



# Obs-TasNet: Online estimation of virtual sensing observation filters for active noise control

Felix Holzmüller\*  and Alois Sontacchi 

Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Graz, Austria

Received 30 December 2025, Accepted 13 March 2026

**Abstract** – Local active noise control (ANC) with adaptive processing requires an accurate residual error signal at the point of cancellation, which is often obtained via virtual sensing. We propose Obs-TasNet, a neural approach that estimates observation filter coefficients for the remote microphone technique (RMT) in time-variant settings. By estimating coefficients during operation, the method eliminates the need for pre-optimizing filters for selected scenarios and subsequent interpolation. Obs-TasNet builds on a modified inter-channel Conv-TasNet. The raw waveform signals from remote microphones, as well as the coordinates of the virtual microphone, are embedded as latent representation using a learnable encoding. A temporal convolutional network (TCN), employing dilated depthwise separable convolutions and an output transformation, predicts the observation filter coefficients. Following a brief hyperparameter search, an ablation study demonstrates that the proposed architectural modifications lead to reduced estimation error while substantially reducing both the number of model parameters and computational cost. In simulation, the proposed ANC system with RMT and Obs-TasNet achieves superior noise reduction over a substantially wider frequency range than a multi-point ANC baseline, validating the effectiveness of observation filter estimation.

**Keywords.** Active noise control, ANC, Virtual sensing, Remote microphone technique, RMT, Sound field estimation, Deep learning, Neural networks, Conv-TasNet

## 1 Introduction

Over the last decades, active noise control (ANC) has been widely established as a method to reduce unwanted acoustic disturbances. The underlying principle is fairly simple – so-called secondary sources emit control signals that resemble the primary disturbances, but with inverse phase [1, 2]. When executed properly, noise is reduced through destructive interference. A special form are *local* ANC systems. These systems are designed to achieve noise reduction at specific *points of cancellation* (PoC), rather than across the entire space. As secondary sources in local ANC systems are usually located relatively close to the PoC, the comparably small plant delay and tight coupling allows control of a broader frequency range while not increasing the overall sound energy across the room significantly [1, 3]. A typical use case of local ANC systems is the application in vehicles such as aircraft or cars [4–7], where the performance of passive noise reduction measures is limited by constraints of weight and size.

An inherent property of local ANC systems is the limited extent of the *zone of quiet* (ZoQ), which describes

the area where at least 10 dB attenuation is achieved [8]. Several authors have investigated the shape and extent of the ZoQ for different primary disturbances and secondary source configurations. Most notably, the ZoQ is spherically shaped with a diameter of 1/10-th of the disturbance's wavelength in a pure tone diffuse sound field with a secondary source in the acoustic far field [8]. This finding has been extended towards multiple PoCs [9], secondary sources in the acoustic near field [10–12], and broadband disturbances [13, 14].

A large number of local ANC approaches are implemented as adaptive filters [1, 15–18]. These usually require the residual error signals at the PoC for adaptation. However, the limited extent of the ZoQ often prevents recording the residual error with physical sensors without disturbing the listener. Instead, *virtual sensing* techniques are used to estimate the sound field with a *virtual microphone* at the PoC, relying on nearby (physical) remote microphones and knowledge about the system and disturbances [19]. Usually, virtual sensing techniques are optimized for a single scenario based on pre-recorded data. However, if a mismatch occurs between the trained and the actual scene – for example when the real and

\*Corresponding author: [holzmueller@iem.at](mailto:holzmueller@iem.at)

expected PoCs differ or the primary disturbances change – the estimation accuracy may deteriorate, leading to reduced performance and stability of the ANC system [20]. To handle time-variant situations, multiple virtual sensing models can be trained for different pre-defined scenarios. During operation, the model for the most similar state is used and switched in a nearest-neighbour fashion if the scene changes [21]. This approach has been shown to increase noise control performance for head movements significantly; however, the virtual sensing models have to be optimized for a dense grid of positions and states to avoid a large mismatch. In more elaborate approaches, such as the moving remote microphone technique, the error signal is calculated for several states in parallel and then interpolated towards the desired target state [22]. While in theory error signal estimation can be improved, this approach requires substantially more computational resources during operation due to the parallel computation for several states. Furthermore, signal estimation accuracy depends on the chosen interpolation algorithm as well [23]. Recently, kernel-based interpolation techniques have been used for ANC tasks [24, 25]. While these techniques perform well in estimating the sound field inside a remote microphone arrangement, prior assumptions about wave propagation and source direction have to be made [24].

To handle and process time-variant scenes with conventional virtual sensing strategies, usually multiple models are optimized for different states and interpolated by some processing logic [19, 21–23]. This requires the development of a scene-detection algorithm, a pre-selection of representative states, and a distinct optimization of virtual sensing models for each state. In contrast, (deep) neural networks are well known to be capable of identifying patterns and classifying acoustic scenes [26, 27]. Neural network-based classification models have been successfully adapted for selecting filters in virtual sensing approaches in active control [28, 29]. However, interpolation and manual filter coefficient calculation for selected scenes and positions may still be necessary. Recently, neural networks have been employed for sound field reconstruction and interpolation [30–33]. While these approaches perform well in estimating sound fields, the networks’ architectures, complexity, and processing latency often prohibit their use in real-world ANC systems. Different tasks that show some resemblance to virtual sensing are acoustic beamforming and multichannel speech enhancement [34], all of which aim to extract a certain target signal by processing signals of a microphone array. However, neural beamformers and speech enhancement algorithms traditionally often operate by applying masks on the spectrograms of the input signals [35–37]. While mask-based approaches provide good performance for speech enhancement, the inherent latency caused by the block processing may exceed requirements for some ANC systems. However, a neural beamformer can be modified to estimate the required filter coefficients for conventional virtual sensing algorithms [38]. The obtained coefficients can then either be applied in

time-domain filtering to achieve low processing delay, or in fast algorithms with block processing to minimize computational load.

In this article, we propose the *Obs-TasNet* to estimate filter coefficients of the observation filter for the remote microphone technique (RMT) [39] with an asynchronously operating neural network. In the proposed estimation approach, recorded signals of the remote microphones are processed in conjunction with the position of the virtual microphone. During operation, no additional information about the disturbances such as the location of the primary source is provided. By using an end-to-end architecture, explicit calculation of virtual sensing weights for a limited number of selected states, scene detection, and interpolation is mitigated. With the neural estimation of the observation filter coefficients outsourced to an external co-processor or a neural processing unit, the proposed method allows for efficient real-time estimation of the error signal. Section 2 explains the filtered-reference least mean squares (FxLMS) algorithm and the RMT as fundamental signal processing techniques for local ANC with virtual sensing. The architecture of the *Obs-TasNet* is described in Section 3. In Section 4, the generated dataset, training setup, and evaluation metrics are described. The model is evaluated in Section 5, before concluding in Section 6. This article is based on previously published pilot studies with simpler architectures in time-invariant scenarios [40, 41]. A PyTorch-implementation of the model is openly accessible<sup>1</sup> [42].

## 2 Signal processing for local ANC

This section summarises the FxLMS algorithm and the RMT as commonly used signal processing methods for ANC in time-discrete systems.

