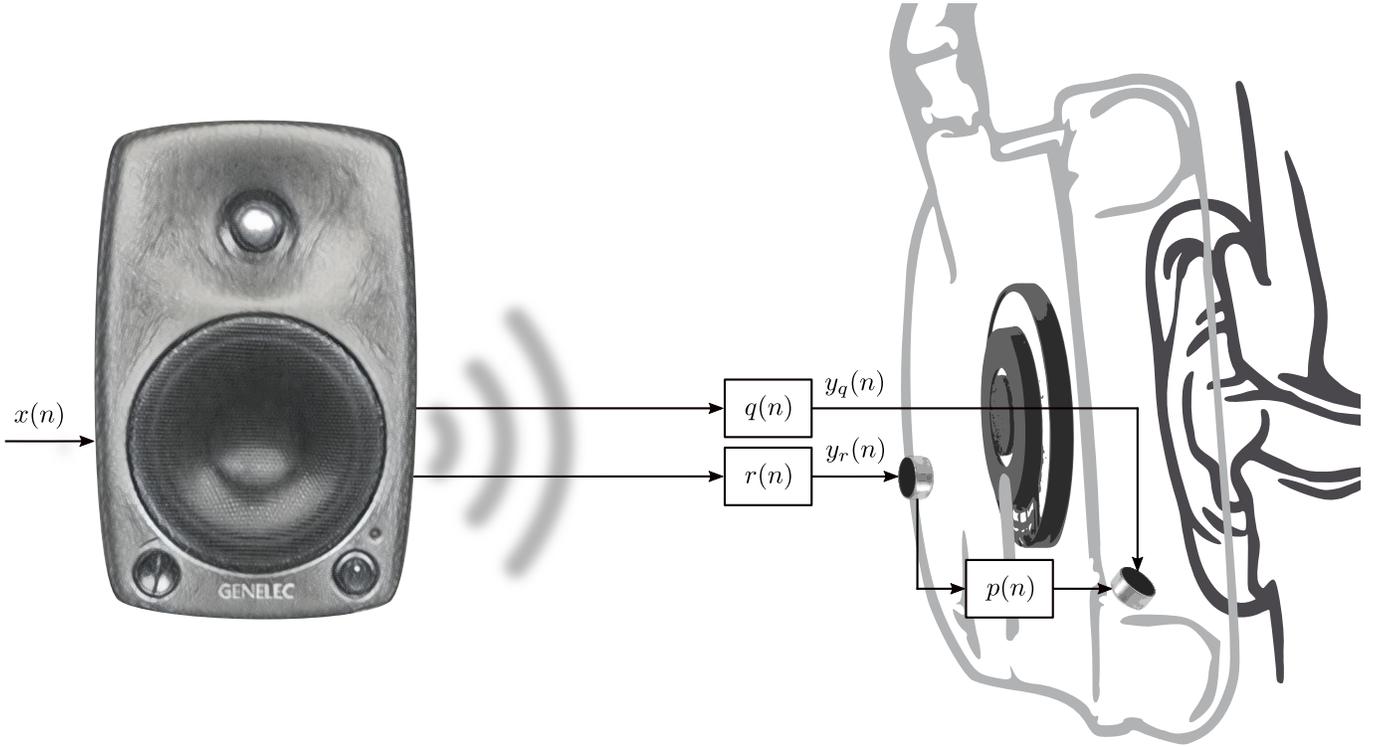


Fig. 1. General physical setup.  $q(n)$  and  $r(n)$  are the auxiliary impulse responses from external loudspeaker to inner and outer microphones used to indirectly derive the transfer function of the primary path,  $p(n)$ .

method (LS) described in [9] for the indirect derivation of the primary path based on auxiliary impulse responses. Its validity and accuracy are evaluated based on the comparison with the standard frequency-domain approach based on cross-power and auto-power spectra (CAPS) used in [10].

In the following sections the theory of the measurement procedure of the auxiliary impulse responses  $q(n)$  and  $r(n)$  is briefly described. This is followed by the description of the Least Squares extension used to indirectly derive an equivalent impulse response,  $p(n)$ , for the primary path. Afterwards, the measurement setup used on a real system is described and the results based on different lengths of  $p(n)$  are presented. At the end, the accuracy of the approach is discussed and conclusions about it are drawn.

