Optimized Learning of Spatial-Fourier Representations from Fast HRIR Recordings

Gerald Enzner
Dept. Medical Physics and Acoustics
Carl von Ossietzky University Oldenburg
26129 Oldenburg, Germany
gerald.enzner@uol.de

Christoph Urbanietz
Inst. of Communication Acoustics
Ruhr-Universität Bochum
44780 Bochum, Germany
christoph.urbanietz@rub.de

Rainer Martin
Inst. of Communication Acoustics
Ruhr-Universität Bochum
44780 Bochum, Germany
rainer.martin@rub.de

***Abstract*—The acquisition of head-related impulse responses (HRIRs) has traditionally been a time-consuming acoustic measurement process. Novel continuous-azimuth recording techniques have dramatically accelerated the acquisition, but conversion into continuous Spatial-Fourier representations (SpaFoR) of HRIRs provides a host of cumbersome implementation challenges. The direct closed-form least-squares approach is unfortunately not practical and we will therefore explore the retrieval of SpaFoR model parameters of HRIR by contemporary machine-learning tools. Specifically, we employ the standard stochastic-gradient learning with Tensorflow on a graphics processing unit (GPU) and compare its performance with previous covariance-based least-squares on the general purpose processor. Apart from the sought simplification and acceleration, our paper is dedicated to hyperparameter optimization in order to make sure the final state of the machine learning approach still attains the accuracy of the optimal least-squares solution. The paper finally applies the proposed method to a real acoustic HRIR recording to illustrate the validity of the system identification obtained by learning.**

***Index Terms*—HRTF acquisition, machine learning**

I. Introduction

The head-related transfer function (HRTF) or its time-domain representation, the HRIR, is an essential component of binaural sound rendering with headphones. While a number of generic (non-individual) HRIR sets are easily available, it was demonstrated that individual HRIRs would contribute to fully immersive virtual reality [1]–[3]. Many routes to individual HRTFs thus have been reported based on acoustic measurements [4]–[6], geometric modeling and numerical computation [7], [8], or the personalization of available HRIR catalogues by perceptual [9], morphological [10], or anthropometrical matching with deep learning tools [11]–[13].

A recent HRIR trend was dedicated to the very fast acoustic measurement per individual using the constrained [14]–[18] or unconstrained [19]–[21] continuous head movement of the subject of interest. The challenge here consists in the retrieval of HRIR from very little data and dynamic measurements. Another important research trend for HRIR has been dedicated to its continuous spatial interpolation with the spatial-Fourier representation [22], [23] in order to support seamless binaural rendering of dynamic acoustic scences. Both trends then have also been united in terms of the spatial-Fourier regression to fast continuous HRIR recordings [24], [25] with yet increased challenges related to the optimal retrieval of SpaFoR parameters in an end-to-end fashion from the limited data set. Specifically, it requires huge computational and memory resources or, conversely, very careful attention to the detailed structure of the data covariance matrix to accomplish the closed-form least-squares (LS) regression.

In this paper we therefore pursue an alternative route to SpaFoR regression using a contemporary machine-learning (ML) toolchain for data-driven optimization. This requires a reformulation of the LS problem such that it can be efficiently solved with parallel hardware. We thus aim to harness the computational benefits of the ML-framework on GPU. The paper specifically derives a convolutional SpaFoR architecture with manageable memory requirements and implements it with Keras in Tensorflow. The GPU then invokes several epochs of simple stochastic gradient descent (SGD) to fit the model to the acoustic recording. While the reformulation significantly simplifies the computational model on theoretical grounds, further quantitative analysis and optimization of hyperparameters is conducted in this paper in order to attain the LS baseline with rapid convergence and desired final accuracy.

The remainder of this paper is organized as follows. Section II describes the continuous measurement setup and the computational demands of the SpaFoR model for HRIR retrieval. Section III then presents the proposed ML approach for SpaFoR computation, followed by a systematic investigation of optimal hyperparameters for training in Section IV. Section V finally reports a comparison of utility with the baseline covariance-based least-squares regression on a real HRIR recording. Section VI draws our conclusions.