### 2.1 FxLMS

A large number of ANC systems are based on the FxLMS algorithm [1]. In a simple single-input-single-output system, the control signal for a secondary source

$$u[n] = \mathbf{w}^T \mathbf{x}[n] \quad (1)$$

with discrete time index  $n$  is generated by processing a reference signal  $\mathbf{x}[n] = [x[n] \ x[n-1] \ \dots \ x[n-H+1]]^T$  with a control filter  $w = [w_0 \ w_1 \ \dots \ w_{H-1}]^T$  of order  $H-1$ . The reference signal must be closely related to the primary disturbances for sufficient noise reduction performance. The adaptive control filter is updated with

$$\mathbf{w} \leftarrow \mathbf{w} - \mu e[n] \mathbf{x}_f[n], \quad (2)$$

where  $e[n]$  refers to the error signal at the PoC,  $\mu$  to the step-size for adaptation, and  $\mathbf{x}_f[n] =$

<sup>1</sup> <https://github.com/fholzm/Obs-TasNet>

$[x_f[n] \ x_f[n-1] \ \dots \ x_f[n-H+1]]^T$  to the filtered reference signal. The latter is generated by convolving the reference signal with the secondary plant response  $\mathbf{g}_e$ , corresponding to the transfer path between secondary source and the PoC. Usually, only an estimate  $\hat{\mathbf{g}}_e = [\hat{g}_{e,0} \ \hat{g}_{e,1} \ \dots \ \hat{g}_{e,I-1}]^T$  with order  $I-1$  is available, which defines the filtered reference signal to

$$x_f[n] = \hat{\mathbf{g}}_e^T \mathbf{x}[n]. \quad (3)$$

In many cases, the error signal  $e[n]$  cannot be recorded directly with a physical microphone at the desired PoC. Instead, the signal can be estimated by virtual sensing techniques such as the RMT.

## 2.2 Remote microphone technique

The RMT [39] is an extensively used state-of-the-art method [6, 21, 43, 44] for estimating the residual error signal at a virtual microphone. A block diagram of the RMT is shown in Figure 1a. For simplicity, only a single virtual microphone is considered in the following.

The error signal at the virtual microphone is defined as

$$e[n] = y_e[n] + d_e[n], \quad (4)$$

where  $d_e[n]$  refers to the primary disturbances and  $y_e[n]$  to the contributions of the secondary source. The latter can be estimated and decomposed to

$$\hat{y}_e[n] = \hat{\mathbf{g}}_e^T \mathbf{u}[n] \quad (5)$$

with the control signal vector for the secondary source  $\mathbf{u}[n] = [u[n] \ u[n-1] \ \dots \ u[n-I+1]]^T$ .

To estimate the signal at the virtual error microphone,  $N_m$  nearby remote microphones are used. The signal at the  $r$ -th remote microphone is, similar to equation (4), given as

$$m_r[n] = y_{m,r}[n] + d_{m,r}[n], \quad (6)$$

which is composed of the primary disturbance  $d_{m,r}[n]$  and the contributions of the secondary sources  $y_{m,r}[n]$  at the respective remote microphone. Similar to equation (5), they are calculated as

$$\hat{y}_{m,r}[n] = \hat{\mathbf{g}}_{m,r}^T \mathbf{u}[n] \quad (7)$$

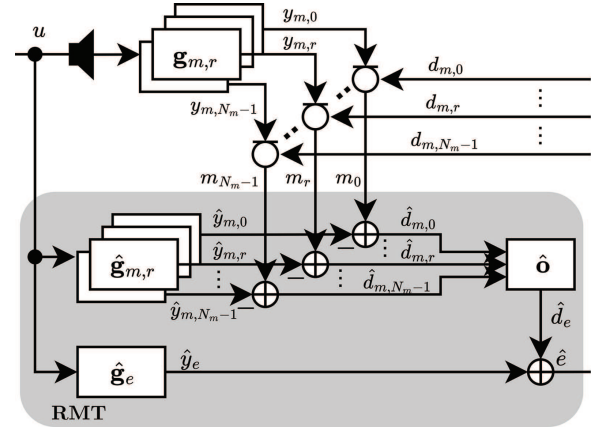
with the estimated  $(J-1)$ -th order secondary plant response  $\hat{\mathbf{g}}_{m,r} = [\hat{g}_{m,r,0} \ \hat{g}_{m,r,1} \ \dots \ \hat{g}_{m,r,J-1}]^T$ .

As the remote microphone signals  $m_r[n]$  and the control signals  $\mathbf{u}[n]$  are typically known, the primary disturbances at the remote microphones can be extracted as

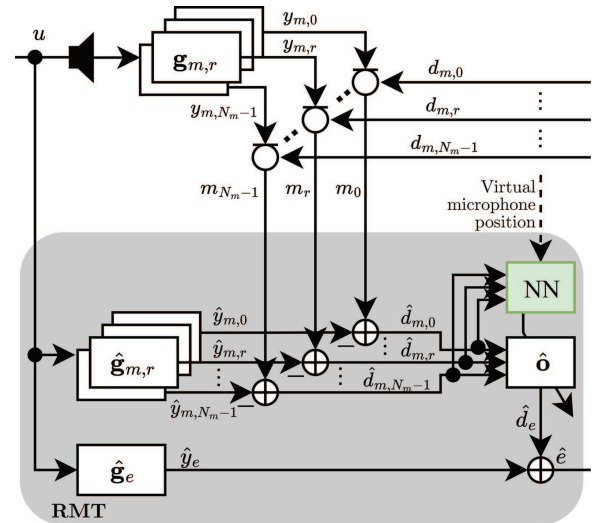
$$\hat{d}_{m,r}[n] = m_r[n] - \hat{y}_{m,r}[n] \quad (8)$$

$$= m_r[n] - \hat{\mathbf{g}}_{m,r}^T \mathbf{u}[n]. \quad (9)$$

With the estimate of the primary disturbances at the *remote microphones*  $\hat{d}_{m,r}[n]$ , an *observation filter*



(a) Conventional remote microphone technique.



(b) Remote microphone technique with neural network (NN) to estimate filter coefficients during operation.

**Figure 1.** Block diagram of the remote microphone technique and the proposed modification for online observation filter coefficient estimation. The discrete time index  $n$  of signals is omitted for easier readability.

$\hat{\mathbf{o}} = [\hat{\mathbf{o}}_0 \ \hat{\mathbf{o}}_1 \ \dots \ \hat{\mathbf{o}}_r \ \dots \ \hat{\mathbf{o}}_{N_m-1}] \in \mathbb{R}^{K \times N_m}$  with  $\hat{\mathbf{o}}_r = [\hat{o}_{r,0} \ \hat{o}_{r,1} \ \dots \ \hat{o}_{r,K-1}]^T$  of order  $K-1$  can be applied to calculate the estimated primary disturbances at the *virtual microphone* as

$$\hat{d}_e[n] = \sum_{r=0}^{N_m-1} \hat{\mathbf{o}}_r^T \hat{\mathbf{d}}_{m,r}[n], \quad (10)$$