## II. THEORY

As extensively described in [6], the exponential sine sweep excitation signal (ESS)

$$x(n) = \sin\left(\frac{\omega_1 \cdot (L_x - 1)}{\ln(\frac{\omega_2}{\omega_1})} \cdot \left(e^{\frac{n}{L_x-1} \cdot \ln(\frac{\omega_2}{\omega_1})} - 1\right)\right), \quad (1)$$

of length  $L_x$ , start angular frequency  $\omega_1$  and stop angular frequency  $\omega_2$ , is generated together with its deconvolution filter [4]

$$x_{inv}(n) = x(L_x - 1 - n) \cdot (\omega_2/\omega_1)^{\frac{-n}{L_x-1}}, \quad (2)$$

to yield the convolution result given by

$$x(n) * x_{inv}(n) = \alpha \cdot \delta(n - L_x), \quad (3)$$

with  $\alpha$  as a scaling factor. Then the system to be measured,  $h(n)$ , may be excited with  $x(n)$

$$y(n) = h(n) * x(n), \quad (4)$$

and its output convolved with  $x_{inv}(n)$  yields

$$y(n) * x_{inv}(n) = h(n) * x(n) * x_{inv}(n) \quad (5)$$

$$y(n) * x_{inv}(n) = \alpha \cdot h(n) * \delta(n - L_x) \quad (6)$$

$$y(n) * x_{inv}(n) = \alpha \cdot h(n - L_x) \quad (7)$$

Applying the above mentioned technique to measure  $r(n)$  and  $q(n)$  simultaneously, the speaker is excited with the input signal  $x(n)$  and subsequently the signals  $y_r(n)$  and  $y_q(n)$  coming from the respective microphones are recorded. Then the approximated time-shifted impulse responses are computed by convolving the measured signals with the deconvolution filter  $x_{inv}(n)$ , as

$$y_q(n) * x_{inv}(n) = \alpha \cdot q(n - L_x) \quad (8)$$

and

$$y_r(n) * x_{inv}(n) = \alpha \cdot r(n - L_x). \quad (9)$$

Finally, the time-shift and scaling factor  $\alpha$  can be compensated to retrieve  $q(n)$  and  $r(n)$ .

### A. Impulse Response Estimation Extended with Least Squares Minimization

The measurement of the impulse responses  $q(n)$  and  $r(n)$  is depicted in Fig. 2. The indirect derivation of  $p(n)$  can be formulated as an error minimization problem by taking  $\alpha \cdot r(n-L_x)$  as its input and the error as the difference between its output and  $\alpha \cdot q(n-L_x)$ . Finally, if time-invariance is applied and the scaling factor  $\alpha$  is compensated, the right side of the diagram (see Fig. 3a) can be simplified to the one depicted in Fig. 3b, and the expression of  $e(n)$  is simplified to

$$e(n) = q(n) - p(n) * r(n). \quad (10)$$

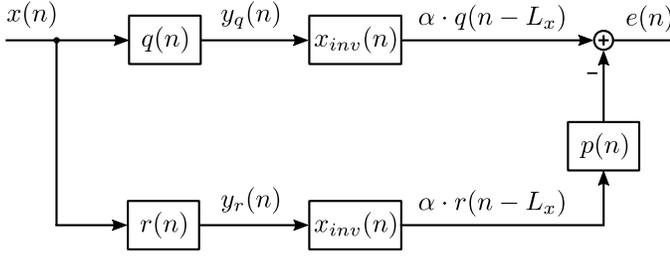


Fig. 2. System block diagram proposed to indirectly derive the transfer function of the primary path, based on ESS measurements of the auxiliary impulse responses  $q(n)$  and  $r(n)$  and a Least Squares approach.

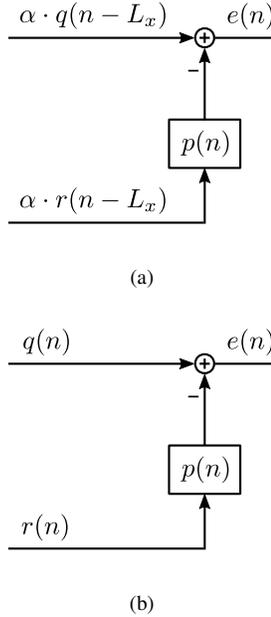


Fig. 3. Least squares error definition refinement: (a) before time-invariance is applied and (b) after time-invariance is applied.

