forum acusticum 2026
8-12 September · Graz, Austria

# ONLINE ESTIMATION OF VIRTUAL SENSING OBSERVATION FILTER COEFFICIENTS WITH REAL-WORLD SOUNDS FOR LOCAL ACTIVE NOISE CONTROL

Felix Holzmüller[1], Paul Armin Bereuter[1], Christian Blöcher[1], Alois Sontacchi[1]

[1]*Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria*



## Abstract

Adaptive local active noise control (ANC) requires accurate signal estimates at the points of cancellation. These are often obtained using nearby microphones via virtual sensing methods such as the remote microphone technique (RMT). Obs-TasNet was recently proposed as a neural network–based method for online estimation of RMT observation filter coefficients. In this work, we evaluate Obs-TasNet's performance in synthesized reverberant environments with stationary and non-stationary sounds sampled from the FSD50K dataset. In reverberant environments, Obs-TasNet exhibits only a slight decrease in estimation performance when moving from stationary to non-stationary noise sources. Estimation error increases as the frequency approaches the microphone array's aliasing frequency. Obs-TasNet consistently outperforms inverse distance weighting and nearest neighbor baselines. The presented results indicate that Obs-TasNet is robust across source types and can handle acoustic conditions that are relevant to practical deployment.

**Keywords:** *Local Active Noise Control, Virtual Sensing, Remote Microphone Technique, Sound Field Estimation, Deep Learning*

## 1. Introduction

Local active noise control (ANC) aims to reduce acoustic disturbances at specific points in space using secondary sources through destructive interference [1]. The control signals for the secondary sources are often rendered by adaptive filters, requiring the residual error signal at the targeted point(s) of cancellation (PoC). However, this signal is in many cases not directly accessible without invading the listener's space. Hence, virtual sensing techniques such as the remote microphone technique (RMT) [2] are utilized to estimate the signal of a virtual microphone at the PoC, using nearby remote microphones and knowledge about the system and the disturbances.

A central component of the RMT is the observation filter, which is applied to the primary disturbance signals at the remote microphones to estimate primary disturbances at the virtual microphone. The optimal coefficients of this observation filter depend on the spectral properties of the primary disturbances and the locations of the primary source(s) and receivers [3]. A mismatch between estimated and real acoustic scene can lead to performance degradation or even system instability [4]. To mitigate this issue, observation filter coefficient sets can be calculated in advance for several acoustic scenes and actively switched [3] or interpolated [5] during operation. However, this requires dedicated measurements and calculations for selected acoustic scenes as well as a scene detection algorithm [6].

Recently, Obs-TasNet [7] has been proposed to estimate observation filter coefficients online using a deep neural network, based on the inter-channel fully-convolutional time-domain audio separation network (IC Conv-TasNet) [8]. It obviates the need for explicit filter calculation and scene detection during operation while facilitating substantially higher noise reduction compared to a multiple error ANC system. However, Obs-TasNet has thus far only been tested for stationary, anechoic sounds. This work extends the previous analysis for reverberant sound fields with both stationary and non-stationary sounds, validating Obs-TasNet's performance in complex acoustic scenes. Section 2 outlines the RMT as well as Obs-TasNet's architecture. In Section 3, the experiment's setup is explained. After presenting the results in Section 4, the most relevant findings are summarized in Section 5.

***Corresponding author:*** *holzmueller@iem.at*

## 2. Background

### 2.1. Remote Microphone Technique

The RMT [2], shown schematically in Figure 1, is a commonly used state-of-the-art method for error signal estimation in local ANC. Assuming a simple setup with just one secondary source and a single PoC, the residual error

$$e[n] = d_e[n] + y_e[n] \tag{1}$$

at time $n$ is composed of the primary disturbances $d_e[n]$ and the contributions of the secondary source $y_e[n]$ at the PoC. The latter corresponds to the control signal $u[n]$ filtered by the secondary plant $G_e(z)$, which describes characteristics of the transducer and sound propagation.