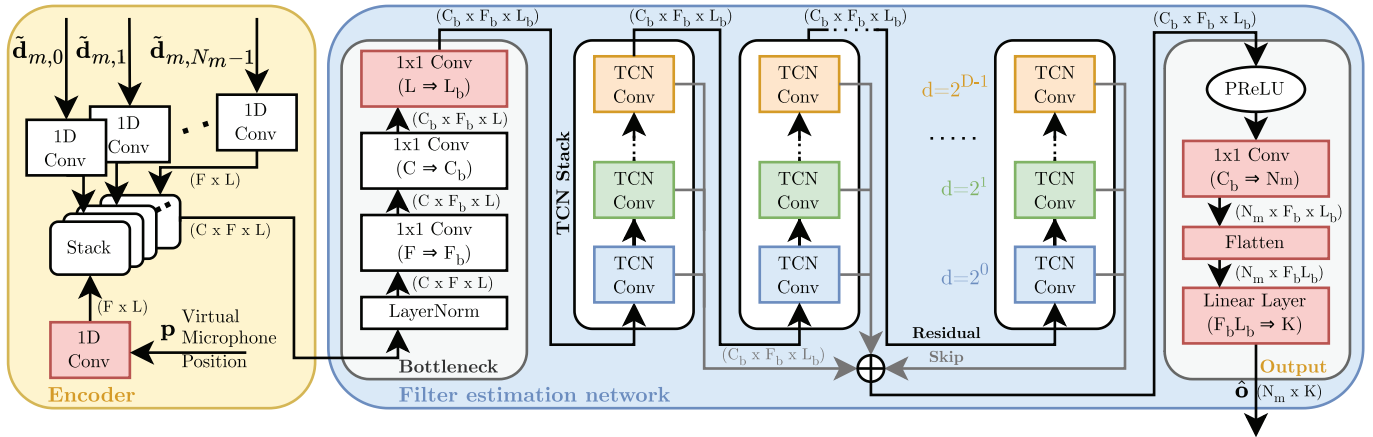
where

$$\hat{\mathbf{d}}_{m,r}[n] = [\hat{d}_{m,r}[n] \ \hat{d}_{m,r}[n-1] \ \dots \ \hat{d}_{m,r}[n-K+1]]^T.$$

By inserting equation (10) in equation (4), the error signal at the virtual microphone is estimated as

$$\hat{e}[n] = \hat{y}_e[n] + \hat{d}_e[n] \quad (11)$$

$$= \hat{y}_e[n] + \sum_{r=0}^{N_m-1} \hat{\mathbf{o}}_r^T \hat{\mathbf{d}}_{m,r}[n]. \quad (12)$$



**Figure 2.** Architecture of the Obs-TasNet to estimate observation filter coefficients. Additional layers compared to the original IC Conv-TasNet [37] 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 colours for different dilation factors  $d$ , *1 x 1 Conv* to a pointwise 2D convolution, and *PReLU* to a parametric rectified linear unit activation function.

To ensure causality of the observation filter, the *delayed RMT* [45] can be used. Instead of estimating the current value of  $\hat{e}[n]$ , a  $\Delta$  samples delayed version

$$\hat{e}[n - \Delta] = \hat{y}_e[n - \Delta] + \hat{d}_e[n - \Delta] \quad (13)$$

$$= \hat{y}_e[n - \Delta] + \sum_{r=0}^{N_m-1} \hat{\mathbf{o}}_{\Delta,r}^T \hat{\mathbf{d}}_{m,r}[n] \quad (14)$$

is calculated. By applying this delay, substantial parts of the observation filter can be shifted into a causal range [5]. Typically, the delay  $\Delta$  should exceed the maximum time difference of arrival of primary disturbances between the remote microphone arrangement and the virtual microphone.

The secondary plant responses  $\hat{\mathbf{g}}_e$  and  $\hat{\mathbf{g}}_{m,r}$  can be measured for various PoCs and interpolated afterwards [7, 23] as they only depend on the position of the virtual microphone. However, the observation filter  $\hat{\mathbf{o}}$  is influenced by the (virtual) error microphone position, primary source position, and spectral properties of the primary sources [21]. Estimating primary disturbances at the virtual error microphone using an observation filter forms an inverse problem that is sensitive to uncertainties and mismatches between pre-calculated filter sets and observed scenes [20, 21].

To mitigate these problems and provide an end-to-end solution to cover several scenarios, we propose an extension of the RMT by a neural network that estimates observation filter coefficients during operation as indicated in Figure 1b. By estimating the filter coefficients asynchronously on a separate thread or co-processor, no additional computational resources are required in the real-time branch of the ANC system compared to conventional processing. The following section details the proposed model architecture.

### 3 Model architecture

To estimate the coefficients of the observation filter, we propose *Obs-TasNet*, a modified version of the inter-channel fully-convolutional time-domain audio separation network (IC Conv-TasNet) [37]. Initially, it was proposed for multichannel speech enhancement, e.g. for processing speech recorded with a microphone array in a noisy environment. The IC Conv-TasNet itself is based on the original Conv-TasNet [46] and its multichannel extension [36], but provides higher performance with a significantly reduced number of parameters [37]. Key components of the (IC) Conv-TasNet are latent representations of the input data learned from the waveform domain and temporal convolutional network (TCN) stacks [47] with dilated depthwise separable convolutions [48].

In its original form, the IC Conv-TasNet estimates a real-valued mask for application on one transformed input channel. Therefore, it cannot be used for virtual sensing, as important phase information is discarded [49] and operation with tight real-time constraints is not possible due to the block processing. However, we modified the architecture to ensure suitability for virtual sensing tasks. A network overview is shown in Figure 2. For easier readability, the most important hyperparameters are described in Table 1. The two main components of the network, the encoder and the actual filter estimation network, are described in the following in detail.

#### 3.1 Encoder

The encoder transforms the input signals, namely the estimated primary disturbances  $\hat{\mathbf{d}}_m[n]$  at the remote microphones, into a latent feature space. A commonly used transform for many signal processing tasks is the short-time Fourier transform (STFT). While the STFT is a versatile method suitable for many problems, it is

**Table 1.** Hyperparameters of the Obs-TasNet and their description.

Param	Description
$W$	Window length for the input signal
$K$	Number of output FIR coefficients
$F$	Number of encoder features
$C$	Number of input channels with $C = N_m + 1$
$L$	Number of input signal blocks
$F_b$	Feature dimension after bottleneck layer
$C_b$	Channel dimension after bottleneck layer
$L_b$	Temporal dimension after bottleneck layer
$S$	Number of TCN stacks
$D$	Number of TCN convolution blocks per stack
$C_{\text{TCN}}$	Channels in the TCN convolution blocks

necessarily not the *optimal* transform for each task [46, 50, 51].

Therefore, TasNet-style networks use a learnable input encoding [37, 46, 52]. The time domain input signals are buffered and split into  $L$  overlapping segments of length  $W$ , where  $\tilde{\mathbf{d}}_{m,r} \in \mathbb{R}^{W \times L}$  refers to the blocked signal of the  $r$ -th input channel  $\hat{d}_{m,r}[n]$ . The audio data is then transformed to feature space as  $F$ -dimensional latent representation by applying a linear transform

$$\mathbf{D}_{m,r} = \mathbf{U}_d \tilde{\mathbf{d}}_{m,r}, \quad (15)$$

where  $\mathbf{D}_{m,r} \in \mathbb{R}^{F \times L}$  is the transformed input data of channel  $r$ , and  $\mathbf{U}_d \in \mathbb{R}^{F \times W}$  a matrix containing the  $F$  encoder basis vectors. The encoder matrix  $\mathbf{U}_d$  is trained end-to-end with the filter estimation network and shared across all input audio data channels.

For the proposed virtual sensing application, coordinates of the virtual microphone are processed as well. For each of the  $L$  input data blocks, Cartesian coordinates of the virtual microphone position are provided. In the same fashion as in equation (15), the matrix  $\mathbf{p} \in \mathbb{R}^{3 \times L}$  of coordinates for the  $L$  segments is transformed

$$\mathbf{P} = \mathbf{U}_p \mathbf{p} \quad (16)$$

with an encoder matrix  $\mathbf{U}_p \in \mathbb{R}^{F \times 3}$  to the encoded position  $\mathbf{P} \in \mathbb{R}^{F \times L}$ .

After encoding, all audio and position encoder outputs are stacked [37], forming a tensor of dimension  $(C \times F \times L)$ , where  $C = N_m + 1$ .

The input encoding in equations (15) and (16) is described as separate blocking and matrix multiplication. However, in practice these operations are implemented as a single convolutional layer with output channel size  $F$ , directly applied on the time-domain signal. This reformulation improves training speed and convergence [46].

