KU LEUVEN

STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Hannes Rosseel and Toon van Waterschoot Title: On Time Delay Interpolation for Improved Acoustic Reflector Localization Journal: Journal of the Acoustical Society of America Pages: 1–20
Archived version	■ Submitted version ☐ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ☐ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ■ EU H2020 ☐ EU Horizon Europe ■ FWO ■ KULeuven Internal Funds ☐ VLAIO ■ AI ☐ Other
Published version	
Abstract	The localization of acoustic reflectors is a fundamental component in various applications, including room acoustics analysis, sound source localization, and acoustic scene analysis. Time Delay Estimation (TDE) is essential for determining the position of reflectors relative to a sensor array. Traditional TDE algorithms generally yield time delays that are integer multiples of the operating sampling period, potentially lacking sufficient time resolution. To achieve subsample TDE accuracy, various interpolation methods, including parabolic, Gaussian, frequency, and sinc interpolation, have been proposed. This paper presents a comprehensive study on time delay interpolation to achieve subsample accuracy for acoustic reflector localization in reverberant conditions. We derive the Whittaker-Shannon interpolation formula from the previously proposed sinc interpolation in the context of short-time windowed TDE for acoustic reflector localization. Simulations show that sinc and Whittaker-Shannon interpolation outperform existing methods in terms of time delay error and positional error for critically sampled and band-limited reflections. Performance is evaluated on real-world measurements from the MYRiAD dataset, showing that sinc and Whittaker-Shannon interpolation consistently provide reliable performance across different sensor-source pairs and loudspeaker positions. These results can enhance the precision of acoustic reflector localization systems, vital for applications such as room acoustics analysis, sound source localization, and acoustic scene analysis.
Author contact	hannes.rosseel@esat.kuleuven.be

On Time Delay Interpolation for Improved Acoustic Reflector Localization

Hannes Rosseel and Toon van Waterschoot
KU Leuven, Dept. of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium[a]

The localization of acoustic reflectors is a fundamental component in various applications, including room acoustics analysis, sound source localization, and acoustic scene analysis. Time Delay Estimation (TDE) is essential for determining the position of reflectors relative to a sensor array. Traditional TDE algorithms generally yield time delays that are integer multiples of the operating sampling period, potentially lacking sufficient time resolution. To achieve subsample TDE accuracy, various interpolation methods, including parabolic, Gaussian, frequency, and sinc interpolation, have been proposed. This paper presents a comprehensive study on time delay interpolation to achieve subsample accuracy for acoustic reflector localization in reverberant conditions. We derive the Whittaker-Shannon interpolation formula from the previously proposed sinc interpolation in the context of short-time windowed TDE for acoustic reflector localization. Simulations show that sinc and Whittaker-Shannon interpolation outperform existing methods in terms of time delay error and positional error for critically sampled and band-limited reflections. Performance is evaluated on real-world measurements from the *MYRiAD* dataset, showing that sinc and Whittaker-Shannon interpolation consistently provide reliable performance across different sensor-source pairs and loudspeaker positions. These results can enhance the precision of acoustic reflector localization systems, vital for applications such as room acoustics analysis, sound source localization, and acoustic scene analysis.

Pages: 1–20

I. INTRODUCTION

Accurately estimating the position of acoustic reflectors is of great importance for many acoustical signal processing applications. These applications include inferring room geometry (Annibale *et al.*, 2012; Antonacci *et al.*, 2012, 2010; El Baba *et al.*, 2017; Tervo and Tossavainen, 2012), analyzing and visualizing room acoustics (Pätynen *et al.*, 2013), and spatial audio encoding for the auralization (Tervo *et al.*, 2013) of concert halls (Lokki *et al.*, 2016), listening environments (Tervo *et al.*, 2014), and car cabins (Camilleri *et al.*, 2019; Kaplanis *et al.*, 2017).

Acoustic reflector localization relies on the time delay information between the arrival of an acoustic wavefront at different sensor positions. The position of an acoustic reflector in 3D space can be estimated by measuring the Time of Arrival (TOA) and Time Difference of Arrival (TDOA) of an acoustic wavefront at a set of spatially separated sensors. The TOA can be estimated using peak-picking methods (Antonacci *et al.*, 2012; Carlo *et al.*, 2021; DeFrance *et al.*, 2008, 2009; Remaggi *et al.*, 2017; Tervo *et al.*, 2012) or maximum-likelihood estimation (Ehrenberg *et al.*, 1978). The TDOA information, which is the time difference between the arrival of a wavefront at two spatially separated sensors, is commonly derived using cross-correlation-based methods, such as the Generalized Cross-Correlation (GCC) method (Knapp and Carter, 1976). The TDOA corresponds to the time instant at which the cross-correlation of two sensor pressure values is maximized, and can be used to determine the Direction of Arrival (DOA) of the acoustic wavefront propagating along the sensors. The combination of TDOA and TOA information can be used to reliably localize acoustic reflectors in a room (Tervo *et al.*, 2012, 2013).

The localization of acoustic reflectors from a set of Room Impulse Responses (RIRs), measured by a set of N spatially separated sensors, can be achieved by estimating the equivalent image source position of each acoustic reflection impinging on the sensor array (Allen and Berkley, 1979). The image sources are localized by dividing the set of measured RIRs into overlapping short-time windows, and applying a localization method to each short-time window (Tervo *et al.*, 2013). The window length should be chosen slightly larger than the time it takes for an acoustic reflection to propagate along the sensor array (Merimaa and Pulkki, 2005; Pulkki and Merimaa, 2006; Tervo *et al.*, 2012, 2013). It is assumed that only one acoustic reflection k is active in each short-time window. This assumption is generally valid for the direct and early acoustic reflections in the measured RIR

[a]Corresponding author: hannes.rosseel@esat.kuleuven.be

if the dimensions of the sensor array are relatively small compared to the dimensions of the room and the sensor array is not positioned near walls or reflective surfaces.

The precision of acoustic source localization is directly impacted by the accuracy of TDOA and TOA estimation. The temporal resolution of TDOA and TOA estimates for individual acoustic reflections is limited by the sampling rates at which measured RIRs are sampled. Estimation errors arise when the actual time delay between two acoustic wavefronts arriving at spatially separated sensors is not an integer multiple of the sampling period. Inaccurate Time Delay Estimation (TDE) may lead to significant localization errors in the estimated position of the acoustic reflector. These errors can occur when the sampling rate is constrained by available hardware, resulting in insufficient temporal resolution to accurately estimate the position of each acoustic reflector. In general, the TDOA and TOA estimate corresponds to the time instant at which a specific TDE function is maximized. For TDOA estimation, this function is the cross-correlation between two acoustic reflections arriving at spatially separated sensors. For TOA estimation, this function is the filter-matched RIRs measured by each sensor. Since the TDE function is discretized in time, the TDE is limited to the nearest sample value, which may not provide sufficient temporal resolution for accurate localization of acoustic reflectors.

Interpolation methods can improve the accuracy of TDE by estimating time delays with subsample precision. Previous studies have proposed various interpolation methods that can be categorized into two groups, namely: time-domain interpolation methods and frequency-domain interpolation methods. Time-domain interpolation methods aim to improve the resolution of the TDE function by fitting a continuous-time function around its peak. Existing interpolation functions are: parabolic functions (Lai and Torp, 1999), Gaussian functions (Zhang and Wu, 2005), or sinc functions (Rosseel and van Waterschoot, 2021). Another time-domain approach involves reconstructing the continuous-time band-limited TDE function from the discretized function by applying Whittaker-Shannon interpolation (Shannon, 1949). Frequency-domain interpolation methods achieve subsample TDE accuracy by fitting a linear function to the weighted phase spectrum of the TDE function (Svilainis, 2019; Svilainis *et al.*, 2013). In Sec. III, we provide a detailed discussion of these interpolation methods and their application to TDE for acoustic reflector localization.

In (Rosseel and van Waterschoot, 2021), the authors demonstrated that improving TDE through sinc interpolation also improves the localization performance of a single acoustic reflector in anechoic conditions. It was shown that sinc interpolation is able to outperform parabolic and Gaussian interpolation in anechoic conditions. However, its performance has not yet been compared to frequency interpolation. Moreover, the authors have not yet verified the effectiveness of applying sinc interpolation for localizing acoustic reflectors in reverberant conditions when a short time window is introduced to isolate each acoustic reflection. This paper seeks to address these limitations by investigating the impact of short-time windowing on the performance of sinc interpolation for TDOA estimation, TOA estimation, and localization of acoustic reflectors. Additionally, this study compares the performance of sinc interpolation to frequency interpolation and proposes a reformulation of sinc interpolation to improve TDE of acoustic reflections in reverberant conditions.

This paper is structured as follows: Sec. II describes the assumed signal model and outlines the methods used for acoustic reflector localization. Sec. III provides a detailed discussion of various interpolation techniques for improving TDE. The effect of short-time windowing on the accuracy of TDE is analyzed in Sec. IV. Subsequently, Sec. V presents a novel reformulation of sinc interpolation for TDE of acoustic reflections in reverberant conditions. A comparison of the performance of different interpolation methods in various simulated scenarios is presented in Sec. VI. In Sec. VII, the performance of the interpolation methods is evaluated on real-world measurements from the *MYRiAD* dataset. Finally, Sec. VIII summarizes the contributions outlined in this paper.

II. PRELIMINARIES AND METHODS

A. Room Impulse Response

The RIR can be modeled as the sum of all individual acoustic events (Tervo *et al.*, 2013) originating from a sound wave propagating from a source position $\mathbf{x}$ to a receiver sensor position $\mathbf{r}_n$, such that

$$h_n(t) \triangleq \sum_{k=1}^{K} h_n^k(t) + w_n(t), \tag{1}$$

where t represents time, n is the sensor index, $k = 0, \ldots, K$ indicates the index of each acoustic event, and $w_n(t)$ is the measurement and model error.