To find an optimal solution for the minimization of the error signal using finite-length impulse responses, an appropriate notation is defined. Let the following example column vectors

$$\mathbf{y}_{L_y \times 1} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{L_y-1} \end{bmatrix}, \mathbf{h}_{L_h \times 1} = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{L_h-1} \end{bmatrix}, \quad (11)$$

and

$$\mathbf{x}_{L_x \times 1} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{L_x-1} \end{bmatrix} \quad (12)$$

be the first  $L_y$ ,  $L_h$ , and  $L_x$  samples of the signals or system impulse responses  $y(n)$ ,  $h(n)$ , and  $x(n)$ , respectively. Then, let  $\mathbf{y}$ ,  $\mathbf{h}$ , and  $\mathbf{x}$  be related by the convolution defined by the scalar product

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{x}, \quad (13)$$

being  $\mathbf{H}$  created by repeating  $\mathbf{h}$   $L_x$  times along the diagonal

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{L_y-1} \end{bmatrix} = \begin{bmatrix} h_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ h_{L_h-1} & & h_0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & h_{L_h-1} \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{L_x-1} \end{bmatrix}. \quad (14)$$

Consequently, their lengths are related by

$$L_y = L_h + L_x - 1. \quad (15)$$

Making use of the introduced notation in the final equivalent system in Fig. 3, Eq. (10) can be re-written as

$$\mathbf{e} = \mathbf{q}_{zp} - \mathbf{R} \cdot \mathbf{p}, \quad (16)$$

with  $\mathbf{q}_{zp}$  being  $\mathbf{q}$  zero-padded to match the length of the convolution between  $\mathbf{r}$  and  $\mathbf{p}$ . Using Eq.(16), one can define the quadratic error as

$$\mathbf{e}^T \mathbf{e} = (\mathbf{q}_{zp} - \mathbf{R} \cdot \mathbf{p})^T (\mathbf{q}_{zp} - \mathbf{R} \cdot \mathbf{p}), \quad (17)$$

which can be differentiated with respect to  $\mathbf{p}$  and set to zero

$$\frac{\partial \mathbf{e}^T \mathbf{e}}{\partial \mathbf{p}} = -2 \cdot (\mathbf{q}_{zp} - \mathbf{R} \cdot \mathbf{p})^T \cdot \mathbf{R} = 0, \quad (18)$$

which finally leads to the optimum solution for the approximation of  $p(n)$  given by

$$\mathbf{p} = (\mathbf{R}^T \cdot \mathbf{R})^{-1} \cdot \mathbf{R}^T \cdot \mathbf{q}_{zp}, \quad (19)$$

with the parameters  $L_q$ ,  $L_r$ , and  $L_p$ .

### III. MEASUREMENT SETUP

A general overview of the measurement setup can be seen in Fig. 4. The Beyerdynamic DT 770 PRO headphone customized with inner and outer microphones presented in Fig. 5 is used as system under test. The microphone signals are pre-amplified to reach line-level with amplification factors of 42.3 dB and 22.3 dB for the inner and outer microphone, respectively. A Neumann KU100 dummy head is used to simulate the presence of the human ear in the enclosed space of the ear-cup. A Genelec 8030B speaker is placed 1 m away from the dummy head and used to generate a 25 sec long ESS going from 10 Hz to 23 kHz at a sampling frequency of 48 kHz. The ESS is faded-in during 62 500 samples and faded-out during 625 samples, to avoid the speaker to pop. Consequently the deconvolution filter is also faded-in and out,



Fig. 4. Measurement setup overview. Dummy head (left) wearing the ANC headphone and external loudspeaker (right) used to generate the auxiliary impulse responses.

but interchanging the lengths, so that it maintains its original relation with the ESS. Additionally, 20 480 samples of silence are post-appended to the signal, to give the system time to settle before the end of the recording. An RME Fireface UCX audio interface is used to generate and record the signals, which afterwards are processed in MATLAB.