### 3.2 Filter estimation network

The main part of this network is the filter estimation network, consisting of multiple TCN stacks [37, 46, 47]

with dilated depthwise convolutions [48]. This way, temporal dependencies are captured similar to networks based on recurrent architectures, but with a smaller number of parameters and less influence of long-term dependencies [46]. Before processing in the actual TCN stacks, data is passed through bottleneck layers; the output of the TCN is transformed subsequently to provide the results in the correct shape of individual filter coefficients.

#### 3.2.1 Bottleneck layers

To reduce complexity, the encoded inputs are first passed through three bottleneck layers, transforming each of the input tensor's dimensions. The bottleneck layers are implemented as pointwise convolutions with kernel size 1, also referred to as  $1 \times 1$  convolutions. As proposed in the original Conv-TasNet [46], the first bottleneck layer maps the  $F$  encoder outputs to  $F_b$  features. The IC-Conv-TasNet introduces a second bottleneck to map the  $C = N_m + 1$  input channels onto  $C_b$  shared channels. For speech enhancement, an individual mask is estimated for each of the encoded input data blocks. However, several blocks are processed jointly in the Obs-TasNet to calculate filter coefficients asynchronously. This allows the introduction of a third bottleneck layer along the temporal dimension to map the  $L$  input blocks to  $L_b$  temporal features. Before passing the signal to the bottleneck, layer normalization [53] is applied<sup>2</sup>.

#### 3.2.2 TCN stacks

Subsequently, data is processed by  $S$  identical TCN stacks, each consisting of  $D$  TCN convolution blocks with increasing dilation to capture different temporal dependencies. The TCN convolution blocks return two tensors – a residual output passed to the next TCN block, as well as a skip output. The summed skip outputs of all blocks are used to compute the observation filter coefficients.

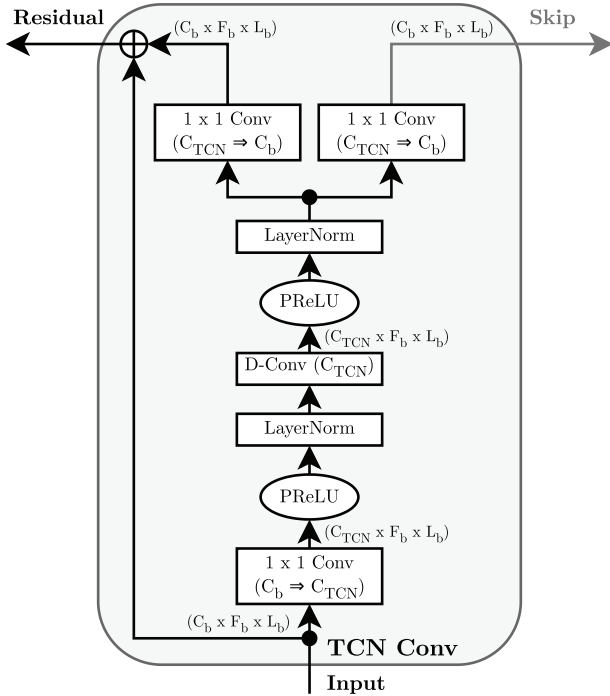
The structure of a TCN convolution block is shown in Figure 3. At first, the input is processed by a pointwise convolution, mapping the  $C_b$  channel features to  $C_{\text{TCN}} > C_b$  internal features [37]. A parametric rectified linear unit (PReLU) [55] is used as activation function, defined as

$$\text{PReLU}(x) = \begin{cases} x, & \text{if } x > 0, \\ ax, & \text{otherwise,} \end{cases} \quad (17)$$

with a learnable parameter  $a \in \mathbb{R}$ .

After layer normalization [53], a depthwise separable convolution ( $S$ -Conv) [48] is performed. The  $S$ -Conv is used in image and audio processing as an alternative to conventional convolution with less parameters. To that end, the convolution is split into a 2D depthwise convolution ( $D$ -Conv) and dedicated pointwise convolutions for residual and skip output. A  $D$ -Conv performs the convolution for each of the  $C_{\text{TCN}}$  channels separately, while

<sup>2</sup> For a more straight-forward implementation, group normalization [54] with just one group is used to act as layer normalization.



**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.  $1 \times 1 \text{ Conv}$  refers to a pointwise 2D convolution,  $D\text{-Conv}$  to a depthwise convolution, and  $PReLU$  to a parametric rectified linear unit activation function.

the pointwise convolution is applied over  $C_{TCN}$  to map the internally used channels to  $C_b$  outputs. Between the  $D\text{-Conv}$  and the pointwise convolutions, layer normalization and a  $PReLU$  activation function is applied. The  $D\text{-Conv}$ s are dilated along their temporal dimension. In each TCN stack, the  $D\text{-Conv}$  in the first TCN convolution block uses a dilation of  $d = 2^0$ , the second TCN convolution block  $d = 2^1$ , until the  $D$ -th block is dilated by  $d = 2^{D-1}$ . Zero-padding is applied to keep tensor dimensions consistent. While the skip connections are returned right after the pointwise convolution, the residual output is summed with the input of the TCN convolution block.

### 3.2.3 Output transform

The summed skip output of the TCN network is further processed to return the matrix of observation filter coefficients  $\hat{\mathbf{o}}$ . After applying a  $PReLU$  activation function, the channel dimension  $C_b$  is mapped to the actually required  $N_m$  microphone channels by a pointwise convolution. The remaining two dimensions corresponding to temporal and feature information are then flattened and processed by a linear layer, mapping the  $F_b L_b$  features to  $K$  observation filter coefficients.

## 4 Experimental setup

### 4.1 Dataset

The network is trained, validated, and tested with synthetic datasets, generated in TASCAR<sup>3</sup> [56, 57]. In all scenarios, a single primary source with random azimuth and elevation and a distance of 1.5 m to 3 m to the centre of the coordinate system is created. Gaussian distributed noise with a spectral density proportional to  $1/f^\beta$  with frequency  $f$  and  $\beta \in [0; 2]$  is used as innovation signal, driving the primary source. Four remote microphones are positioned at the Cartesian coordinates  $(0.1, 0.1, 0.1)$ ,  $(0.1, -0.1, -0.1)$ ,  $(-0.1, 0.1, -0.1)$ ,  $(-0.1, -0.1, 0.1)$  m, forming a tetrahedral arrangement around the coordinate system's centre. It should be noted, that sound propagation delay in TASCAR is by default rounded to integer samples, introducing a bias of up to  $\pm 31.25 \mu\text{s}$  with the selected sample rate.

Two different datasets are created: one with time-invariant (*static* dataset), the other with time invariant and time-variant virtual microphone position (*mixed* dataset). In the former, the microphone position within a sphere with 10 cm diameter around the coordinate system's centre is drawn from a uniform distribution. For the second dataset, the virtual microphone position is moved up to two times in each scene, where each movement follows a linear trajectory with a duration of 0.5 s to 2 s and a pause of 2 s to 6 s between movements.

In total, each dataset consists of 80 000 scenes with 20 s duration, sampled at a rate of 16 kHz. Training and validation data are split with a 80/20% ratio. Additionally, two test sets with 1000 scenes are created, containing either just static or just time-variant virtual microphone positions. A uniformly randomized gain is applied to all scenes, so that a root mean square (RMS) level of  $-40$  dB to  $-16$  dB is achieved at the virtual error microphone to avoid clipping.