II. Measurement Setup and SpaFoR Model

We rely on the continuous HRTF measurement setup with head revolution time T_{360} as described in [16] and shown

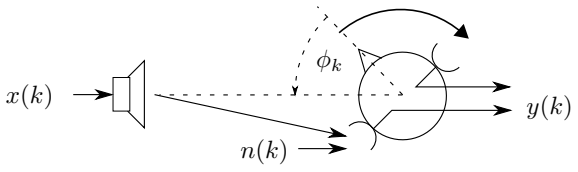

Fig. 1. Continuous measurement of an HRIR with rotating head.

in Fig. 1 for convenience. A white-noise probe signal $x(k)$ is applied and the received signal $y(k)$ is described by the "rotating-head" convolutive model [26]

$$y(k) = \sum_{l=0}^{N-1} h(l,\phi_k)\, x(k-l) + n(k)\,, \tag{1}$$

where $h(l,\phi_k)$ is the HRIR at impulse response index l and relative azimuth ϕ_k between head and source at discrete time k. Reliant on constant angular velocity $\omega = 2\pi/T_{360}$, we easily have $\phi_k = \omega k T_s$ with $T_s = 1/f_s$ the sampling time interval. The signal $n(k)$ represents independent measurement noise.

A spatial-Fourier model of the HRIR refers to a linear expansion

$$h(l,\phi) = \sum_{q=-Q}^{Q} a_{q,l}^*\, c_q(\phi) \tag{2}$$

of the entire angular trajectory of each HRIR coefficient in terms of the continuous Fourier basis $c_q(\phi) = \exp(jq\phi)$. The spatial Fourier coefficients $a_{q,l}$ are independent of the azimuth ϕ. The model is nicely compact in that the necessary range of spatial frequencies (i.e., model order Q) is well bandlimited to, e.g., $Q = 39$ at $f_s = 48\,$kHz sampling [4], [22].

Using (2), an expansion model of the received signal in (1) at the left or right ear is then given by

$$\widehat{y}(k) = \sum_{l=0}^{N-1} \sum_{q=-Q}^{Q} a_{q,l}^*\, c_q(\phi_k)\, x(k-l) \tag{3}$$

$$= \sum_{q=-Q}^{Q} \boldsymbol{a}_q^H\, \boldsymbol{x}(k)\, c_q(\phi_k)\,, \tag{4}$$

where in the second step the convolution is expressed as an inner product of coefficients $\boldsymbol{a}_q = (a_{q,0},\ldots,a_{q,N-1})^T$ and signal vectors $\boldsymbol{x}(k) = (x(k), x(k-1), \ldots x(k-N+1))^T$. Note that the expansion models in (2) and (4) each comprise overall $U = N(2Q+1)$ unknown parameters $a_{q,l}$.

These HRIR model parameters can now be determined by fitting the model output $\widehat{y}(k)$ to a real measurement $y(k)$ of L samples. In [24], [25] this has been achieved by minimizing the least-squares (LS) error

$$J(\boldsymbol{a}) = \sum_{k=0}^{L-1} |y(k) - \widehat{y}(k)|^2 = \sum_{k=0}^{L-1} \left|y(k) - \boldsymbol{a}^H \boldsymbol{z}(k)\right|^2 \tag{5}$$

with respect to a stacked coefficient vector $\boldsymbol{a}$ across all L measurement samples, where $\boldsymbol{z}(k)$ comprehensively combines input signal $x(k)$ with basis functions $c_q(\phi_k)$, i.e.,

$$\boldsymbol{a} = \begin{pmatrix} \boldsymbol{a}_{-Q} \\ \boldsymbol{a}_{-Q+1} \\ \vdots \\ \boldsymbol{a}_{Q} \end{pmatrix} \quad \text{and} \quad \boldsymbol{z}(k) = \begin{pmatrix} c_{-Q}(\phi_k)\, \boldsymbol{x}(k) \\ c_{-Q+1}(\phi_k)\, \boldsymbol{x}(k) \\ \vdots \\ c_{Q}(\phi_k)\, \boldsymbol{x}(k) \end{pmatrix}. \tag{6}$$