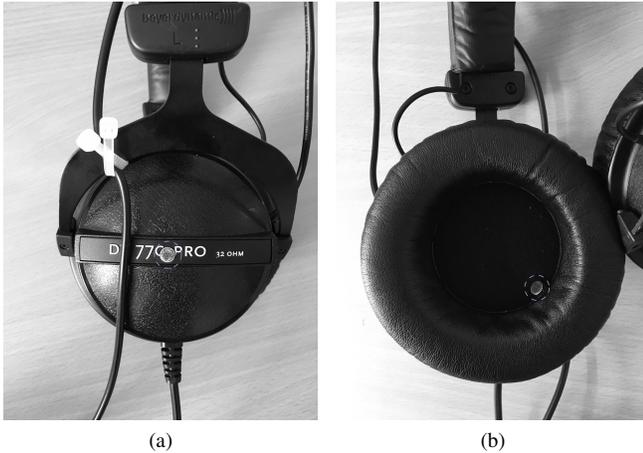


Fig. 5. Utilized Headphones with (a) outer reference microphone and (b) inner error microphone

#### IV. RESULTS AND EVALUATION

Using the measurement setup presented in the previous section and the impulse response estimation technique described in Section II, the impulse responses  $q(n)$  and  $r(n)$  are measured (see Fig. 6a and Fig. 7a). The time-shift they have in common, which is introduced by the measurement chain, is partially compensated and their lengths are shortened to  $L_q = 25\,000$  and  $L_r = 25\,000$  samples to be able to work with them also in time-domain. Their discrete-time Fourier transform magnitude responses between 10 - 23 000 Hz are presented in Fig. 6b and Fig. 7b. A common high-pass characteristic with a cut-off frequency of around 60 Hz can

be seen in both magnitude responses, which is in accordance with the technical specifications of the external loudspeaker. Looking at the magnitude response of  $r(n)$ , it is interesting to see a relatively flat response in comparison with the magnitude response of  $q(n)$ , with moderate influence of the modes of the room where the measurement was done. In contrast, but as expected, the frequency response of  $q(n)$  is the one of a low-pass filter with a cut-off frequency roughly at 200 Hz. Above 3 kHz some significant zeros are introduced by the volume of air enclosed by the dummy head and the ear-cup.

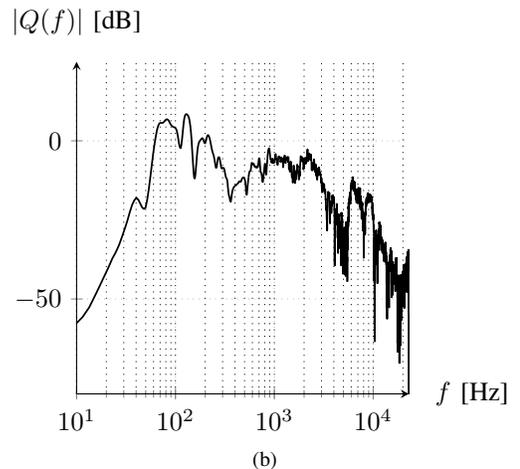
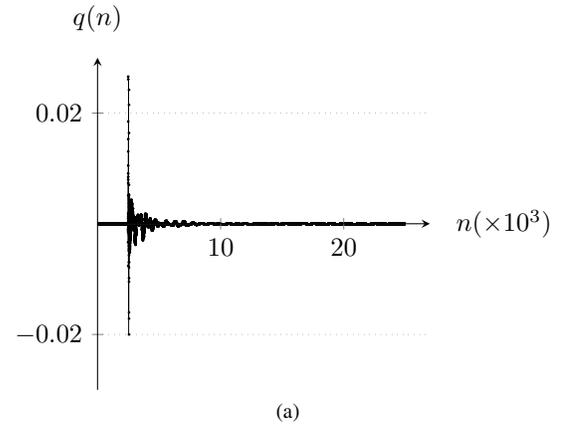
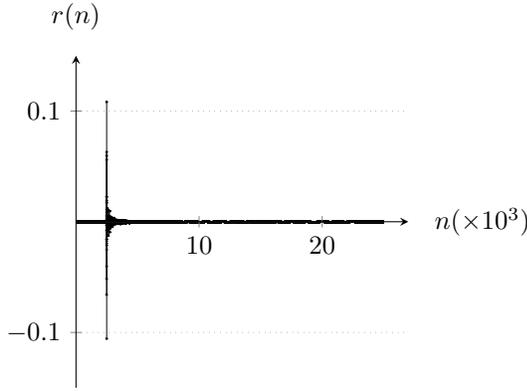