A RIR is typically measured by exciting an enclosed space with a known broadband excitation signal and subsequently recording the resulting sound pressure at a receiver position in the space. The RIR can be extracted from the measured recording and the excitation signal through a deconvolution process. A common excitation signal for measuring RIRs is the sine sweep (Farina, 2000; Müller and Massarani, 2001; Novak *et al.*, 2015). The measured RIR is influenced by the characteristics of the source and the sensor that are used in the measurement process. The RIR measured by a sensor can be modeled as

$$\tilde{h}_n(t) = h_n(t) * m_n(t), \tag{2}$$

where n is the sensor index, $\tilde{h}_n(t)$ denotes the measured RIR, and $m_n(t)$ is the combined impulse response, which describes the characteristics of the source and sensor that were used. It is possible to compensate for the characteristics of the measurement equipment by applying a

matched filter to the measured RIR (Antonacci *et al.*, 2012; Ehrenberg *et al.*, 1978). This matched filter is dependent on the direction of the source relative to the sensor position when the source has a directional radiation patterns. However, in this paper, we will assume that the source behaves as an omnidirectional point source, and the matched filter is independent of the direction of the source. Moreover, it is assumed that the matched filter is known a priori. In practice however, the matched filter can be found by measuring the free-field impulse response between the excitation source and each sensor. Alternatively, if this free-field impulse response is not available, $m_n(t)$ can be estimated from the measured RIR by isolating the direct-path component using a windowing function (Antonacci *et al.*, 2012; Carlo *et al.*, 2021). The RIR $h_n(t)$ can be found by cross-correlating the matched filter $m_n(t)$ with the measured RIR $\tilde{h}_n(t)$ as

$$h_n(t) = \int_{-\infty}^{\infty} \tilde{h}_n(\tau) m_n(\tau + t) d\tau, \tag{3}$$

The definition for a RIR described in (1) is commonly used in the literature on acoustic source localization (Remaggi *et al.*, 2017; Tervo *et al.*, 2012, 2013; Tervo and Tossavainen, 2012). In this paper, we specifically consider rooms with rigid walls, allowing the assumption of ideal specular reflections. Consequently, the phase response from the source to each specular reflection picked up by a sensor is considered frequency-independent. Additionally, the sound source is assumed to behave as an omnidirectional point source, exciting the room with an ideal spherical wave pattern. These assumptions allow the use of the image-source model (Allen and Berkley, 1979) to represent the RIR as the sum of individual acoustic reflections, modeled as image sources at positions $\mathbf{x}_k$, such that

$$h_n^k(t) = \frac{\delta(t - t_n^k)}{4\pi \left\|\mathbf{r}_n - \mathbf{x}_k\right\|_2}, \tag{4}$$

where $\delta(\cdot)$ denotes the Dirac delta function, and t_n^k denotes the TOA or propagation time of a spherical sound wave originating at image-source position $\mathbf{x}_k$ and arriving at sensor position $\mathbf{r}_n$. The TOA t_n^k is defined as

$$t_n^k \triangleq \left\|\mathbf{r}_n - \mathbf{x}_k\right\|_2 c^{-1}, \tag{5}$$

where c denotes the sound propagation speed in air, estimated to be around $343\,\mathrm{m\,s^{-1}}$ at a temperature of $20\,^{\circ}\mathrm{C}$.

Up to this point, we have assumed that the RIR measured at the receiver is not band-limited. In practice however, the measurement equipment that is used exhibits finite bandwidth due to their inherent frequency response characteristics. Additionally, the sampling operation at the receiver position discretizes the RIR at a fixed sampling interval $T = \frac{1}{f_s}$, where f_s is the sampling rate. This sampling operation introduces a bandlimit on the measured signal at the receiver position in accordance with the Nyquist-Shannon sampling theorem (Shannon, 1949). As a result, when a bandlimit equal to the Nyquist frequency $\frac{f_s}{2}$ is imposed, the individual acoustic reflections described in (4) can be modeled as

$$\hat{h}_n^k(t) = \frac{\operatorname{sinc}\left(f_s\left(t - t_n^k\right)\right)}{4\pi \left\|\mathbf{r}_n - \mathbf{x}_k\right\|_2}, \tag{6}$$

where $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$ is the normalized sinc function, and $t = \kappa T, \kappa \in \mathbb{Z}$ when the sampling operation is applied.

B. Localization of individual image-sources

The choice of localization method depends on the assumption of the sound propagation model and the measurement conditions. Popular methods employ both TOA and TDOA information (Tervo *et al.*, 2012) to estimate the position of an acoustic reflection. For now, we will assume that all acoustic reflections can be perfectly separated in time. In Sec. IV, the separation of individual acoustic reflections from a measured RIR is discussed in more detail.

The TOA denotes the precise time at which an acoustic reflection arrives at a sensor position $\mathbf{r}_n$, and is defined in (5). The TDOA is the relative time difference between the arrival of an acoustic reflection originating from an image-source position $\mathbf{x}_k$ at two sensor positions $\{\mathbf{r}_n, \mathbf{r}_m\}$. It can be noted that the TDOA can be expressed in terms of the TOA information as

$$\tau_{m,n}^k = t_n^k - t_m^k = \frac{||\mathbf{r}_n - \mathbf{x}_k||_2 - ||\mathbf{r}_m - \mathbf{x}_k||_2}{c}. \tag{7}$$

The TDOA information between two RIRs is estimated using cross-correlation-based methods, such as the GCC (Knapp and Carter, 1976). For an acoustic reflection k, measured at two distinct sensor positions $\mathbf{r}_m$ and $\mathbf{r}_n$, the GCC $r_{m,n}^k(\tau)$ is defined as:

$$r_{m,n}^k(\tau) = \int_{-\infty}^{\infty} \Psi(\omega) G_{m,n}^k(\omega) e^{j\omega\tau} d\omega, \tag{8}$$

where $G_{m,n}^k(\omega)$ denotes the Cross-Spectral Density (CSD) between the k^{th} acoustic reflection observed by the m and n^{th} sensor, respectively. The general frequency weighting is indicated by $\Psi(\omega)$ (Knapp and Carter, 1976). Since acoustic reflections are assumed to exhibit broadband frequency spectra, the general frequency weighting can be set to $\Psi(\omega) \equiv 1$, reducing the GCC to regular cross-correlation. In Rosseel and van Waterschoot (2021), it was shown that the cross-correlation between two sinc functions, as defined for acoustic reflections in (6), is a scaled and time-delayed sinc function of the form

$$r_{m,n}^k(\tau) = \frac{\operatorname{sinc}\left(f_s\left(t_m^k - t_n^k + \tau\right)\right)}{16\pi^2 f_s\, c^2\, t_n^k\, t_m^k}. \tag{9}$$

The TDOA between two acoustic reflections measured at different points in space, can then be found as the time lag for which $r_{m,n}^k(\tau)$ is maximized, i.e.,

$$\hat{\tau}^k_{m,n} = \arg\max_{\tau} r^k_{m,n}(\tau) = t^k_n - t^k_m. \tag{10}$$

Moreover, when image-sources are assumed to be located in the far field relative to the sensor array position, the plane wave propagation model can be assumed for the localization of the image-sources. This assumption generally leads to a more efficient method for image-source localization, and makes it possible to rewrite the TDOA information using the Slowness Vector (SV) $\mathbf{s}^k$ relating to an image-source with index k (Pirinen, 2009), such that

$$\tau^k_{m,n} = \mathbf{v}^T_{m,n}\mathbf{s}^k, \tag{11}$$

where $\mathbf{v}_{m,n}$ is the sensor position difference vector, defined as

$$\mathbf{v}_{m,n} = \mathbf{r}_n - \mathbf{r}_m. \tag{12}$$

The SV fully describes the propagation characteristics of a plane wave, and is defined as the vector that points in the direction of the acoustic reflection k with a magnitude equal to the inverse of the speed of sound c in the medium. By estimating the TDOA information using (10) for all sensor pairs in the array, the estimated SV $\hat{\mathbf{s}}^k$ can be found from the least squares solution (Pirinen, 2009) given by

$$\hat{\mathbf{s}}^k = \mathbf{V}^+\hat{\boldsymbol{\tau}}^k \tag{13}$$

where $(.)^+$ denotes the Moore-Penrose pseudo-inverse, $\mathbf{V}$ is the matrix containing the sensor position difference vectors $\mathbf{v}_{m,n}$ for all pairs across the N sensors in the array,

$$\mathbf{V} = \left[\mathbf{v}_{1,2}, \mathbf{v}_{1,3}, \ldots, \mathbf{v}_{N-1,N}\right]^T, \tag{14}$$

and $\hat{\boldsymbol{\tau}}^k$ is the corresponding vector of TDOA estimates,

$$\hat{\boldsymbol{\tau}}^k = \left[\hat{\tau}^k_{1,2}, \hat{\tau}^k_{1,3}, \ldots, \hat{\tau}^k_{N-1,N}\right]^T. \tag{15}$$

From the estimated SV, the position of each image-source k can be found relative to the position of the sensor array's center $\mathbf{r}_c$, as

$$\hat{\mathbf{x}}_k = \mathbf{r}_c - \frac{\hat{\mathbf{s}}^k\|\mathbf{r}_c - \mathbf{x}_k\|_2}{\|\hat{\mathbf{s}}^k\|_2}, \tag{16}$$

where $\mathbf{r}_c = \frac{1}{N}\sum_{n=1}^{N}\mathbf{r}_n$ (Pirinen, 2009). The distance $\|\mathbf{r}_c - \mathbf{x}_k\|_2$ between the sensor array center and the k^{th} image-source can be found from the TOA information t^k_c at the center of the sensor array. If no central sensor is present in the array, the center TOA information can be estimated by averaging the TOA information from all sensors in the array, i.e.,

$$t^k_c = \frac{1}{N}\sum_{n=1}^{N} t^k_n. \tag{17}$$

C. Localization limitations

Various factors can impact the performance of image-source localization methods (Amengual Garí *et al.*, 2021; Tervo *et al.*, 2012, 2013), including the presence of measurement noise, the type of excitation signal used for the RIR measurement, the type of sensor-array, the sampling rate, and the size of the analysis window. In this section, we will focus on two of these factors, namely the analysis window size, and the sampling rate.