In this form of the problem statement, its theoretical solution is easily found in closed form as [27], [28]

$$\widehat{\boldsymbol{a}} = \boldsymbol{R}_{zz}^{-1}\, \boldsymbol{r}_{yz}\,, \tag{7}$$

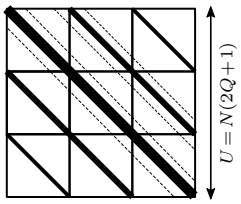

Fig. 2. Block-Toeplitz shape of the data covariance matrix $\boldsymbol{R}_{zz}$ of the closed-form least-squares approach with model order $Q = 1$.

with complex sample correlation $\boldsymbol{r}_{yz} = \sum_{k=0}^{L-1} y(k)\, \boldsymbol{z}(k)$ and sample covariance $\boldsymbol{R}_{zz} = \sum_{k=0}^{L-1} \boldsymbol{z}(k)\, \boldsymbol{z}^H(k)$ of size $U \times U$. The latter exhibits block-Toeplitz structure, as shown in Fig. 2, due to the stacked definition of the regressor $\boldsymbol{z}(k)$. Considering realistic HRIR model dimensions of $N = 128$, $Q = 39$ and a recording duration of $T_{360} = 60\,$s at $f_s = 48\,$kHz, the computation time for covariance accumulation in the order of $\mathcal{O}(N^2 Q^2 f_s\, T_{360})$ turns out to be prohibitive. In order to still accomplish the LS solution, a sophisticated approach including the following procedures has been developed [24], [25]:

- acquisition of the first block row of $\boldsymbol{R}_{zz}$,
- submatrix approximation in Toeplitz shape,
- steepest descent iteration on $\boldsymbol{R}_{zz}$ submatrices.

While this solution of (7) requires 7 min on a Desktop-Matlab Core-i7 system, it is inconvenient for further generalization.

III. Model Identification by Machine Learning

The motivation for the ML approach to be pursued here therefore consists in a data-driven minimization of the cost function in (5) while circumventing costly covariance matrix accumulation and necessary attention to its detail. To this end, we may just hold the sought coefficients $a_{q,l}$ in a linear feed-forward neural-network layer and feed $c_q(\phi_k)x(k-l)$ products according to (3) as the respective input data. Specifically, this size U data $c_q(\phi_k)x(k-l)$ at each time instant k could be arranged line-by-line as training samples in a matrix of size $U \times L$ and fitted to each sample of a recording $y(k)$ as shown in Fig. 3. With typical numbers of the previous section and using `float32`, this amounts to a 240 GB complex-valued representation of the data – way too much for concurrent desktop computer RAM and entirely unbalanced with a 60 s HRIR sound recording (even if "data store" concepts on the harddrive were considered). However, the predictors $c_q(\phi_k)x(k-l)$ do exhibit a subtle structure that is to be described and epxloited in the sequel to save memory resources.

A. Convolutional Representation

We may rewrite the spatial-Fourier basis functions in (3) as $c_q(\phi_k) = c_q(\phi_{k-l})\, \exp(jq\phi_{\Delta l})$, where $\phi_{\Delta l} = \phi_k - \phi_{k-l}$ is the head movement during time lT_s. Moreover, it turns out that $\phi_{\Delta l} = 2\,\pi\, l\, T_s / T_{360}$ is independent of time k for the assumed constant speed of rotation. The signal model from (3) is then restated with equality as

$$\widehat{y}(k) = \sum_{l=0}^{N-1} \sum_{q=-Q}^{Q} \widetilde{a}_{q,l}^*\, c_q(\phi_{k-l})\, x(k-l) \tag{8}$$

using modulated SpaFoR coefficients $\widetilde{a}_{q,l}^* = a_{q,l}^* \exp(jq\phi_{\Delta l})$. The time indices of angle ϕ and waveform x are now both