### 4.2 Training

If not specified otherwise, observation filters coefficients for a delayed RMT [45] with  $K = 257$  taps and a delay of  $\Delta = 64$  samples are predicted. Obs-TasNet is inferred every 512 ms, corresponding to 32 frames of input data with  $W = 512$ . The models are trained by minimizing the mean squared error (MSE) between estimated primary disturbances  $\hat{d}_e[n]$  at the virtual microphone and a ground truth  $d_e[n]$ . For a level invariant loss [58], the signals are normalized using the ground truth's RMS value. By calculating the loss on the time domain signals and for example not on magnitude spectra, phase distortions are penalized as well. The estimated primary disturbances  $\hat{d}_e[n]$  are calculated by convolving the obtained observation filter coefficients  $\hat{\mathbf{o}}$  with the primary disturbances  $\mathbf{d}_{m,r}[n]$  as described in equation (10).

<sup>3</sup> <https://www.tascar.org>

For faster training and evaluation, filtering is performed using the overlap-save method [59, Sect. 8.7.3].

The models are trained in two stages – at first 50 epochs with the static dataset, followed by 50 epochs with the mixed set. The initial training with the static dataset is carried out to properly initialize model weights before training with the mixed dataset. The Adam optimiser [60] is used in all cases – for the static dataset with a fixed learning rate of  $1 \times 10^{-4}$ , for the mixed dataset with exponentially decaying learning rate starting at  $1 \times 10^{-4}$  until reaching  $1 \times 10^{-5}$  after 50 epochs. Learning rate, scheduling, and the training duration of 50 epochs have been determined in preliminary trainings for optimal convergence while avoiding overfitting. All trainings are performed using PyTorch with a batch size of 100 on either a NVIDIA RTX Pro 6000 Blackwell Max-Q or a NVIDIA RTX 3080 GPU.

### 4.3 Metrics

A central validation metric is the normalized MSE (NMSE)

$$\text{NMSE} = 10 \log_{10} \left( \frac{\sum_{n=0}^{\infty} \epsilon[n]^2}{\sum_{n=0}^{\infty} d_e[n]^2} \right) \quad (18)$$

with

$$\epsilon[n] = d_e[n] - \hat{d}_e[n], \quad (19)$$

providing a single scalar for the broadband performance. Closely related to the NMSE is the estimation error [21]

$$E(f) = 10 \log_{10} \left( \frac{\hat{S}_{\epsilon\epsilon}(f)}{\hat{S}_{d_e d_e}(f)} \right), \quad (20)$$

where  $\hat{S}_{\epsilon\epsilon}(f)$  and  $\hat{S}_{d_e d_e}(f)$  with frequency  $f$  refer to an estimate of the power spectral density (PSD) of  $\epsilon[n]$  and  $d_e[n]$ , respectively, calculated using Welch's method [61]. The estimation error  $E(f)$  can be interpreted as the NMSE in frequency domain.

A metric for the evaluation of the overall performance of an ANC system is the noise reduction

$$\text{NR}(f) = 10 \log_{10} \left( \frac{\hat{S}_{d_e d_e}(f)}{\hat{S}_{\epsilon\epsilon}(f)} \right), \quad (21)$$

where the estimated PSD  $\hat{S}_{\epsilon\epsilon}(f)$  of the error signal  $e[n]$  is compared to the estimated PSD of primary disturbances  $d_e[n]$ .

## 5 Experiments

Several experiments are conducted to assess the proposed Obs-TasNet. They can be grouped into four tasks, namely

1. a hyperparameter search to find suitable combinations,

2. an ablation study regarding the additional bottleneck along the temporal dimension
3. an assessment of the estimation performance based on the virtual microphone position, and
4. an evaluation as part of an ANC system, compared to a multi-point ANC system using the error signals at the remote microphones.

### 5.1 Hyperparameter search

A brief search for a suitable combination of the hyperparameters listed in Table 1 has been conducted. While an extensive grid search is not feasible, selected hyperparameters are varied sequentially. The tested combinations, corresponding model parameter count, computational complexity, and resulting mean validation NMSE are shown in Table 2.

Similar hyperparameters as in the original IC Conv-TasNet [37] are used as initial configuration. A selection of  $S = 3$  TCN stacks with  $D = 6$  blocks shows best performance for the selected dataset and parameter configuration. Increasing the number of input signal blocks from  $L = 32$  to  $L = 64$  does increase complexity slightly, but does not influence the mean NMSE. Similarly, increasing or decreasing the number of encoded features  $F$  has little influence, presumably due to the used Gaussian primary disturbances. Changes in the output dimensions of the bottleneck layers have larger impact on the complexity and number of parameters, yet the initial selection of  $L_b = 8$ ,  $F_b = 128$ , and  $C_b = 64$  yields in lowest mean NMSE. However, increasing the number of channels in the TCN convolution blocks to  $C_{\text{TCN}} = 512$  can further reduce mean NMSE by 0.27 dB, but almost doubles computational complexity. A reduction of the number of coefficients to  $K = 129$  affects only the linear layer at the output, hence mainly the number of parameters is influenced while computational complexity barely changes. As the filter order still exceeds the sound propagation delay of the microphone aperture, only a minor decrease of mean NMSE can be observed.

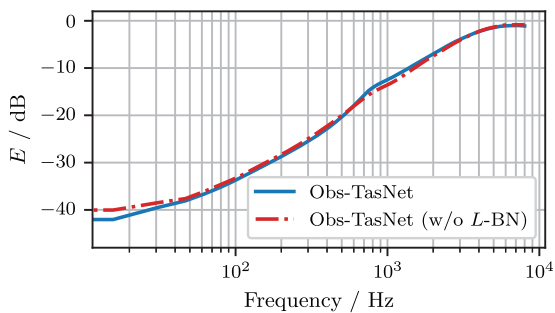
In general, the mean NMSE values in Table 2 differ only slightly across most hyperparameter configurations. This is expected because the broadband metric evaluates the entire frequency range of signals sampled at 16 kHz. As shown in Figure 4, even for the best-performing model the estimation error tends toward 0 dB at higher frequencies, which disproportionately increases the overall error relative to lower frequencies where ANC systems are primarily applied.

### 5.2 Ablation study – Temporal bottleneck layer

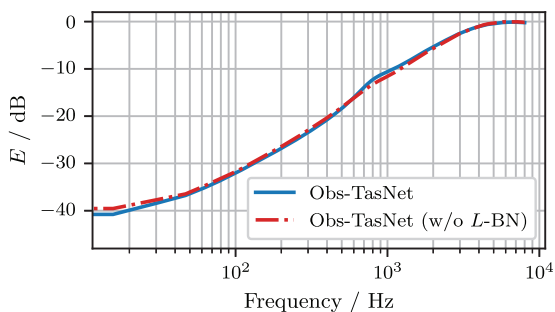
In order to reduce computational load and jointly process temporal dependencies, the Obs-TasNet features compared to the IC Conv-TasNet an additional bottleneck, mapping the  $L$  input data blocks to  $L_b$  latent representations. To demonstrate the effectiveness, a version of the Obs-TasNet based on hyperparameter configuration 2 in Table 2, but without the temporal bottleneck

**Table 2.** Mean validation NMSE for different hyperparameter combinations from Table 1. Varied parameters are highlighted. The number of model parameters is shown in millions, the computational complexity as  $10^9$  multiply-accumulate operations (GMACs).