1. Analysis window size

As briefly mentioned in the introduction, for acoustic reflector localization in a reverberant environment, the analysis window size should be chosen greater than the time it takes for an acoustic reflection wavefront to propagate through the sensor array, i.e.,

$$L > \frac{d_{\max}}{c} f_s, \tag{18}$$

where L is the analysis window length in samples and $d_{\max}$ is the maximum pairwise distance across all sensors in the array. The minimum window length L ensures that an acoustic reflection will be captured by all sensors in the array, enabling the estimation of TDOA information between all pairs of sensors. In the literature, the minimum window length is often defined as $L > 2\frac{d_{\max}}{c} f_s$, to ensure that a sufficient number of samples is available at every sensor to perform acoustic reflector localization (Puomio, 2021; Tervo *et al.*, 2013). Whereas (18) provides a valid lower bound, an increase in the number of samples within the analysis window will result in an improvement in localization performance. However, this also increases the likelihood of two acoustic reflections occurring within a single window, which will produce inaccurate time delay estimates leading to poorer localization performance. The choice of window size is therefore a trade-off between localization accuracy and the number of reflections that can be accurately localized in a single RIR measurement (Tervo *et al.*, 2013). A maximum window size can heuristically be derived based on the echo density of the room (Kuttruff, 2017; Tervo *et al.*, 2012).

2. Sampling rate

The accuracy of image-source localization, as described by (13) and (16), is dependent upon the accuracy of TOA and TDOA estimation. The TOA of acoustic reflections is typically determined by applying a peak-picking algorithm to the RIR (Antonacci *et al.*, 2012; Carlo *et al.*, 2021; DeFrance *et al.*, 2008, 2009; Remaggi *et al.*, 2017; Tervo *et al.*, 2012), while the TDOA estimates are obtained by identifying the lag corresponding to the maximum sample value of the cross-correlation function $r^k_{m,n}(\tau)$, as per (10). The accuracy of these estimates is therefore directly related to the sampling period $T = \frac{1}{f_s}$ that is used during the RIR measurements.

In both TOA and TDOA estimation, the true time delay typically does not coincide with an integer multi-

ple of the sampling period T. Instead, it lies within the time interval $[(\kappa_0-1)T, (\kappa_0+1)T]$, where κ_0 denotes the index of the sample value that corresponds to the closest integer multiple of the true time delay. Consequently, the estimated time delays are subject to a maximum discretization error of T seconds.

To decrease the discretization error that is made during TOA and TDOA estimation, the sampling rate f_s can be increased, resulting in a smaller sampling period T. However, hardware limitations during RIR measurements often restrict the achievable sampling rate. In such cases, interpolation schemes can be applied to enable TOA and TDOA estimation with subsample accuracy. The following section will review several existing interpolation schemes for improving TDE with subsample accuracy.

III. TIME-DELAY INTERPOLATION

The discretization error in TDE can be mitigated by applying time-interpolation methods to a discretized TDE function. For TOA estimation of acoustic reflections, this function corresponds to the filter-matched RIR, while for TDOA estimation, it corresponds to the cross-correlation function between windowed RIRs measured at two spatially-separated sensors. This section will discuss existing interpolation methods, including parabolic interpolation (Lai and Torp, 1999), Gaussian interpolation (Zhang and Wu, 2005), frequency interpolation (Svilainis, 2019; Svilainis *et al.*, 2013), and sinc interpolation (Rosseel and van Waterschoot, 2021), which was proposed by the authors. Finally, Whittaker-Shannon interpolation is discussed as a generalization of sinc interpolation for TDE in reverberant conditions.

A. Parabolic Interpolation

Parabolic interpolation is a three-point curve-fitting interpolation method used to improve TDE with subsample accuracy (Lai and Torp, 1999). This method involves finding a quadratic function which passes through the maximum sample value of the TDE function and its two neighboring points. The quadratic function has the form $y_p(\tau) = a\tau^2+b\tau+c$. The improved time delay estimate $\tilde{\tau}_p$ can be found at the vertex of the fitted parabolic curve. The symbol τ is defined as a generalized time symbol τ, which represents time for TOA estimation, and time lag for TDOA estimation. For a discretized function $f[\kappa]$, with a maximum sample value at the discrete index κ_0, the vertex position can be found at (Lai and Torp, 1999),

$$\tilde{\tau}_p = \frac{f[\kappa_0-1] - f[\kappa_0+1]}{2\left(f[\kappa_0-1] - f[\kappa_0] + f[\kappa_0+1]\right)}. \tag{19}$$

Parabolic fitting is a popular interpolation method for subsample TDE because of its simplicity and computational efficiency. However, a major drawback is the high estimation bias it produces (Céspedes *et al.*, 1995). Estimation bias refers to the systematic error in the estimated time delay. The magnitude and direction of this bias depend on the exact position of the true time delay relative to the sampling period $T_s = 1/f_s$, i.e., where the true time delay falls within T_s (Céspedes *et al.*, 1995).

B. Gaussian Interpolation

Alternatively, an improved time delay estimate can be found by fitting the TDE function with an exponential function of the form $y_g(\tau) = \alpha \cdot e^{-\beta(\tau-\tilde{\tau}_g)^2}$. The vertex position of this function depends only on parameter $\tilde{\tau}_g$, which can be found from the maximum sample value of the discretized function and its neighboring points (Zhang and Wu, 2005) as,

$$\tilde{\tau}_g = \frac{\ln(f[\kappa_0+1]) - \ln(f[\kappa_0-1])}{4\ln(f[\kappa_0]) - 2\ln(f[\kappa_0+1]) + \ln(f[\kappa_0-1])}, \tag{20}$$

where it is required that $f[\kappa] > 0$, for $\kappa \in \{\kappa_0-1, \kappa_0, \kappa_0+1\}$. If this condition is not met, as can happen with critically sampled functions, (Rosseel and van Waterschoot, 2021) proposes to estimate $\tilde{\tau}_g$ by first offsetting the values $f[\kappa_0-1], f[\kappa_0], f[\kappa_0+1]$, by a common scalar constant to ensure that they are strictly positive.

This interpolation approach is utilized in the Spatial Decomposition Method (SDM) (Meyer-Kahlen *et al.*, 2022; Tervo *et al.*, 2013) to enhance the localization accuracy of acoustic reflectors and facilitate spatial sound field analysis. Compared to parabolic interpolation, Gaussian interpolation generally produces a lower TDOA estimation error, but remains susceptible to estimation bias (Céspedes *et al.*, 1995).

C. Frequency Interpolation

While the previous interpolation methods are performed in the time-domain, and can be applied to both TOA and TDOA estimation, frequency interpolation is specifically designed for improving TDOA estimation in the frequency domain (Svilainis, 2019; Svilainis *et al.*, 2013). Frequency interpolation relies on the observation that the phase spectrum of the CSD function, which is the frequency-domain transformation of the cross-correlation function obtained through a Discrete-Time Fourier Transform (DTFT), is linear due to the complex conjugation operation of the CSD. The method works by first eliminating phase-wrapping by circularly shifting the cross-correlation function in the time-domain by a rough estimate κ_0 of the time delay (Svilainis *et al.*, 2013).

$$\hat{f}[\kappa] = \text{circshift}(f[\kappa], \kappa_0). \tag{21}$$

The shifted version of the TDE function $\hat{f}[k]$ can now be transformed to the frequency domain to obtain the CSD $\hat{F}[\omega]$ for TDOA estimation. The subsample TDOA estimate $\tilde{\tau}_f$ corresponds to the slope of a linear regression fit on the phase values. It was shown in (Svilainis, 2019), that minimizing the L2 norm of the difference between the CSD and a linear function of the form $\alpha\omega$ exhibited

the best performance. The improved TDOA estimate using frequency interpolation can be found as,

$$\tilde{\tau}_{\mathrm{f}} = \kappa_0 + \min_{\alpha} \left\| |\hat{F}[\omega]|^2 \left(\varphi(\hat{F}[\omega]) - \alpha\omega \right) \right\|_2, \tag{22}$$

where $\varphi(\cdot)$ denotes the phase operator and ω denotes frequency.

Frequency interpolation exhibits very low estimation bias (Svilainis, 2019). However, its accuracy is dependent upon sufficiently high Signal-to-Noise Ratio (SNR) conditions. This dependency can become a limitation when localizing acoustic reflections from distant image-sources relative to the sensor array, as the SNR of such reflections diminishes according to the inverse square law of sound propagation. Additionally, since frequency interpolation depends on the linearity of the phase in the frequency-domain TDE function, it is unsuitable for TOA estimation with subsample precision. However, this method can provide reasonable accuracy for TOA estimation when the phase of the TDE function remains approximately linear. In Sec. VI D, the performance of frequency interpolation is evaluated for TDOA and TOA estimation of acoustic reflections in reverberant conditions.