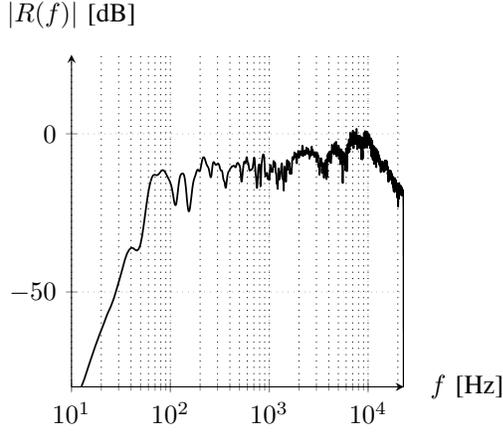
Fig. 6. (a) Impulse response  $q(n)$  and (b) the magnitude response of its discrete-time Fourier transform.

Based on the measured  $r(n)$  and  $q(n)$ , the primary path is then derived using both approaches with lengths starting from 10 240 and increasing in steps of 1 024 up to 20 480 samples. Because of the different pre-amplification gains, the noise floor level coming from the inner microphone is higher than the one from the outer microphone. This produces a boost in magnitude outside of the range 10 - 23 000 Hz in the derived  $p(n)$ . To get rid of this undesired effect, the derived primary path is additionally zero-phase band-pass filtered in this frequency range before evaluating it.

To generate a metric for the accuracy of each derivation of  $p(n)$ , and to be able to compare them, the system depicted in Fig. 8 is used. With it, the desired behaviour of the impulse



(a)



(b)

Fig. 7. (a) Impulse response  $r(n)$  and (b) the magnitude response of its discrete-time Fourier transform.

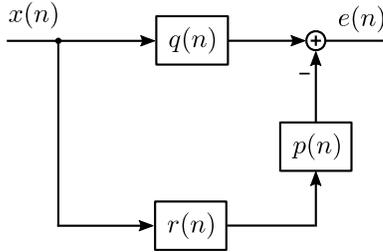


Fig. 8. System block diagram used for the evaluation. Q and R as the systems based on the measured impulse responses  $q(n)$  and  $r(n)$ , and P as the estimated primary path transfer function under evaluation.

response  $q(n)$  is compared sample-by-sample with the one of  $r(n) * p(n)$ . The input  $x(n)$  used is a uniformly distributed white noise signal with zero mean, which is first band-passed between 10 and 23 000 Hz to match the valid frequency range of the measured impulse responses. All simulations are run for 10 sec and at the end the mean quadratic error is calculated and stored. The results of the evaluation routine are resumed and presented in Table I. The table has 4 columns: the first one on the left is the length of the impulse response used for the derivation of  $p(n)$  on the respective row; the second

TABLE I  
MEAN SQUARED ESTIMATION ERROR OF BOTH APPROACHES FOR DIFFERENT LENGTHS  $L_p$ .

$L_p$	LS [dB]	CAPS [dB]	Difference [dB]
10 240	-59.0779	-48.1467	-10.9312
11 264	-63.0861	-48.0708	-15.0153
12 288	-64.2488	-48.1622	-16.0866
13 312	-61.4445	-48.1679	-13.2765
14 336	-66.6302	-48.1760	-18.4542
15 360	-71.7577	-48.1980	-23.5597
16 384	-67.7916	-48.1995	-19.5921
17 408	-67.4587	-48.1985	-19.2602
18 432	-70.3107	-48.228	-22.0827
19 456	-71.3243	-48.2251	-23.0992
20 480	-71.2544	-48.2498	-23.0047