In a similar manner, the signals $\boldsymbol{m}[n]$ at $N_m$ nearby remote microphones

$$\boldsymbol{m}[n] = \boldsymbol{d}_m[n] + \boldsymbol{y}_m[n] \tag{2}$$

can be described, where $\boldsymbol{d}_m[n]$ are primary disturbances and $\boldsymbol{y}_m[n]$ the contributions of the secondary source at the remote microphone positions, filtered by the plant $\boldsymbol{G}_m(z)$.

For the RMT, it is assumed that models $\hat{\boldsymbol{G}}_m(z)$ and $\hat{G}_e(z)$ of the secondary plants as well as the control signal $u[n]$ are known. Hence, the estimated contributions $\hat{\boldsymbol{y}}_m[n]$ of the secondary sources can be calculated and subtracted to estimate the primary disturbances at the remote microphone positions

$$\hat{\boldsymbol{d}}_m[n] = \boldsymbol{m}[n] - \hat{\boldsymbol{y}}_m[n]. \tag{3}$$

The primary disturbances $\hat{d}_e[n]$ at the virtual microphone are estimated by applying an *observation filter* $\hat{\boldsymbol{O}}(z)$ onto $\hat{\boldsymbol{d}}_m[n]$, before adding the estimated contributions $\hat{y}_e[n]$ of the secondary source.

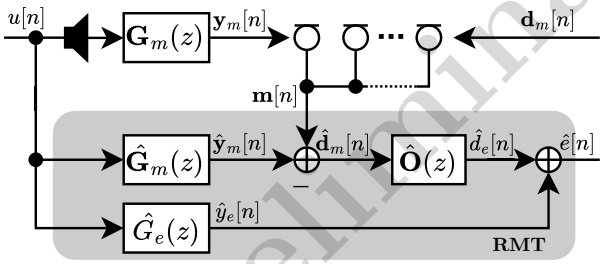



**Figure 1:** Block diagram of the remote microphone technique.

While the plant responses $\hat{\boldsymbol{G}}_m(z)$ and $\hat{G}_e(z)$ can be modeled or measured for certain positions, the observation filter $\hat{\boldsymbol{O}}(z)$ is usually calculated using recordings of the primary sound field. The optimal observation filter coefficients do not only depend on the position of the virtual microphone, but also on the locations and signals of the primary sources. However, instead of calculating a large number of filter coefficient sets in advance, neural networks can be employed to estimate the coefficients during runtime based on the captured sound field and target position information.

### 2.2. Model architecture

In this work, we employ the Obs-TasNet [7] architecture, shown in Figure 2. The Obs-TasNet itself is based on the Conv-TasNet [9], more specifically its inter-channel variant [8]. These architectures are typically used in the context of mask-based speech enhancement. Although the tasks may seem different at first, the objective of extracting a target signal from a dislocated microphone array essentially remains the same. Conv-TasNet and Obs-TasNet share much of their architecture. However, the latter incorporates the virtual microphone's coordinates and processes temporal dependencies across a longer time frame in order to estimate FIR filter coefficients, rather than calculating spectral masks for each audio block individually. Only central aspects of the architecture are outlined in this manuscript; for a more detailed description we refer to [7].

#### *2.2.1. Encoder*

The raw signals of the $N_m$ remote microphones and the Cartesian coordinates of the virtual microphone are transformed to a latent representation using a learnable encoder. The raw microphone data is split into $L$ overlapping segments of length $W$. By using a learnable approach, the encoder can adapt to the input signals for optimal transform of the data, outputting $F$ features for each block $L$. Finally, encoder outputs for the $N_m = C - 1$ microphones and the position encoder output are stacked, forming a $C \times F \times L$ tensor.

#### *2.2.2. Bottleneck*

After applying layer normalization [10], the encoder outputs are processed by bottleneck convolutional layers across all dimensions. To that end, pointwise (or also called $1 \times 1$) convolutions are utilized. These bottlenecks map $F$ to $F_b$ features, $C$ to $C_b$ channels, and $L$ to $L_b$ temporal frames in a latent representation.