D. Sinc Interpolation

In (Rosseel and van Waterschoot, 2021), the authors proposed an alternative interpolation method for improving TDE of acoustic reflections. This method finds its basis in the Nyquist-Shannon sampling theory and takes advantage of the assumption that the TDE function for TOA and TDOA estimation of a single acoustic reflection can be modeled as a sinc function. For TOA estimation, this is evident from (6). For TDOA estimation, it was shown in (9) that the cross-correlation function between two sinc functions is a scaled and time-delayed sinc function with a time delay equal to the TDOA. From this observation, it can be inferred that sinc interpolation is the most suitable interpolation method for TDOA estimation of acoustic reflections.

Sinc interpolation consists of fitting a critically sampled and truncated sinc function around the maximum sample value of the normalized TDE function. The fractional part of the time delay can be found by minimizing the following objective function in the discrete-time domain,

$$\kappa_{\mathrm{s}} = \arg\min_{\kappa_{\mathrm{s}}} \sum_{\kappa=\kappa_0-S}^{\kappa_0+S} \left| \mathrm{sinc}(\pi f_{\mathrm{s}}(\kappa T - \kappa_{\mathrm{s}} T_{\mathrm{i}})) - \frac{f[\kappa]}{f[\kappa_0]} \right|^2, \tag{23}$$

where κ_{s} is the fractional part of the time delay, S is a tunable parameter determining the truncation point of the sinc function, and T_{i}, where $T_{\mathrm{i}} < T$, represents the interpolation sampling period which determines the resolution of the interpolation. The time delay with subsample accuracy is found as $\tilde{\tau}_{\mathrm{s}} = \kappa_0 T + \kappa_{\mathrm{s}} T_{\mathrm{i}}$.

It was shown in (Rosseel and van Waterschoot, 2021) that this method outperforms both parabolic and Gaussian interpolation for TDOA estimation of acoustic reflections in the free-field. However, the performance of this method has not yet been evaluated for TOA estimation and acoustic reflector localization in reverberant conditions, where the measured RIRs are typically windowed to isolate each acoustic event. Moreover, sinc interpolation has not been compared to frequency interpolation for TDOA and TOA estimation. In the following section, an adaptation of sinc interpolation is presented for acoustic reflector localization in reverberant conditions, where the acoustic events are isolated using an analysis window.

IV. APPLYING A WINDOWING FUNCTION

In the previous sections, it was assumed that each acoustic reflection $h_n^k(t)$ could be perfectly separated in time. In practice however, we only have access to the RIR $h_n(t)$, which is a superposition of all acoustic reflections arriving at the sensor array. Isolating individual acoustic reflections is not trivial and poses significant challenges.

To distinguish between individual acoustic reflections in the RIR, a sliding short-time windowing function can be applied to $h_n(t)$ (Tervo *et al.*, 2012). This analysis window, denoted by $\gamma(t, v)$, is only non-zero within the interval $t \in [v - \frac{L}{2} \ldots v + \frac{L}{2}]T$, where v is the window position and L denotes the window length in samples. The window length is chosen based on the physical dimensions of the sensor array, see section II C 1. By applying the analysis window at every discrete time instant vT, with $v \in \mathbb{Z}$, a set of short-time windowed RIRs $\tilde{h}_n(t, v)$ is obtained which partially overlap in time. It is assumed that there is at most a single acoustic reflection k present in each window (Tervo *et al.*, 2013). This assumption generally holds for the direct path and early reflections of the RIR $h_n(t)$, provided the sensor positioning avoids geometric symmetries with respect to the sound source and surrounding room boundaries. The short-time windowed RIRs are given by,

$$\tilde{h}_n(t, v) = h_n(t)\gamma(t, v) \tag{24}$$

where v will be referred to as the window number.

The TOA and TDOA of the acoustic reflections can be estimated by analyzing the short-time windowed RIRs. The TOA can be estimated from each discretized short-time windowed RIR $\tilde{h}_n[\kappa, v]$ by finding the sample-index of the maximum absolute amplitude. The TOA $\hat{t}_n(v)$ of the acoustic reflection in window v is estimated by,

$$\hat{t}_n(v) = \frac{v + \arg\max_{\kappa} |\tilde{h}_n[\kappa, v]|}{f_{\mathrm{s}}}. \tag{25}$$

Similarly, the TDOA $\hat{\tau}_{m,n}(v)$ of an acoustic reflection in window v is estimated as the time-index which maximizes the cross-correlation between two short-time windowed RIRs $\tilde{h}_n(t, v)$ and $\tilde{h}_m(t, v)$, where m denotes the reference sensor. The TDOA $\hat{\tau}_{m,n}(v)$ is estimated by,

$$\hat{\tau}_{m,n}(v) = \frac{\arg\max_{\kappa} r_{m,n}[\kappa, v]}{f_{\mathrm{s}}}, \tag{26}$$

where $r_{m,n}[\kappa, v]$ denotes the cross-correlation between the discretized short-time windowed RIRs $\tilde{h}_m[\kappa, v]$ and $\tilde{h}_n[\kappa, v]$.

V. REFORMULATING SINC INTERPOLATION

By applying an analysis window to the measured RIRs, the cross-correlation between two windowed acoustic reflections deviates from the scaled sinc function previously described in (9). This deviation motivates the need for a reformulation of sinc interpolation to accommodate TDE of windowed acoustic reflections with subsample accuracy. The performance implications of this windowing operation on the previously proposed sinc interpolation are examined in Sec. VI D. This section presents a reformulation of sinc interpolation for short-time windowed TOA and TDOA estimation, resulting in a method known as Whittaker-Shannon interpolation which has not been previously applied to TDE of acoustic reflections.

It can be shown that Whittaker-Shannon interpolation can be derived from the previously proposed sinc interpolation. The objective function in (23) can be adapted by using the windowed TDE function $f(\kappa T, v)$, which is the cross-correlation function of the windowed RIRs for TDOA estimation between two sensors, and the windowed RIR for TOA estimation. The resulting objective function is given by

$$\kappa_{\mathrm{ws}} = \arg\min_{\kappa_{\mathrm{ws}}} \sum_{\kappa=-\infty}^{\infty} \left|\mathrm{sinc}(\pi f_{\mathrm{s}}(\kappa T - \kappa_{\mathrm{ws}} T_{\mathrm{i}})) - f(\kappa T, v)\right|^2. \tag{27}$$

No truncation is applied to the objective function and $f(\kappa T, v)$ is not normalized by the maximum value of the TDE function, as in (23). The objective function can be reduced by completing the square, as

$$\begin{aligned}\kappa_{\mathrm{ws}} = \arg\min_{\kappa_{\mathrm{ws}}} \Bigg(& \sum_{\kappa=-\infty}^{\infty} \mathrm{sinc}\left(\pi f_{\mathrm{s}}\left(\kappa T - \kappa_{\mathrm{ws}} T_{\mathrm{i}}\right)\right)^2 \\ & - 2 \sum_{\kappa=-\infty}^{\infty} f(\kappa T, v)\, \mathrm{sinc}\left(\pi f_{\mathrm{s}}\left(\kappa T - \kappa_{\mathrm{ws}} T_{\mathrm{i}}\right)\right) \\ & + f(\kappa T, v)^2 \Bigg),\end{aligned} \tag{28}$$

where it can be shown that the first term is equal to 1, and the third term is independent of κ_{ws}, leading to the following reduced objective function:

$$\begin{aligned}\kappa_{\mathrm{ws}} &= \arg\min_{\kappa_{\mathrm{ws}}} \sum_{\kappa=-\infty}^{\infty} -f(\kappa T, v)\ \mathrm{sinc}(\pi f_{\mathrm{s}}(\kappa T - \kappa_{\mathrm{ws}} T_{\mathrm{i}})) \\ &= \arg\max_{\kappa_{\mathrm{ws}}} \sum_{\kappa=-\infty}^{\infty} f(\kappa T, v)\ \mathrm{sinc}(\pi f_{\mathrm{s}}(\kappa_{\mathrm{ws}} T_{\mathrm{i}} - \kappa T)),\end{aligned} \tag{29}$$

where $f(\kappa T, v)$ is only non-zero in the interval $\kappa \in \mathcal{L}$. The objective function can be further reduced by changing the bounds of the summation to

$$\kappa_{\mathrm{ws}} = \arg\max_{\kappa_{\mathrm{ws}}} \sum_{\kappa\in\mathcal{L}} f(\kappa T, v)\ \mathrm{sinc}(\pi f_{\mathrm{s}}(\kappa_{\mathrm{ws}} T_{\mathrm{i}} - \kappa T)), \tag{30}$$

where $\mathcal{L} = [-\frac{L}{2}, \frac{L}{2}]$ for TOA estimation, and $\mathcal{L} = [-L+1, L-1]$ for TDOA estimation.

This result shows that sinc interpolation can be reformulated to a discrete convolution of the TDE function $f(\kappa T, v)$ and a critically sampled sinc function by assuming infinite support of the sinc function over discrete time. This finding can be related back to Whittaker-Shannon interpolation (Marks, 1991), which is derived from the Nyquist-Shannon sampling theorem (Shannon, 1949). The Whittaker-Shannon interpolation method relies on the principle that a band-limited signal can be reconstructed from its discrete samples if the sampling frequency f_{s} is at least twice the highest frequency present in the signal (Shannon, 1949). In this case, the sinc function acts as an interpolation kernel that reconstructs the continuous-time signal from its discrete samples.