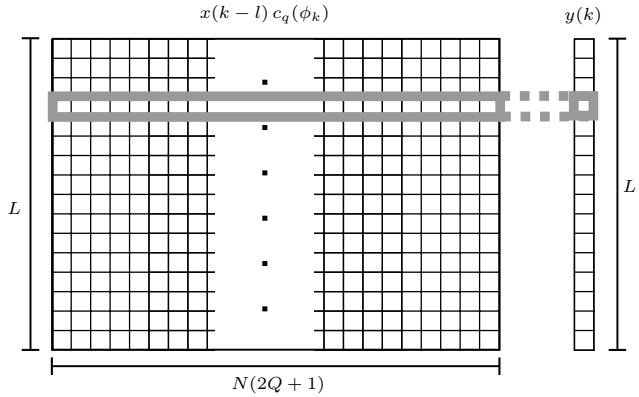

Fig. 3. Illustration of data/network configuration according to Eq. (3).

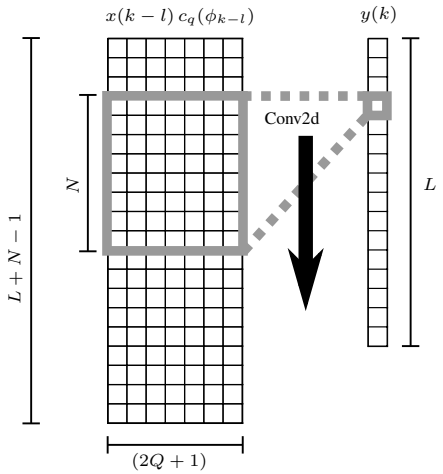

Fig. 4. Illustration of data/network configuration according to Eq. (8).

aligned to $k-l$ and the entire model in (8) therefore appears as a convolution along time k, where l is the convolution index and $c_q(\phi_{k-l})\,x(k-l)$ are modified predictors.

As a major consequence, we can feed the modified predictors to a convolutional neural-network layer as shown in Fig. 4. The convolutional layer in this case holds the same number U of modulated coefficients $\widetilde{a}^*_{q,l}$ as before, however, the computational mapping operates with a sliding window of size $N\times(2Q+1)$ with stride 1 to fit the $y(k)$ sequence. The exorbitant size of the input matrix mentioned before is thus reduced to $2Q\times(L+N)$ elements, which amounts to a manageable 2 GB with former dimensions of the HRIR model. After model fitting, a simple fixed matrix operation delivers the desired weights $a_{q,l}$ from the modulated weights $\widetilde{a}^*_{q,l}$.

Merely for curiosity, note that (8) refers to a "rotating loudspeaker" convolutive signal model (instead of a rotating head) according to physical considerations in [26]. We would in the actual case of a rotating loudspeaker as well use the modified predictors and directly obtain sought coefficients $a_{q,l}$ and even save the fixed-matrix conversion. A measurement apparatus with a loudspeaker rotating silently around the listener, however, would be difficult to set up.

B. Real-Valued Implementation

For actual model fitting we rely on Tensorflow/Keras 2.3.0 on a Core-i7 Ubuntu Desktop PC with NVidia RTX 2070 Super GPU. In order to interface with its real-valued computations, the convolutional model (8) can be rewritten by Euler's equation $c_q(\phi)=\cos(q\phi)+j\,\sin(q\phi)$ in terms of real-valued functions. Considering the real-valued nature of the HRIR system in (1), we can further make use of the conjugate symmetry $\widetilde{a}^*_{q,l}=\widetilde{a}_{-q,l}=\mathrm{Re}(\widetilde{a}_{q,l})+j\,\mathrm{Im}(\widetilde{a}_{q,l})$ to express

$$\widehat{y}(k)=\sum_{l=0}^{N-1}\sum_{q=-Q}^{Q} w_{q,l}\,r_q(\phi_{k-l})\,x(k-l) \tag{9}$$

with all real-valued basis functions and weights,

$$r_q(\phi)=\begin{cases}\cos(q\phi) & \text{for } q>0\\ 1 & \text{for } q=0\\ \sin(q\phi) & \text{for } q<0\,,\end{cases}\quad w_{q,l}=\begin{cases}2\,\mathrm{Re}(\widetilde{a}_{q,l}) & \text{for } q>0\\ \widetilde{a}_{q,l} & \text{for } q=0\\ 2\,\mathrm{Im}(\widetilde{a}_{q,l}) & \text{for } q<0\,,\end{cases}$$

and weights $\widetilde{a}_{q,l}$ are retrieved easily from $w_{q,l}$ after fitting.