#### *2.2.3. TCN network*

The central component of the architecture is a temporal convolutional network (TCN) [11], arranged as several identical TCN stacks. These consist of multiple dilated depthwise separable convolutions [12], shown schematically in Figure 3. Before applying the depthwise separable convolution, the channel dimension is expanded from $C_b$ to $C_{\text{TCN}}$ via a pointwise convolution. To capture mid- and long-term temporal dependencies in the signals, each convolution in a TCN stack exhibits increasing dilation across the latent temporal dimension. While each block receives the residual output of the previous block as input, the output of the whole TCN is generated by summing the skip outputs of all depthwise separable convolutions. Parametric rectified linear units (PReLU) [13] are used as activation functions.

#### *2.2.4. Output transform*

An output transform is applied to map the TCN's output of dimension $C_b \times F_b \times L_b$ to the required

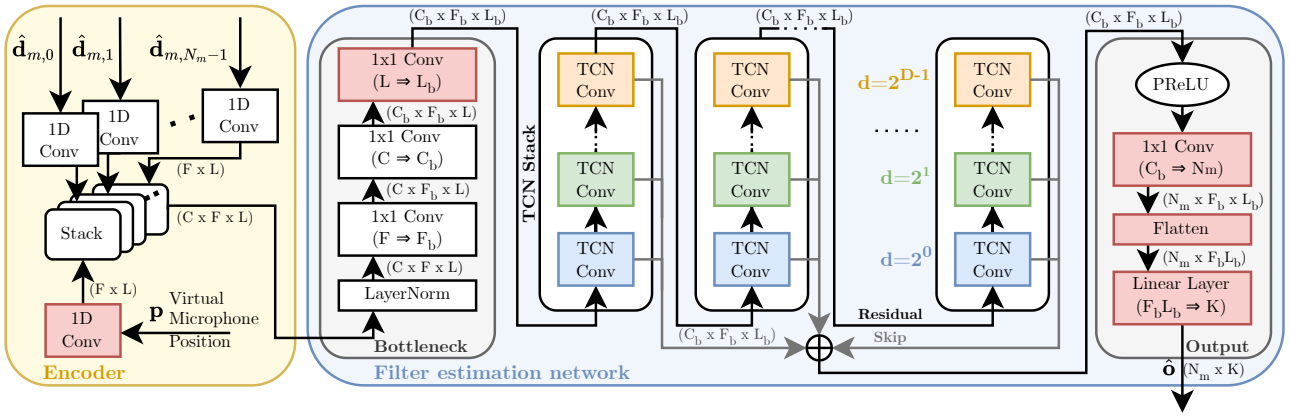



**Figure 2:** Architecture of the Obs-TasNet to estimate observation filter coefficients. Additional layers compared to the original IC Conv-TasNet [8] are displayed red. Expressions in parentheses at signal lines indicate tensor size, parentheses in layers correspond to a transform of dimensions. 1D Conv refers to a standard 1D convolutional layer, TCN Conv to a TCN convolution block with varying colors for different dilation factors $d$, $1 \times 1$ Conv to a pointwise 2D convolution, and PReLU to a parametric rectified linear unit activation function.

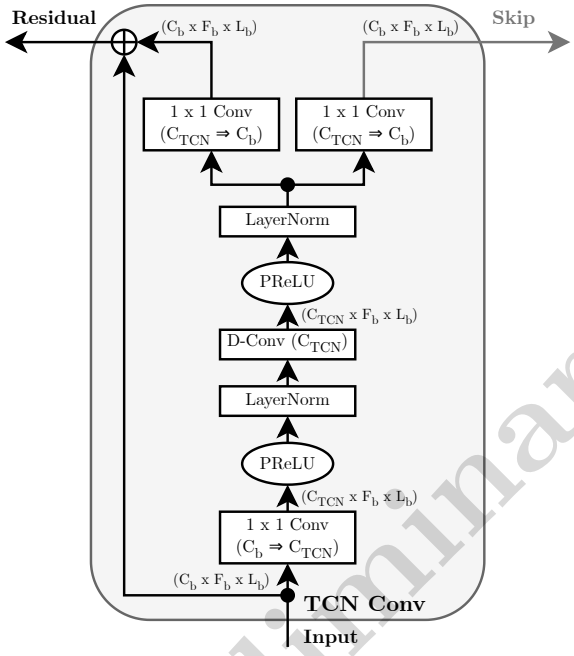



**Figure 3:** Block diagram of a TCN convolution block. Expressions in parentheses at signal lines denote tensor size, parentheses in layers correspond to a transform of dimensions. $1 \times 1$ Conv refers to a pointwise 2D convolution, D-Conv to a depthwise convolution, and PReLU to a parametric rectified linear unit activation function.