The Whittaker-Shannon interpolation can be modified to only consider S samples around the maximum value of the TDE function, as done in (23). The resulting optimization problem is given by

$$\kappa_{\mathrm{ws}} = \arg\max_{\kappa_{\mathrm{ws}}} \sum_{\kappa=\kappa_0-S}^{\kappa_0+S} f(\kappa T, v)\ \mathrm{sinc}(\pi f_{\mathrm{s}}(\kappa_{\mathrm{ws}} T_{\mathrm{i}} - \kappa T)). \tag{31}$$

The interpolation methods presented in this section are illustrated in the time domain in Fig. 1. In this figure, an acoustic reflection is simulated using a Thiran fractional delay filter, which is described in greater detail in Sec. VI. A band-limit of $B = \frac{2f_{\mathrm{s}}}{5}$ is imposed on the simulated reflection, after which various interpolation techniques are applied. For weighted frequency-domain interpolation, only the estimated peak value is shown, as this method does not operate directly in the time domain. The dotted vertical line marks the true time delay, while the marker indicates the location of the maximum value estimated by each interpolation method.

In the following section, the performance of the different interpolation methods is evaluated using simulated data. The results are presented in Sec. VI D.

VI. SIMULATIONS

To objectively assess the performance of the various interpolation methods for acoustic reflection localization through subsample TOA and TDOA estimation, a series of simulations were developed in Python (Rosseel and van Waterschoot, 2025).

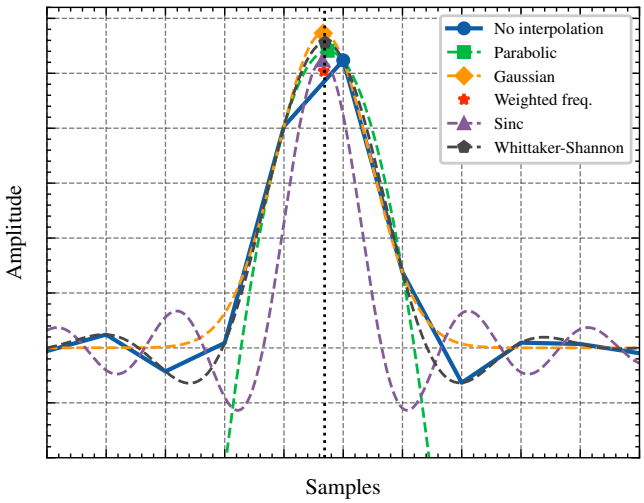


FIG. 1. Time-domain visualization of interpolation methods applied to a simulated acoustic reflection, modeled using a Thiran fractional delay filter with a band-limit of $B = \frac{2f_s}{5}$. The dotted vertical line represents the true time delay, and the marker shows the estimated maximum value obtained using each interpolation method. For weighted frequency-domain interpolation, only the peak value is included, as the method does not operate on the time-domain signal.

A. Setup

The simulations were designed to closely mimic real-world conditions while retaining the ability to objectively measure subsample TOA and TDOA performance using different interpolation methods. For simplicity, the simulations were carried out in two dimensions only. This limitation in dimensionality does not impact the performance of the interpolation methods for TOA and TDOA estimation.

The simulation setup consisted of a six-channel circular sensor array with a radius of 50 mm, positioned at the origin of the two-dimensional simulated space. The propagation speed of sound was set to $c = 343\ \mathrm{m\,s^{-1}}$, and the default sampling rate was set to $f_s = 8$ kHz. The signals simulated at each sensor position were corrupted with white noise to simulate measurement noise and had an SNR of 40 dB. To ensure that the SNR remained constant for all acoustic reflections in the simulation, the reflections were not scaled according to the distance between each sensor and the simulated image source. The analysis window had a duration of 4 ms, which at the default sampling rate results in a window length of $L = 32$ samples. The time-resolution with which sinc and Whittaker-Shannon interpolation methods were applied was determined by the interpolation factor $i = T/T_i$, where T is the operating sampling period and T_i is the interpolation sampling period used in sinc and Whittaker-Shannon interpolation. The default interpolation factor was set to $i = 200$ during the simulations.

It was ensured that, for each window of length L, only a single acoustic wavefront originating from an image source was present. This was accomplished by positioning K image sources around a sensor array in two-dimensional space such that the spherical wave originating from each image source arrived at the center of the sensor array at different time instants with a time lag of L/f_s seconds between each arrival. The position of the image sources can be described in two-dimensional space by the following equation:

$$\mathbf{x}_k = \mathbf{r}_c + \begin{bmatrix} \cos(\theta_k) \\ \sin(\theta_k) \end{bmatrix} \cdot k\frac{cL}{f_s}, \tag{32}$$

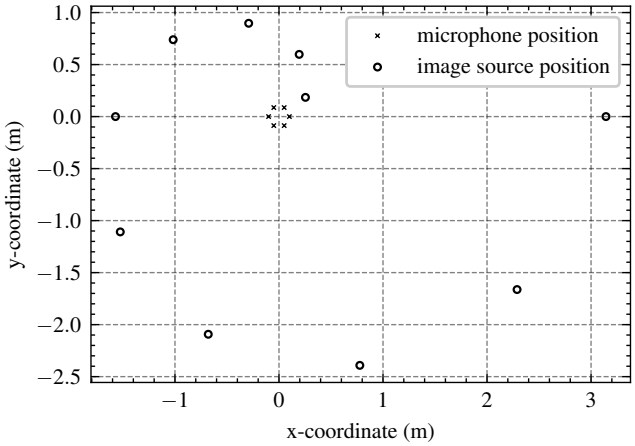


FIG. 2. Positioning of ten image sources around a sensor array such that the time of arrival in the center of the sensor array is equal to $t_k = k \cdot L/f_s$. For illustration purposes, the angular direction θ_k of each image source was set to $\theta_k = k \cdot 2\pi/K$.

where θ_k is the angle of the kth image source relative to the center of the sensor array. This is shown in Fig. 2, where ten image sources are placed around a sensor array with a TOA at the center of the sensor array equal to $t_k = k \cdot L/f_s$. In the simulations, the angular direction θ_k of each image source relative to the center of the sensor array was randomly distributed between $\theta_k \in [0, 2\pi]$ according to a uniform distribution. The number of image sources was set to $K = 1000$.

A number of independent parameters were varied to assess their influence on the performance of the interpolation methods for TDE with subsample accuracy. The independent parameters included the sampling frequency, the interpolation factor i, the SNR, and the analysis window length. When varying each parameter, all other parameters were kept constant. The performance of the interpolation methods was evaluated in terms of the TOA and TDOA error, which was calculated as the absolute difference between the estimated and true TOA and TDOA, respectively. The TDOA error was averaged over all sensor pairs and image-sources, while the TOA error was averaged over all sensors and image-sources.

The sampling frequency was varied from 2 kHz to 48 kHz, which allowed for an analysis into the effects of different sampling rates on the accuracy of TDE using different interpolation factors. The interpolation factor i was varied from 1 to 200 to analyse the performance of the sinc and Whittaker-Shannon interpolation methods under different time-resolution conditions. The robustness of each interpolation method under varying noise conditions was investigated by varying the SNR from -10 dB up to 60 dB. Finally, to determine the influence of the analysis window on TDOA estimation, the window duration was varied from 1 ms up to 16 ms, which at a sampling rate of $f_s = 8$ kHz corresponds to a window length of $L = 8$ samples up to $L = 128$ samples.

For each simulation, the time delays were calculated at the time instants when an acoustic reflection arrived at the middle of the sensor array. This ensures that the time delay estimates were calculated at the time instant when each windowed sensor signal contains at most a single acoustic reflection. The simulation results are discussed in section VI D.

B. Modeling Fractional Delays

The acoustic waves emanating from each image source towards the sensors were modeled using spherical wave propagation. As the TOA of each acoustic wavefront is not necessarily an integer multiple of the sampling period, a fractional delay filter was employed to simulate the true TOA of each acoustic wave with subsample accuracy. Ideally, a shifted sinc function serves as the optimal filter for modeling fractional delays (Laakso *et al.*, 1996). However, its use in the simulations would introduce an advantage to sinc and Whittaker-Shannon interpolation methods. To ensure a fair evaluation across various interpolation techniques, the Thiran all-pass filter (Laakso *et al.*, 1996) was selected. Thiran all-pass filters are a type of IIR fractional delay filter that exhibit a maximally flat group delay response at DC. The transfer function of a P^{th} order Thiran filter is given by

$$\begin{aligned} H(z) &= \frac{z^{-P} A(z^{-1})}{A(z)} \\ &= \frac{a_P + a_{P-1} z^{-1} + \cdots + a_1 z^{-(P-1)} + z^{-P}}{1 + a_1 z^{-1} + \cdots + a_{P-1} z^{-(P-1)} + a_P z^{-P}}, \end{aligned} \tag{33}$$

where for a desired fractional delay of δ samples, resulting in a total filter delay $\Delta = P + \delta$ samples, the coefficients a_p for $p = 0, 1, 2, \ldots, P$ are given by

$$a_p = (-1)^p \binom{P}{p} \prod_{k=0}^{P} \frac{\delta - P + k}{\delta - P + p + k}. \tag{34}$$

The precise modeling of fractional delays is crucial for evaluating the performance of the various interpolation methods. As demonstrated in (Välimäki, 1995), the fractional delay modeling error of the Thiran filter is minimized for $P - \frac{1}{2} \leq \delta < P + \frac{1}{2}$. In order to ensure a minimal modeling error for arbitrary fractional delays δ, the Thiran fractional delay filter is implemented in such a way that the desired fractional delay δ is independent of the filter order P, while also minimizing the modeled error. This is accomplished by decomposing the total filter delay Δ into an integer delay Δ_i and a fractional delay Δ_f, as follows:

$$\Delta_i = \left\lfloor \Delta + \frac{1}{2} \right\rfloor, \tag{35}$$

$$\Delta_f = \Delta - \Delta_i. \tag{36}$$

When $\Delta_i \geq P$, the total filter delay Δ can be implemented using a Thiran all-pass filter of order P. This is achieved by first applying an integer delay of $\Delta_i - P$ to the signal, followed by the Thiran all-pass filter with a desired fractional delay of $\delta = \Delta_f$ samples. This ensures that the output is delayed by exactly Δ samples. In the case where $\Delta_i < P$, the total filter delay Δ can be implemented by first applying a fractional delay of $\delta = \Delta_f$ samples using a P^{th} order Thiran all-pass filter, after which the first $P - \Delta_i$ samples of the output are removed.