ID	$D$	$S$	$L$	$L_b$	$F$	$F_b$	$C_b$	$C_{\text{TCN}}$	$K$	Params	Complexity	Mean NMSE
1	4	3	32	8	256	128	64	256	257	1.07 M	0.69 GMACs	-14.56 dB
2	6	3	32	8	256	128	64	256	257	1.39 M	1.03 GMACs	-15.32 dB
3	8	3	32	8	256	128	64	256	257	1.70 M	1.36 GMACs	-14.73 dB
4	6	2	32	8	256	128	64	256	257	1.07 M	0.69 GMACs	-14.53 dB
5	6	4	32	8	256	128	64	256	257	1.70 M	1.36 GMACs	-15.13 dB
6	6	3	64	8	256	128	64	256	257	1.39 M	1.05 GMACs	-15.32 dB
7	6	3	32	8	128	128	64	256	257	1.30 M	1.01 GMACs	-14.72 dB
8	6	3	32	8	512	128	64	256	257	1.55 M	1.05 GMACs	-15.01 dB
9	6	3	32	4	256	128	64	256	257	1.25 M	0.52 GMACs	-14.64 dB
10	6	3	32	16	256	128	64	256	257	1.65 M	1.63 GMACs	-14.95 dB
11	6	3	32	8	256	64	64	256	257	1.24 M	0.52 GMACs	-14.41 dB
12	6	3	32	8	256	256	64	256	257	1.68 M	2.03 GMACs	-15.16 dB
13	6	3	32	8	256	128	32	256	257	0.94 M	0.57 GMACs	-14.14 dB
14	6	3	32	8	256	128	128	256	257	1.94 M	2.27 GMACs	-15.09 dB
15	6	3	32	8	256	128	64	128	257	0.91 M	0.53 GMACs	-15.05 dB
16	6	3	32	8	256	128	64	512	257	2.34 M	2.02 GMACs	-15.59 dB
17	6	3	32	8	256	128	64	256	129	1.25 M	1.02 GMACs	-15.11 dB



(a) Static



(b) Moving

**Figure 4.** Estimation error  $E$  of the Obs-TasNet and a modified version without the temporal bottleneck layer (Obs-TasNet (w/o  $L$ -BN)) in scenes with (a) static and (b) moving virtual microphone.

layer, is trained. The performance is evaluated and compared for the two test sets, consisting exclusively of either time-invariant and time-variant scenes, respectively.

Results of this brief ablation study are shown in Table 3 and Figure 4. Most obviously, the proposed architecture uses about 40% fewer parameters, and reduces computational complexity by factor 4. At the same time, NMSE is decreased by approximately 0.5 dB for scenes with static and 0.4 dB for scenes with moving virtual microphone. The reduction of the estimation error with additional bottleneck layer is mainly visible towards low frequencies.

### 5.3 Position-dependent performance

Using hyperparameter combination 16 in Table 2, we briefly evaluated the static test set regarding the distance  $a$  of the virtual microphone from the centre of the arrangement. In line with [41], the mean NMSE reported in Table 4 tends to increase with  $a$ , though not strictly monotonic. Figure 5 further shows that the estimation error increases steadily with  $a$  in the lower-frequency region.

### 5.4 ANC simulation

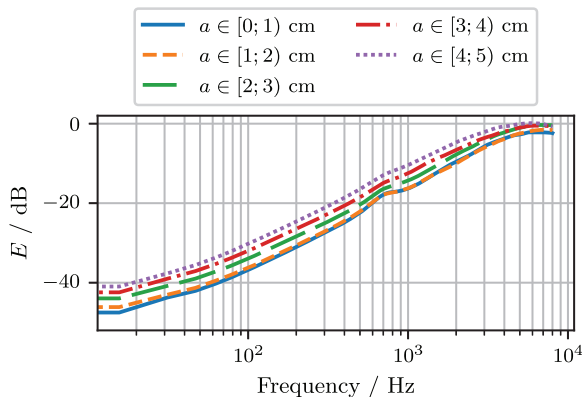
The proposed approach for error signal estimation is tested as part of a conventional ANC system as described in Section 2. As primarily the performance of the observation filter estimation is of interest, only the test set with static scenes is used here. This way, the secondary plant responses are time-invariant and can be modelled accurately without artefacts of secondary path interpolation. However, each scene uses randomized primary sources and virtual microphone positions as described in Section 4.1. A comparison to the conventional RMT or other commonly utilized

**Table 3.** Mean test NMSE for the proposed architecture (Obs-TasNet) and a modified version without the temporal bottleneck layer (Obs-TasNet (w/o  $L$ -BN)). The number of model parameters is shown in millions, the computational complexity as  $10^9$  multiply-accumulate operations (GMACs).

Model	Params	Complexity	Mean NMSE	
			Static	Moving
Obs-TasNet	<b>1.39 M</b>	<b>1.03 GMACs</b>	<b>-16.58 dB</b>	<b>-15.37 dB</b>
Obs-TasNet (w/o $L$ -BN)	2.18 M	4.02 GMACs	-16.04 dB	-14.95 dB

**Table 4.** Mean test NMSE for different distance  $a$  of the virtual microphone from the centre of the arrangement.

Position	Mean NMSE
$a \in [0; 1)$ cm	-20.54 dB
$a \in [1; 2)$ cm	-16.53 dB
$a \in [2; 3)$ cm	-14.58 dB
$a \in [3; 4)$ cm	-16.91 dB
$a \in [4; 5)$ cm	-15.21 dB

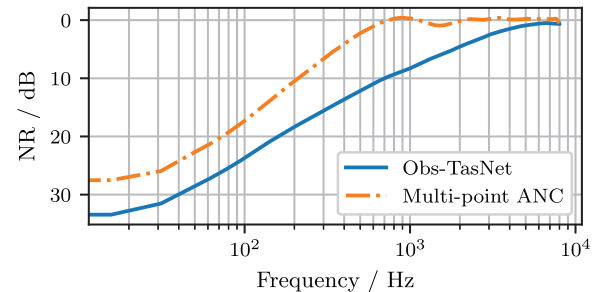


**Figure 5.** Estimation error  $E$  for different regions  $a$  of the virtual microphone from the centre of the arrangement.

techniques such as the additional filter method [62] is not viable. In these techniques, filters are optimized separately for each scenario with different virtual microphone position and primary disturbance direction and must be selected during operation, requiring additional metadata about the primary disturbances [21]. Instead, we used a multi-point ANC system [63] as benchmark. This approach does not process the error signal at the PoC for adaptation, but the signals of several nearby physical microphones – in our case the signals of the four remote microphones that are also processed by Obs-TasNet.

In each of the 1000 test-scenarios, a control filter with  $H = 256$  taps is adapted with a step-size of  $\mu = 0.1$ . The innovation signal driving the primary sources is used as reference signal  $x[n]$  as well. The secondary source is placed at Cartesian coordinates (1, 0, 0) m and modelled with  $I = 64$  taps.

Figure 6 shows the average steady-state noise reduction of an ANC system with error signal estimation



**Figure 6.** Averaged steady-state noise reduction (NR) after 15s adaptation over 1000 realizations of an ANC system with the proposed Obs-TasNet for error signal estimation and a multi-point ANC system using the four (physical) remote microphones.

using the proposed Obs-TasNet compared to a multi-point ANC system after 15s adaptation. Only realizations where convergence with the actual error signal  $e[n]$  is possible are taken into account. The proposed algorithm outperforms the multi-point ANC over the whole assessed frequency range. Major differences in performance are visible especially towards higher frequencies. While the multi-point ANC system reaches 10 dB noise reduction only up to 200 Hz, the system with Obs-TasNet expands this frequency range up to 700 Hz.

## 6 Conclusion

We presented Obs-TasNet, a neural network for online estimation of observation filter coefficients in the RMT. Built on the IC Conv-TasNet, the model processes raw waveforms from remote microphones together with the virtual microphone’s coordinates, while not requiring additional metadata about the acoustic scene. A learnable encoder maps the inputs into a shared latent space; cross-channel processing in bottleneck layers is followed by a TCN with subsequent output layers to predict the observation filter coefficients. Because the network outputs only filter coefficients – without end-to-end audio processing – the error signal computation in the real-time processing branch is unchanged compared to the conventional RMT, incurring no additional runtime load. The asynchronous design further allows model inference to be offloaded to a co-processor or neural processing unit.