matrix of $N_m \times K$ filter coefficients. To that end, the internally used $C_b$ channels are mapped to $N_m$ channels through a pointwise convolution, before the remaining dimensions are flattened and mapped to $K$ filter coefficients using a linear layer.

## 3. Experimental setup

### 3.1. Dataset

This study assesses Obs-TasNet's performance for a large variety of signals in both anechoic and simulated reverberant scenarios. Similar to the original study, remote microphones are arranged as a regular tetrahedron with 28.3 cm edge length, resulting in a spatial aliasing frequency of 606 Hz [14, p.26]. The virtual error microphone's position is selected randomly within a sphere of 10 cm diameter, located at the tetrahedron's center position.

In the anechoic case, a random primary source position with 0.75 m to 3 m distance from the center of the remote microphone tetrahedron is selected. For reverberant scenes, a room with edge lengths between 2 m and 4 m and a reverberation time between 0.2 s and 0.8 s is simulated using `pyroomacoustics`. Remote microphones and primary sources are placed at least 0.75 m apart.

Either colored noise signals with energy proportional to $\frac{1}{f^\beta}$ with frequency $f$ and $\beta \in [0, 2)$ or real-world recordings of disturbances are used as primary source inputs. The recordings are sampled from the FSD50K dataset [15], where only items of domestic and traffic-related sounds with at least 12 s length and 75 % non-zero values are extracted. Exclusively items from the development set are used for training, whereas the items from the evaluation set are split 50/50 into validation and test samples. In sum, 2990 items are pre-selected for training and 776 for validation and testing each.

In total, 50 000 training samples as well as 10 000 validation and test samples with 10 s duration are generated for each of the four datasets at a sample rate of 8 kHz. The gain of the primary input signals is set randomly so that the maximum value lies in the range between −40 dBFS and −16 dBFS to avoid clipping. All samples are truncated to remove the initial settling phase in reverberant scenes. Additionally, the FSD50K samples in the training set are spectrally augmented by a 2nd-order IIR filter with randomized coefficients $a_1, a_2, b_1, b_2 \in \left[-\frac{3}{8}, \frac{3}{8}\right)$ and $a_0 = b_0 = 1$.

### 3.2. Training

Models were trained for each of the four datasets (anechoic colored noise, reverberant colored noise, anechoic FSD, reverberant FSD), configured with the hyperparameters in Table 1 chosen similarly to [7]. Inference is performed every 16 input signal blocks (approximately once per second). Additionally, coefficients are predicted with a 64-tap delay for increased causality with a delayed RMT [16]. For faster calculation, filter coefficients are applied with the overlap-save method. However, processing in the time domain for lower latency in deployed systems is viable. A level invariant [17] mean squared error (MSE) loss between estimated error signal and the ground truth in the time domain is used as training objective.

**Table 1:** Hyperparameters and their descriptions.

| Param | Description | Value |
|---|---|---|
| $W$ | Window length of input signal (50 % overlap) | 1024 |
| $K$ | Number of output FIR coefficients | 513 |
| $F$ | Number of encoder features | 256 |
| $C$ | Number of input channels with $C = N_m + 1$ | 5 |
| $L$ | Number of input signal blocks | 16 |
| $F_b$ | Feature dimension after bottleneck layer | 128 |
| $C_b$ | Channel dimension after bottleneck layer | 64 |
| $L_b$ | Temporal dimension after bottleneck layer | 8 |
| $C_{\mathrm{TCN}}$ | Channels in the TCN convolution | 512 |
| $S$ | Number of TCN stacks | 3 |
| $D$ | Number of TCN convolutions per stack | 6 |

The models are trained for 100 epochs with a batch size of 100 using the Adam optimizer [18]. An initial learning rate of $1 \times 10^{-4}$ is selected. After 50 epochs an exponential learning rate decay is applied, so that a learning rate of $1 \times 10^{-5}$ is reached at epoch 100. All trainings are performed on an NVIDIA RTX 4090 GPU using PyTorch.