C. Design Parameter *S*

The design parameter S specifies how many samples are included in sinc interpolation (23) and Whittaker-Shannon interpolation (31). The effect of this design parameter on the performance of the interpolation methods was assessed under two different bandwidth conditions: a critically sampled condition, i.e., $B = \frac{f_s}{2}$, and one where the bandwidth is limited to $B = \frac{2f_s}{5}$. For TDOA estimation, the design parameter S was varied from 1 sample up to L samples for each bandwidth condition. For TOA estimation, the design parameter S was varied from 1 sample up to $\frac{L}{2}$ samples for each bandwidth condition. The results of this analysis are shown for TDOA estimation in Fig. 3, and for TOA estimation in Fig. 4.

Critically Sampled Conditions

In critically sampled conditions, where the bandwidth of the acoustic impulses arriving at the sensors is equal to the Nyquist frequency, i.e., $B = \frac{f_s}{2}$, the performance of sinc interpolation is optimal for small values of S. This is because, for both TOA and TDOA interpolation of critically sampled acoustic impulses, the TDE function is again a critically sampled sinc function. As a result, the best performance can be achieved when the interpolation is performed only on samples that are close to the maximum value of the estimation function. On the other hand, Whittaker-Shannon interpolation achieves optimal performance for TDOA and TOA interpolation when parameter S is set to a higher value. This is due to the fact that a critically sampled sinc function does not have sufficient samples available around the main lobe of the function to achieve an accurate interpolation using

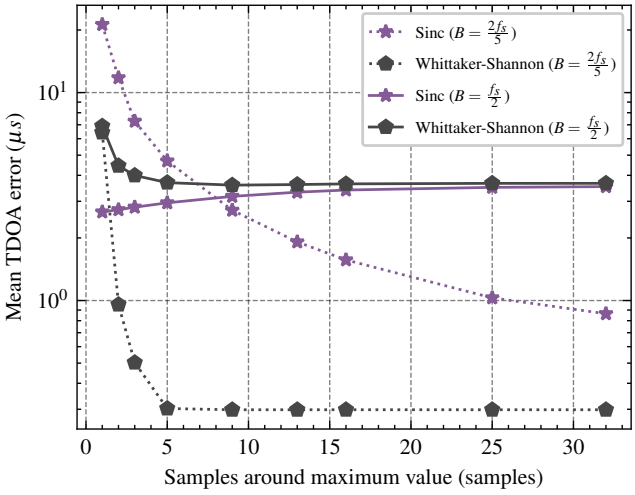


FIG. 3. The effect of varying parameter S from 1 sample up to L samples around the maximum value on the performance of sinc and Whittaker-Shannon interpolation for TDOA interpolation. This effect is examined under two bandwidth conditions: a critically sampled condition, $B = \frac{f_s}{2}$, and one where the bandwidth is limited to $B = \frac{2f_s}{5}$. The non-varying parameters were set to $f_s = 8$ kHz, $i = 200$, SNR = 40 dB, and $L = 32$ samples.

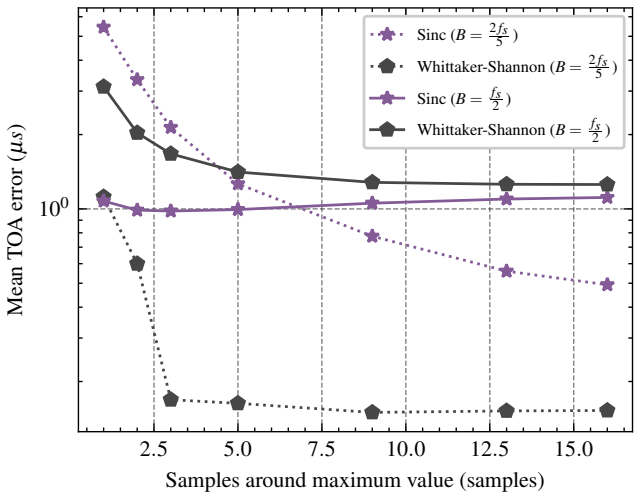


FIG. 4. The effect of varying parameter S from 1 sample up to $\frac{L}{2}$ samples around the maximum value on the performance of sinc and Whittaker-Shannon interpolation for TOA interpolation. This effect is examined under two bandwidth conditions: a critically sampled condition, i.e., $B = \frac{f_s}{2}$, and one where the bandwidth is limited to $B = \frac{2f_s}{5}$. The non-varying parameters were set to $f_s = 8$ kHz, $i = 200$, SNR = 40 dB, and $L = 32$ samples.

Whittaker-Shannon interpolation. Therefore, the best performance is achieved when additional samples around the maximum value are included in the interpolation.

Bandwidth Limited Conditions

In band-limited conditions, where the bandwidth of the acoustic impulses arriving at the sensors is $B = \frac{2f_s}{5}$, sinc interpolation achieves optimal performance for TDOA and TOA estimation when all samples in the window are included in the interpolation. In contrast, Whittaker-Shannon interpolation achieves optimal performance for both TDOA and TOA estimation when S is set to around half of the window size. Notably, the performance of both interpolation methods significantly deteriorates when S is set to a small value.

The increase in performance of Whittaker-Shannon interpolation over sinc interpolation for TOA and TDOA estimation in band-limited conditions can be attributed to the oversampling of the TDE function of band-limited acoustic reflections. This oversampling provides Whittaker-Shannon interpolation with a greater number of samples around the main lobe, allowing it to achieve more accurate TOA and TDOA estimates.

A significant degradation in the performance of TOA and TDOA estimation is observed in critically sampled conditions compared to band-limited conditions. This deterioration results from a limited availability of samples around the main lobe of a critically sampled sinc function, which negatively impacts the accuracy of TOA and TDOA estimates. As a result, the performance of interpolation methods degrades in critically sampled conditions.

D. Simulation Results

This section presents the simulation results, which demonstrate the impact of varying parameters on the performance of interpolation methods for TOA and TDOA estimation. The effects of sampling rate, interpolation factor, SNR, and analysis window are discussed, with results illustrated in Fig. 5 and Fig. 6 for TDOA estimation with critically sampled and band-limited acoustic reflections, respectively. The results for TOA estimation are presented in Fig. 7 and Fig. 8 for critically sampled and band-limited reflections, respectively.

1. Sampling Rate

The influence of sampling rate on TDOA estimation performance is shown in Fig. 5(a) for critically sampled acoustic reflections and in Fig. 6(a) for band-limited reflections. The results show that TDOA estimation accuracy increases as sampling rates increase across all interpolation methods. This enhancement can be attributed to the higher time resolution achieved as the sampling period decreases.

For critically sampled reflections, sinc interpolation outperforms other methods. In contrast, Whittaker-Shannon interpolation excels for band-limited reflections at lower sampling frequencies. Notably, at very high