IV. Hyperparameter Optimization

We now optimize the configuration of the ML toolchain in order to guide its rapid and stable convergence towards the optimal LS solution. The analysis makes use of a Gaussian simulation with unit-variance $x(k)$ and HRIR model dimensions as described before the end of Sec. II at SNR=30 dB (if not indicated otherwise) to resemble a typical acoustic recording setup. The network training uses stochastic gradient (SGD) optimization to minimize the mean-squared error (MSE) loss in order to support the optimal LS solution and a related minimum system distance $\mathrm{SD}=||\widehat{\boldsymbol{a}}-\boldsymbol{a}||^2/||\boldsymbol{a}||^2$ of estimated SpaFoR parameters $\widehat{\boldsymbol{a}}$ at convergence. Other relevant optimization criteria are the maximum learning rate μ, the minimum data-to-parameter ratio $\mathrm{DPR}=L/U$, and the robustness to potential SNR variations.

A. Speed of Convergence

With an increasing number of training epochs, the system distance SD between true and estimated HRIR coefficients ideally decays from a starting point in the order of 0 dB to the low negative dB range of the optimal LS solution where it then saturates. We can thus define the speed of convergence as the slope of the system distance SD [dB] in the first epochs of the training process. Fig. 5 depicts the resulting speed of convergence as a function of DPR and learning rate μ for the proposed convolutional system. Except for small DPR, it can be seen that DPR hardly affects the speed of convergence. The clear trend of SD improvement per epoch [dB] seen with the learning rate μ approximately follows a $10\log_{10}(1-\mu(2-\mu))$ law. This quantification is here inspired by similar results in stochastic gradient theory for sample-wise online system identification in the large SNR regime [29, Section 13.5]. However, we still have to determine how much the learning rate $0<\mu<1$ can be exploited for stable convergence.

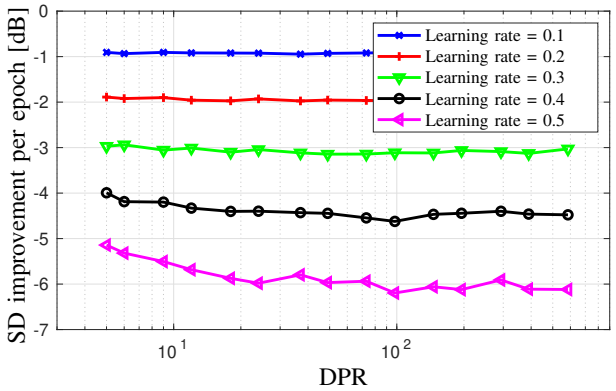

Fig. 5. Convergence speed for various learning rates over DPR.

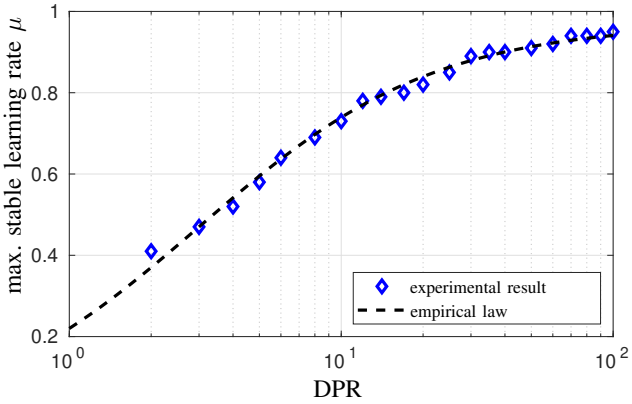

Fig. 6. Maximum stable learning rate vs. data-to-parameter ratio.