### 3.3. Metrics

To evaluate the performance of the trained models in this study we utilize the normalized MSE (NMSE) of primary disturbances

$$\mathrm{NMSE} = 10 \log_{10} \left( \frac{\sum_{n=0}^{N-1} \epsilon[n]^2}{\sum_{n=0}^{N-1} d_e[n]^2} \right) \tag{4}$$

with

$$\epsilon[n] = d_e[n] - \hat{d}_e[n] \tag{5}$$

over each scene with duration $N$. Additionally, the NMSE is evaluated below the array's spatial aliasing frequency of 606 Hz by applying an 8th-order Butterworth lowpass filter before calculation.

To assess performance in the frequency domain, we calculate the estimation error [3] with

$$E(f) = 10 \log_{10} \left( \frac{\hat{S}_{\epsilon\epsilon}(f)}{\hat{S}_{d_e d_e}(f)} \right), \tag{6}$$

where $\hat{S}_{\epsilon\epsilon}(f)$ and $\hat{S}_{d_e d_e}(f)$ are estimates for the power spectral density (PSD) of $\epsilon[n]$ and $d_e[n]$, respectively.

### 3.4. Baseline methods

In this study, Obs-TasNet is benchmarked against a nearest neighbor method (using the signal of the nearest remote microphone) and inverse distance weighting (IDW) [19]. The latter aims to estimate the signal at the virtual microphone via the weighted and summed remote microphone signals with

$$\hat{e}_{\mathrm{IDW}} = \begin{cases} \frac{\sum_{i=0}^{N_m-1} r_i^{-\gamma} m_i[n]}{\sum_{i=0}^{N_m-1} r_i^{-\gamma}} & , \text{if } \; r_i \neq 0 \forall i \\ m_i[n] & , \text{if } \; \exists i : r_i = 0, \end{cases} \tag{7}$$

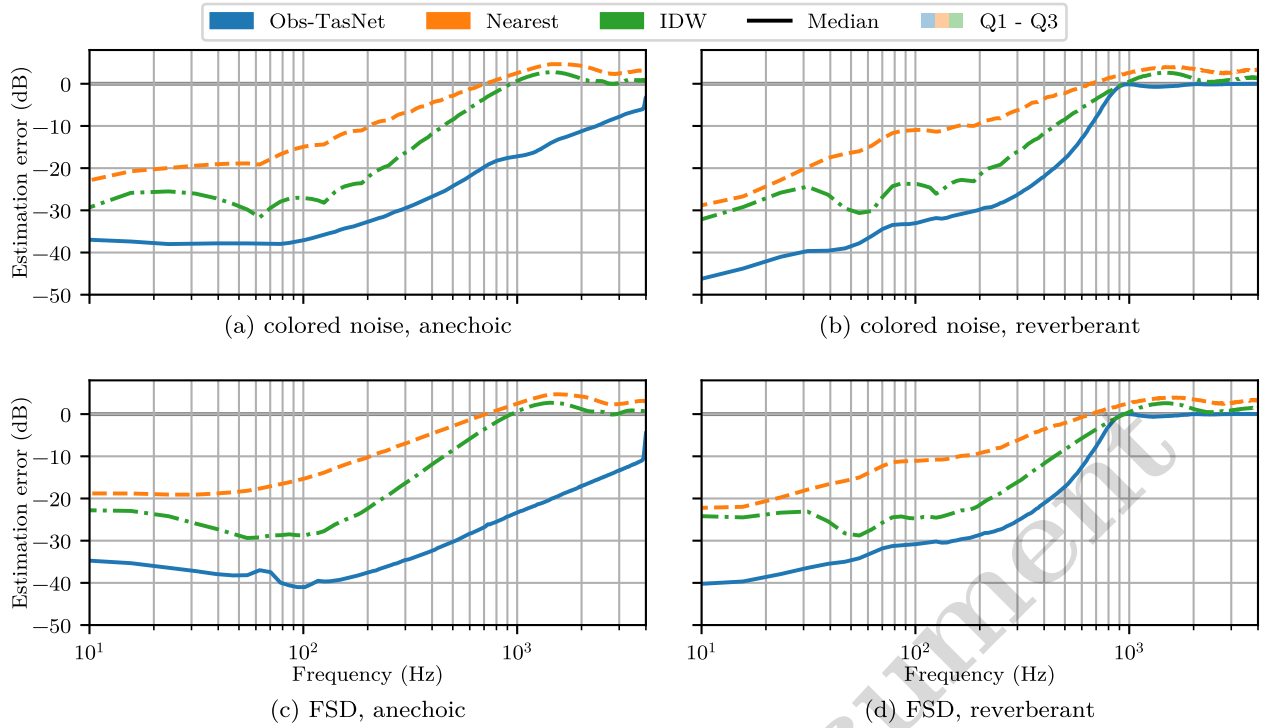
where $r_i$ is the Euclidean distance between the positions of the virtual error microphone and the $i$-th remote sensor with signal $m_i[n]$. The exponent is set to $\gamma = 4$ as suggested in [19].