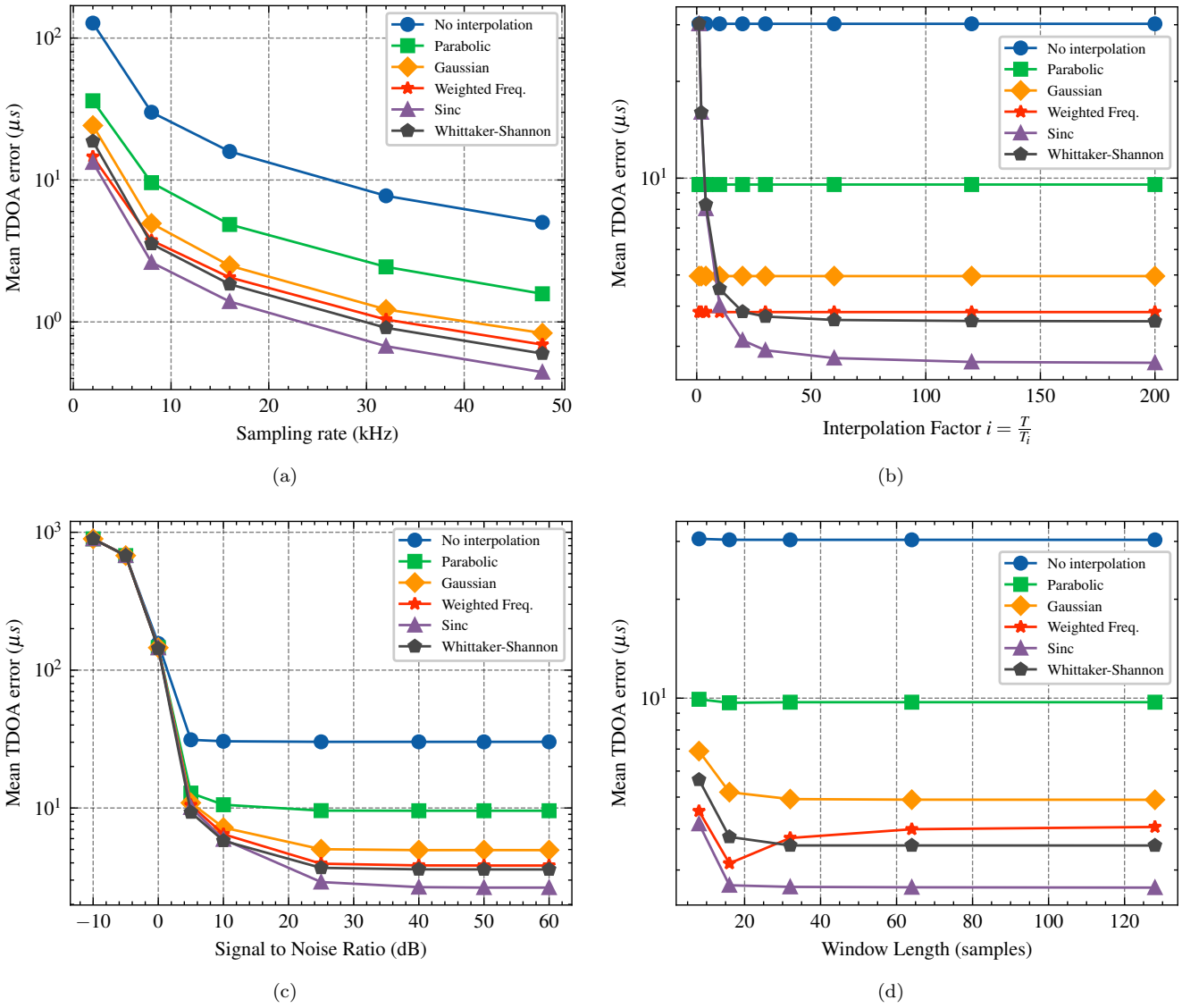


FIG. 5. The effect of varying (a) the sampling frequency f_s, (b) the interpolation factor i, (c) the SNR, and (d) the analysis window length L on interpolation performance for TDOA estimation of critically sampled acoustic reflections. The non-varying parameters were set to $f_s = 8$ kHz, $i = 200$, SNR $= 40$ dB, $L = 32$ samples, $S_{\text{sinc}} = 1$ sample, and $S_{\text{ws}} = 9$ samples.

sampling frequencies, weighted frequency interpolation becomes the top performer for band-limited reflections. This is because weighted frequency interpolation can effectively utilize the increased number of samples available within a fixed window duration at higher sampling rates, as observed in Fig. 6(d).

For TOA estimation, the performance of the interpolation methods is consistent with the results for TDOA estimation. Sinc interpolation outperforms other methods for critically sampled reflections, while Whittaker-Shannon interpolation excels for band-limited reflections.

2. Interpolation factor

The impact of varying the interpolation factor on TDOA estimation is shown in Fig. 5(b) and Fig. 6(b) for critically sampled and band-limited acoustic reflections, respectively. It is shown that for critically sampled acoustic reflections, interpolation factors of $i \geq 20$ allow both sinc and Whittaker-Shannon interpolation to outperform other interpolation methods. Sinc interpolation achieves optimal performance for critically sampled reflections.

For band-limited reflections, it can be observed that Whittaker-Shannon interpolation outperforms weighted frequency interpolation from interpolation factors of $i \geq 60$. In both cases, as the interpolation factor increases, the performance of the interpolation methods increases.

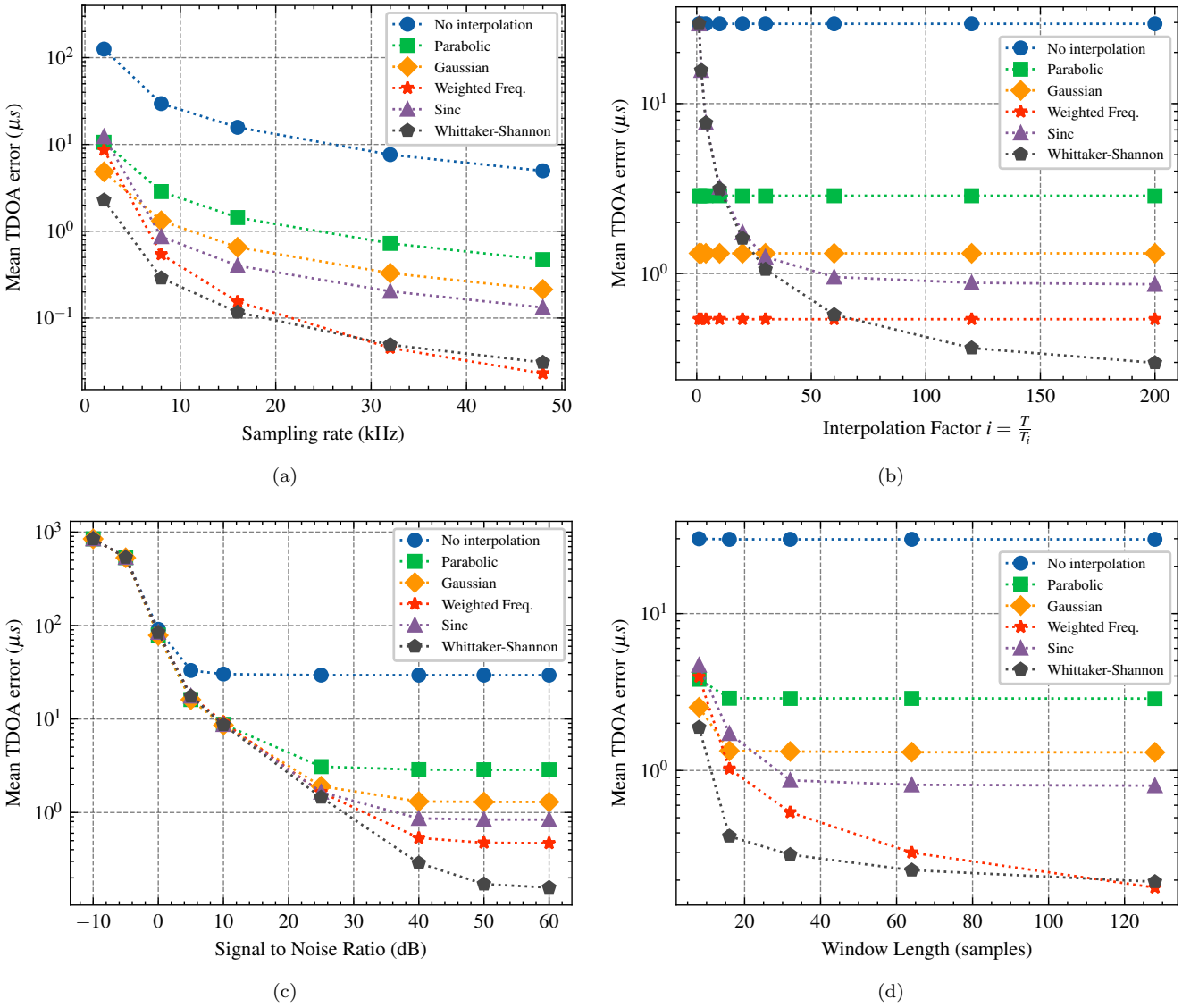


FIG. 6. The effect of varying (a) the sampling frequency f_s, (b) the interpolation factor i, (c) the SNR, and (d) the analysis window length L on interpolation performance for TDOA estimation of band-limited acoustic reflections. The bandlimit was set to $B = \frac{2f_s}{5}$. The non-varying parameters were set to $f_s = 8$ kHz, $i = 200$, SNR = 40 dB, $L = 32$ samples, $S_{\text{sinc}} = 32$ samples, and $S_{\text{ws}} = 9$ samples.

This is consistent with previous findings (Rosseel and van Waterschoot, 2021).

For TOA estimation, the performance of the interpolation methods is consistent with the results for TDOA estimation. Sinc interpolation outperforms other methods for critically sampled reflections, while Whittaker-Shannon interpolation excels for band-limited reflections.

3. Signal-to-Noise Ratio

Varying SNR significantly affects the performance of the interpolation methods for both TOA and TDOA estimation. The influence of SNR on TDOA estimation performance is demonstrated in Fig. 5(c) and Fig. 6(c) for critically sampled and band-limited acoustic reflections, respectively. At low SNR, TDOA estimation performance suffers significantly across all methods due to the noise-induced degradation of the cross-correlation function. In contrast, at higher SNR values, the performance trends observed earlier are reaffirmed. Specifically, sinc interpolation maintains its superiority over Whittaker-Shannon interpolation for critically sampled reflections, whereas Whittaker-Shannon interpolation achieves better performance for band-limited reflections.

The performance of TOA estimation when varying SNR is shown in Fig. 7(c) and Fig. 8(c) for critically sampled and band-limited reflections, respectively. The results are similar to those observed for TDOA estimation.