B. Maximum Stable Learning Rate

We may therefore define the maximum stable learning rate in an experimental sense as the value of parameter μ that guarantees five consecutive and successful training runs to convergence. Owing to different random initializations of each training, this procedure avoid rash generalization of training singularities. The result of the optimization in Fig. 6 is once more depicted as a function of DPR. Here we find a strong dependence of the maximum stable learning rate, which, according to the diagram, is well described by a simple $\mu_{\max} = \mathrm{DPR}/(\mathrm{DPR}+3)$ empirical law. This dependence will guide our choice of μ for optimally fast convergence to the LS solution when HRIR data and its DPR are given.

C. Final System Distance

For the prediction of the expected minimum SD [dB] at convergence, we resort to LS theory, e.g., [27, Eq. (8.57)]. The estimation error covariance of weights $\boldsymbol{a}$ subject to the general cost function (5) is there exposed as

$$\mathrm{cov}(\widehat{\boldsymbol{a}}) = E\{(\widehat{\boldsymbol{a}} - \boldsymbol{a})(\widehat{\boldsymbol{a}} - \boldsymbol{a})^H\} = \sigma_n^2 \boldsymbol{R}_{zz}^{-1} \qquad (10)$$

with σ_n^2 the noise power. The next goal is to express this general statement of estimation error covariance and the related final SD with parameters of the HRIR problem at hand.

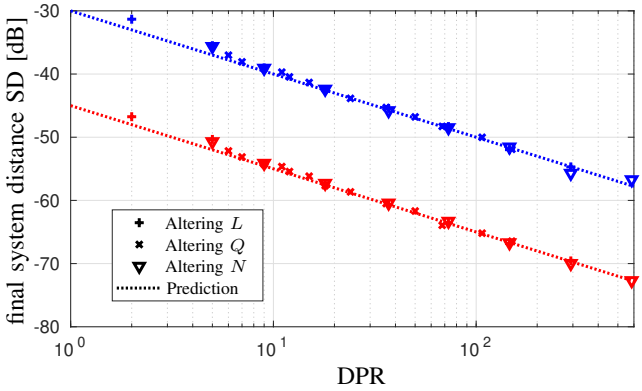

Fig. 7. Final convergence for various configurations. Blue: SNR = 30 dB, Red: SNR = 45 dB. Varied L, N or Q yields the respective DPR.

Having white noise $x(k)$ and orthogonal functions $c_q(\phi_k)$, we may approximate a scaled identity $\boldsymbol{R}_{zz} \approx \sigma_z^2 \, L \, \mathbf{I}_{U\times U}$ and

$$\mathrm{cov}(\widehat{\boldsymbol{a}}) = \frac{\sigma_n^2}{\sigma_z^2 \, L} \, \boldsymbol{I}_{U\times U} = \frac{\sigma_n^2 \, E_a}{\sigma_{\widehat{y}}^2 \, L} \, \boldsymbol{I}_{U\times U} \, , \qquad (11)$$

where in the last step we used the relationship $\sigma_{\widehat{y}}^2 = E_a \, \sigma_z^2$ of the input-related variance σ_z^2, the system energy $E_a = ||a||^2$ of the model coefficients, and the resulting variance $\sigma_{\widehat{y}}^2$ of our noise-free signal model (3). The SD is then predicted as

$$\mathrm{SD} = \frac{\mathrm{Tr}(\mathrm{cov}(\widehat{\boldsymbol{a}}))}{E_a} = \frac{U}{L}\mathrm{SNR}^{-1} = \frac{1}{\mathrm{DPR}\cdot\mathrm{SNR}} \, , \qquad (12)$$

where $\mathrm{SNR} = \sigma_{\widehat{y}}^2/\sigma_n^2$ essentially refers to the microphone SNR of the measurement setup. Thus, higher SNR and DPR both can effectively bring down the minimum system distance.

Fig. 7 depicts our theoretical prediction (12) of LS performance along with markers of corresponding network performance in terms of SD for various model dimensions and data sizes. Here, the expected final SD and the maximum stable learning rate for a given DPR, in conjunction with the related speed of convergence, have guided the number of training epochs required for reaching the LS performance.

V. Further Comparison with Direct Least-Squares

Besides the increased practicality of the ML implementation over the direct LS approach, and otherwise equivalent final SD, we briefly discuss two further aspects.