## 4. Results and Discussion

Table 2 to Table 4 show that Obs-TasNet outperforms the baselines in every tested case. The broadband performance of the trained models is displayed in Table 2, where a 13 dB to 21 dB lower median NMSE for Obs-TasNet is evident in the anechoic case. Interestingly, the median NMSE of Obs-TasNet on the FSD50K dataset is lower than on colored noise. This indicates sufficient filter estimation performance even for time-variant or band-limited signals. When looking at reverberant scenarios, performance of the baseline methods remains consistent compared to the anechoic case whereas Obs-TasNet exhibits a substantial increase in NMSE. However, a consistently low estimation error in Figure 4 towards low frequencies can be observed. This shows that the broadband NMSE in reverberant scenes is dominated by the high frequency error above the microphone array's spatial aliasing frequency. This finding is further confirmed by the NMSE below the array's spatial aliasing frequency listed in Table 3. Here, both IDW and Obs-TasNet exhibit a median NMSE below $-20$ dB with colored noise as primary disturbances, with the latter still outperforming IDW by 8 dB to 12 dB. Even larger performance differences can be observed for the FSD50K dataset. However, Obs-TasNet reacts more sensitively to reverberation than the two baseline methods on the FSD dataset. Whereas the median NMSE in Table 3 even decreases for IDW in reverberant scenes, an increase of almost 12 dB can be observed for Obs-TasNet and the FSD dataset.

In general, the inter-quartile range (IQR) of the estimation error shown in Figure 4(a)-(c) remains consistent for Obs-TasNet across the assessed frequency range in each scenario. Only for reverberant scenes with the FSD dataset in Figure 4(d), a larger IQR towards lower frequencies is observed. However, this behavior is evident for the baseline methods as well.

Table 4 reveals that the performance of all methods, including Obs-TasNet, is relatively robust regarding reverberation time $\mathrm{RT}_{60}$. The chosen length of the observation filter with $K = 513$ (approx. 64 ms) may be a limiting factor for reverberant scenes, even when the direct sound dominates the signal.




(a) colored noise, anechoic

(b) colored noise, reverberant

(c) FSD, anechoic

(d) FSD, reverberant

**Figure 4:** Comparison of the median estimation error for the different datasets with Obs-TasNet, usage of the nearest remote microphone, and inverse distance weighting (IDW). The shaded areas mark the range between first and third quartile.