We conducted a brief hyperparameter search to find a performant yet efficient configuration. An ablation study

demonstrated the effectiveness of the added temporal bottleneck layer, reducing number of parameters by 40% and the computational complexity by a factor of 4, while simultaneously lowering the NMSE. In simulation, an ANC system using Obs-TasNet for error-signal estimation substantially improved noise reduction: whereas on average a multi-point ANC baseline achieved about 10 dB only up to 200 Hz, the RMT-based system with Obs-TasNet extended effective attenuation up to 700 Hz in the tested scenarios.

This work demonstrates Obs-TasNet as a proof of concept for online estimation of observation filter coefficients. Future developments may incorporate additional scene metadata, extensions for multiple points of cancellation with greater degrees of freedom of motion, and a validation of robustness in complex, real-world acoustic environments.

### Acknowledgments

We thank the anonymous reviewers for their insightful comments and constructive critiques, which helped improve this paper. We are also grateful to Paul Armin Bereuter and Christian Blöcher for their valuable feedback during manuscript preparation.

### Funding

Parts of the computing infrastructure 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.

### Conflicts of interest

The authors declare no conflict of interest.

### Data availability statement

Code and model weights associated with this article are available in Zenodo, under the reference [42].

### Author contribution statement

**Felix Holzmüller:** Conceptualization, methodology, software, data curation, visualization, investigation, writing; **Alois Sontacchi:** Conceptualization, supervision, resources, writing – review & editing.

### References

1. S.J. Elliott: Signal Processing for Active Control, 1st edn. Signal Processing and Its Applications. Academic Press, 2001.
2. H.F. Olson, E.G. May: Electronic sound absorber. *Journal of the Acoustical Society of America* 25, 6 (1953) 1130–1136.
3. C. Boucher, S.J. Elliott, P.A. Nelson: Effect of errors in the plant model on the performance of algorithms for adaptive feedforward control. *IEE Proceedings F (Radar and Signal Processing)* 138, 4 (1991) 313.
4. T.J. Sutton, S.J. Elliott, A.M. McDonald, T.J. Saunders: Active control of road noise inside vehicles. *Noise Control Engineering Journal* 42, 4 (1994) 137–147.
5. W. Jung, S.J. Elliott, J. Cheer: Local active control of road noise inside a vehicle. *Mechanical Systems and Signal Processing* 121 (2019) 144–157.
6. J. Buck, D. Sachau: Active headrests with selective delayless subband adaptive filters in an aircraft cabin. *Mechanical Systems and Signal Processing* 148 (2021) 107164.
7. J.Y. Oh, H.W. Jung, M.H. Lee, K.H. Lee, Y.J. Kang: Enhancing active noise control of road noise using deep neural network to update secondary path estimate in real time. *Mechanical Systems and Signal Processing* 206 (2024) 110940.
8. S.J. Elliott, P. Joseph, A. Bullmore, P.A. Nelson: Active cancellation at a point in a pure tone diffuse sound field. *Journal of Sound and Vibration* 120, 1 (1988) 183–189.
9. S.J. Elliott, J. Garcia-Bonito: Active cancellation of pressure and pressure gradient in a diffuse sound field. *Journal of Sound and Vibration* 186, 4 (1995) 696–704.
10. P. Joseph, S.J. Elliott, P.A. Nelson: Near field zones of quiet. *Journal of Sound and Vibration* 172, 5 (1994) 605–627.
11. J. Garcia-Bonito, S.J. Elliott, C.C. Boucher: Generation of zones of quiet using a virtual microphone arrangement. *Journal of the Acoustical Society of America* 101, 6 (1997) 3498–3516.
12. S.J. Elliott, J. Cheer: Modeling local active sound control with remote sensors in spatially random pressure fields. *Journal of the Acoustical Society of America* 137, 4 (2015) 1936–1946.
13. B. Rafaely, S.J. Elliott, J. Garcia-Bonito: Broadband performance of an active headrest. *Journal of the Acoustical Society of America* 106, 2 (1999) 787–793.
14. B. Rafaely: Zones of quiet in a broadband diffuse sound field. *Journal of the Acoustical Society of America* 110, 1 (2001) 296–302.
15. F. Jiang, H. Tsuji, H. Ohmori, A. Sano: Adaptation for active noise control. *IEEE Control Systems Magazine* 17, 6 (1997) 36–47.
16. A. Wang, W. Ren: Convergence analysis of the multi-variable filtered-X LMS algorithm with application to active noise control. *IEEE Transactions on Signal Processing* 47, 4 (1999) 1166–1169.
17. J. Cheer, S. Daley: An investigation of delayless subband adaptive filtering for multi-input multi-output active noise control applications. *IEEE Transactions on Audio, Speech and Language Processing* 25, 2 (2017) 359–373.
18. L. Yin, Z. Zhang, M. Wu, Z. Wang, C. Ma, S. Zhou, J. Yang: Adaptive parallel filter method for active cancellation of road noise inside vehicles. *Mechanical Systems and Signal Processing* 193 (2023) 110274.
19. D. Moreau, B. Cazzolato, A. Zander, C. Petersen: A review of virtual sensing algorithms for active noise control. *Algorithms* 1, 2 (2008) 69–99.
20. S.J. Elliott, W. Jung, J. Cheer: Causality and robustness in the remote sensing of acoustic pressure, with application to local active sound control, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Brighton, United Kingdom, 2019, pp. 8484–8488.
21. W. Jung, S.J. Elliott, J. Cheer: Combining the remote microphone technique with head-tracking for local active sound control. *Journal of the Acoustical Society of America* 142, 1 (2017) 298–307.
22. C.D. Petersen, B.S. Cazzolato, A.C. Zander, C.H. Hansen: Active noise control at a moving location using virtual sensing, in: *Proceedings of the 13th International Congress of Sound and Vibration (ICSV13)*. ICSV. Vol. 13, Vienna, Austria, 2006.