A. Computation Time

As a baseline, we refer to the 7 min computation time of the aforementioned LS implementation with sophisticated block-Toeplitz and sub-Toeplitz architecture [24], [25], model dimensions $Q = 39$, $N = 128$, and recording time $T_{360} = 60$ s at $f_s = 48$ kHz. The related data-to-parameter ratio in this case is $\mathrm{DPR} = 280$. In contrast, the proposed ML approach then uses 39 s to prepare the training tensors in a single-threaded Python environment and another 15 s for 10 training epochs for reliable convergence with a learning rate of $\mu = 0.9$. This amounts to a total of only 54 s compared to about 7 min with the baseline on the same computer system. Also, the SD performance is equal to about -55 dB for both methods.

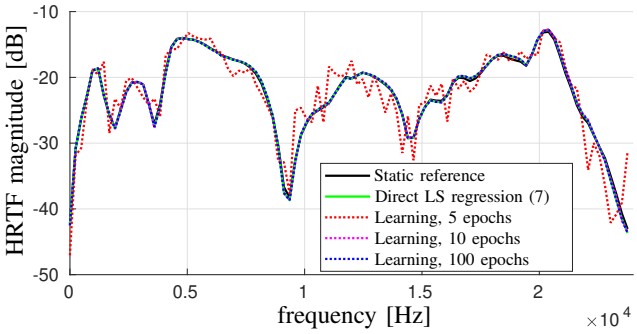

Fig. 8. Frequency responses of reconstructed HRIRs at $\phi = 105°$. All curves visually coincide, except the one for the early-stopped learning (i.e., 5 epochs).

B. Illustration on Real HRIR Measurement

For brief illustration and plausibility of the HRIR system identification of this paper, a real continuous HRIR recording is acquired with a dummy head according to Fig. 1 and processed. During training of the SpaFoR coefficients, a selected HRIR at angle $\phi = 105°$ is reconstructed after 5, 10, and 100 epochs learning and benchmarked in the frequency domain in Fig. 8 against the direct LS regression and a static reference measurement at the same angle. With an insufficient number of 5 training epochs, the result obviously deviates from the reference and from LS. Otherwise, it is indistinguishable.

VI. Conclusion

The spatial-Fourier retrieval of HRIRs from fast continuous measurements has been converted from the former covariance-based closed-form LS regression to an ML framework. The proposed method has been optimized to approach the optimal LS solution fast and reliably. Apart from identical system identification performance, two simultaneous advantages of the ML toolchain then stand out: increased convenience of the data-driven ML implementation and considerably reduced computation time, hence, improving the overall utility of the HRIR acquisition and SpaFoR computation system.

References

[1] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123, July 1993.

[2] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?," *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, June 1996.

[3] M. Geronazzo, S. Spagnol, and F. Avanzini, "Do We Need Individual Head-Related Transfer Functions for Vertical Localization? The Case Study of a Spectral Notch Distance Metric," *IEEE/ACM Trans. Audio, Speech, Lang. Proces.*, vol. 26, no. 7, pp. 1247–1260, July 2018.

[4] G. Enzner, C. Antweiler, and S. Spors, "Trends in Acquisition of Individual Head-Related Transfer Functions," in *The Technology of Binaural Listening*, Jens Blauert, Ed., Modern Acoustics and Signal Processing, pp. 57–92. Springer, Berlin, Heidelberg, 2013.

[5] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, "Inter-Laboratory Round Robin HRTF Measurement Comparison," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 895–906, Aug. 2015.

[6] S. Li and J. Peissig, "Measurement of Head-Related Transfer Functions: A Review," *Appl. Sci.*, vol. 10, no. 14:5014, Jan. 2020.

[7] H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," *J. Acoust. Soc. Am.*, vol. 138, no. 1, pp. 208–222, July 2015.

[8] M. Pollow, K. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, and M. Noisternig, "Calculation of Head-Related Transfer Functions for Arbitrary Field Points Using Spherical Harmonics Decomposition," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 72–82, Jan. 2012.