**Table 2:** Fullband normalized mean squared error (NMSE) in dB: median (Q1, Q3).

| Method | Anechoic | | Reverberant | |
|---|---|---|---|---|
| | Colored Noise | FSD | Colored Noise | FSD |
| Obs-TasNet | **−20.04** (−31.48, −13.05) | **−22.59** (−28.70, −16.46) | **−8.20** (−22.50, 0.06) | **−3.21** (−9.82, 2.95) |
| Nearest | −6.47 (−18.58, 1.19) | −1.16 (−7.38, 2.12) | −4.64 (−17.26, 0.77) | −0.48 (−4.67, 1.94) |
| IDW | −6.12 (−22.05, 1.93) | −0.66 (−6.94, 3.90) | −6.23 (−20.33, 1.56) | −0.44 (−7.01, 4.27) |

**Table 3:** Normalized mean squared error (NMSE) below the spatial aliasing frequency of 606 Hz in dB: median (Q1, Q3).

| Method | Anechoic | | Reverberant | |
|---|---|---|---|---|
| | Colored Noise | FSD | Colored Noise | FSD |
| Obs-TasNet | **−32.53** (−36.48, −28.60) | **−34.61** (−37.73, −30.80) | **−29.24** (−40.08, −23.09) | **−23.01** (−28.10, −18.61) |
| Nearest | −16.30 (−23.09, −8.36) | −9.37 (−15.82, −4.77) | −13.51 (−22.65, −8.30) | −7.56 (−11.59, −4.67) |
| IDW | −20.24 (−29.02, −13.49) | −14.20 (−19.59, −10.02) | −21.06 (−29.38, −14.74) | −14.50 (−20.21, −10.00) |

**Table 4:** Normalized mean squared error (NMSE) below the spatial aliasing frequency of 606 Hz in dB by reverberation time $RT_{60}$: median (Q1, Q3).

| Signal | Method | Reverberation time $RT_{60}$ | | |
|---|---|---|---|---|
| | | 0.2–0.4 s | 0.4–0.6 s | 0.6–0.8 s |
| Colored Noise | Obs-TasNet | **−31.04** (−43.33, −24.00) | **−29.23** (−40.09, −22.90) | **−27.85** (−37.27, −22.52) |
| | Nearest | −15.51 (−25.25, −9.14) | −13.47 (−22.47, −8.14) | −12.19 (−20.33, −7.78) |
| | IDW | −23.00 (−30.89, −15.57) | −21.03 (−29.23, −14.57) | −19.85 (−27.86, −14.27) |
| FSD | Obs-TasNet | **−23.52** (−28.72, −19.03) | **−22.93** (−27.85, −18.52) | **−22.70** (−27.60, −18.39) |
| | Nearest | −8.00 (−12.14, −4.87) | −7.51 (−11.40, −4.63) | −7.26 (−11.11, −4.59) |
| | IDW | −14.92 (−20.76, −10.26) | −14.44 (−20.03, −9.94) | −14.17 (−19.83, −9.84) |

## 5. Conclusion

In this study, we assess the performance of Obs-TasNet, a neural network for online estimation of observation filter coefficients for local ANC in four evaluation scenarios. We trained and tested the NN-based approach in scenarios that include reverberation and real-world noise sources. Compared to a nearest neighbor approach and inverse distance weighting as baseline methods, Obs-TasNet achieves a substantially lower NMSE for anechoic scenes for both colored noise and non-stationary real-world recordings. This behavior can also be observed for reverberant scenes towards low frequencies, although the presence of reverberation deteriorates performance above the array's spatial aliasing frequency. However, the reverberation time itself has little influence on the estimation performance. The findings of this study indicate that Obs-TasNet is well suited for operation in real-world scenarios. Code and data of this article are openly accessible.[1]


## Funding

Computing facilities are funded by the (digital) research infrastructure project "Interactive audiovisual digital twins of performance venues" by the Austrian Federal Ministry of Women, Science and Research.


## References


[1] S. J. Elliott, *Signal Processing for Active Control*, 1st ed. in Signal Processing and Its Applications. Academic Press, 2001. doi: 10.1016/B978-0-12-237085-4.X5000-5.