23. F. Veronesi, C.K. Lai, J. Cheer: Interpolation between plant responses in a head-tracked local active noise control headrest system. *Mechanical Systems and Signal Processing* 240 (2025) 113401.
24. S. Koyama, J. Brunnström, H. Ito, N. Ueno, H. Saruwatari: Spatial active noise control based on kernel interpolation of sound field. *IEEE Transactions on Audio, Speech and Language Processing* 29 (2021) 3052–3063.
25. C.K. Lai, J. Cheer: A comparison between spatial interpolation approaches in a head-tracked active headrest system, in: *Proceedings of the 11th Convention of the European Acoustics Association Forum Acusticum/EuroNoise 2025*. European Acoustics Association, Málaga, Spain, 2025, pp. 23–30.
26. Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, M.D. Plumbley: PANNs: large-scale pretrained audio neural networks for audio pattern recognition. *IEEE Transactions on Audio, Speech and Language Processing* 28 (2020) 2880–2894.
27. DCASE: DCASE2025 Challenge – DCASE, <https://dcase.community/challenge2025/index>.
28. R. Xie, A. Tu, C. Shi, S. Elliott, H. Li, L. Zhang: Cognitive virtual sensing technique for feedforward active noise control, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, Korea, 2024, pp. 981–985.
29. B. Wang, D. Shi, Z. Luo, X. Shen, J. Ji, W.-S. Gan: Transferable selective virtual sensing active noise control technique based on metric learning, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hyderabad, India, 2025, pp. 1–5.
30. E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, P. Gerstoft: Generative models for sound field reconstruction. *Journal of the Acoustical Society of America* 153, 2 (2023) 1179–1190.
31. V.S. Paul, N. Hahn, P.A. Nelson: Learning from data-driven sound field estimation using complex-valued neural networks, in: *Proceedings of the 11th Convention of the European Acoustics Association Forum Acusticum/EuroNoise 2025*. European Acoustics Association, Málaga, Spain, 2025, pp. 4343–4350.
32. S. Koyama, J.G.C. Ribeiro, T. Nakamura, N. Ueno, M. Pezzoli: Physics-informed machine learning for sound field estimation: fundamentals, state of the art, and challenges. *IEEE Signal Processing Magazine* 41, 6 (2024) 60–71.
33. Y.A. Zhang, F. Ma, T.D. Abhayapala, P.N. Samarasinghe, A. Bastine: An active noise control system based on soundfield interpolation using a physics-informed neural network, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, Korea, 2024, pp. 506–510.
34. S.-C. Huang, C.-H. Ma, Y.-C. Hsu, M.R. Bai: Feedforward active noise global control using a linearly constrained beamforming approach. *Journal of Sound and Vibration* 537 (2022) 117190.
35. X. Xiao, S. Watanabe, H. Erdogan, L. Lu, J. Hershey, M.L. Seltzer, G. Chen, Y. Zhang, M. Mandel, D. Yu: Deep beamforming networks for multi-channel speech recognition, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 5745–5749.
36. R. Gu, J. Wu, S.-X. Zhang, L. Chen, Y. Xu, M. Yu, D. Su, Y. Zou, D. Yu: End-to-end multi-channel speech separation, 2019, <http://arxiv.org/abs/1905.06286>.
37. D. Lee, S. Kim, J.-W. Choi: Inter-channel Conv-TasNet for multichannel speech enhancement, 2021, <http://arxiv.org/abs/2111.04312>.
38. J. Kim: Remote microphone sound-field virtual sensing method using neural network for active noise control system, Master’s thesis, Purdue University, West Lafayette, IN, USA, 2024.
39. A. Roure, A. Albarrazin: The remote microphone technique for active noise control, in: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Fort Lauderdale, FL, USA, 1999, pp. 1233–1244.
40. F. Holzmüller, A. Sontacchi: Pilot study on virtual sensing for active noise control, in: *Proceedings of DAS/DAGA 2025*. Vol. 51. DAGA e.V., Copenhagen, Denmark, 2025, pp. 744–747.
41. F. Holzmüller, A. Sontacchi: Deep observation filter for virtual sensing in local active noise control, in: *Proceedings of the 11th Convention of the European Acoustics Association Forum Acusticum/EuroNoise 2025*. European Acoustics Association, Málaga, Spain, 2025, pp. 105–112.
42. F. Holzmüller, A. Sontacchi: Obs-TasNet: code, checkpoints, and results, Zenodo, 2026, <https://zenodo.org/records/18872900>.
43. C. Antoñanzas, M. Ferrer, M. de Diego, A. Gonzalez: Remote microphone technique for active noise control over distributed networks. *IEEE Transactions on Audio, Speech and Language Processing* 31 (2023) 1522–1535.
44. S. Kim, M.E. Altinsoy: A complementary effect in active control of powertrain and road noise in the vehicle interior. *IEEE Access* 10 (2022) 27 121–27 135.
45. D. Treyer, S. Gaulocher, S. Germann, E. Curiger: Towards the implementation of the noise-cancelling office chair: algorithms and practical aspects, in: *Proceedings of the 23rd International Congress on Sound and Vibration (ICSV23)*, Athens, Greece, 2016.
46. Y. Luo, N. Mesgarani: Conv-TasNet: surpassing ideal time–Frequency magnitude masking for speech separation. *IEEE Transactions on Audio, Speech and Language Processing* 27, 8 (2019) 1256–1266.
47. C. Lea, R. Vidal, A. Reiter, G.D. Hager: Temporal convolutional networks: a unified approach to action segmentation, in: G. Hua, H. Jégou, Eds. *Proceedings of the European Conference on Computer Vision (ECCV)*. Vol. 14. Springer, Amsterdam, The Netherlands, 2016, pp. 47–54.
48. F. Chollet: Xception: deep learning with depthwise separable convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Honolulu, HI, USA, 2017, pp. 1800–1807.
49. D.S. Williamson, Y. Wang, D. Wang: Complex ratio masking for monaural speech separation. *IEEE Transactions on Audio, Speech and Language Processing* 24, 3 (2016) 483–492.
50. T.N. Sainath, R.J. Weiss, K.W. Wilson, A. Narayanan, M. Bacchiani, A. Senior: Speaker location and microphone spacing invariant acoustic modeling from raw multichannel waveforms, in: *IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, Scottsdale, AZ, USA, 2015, pp. 30–36.
51. D. Haider, F. Perfler, V. Lostanlen, M. Ehler, P. Balazs: Hold me tight: stable encoder-decoder design for speech

- enhancement, in: Interspeech 2024. ISCA, Kos Island, Greece, 2024, pp. 5013–5017.
52. Y. Luo, N. Mesgarani: TaSNet: time-domain audio separation network for real-time, single-channel speech separation, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AL, Canada, 2018, pp. 696–700.
  53. J.L. Ba, J.R. Kiros, G.E. Hinton: Layer normalization, 2016, <http://arxiv.org/abs/1607.06450>.
  54. Y. Wu, K. He: Group normalization, in: Proceedings of the European Conference on Computer Vision (ECCV). Vol. 15. Springer, Munich, Germany, 2018.
  55. K. He, X. Zhang, S. Ren, J. Sun: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification, in: IEEE International Conference on Computer Vision (ICCV). IEEE, Santiago, Chile, 2015, pp. 1026–1034.
  56. G. Grimm, J. Luberadzka, V. Hohmann: A toolbox for rendering virtual acoustic environments in the context of audiology. Acta Acustica United with Acustica 105, 3 (2019) 566–578.
  57. P. Guiraud, S. Hafezi, P.A. Naylor, A.H. Moore, J. Donley, V. Tourbabin, T. Lunner: An introduction to the speech enhancement for augmented reality (spear) challenge, in: IEEE International Workshop on Acoustic Signal Enhancement (IWAENC). Vol. 17. IEEE, Bamberg, Germany, 2022.
  58. S. Braun, I. Tashev: Data augmentation and loss normalization for deep noise suppression, in: A. Karpov, R. Potapova, Eds. Speech and Computer. Springer International Publishing, St. Petersburg, Russia, 2020, pp. 79–86.
  59. A.V. Oppenheim, R.W. Schaffer: Discrete-Time Signal Processing, 3rd edn. Pearson, Upper Saddle River, NJ, USA, 2010.
  60. D.P. Kingma, J. Ba: Adam: a method for stochastic optimization, in: International Conference on Learning Representation (ICLR). Vol. 3, San Diego, CA, USA, 2015.
  61. P.D. Welch: The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics 15, 2 (1967) 70–73.
  62. J. Zhang, S.J. Elliott, J. Cheer: Robust performance of virtual sensing methods for active noise control. Mechanical Systems and Signal Processing 152 (2021) 107453.
  63. S. Elliott, I. Stothers, P. Nelson: A multiple error LMS algorithm and its application to the active control of sound and vibration. IEEE Transactions on Acoustics, Speech, and Signal Processing 35, 10 (1987) 1423–1434.

**Cite this article as:** Holzmüller F. & Sontacchi A. 2026. Obs-TasNet: Online estimation of virtual sensing observation filters for active noise control. Acta Acustica, 10, 31. <https://doi.org/10.1051/aacus/2026027>.