[9] B. Katz and G. Parseihian, "Perceptually based head-related transfer function database optimization," *J. Acoust. Soc. Am.*, vol. 131, no. 2, pp. EL99–EL105, 2012.

[10] A. Politis, M. R. P. Thomas, H. Gamper, and I. J. Tashev, "Applications of 3D spherical transforms to personalization of head-related transfer functions," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2016, pp. 306–310.

[11] H. Hu, L. Zhou, H. Ma, and Z. Wu, "HRTF personalization based on artificial neural network in individual virtual auditory space," *Applied Acoustics*, vol. 69, no. 2, pp. 163–172, 2008.

[12] T. Chen, T. Kuo, and T. Chi, "Autoencoding HRTFs for DNN Based HRTF Personalization Using Anthropometric Features," in *IEEE Int. Conf. Acoust., Speech and Signal Process.*, May 2019, pp. 271–275.

[13] J. Xi, W. Zhang, and T. Abhayapala, "Magnitude modelling of individualized HRTFs using DNN based spherical harmonic analysis," in *IEEE Workshop on Appl. of Signal Process. to Audio and Acoustics*, 2021, pp. 266–270.

[14] T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic measurement of room impulse responses using a moving microphone," *J. Acoust. Soc. Am.*, vol. 122, no. 3, pp. 1636–1645, 2007.

[15] K. Fukudome, T. Suetsugu, T. Ueshin, R. Idegami, and K. Takeya, "The Fast Measurement of Head Related Impulse Responses for All Azimuthal Directions Using the Continuous Measurement Method with a Servo-Swiveled Chair," *Applied Acoust.*, vol. 68, no. 8, pp. 864–884, Aug. 2007.

[16] G. Enzner, "Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2008, pp. 393–396.

[17] G. Enzner, "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering," in *IEEE Workshop Appl. Signal Process. Audio, Acoust.*, Oct. 2009, pp. 325–328.

[18] J. Richter and J. Fels, "On the Influence of Continuous Subject Rotation During High-Resolution Head-Related Transfer Function Measurements," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 4, pp. 730–741, Apr. 2019.

[19] S. Li and J. Peissig, "Fast estimation of 2D individual HRTFs with arbitrary head movements," in *Int. Conf. on Digital Signal Process. (DSP)*, Aug. 2017, pp. 1–5.

[20] J. He, R. Ranjan, W. Gan, N. K. Chaudhary, N. D. Hai, and R. Gupta, "Fast Continuous Measurement of HRTFs with Unconstrained Head Movements for 3D Audio," *J. Audio Eng. Soc.*, vol. 66, no. 11, pp. 884–900, Nov. 2018.

[21] S. Nagel, T. Kabzinski, S. Kühl, C. Antweiler, and P. Jax, "Acoustic Head-Tracking for Acquisition of Head-Related Transfer Functions with Unconstrained Subject Movement," in *AES Int. Conf. on Audio for Virtual and Augmented Reality*, Aug. 2018.

[22] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Sound Field Analysis along a Circle and Its Applications to HRTF Interpolation," *J. Audio Eng. Soc.*, vol. 56, no. 3, pp. 156–175, Mar. 2008.

[23] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *J. Acoust. Soc. Am.*, vol. 127, no. 4, pp. 2347–2357, 2010.

[24] C. Urbanietz and G. Enzner, "Spatial-Fourier Retrieval of Head-related Impulse Responses from Fast Continuous-Azimuth Recordings in the Time-Domain," in *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, May 2019, pp. 950–954.

[25] C. Urbanietz and G. Enzner, "Direct Spatial-Fourier Regression of HRIRs from Multi-Elevation Continuous-Azimuth Recordings," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 1129–1142, 2020.

[26] C. Urbanietz and G. Enzner, "Binaural Rendering of Dynamic Head and Sound Source Orientation Using High-Resolution HRTF and Retarded Time," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2018, pp. 566–570.

[27] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 2001.

[28] C.M. Bishop, *Pattern Recognition & Machine Learning*, Springer, 2006.

[29] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*, John Wiley & Sons, 2006.