[2] A. Roure and A. Albarrazin, "The Remote Microphone Technique for Active Noise Control," in *InterNoise '99*, Fort Lauderdale, FL, USA, Dec. 1999, pp. 1233–1244.

[3] W. Jung, S. J. Elliott, and J. Cheer, "Combining the Remote Microphone Technique with Head-Tracking for Local Active Sound Control," *J. Acoust. Soc. Am.*, vol. 142, no. 1, pp. 298–307, July 2017, doi: 10.1121/1.4994292.

[4] S. J. Elliott, W. Jung, and J. Cheer, "Causality and Robustness in the Remote Sensing of Acoustic Pressure, with Application to Local Active Sound Control," in *ICASSP 2019*, Brighton, United Kingdom, May 2019, pp. 8484–8488. doi: 10.1109/ICASSP.2019.8682474.

[5] C. D. Petersen, B. S. Cazzolato, A. C. Zander, and C. H. Hansen, "Active Noise Control at a Moving Location Using Virtual Sensing," in *ICSV13*, Vienna, Austria, June 2006.

[6] R. Xie, A. Tu, C. Shi, S. Elliott, H. Li, and L. Zhang, "Cognitive Virtual Sensing Technique for Feedforward Active Noise Control," in *ICASSP 2024*, Seoul, Korea, Apr. 2024, pp. 981–985. doi: 10.1109/ICASSP48485.2024.10446463.

[7] F. Holzmüller and A. Sontacchi, "Obs-TasNet: Online Estimation of Virtual Sensing Observation Filters for Active Noise Control," *Acta Acust.*, vol. 10, 2026, doi: 10.1051/aacus/2026027.

[8] D. Lee, S. Kim, and J.-W. Choi, "Inter-Channel Conv-TasNet for Multichannel Speech Enhancement," no. arXiv:2111.04312. Nov. 2021. doi: 10.48550/arXiv.2111.04312.

[9] Y. Luo and N. Mesgarani, "Conv-TasNet: Surpassing Ideal Time–Frequency Magnitude Masking for Speech Separation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 27, no. 8, pp. 1256–1266, Aug. 2019, doi: 10.1109/TASLP.2019.2915167.

[10] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer Normalization," no. arXiv:1607.06450. July 2016. doi: 10.48550/arXiv.1607.06450.

[11] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal Convolutional Networks: A Unified Approach to Action Segmentation," in *ECCV 2016*, Amsterdam, The Netherlands, 2016, pp. 47–54. doi: 10.1007/978-3-319-49409-8_7.

[12] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *CVPR 2017*, Honolulu, HI, USA, July 2017, pp. 1800–1807. doi: 10.1109/CVPR.2017.195.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *ICCV 2015*, Santiago, Chile, Dec. 2015, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.

[14] M. Brandstein, D. Ward, A. Lacroix, and A. Venetsanopoulos, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. doi: 10.1007/978-3-662-04619-7.

[15] E. Fonseca, X. Favory, J. Pons, F. Font, and X. Serra, "FSD50K: An Open Dataset of Human-Labeled Sound Events," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 30, pp. 829–852, 2022, doi: 10.1109/TASLP.2021.3133208.

[16] D. Treyer, S. Gaulocher, S. Germann, and E. Curiger, "Towards the Implementation of the Noise-Cancelling Office Chair: Algorithms and Practical Aspects," in *ICSV23*, Athens, Greece, July 2016.

[17] S. Braun and I. Tashev, "Data Augmentation and Loss Normalization for Deep Noise Suppression," in *Speech and Computer*, A. Karpov and R. Potapova, Eds., St. Petersburg, Russia: Springer International Publishing, 2020, pp. 79–86. doi: 10.1007/978-3-030-60276-5_8.

[18] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *ICLR 2015*, San Diego, CA, USA, 2015.

[19] F. Veronesi, C. K. Lai, and J. Cheer, "Interpolation between Plant Responses in a Head-Tracked Local Active Noise Control Headrest System," *Mech. Syst. Signal Process.*, vol. 240, p. 113401, Nov. 2025, doi: 10.1016/j.ymssp.2025.113401.


[1] https://zenodo.org/records/21381469