and third are the mean squared error values in dB, reached after the simulation time by each one of the derivations; and the last column is the difference between the values of the second and third column on the same row. The first thing to notice by looking at the second and third columns is the expected reduction of the mean square error as the length of  $p(n)$  increases. In the case of the CAPS method the reduction is steady but almost marginal, since the lengths of  $q(n)$  and  $r(n)$  are maintained constant and, therefore, the solution has one fixed length, from which the first  $L_p$  samples are extracted to form  $p(n)$ . In contrast, the LS method shows to be very sensitive to the number of coefficients for approximating the optimum  $p(n)$ , which interestingly not always introduces an improvement in accuracy. A good example of this is the case with 15 360 samples, which, although far from being at the bottom of the table, generates the smallest mean squared error. Nevertheless, if  $p(n)$  is not zero-phase band-pass filtered between 10 and 23 000 Hz before the evaluation, the improvement shows to be constant. The results show that the mean squared estimation error of the LS method is roughly between three and ten times smaller than the one produced by the CAPS method, depending on the chosen  $L_p$ . To compare both derivation approaches in more detail, the case with  $L_p = 20 480$  is chosen. The estimation accuracy can be qualitatively evaluated by comparing the magnitude response deviations of  $R(f) \cdot P_{CAPS}(f)$  and  $R(f) \cdot P_{LS}(f)$  from  $Q(f)$ . It can be seen in Fig. 9a, that the CAPS approach deviates from the desired magnitude response of  $Q(f)$  in the high frequencies and between 10 and 500 Hz, and approaches the best in the mid frequencies. Comparatively, it can be seen in Fig. 9b, that the LS approach shows no appreciable deviations beyond 30 Hz. This observation indicates an appreciable accuracy improvement in low and high frequencies by using the LS method.

To further corroborate the improvements introduced by the LS method, the error signals from both derivations are transformed to the frequency-domain and their magnitude responses are presented along with the ratio between them in Fig. 10. It can be seen that the LS estimation error is roughly 23 dB smaller than the CAPS estimation error, with the exception of frequencies below 50 Hz, as calculated and

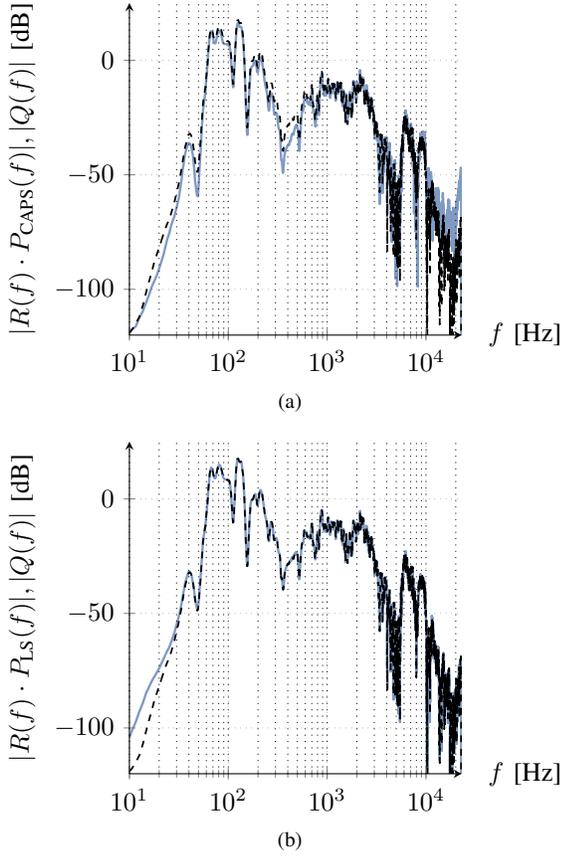


Fig. 9. Magnitude response of (a)  $R(f) \cdot P_{CAPS}(f)$  (in light blue) and  $Q(f)$  (in black dashes), and (b)  $R(f) \cdot P_{LS}(f)$  (in light blue) and  $Q(f)$  (in black dashes), for the case  $L_p = 20\,480$ ,  $L_q = 25\,000$ , and  $L_r = 25\,000$ , between 10 - 23 000 Hz.

presented in Table I. However, the ratio between the estimation errors is not constant throughout the entire frequency range. Moreover, the difference shows a major improvement of estimation accuracy between 10 000 and 23 000 Hz based on the LS method, reaching a peak of almost 35 dB at 20 270 Hz.

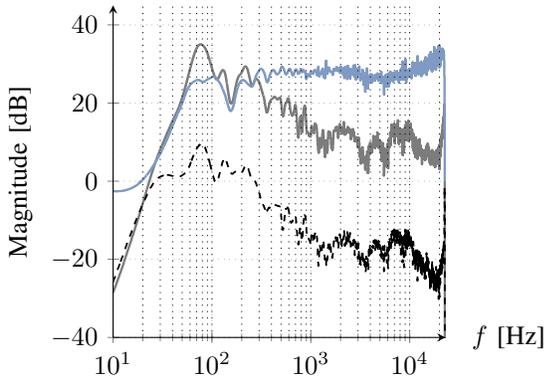


Fig. 10. Discrete-time Fourier transforms of the estimation error signals for the case  $L_p = 20\,480$ ,  $L_q = 25\,000$ , and  $L_r = 25\,000$ , between 10 - 23 000 Hz. In black dashes, the estimation error from the LS method,  $|E_{LS}(f)|$ ; in solid gray, the estimation error from the CAPS method,  $|E_{CAPS}(f)|$ ; and in solid light blue, the ratio  $|E_{CAPS}(f)|/|E_{LS}(f)|$ .

Finally, the estimated primary paths derived with CAPS and LS methods are presented in Fig. 11a and Fig. 12a, respectively. Additionally, their discrete-time Fourier transforms together with  $|Q(f)|$  and  $|R(f)|$  are shown in Fig. 11b and Fig. 12b to show the relation between the transfer functions. It can be seen that in both cases the behaviour of the estimated primary path resembles a low-pass filter indicating that the ear-cup performs passive attenuation in the mid to high frequency range, as mentioned in the Introduction. Additionally, it can be seen in Fig. 11a and Fig. 12a, that  $L_p = 20\,480$  used for both derivations may be sufficient to represent most of the behaviour of the system, based on the observation that the impulse responses have almost settled.

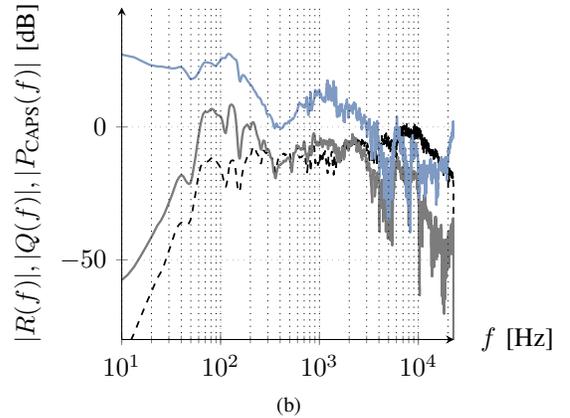
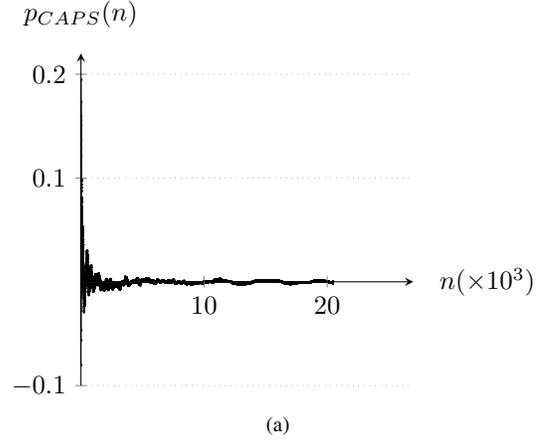


Fig. 11. (a) Estimated primary path based on CAPS method and (b) the magnitude response of  $P_{CAPS}(f)$  (in solid light blue),  $Q(f)$  (in solid gray), and  $R(f)$  (in black dashes).

## V. CONCLUSIONS

The ad hoc adapted impulse response derivation technique based on Least Squares proposed in this work has been studied. Its performance has been compared with the known frequency-domain approach based on cross-power and auto-power spectra. The results based on measurements done over a real ANC headphone using different lengths for  $p(n)$  have shown an overall improvement of at least 10 and up to 23 dB in mean squared estimation error. A major improvement in

## REFERENCES

- [1] K. T, *Adaptive Feed-Forward Control of Low Frequency Interior Noise*, ser. Intelligent Systems, Control and Automation: Science and Engineering ., Dordrecht :: Springer Netherlands., 2012. [Online]. Available: <http://dx.doi.org/10.1007/978-94-007-2537-9>
- [2] M. S and M. P, "Transfer-Function Measurement with Sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, 2001. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=10189>
- [3] S. GB, E. JJ, and A. D, "Comparison of Different Impulse Response Measurement Techniques," *J. Audio Eng. Soc.*, vol. 50, no. 4, pp. 249–262, 2002. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=11083>
- [4] H. M, C. T, and Z. U, "Impulse Response Measurement Techniques and their Applicability in the Real World," in *Proc. of the 12th Int. Conference on Digital Audio Effects [DAFx-09]*, September 2009.
- [5] G. D, "Beyond MLS-Occupied Hall Measurement with FFT Techniques," in *Audio Engineering Society Convention 101*, Nov 1996. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=7376>
- [6] F. A, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *Audio Engineering Society Convention 108*, Feb 2000. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=10211>
- [7] —, "Advancements in Impulse Response Measurements by Sine Sweeps," in *Audio Engineering Society Convention 122*, May 2007. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=14106>
- [8] T.-R. A and J. F, "A New Interpretation of Distortion Artifacts in Sweep Measurements," *J. Audio Eng. Soc.*, vol. 59, no. 5, pp. 283–289, 2011. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=15929>
- [9] M. J, C. P, and H. J, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82.*, vol. 7, May 1982, pp. 1858–1861.
- [10] V. T and B. F, "Transfer Function and Subjective Quality of Headphones: Part 1, Transfer Function Measurements," in *Audio Engineering Society Conference: 11th International Conference: Test & Measurement*, May 1992. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=6267>

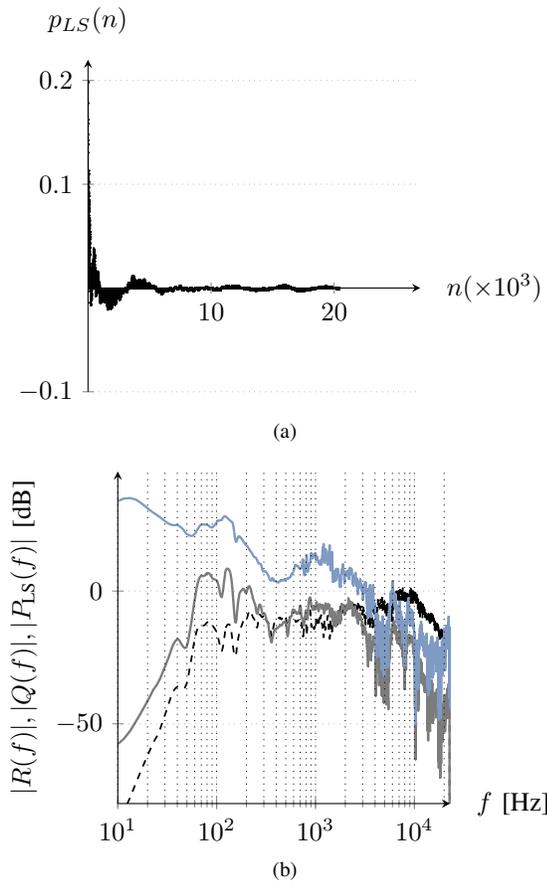


Fig. 12. (a) Estimated primary path based on LS method and (b) the magnitude response of  $P_{LS}(f)$  (in solid light blue),  $Q(f)$  (in solid gray), and  $R(f)$  (in black dashes).

the 10 000-23 000 Hz range, with a peak of 35 dB in the case  $L_p = 20\,480$  has been observed. The sensitivity of the results to the length of  $p(n)$  suggests that further investigation has to be conducted to understand how the minimization of the mean squared estimation error can be improved through the increment of  $L_p$ . However, the computational complexity and memory requirements of the method proposed may be prohibitive for standard personal computers, if lengths longer than 20 480 samples are to be studied.