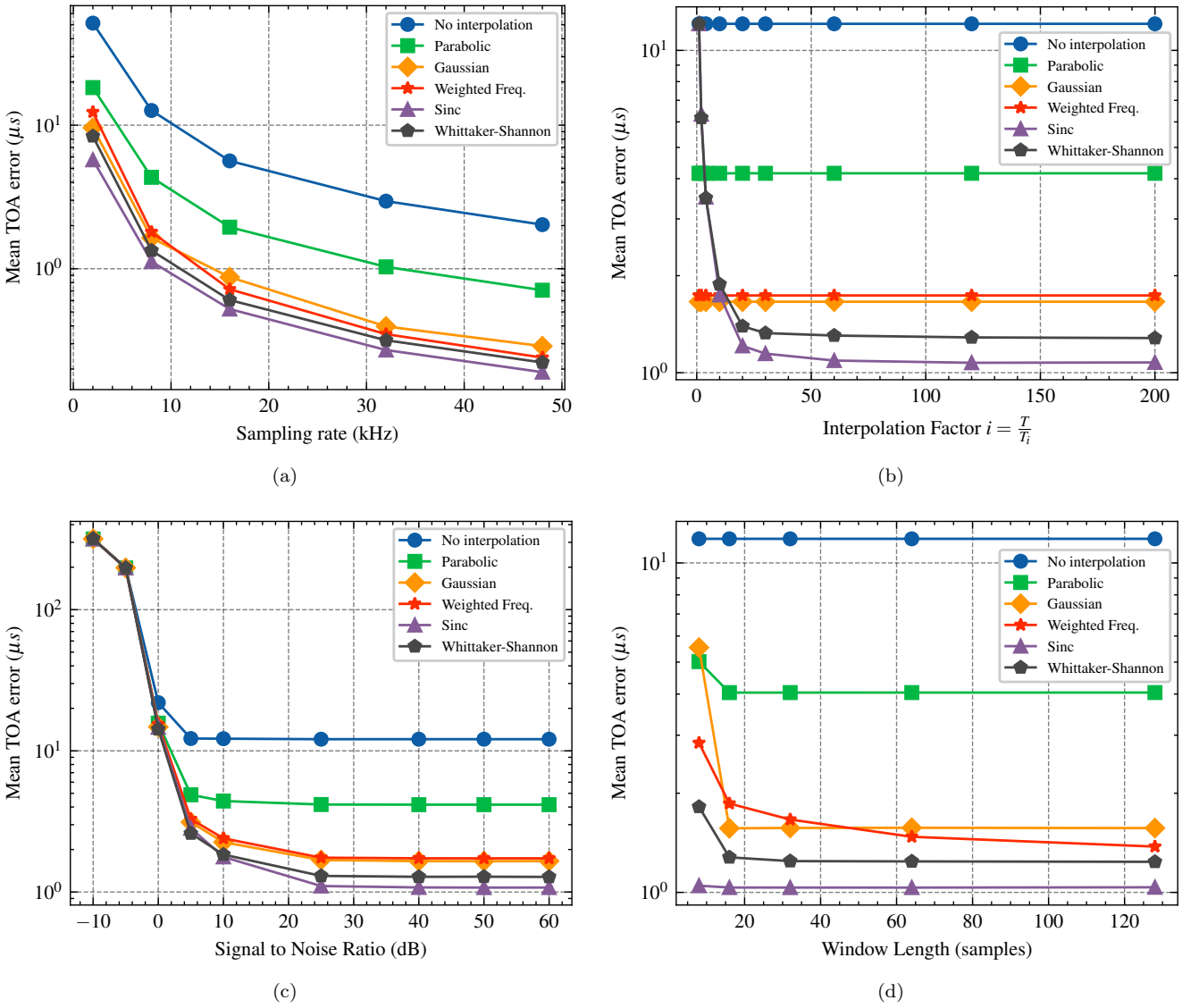


FIG. 7. The effect of varying (a) the sampling frequency f_s, (b) the interpolation factor i, (c) the SNR, and (d) the analysis window length L on interpolation performance for TOA estimation of critically sampled acoustic reflections. The non-varying parameters were set to $f_s = 8$ kHz, $i = 200$, SNR $= 40$ dB, $L = 32$ samples, $S_{\text{sinc}} = 1$ sample, and $S_{\text{ws}} = 9$ samples.

When SNR is sufficiently high, sinc interpolation outperforms other methods for critically sampled reflections, while Whittaker-Shannon interpolation achieves better performance for band-limited reflections. At low SNR, the performance of all interpolation methods deteriorates significantly.

4. Analysis Window Length

The performance of TDOA estimation when varying the analysis window length is shown in Fig. 5(d) and Fig. 6(d) for critically sampled and band-limited acoustic reflections, respectively. It can be observed that for very small window lengths, the performance of TDOA estimation degrades slightly for all interpolation methods due to insufficient information in the window. However, as the window length increases, the performance of TDOA estimation converges for all interpolation methods. Notably, weighted frequency interpolation has an optimal window length for critically sampled reflections, but its performance deteriorates as the window length grows. In contrast, for band-limited reflections, the performance of weighted frequency interpolation increases with increasing window length, allowing it to outperform Whittaker-Shannon interpolation for very long window lengths.

The performance of TOA estimation when varying the window length is shown in Fig. 7(d) and Fig. 8(d) for critically sampled and band-limited reflections, respectively. The results are consistent with those observed for TDOA estimation. For critically sampled re-

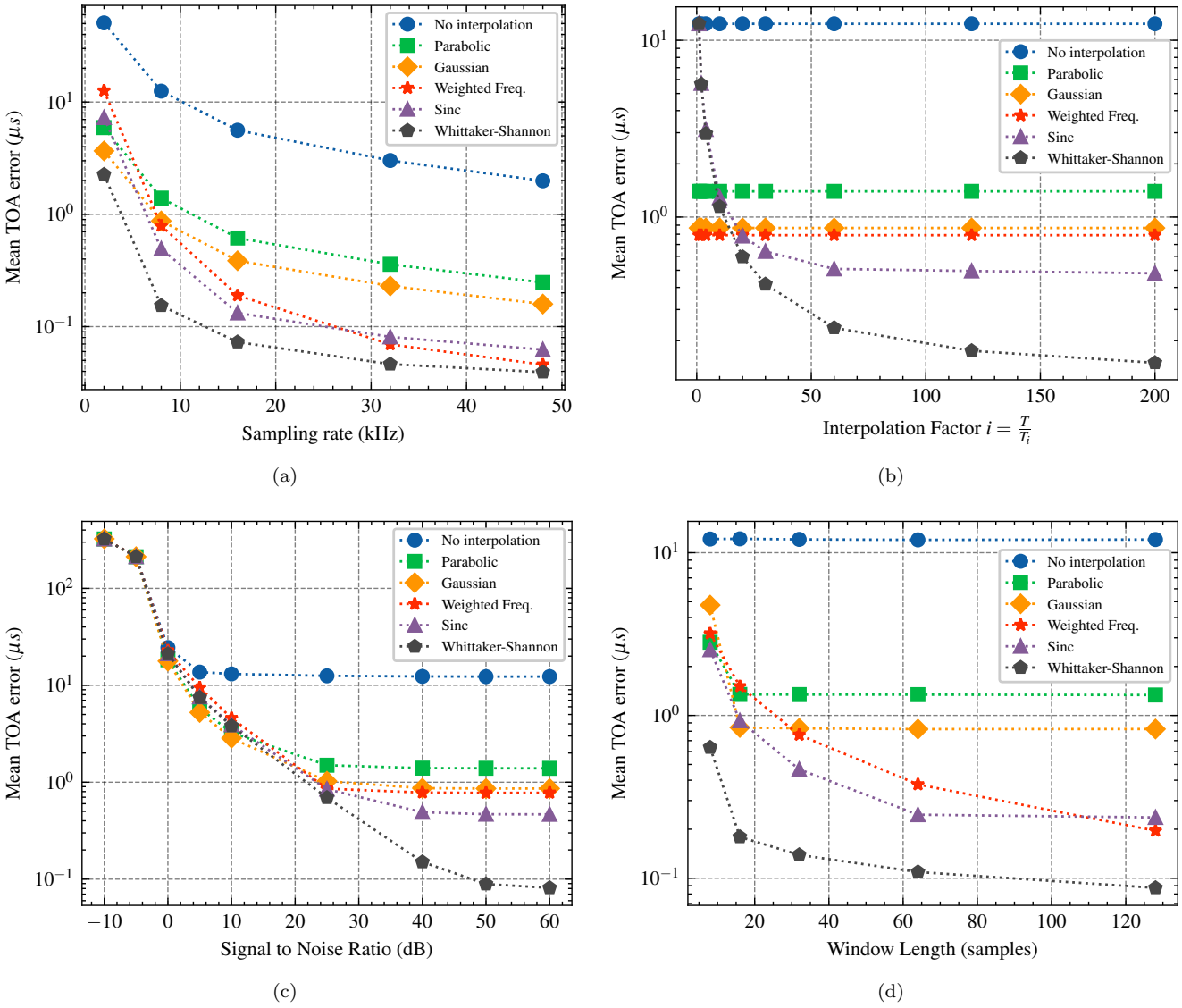


FIG. 8. The effect of varying (a) the sampling frequency f_s, (b) the interpolation factor i, (c) the SNR, and (d) the analysis window length L on interpolation performance for TOA estimation of band-limited acoustic reflections. The bandlimit was set to $B = \frac{2f_s}{5}$. The non-varying parameters were set to $f_s = 8$ kHz, $i = 200$, SNR = 40 dB, $L = 32$ samples, $S_{\text{sinc}} = 16$ samples, and $S_{\text{ws}} = 9$ samples.

flections, sinc interpolation outperforms other methods, while Whittaker-Shannon interpolation excels for band-limited reflections.

E. Localization performance

The acoustic source localization performance was assessed by estimating the position for a set of image sources $\hat{\mathbf{x}}_k$ using (16), where the estimated TOA and TDOA values were obtained using the different interpolation methods. The localization performance was evaluated in terms of the mean localization error ϵ_{pos}, calculated using a normalized Euclidean distance between the estimated and actual position of the image source. The normalization was done with respect to the distance between the center of the sensor array and the image source. The mean localization error is defined as

$$\epsilon_{\text{pos}} = \frac{1}{K} \sum_{k=0}^{K-1} \frac{\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|_2}{\|\mathbf{r}_c - \mathbf{x}_k\|_2}. \tag{37}$$

The results of this analysis can be seen in Fig. 9 and Fig. 10 for critically sampled and band-limited acoustic reflections, respectively. The results show that the localization performance increases with increasing sampling rates, interpolation factors, and SNRs. The performance of the interpolation methods for localization is consistent with the results observed for TOA and TDOA es-

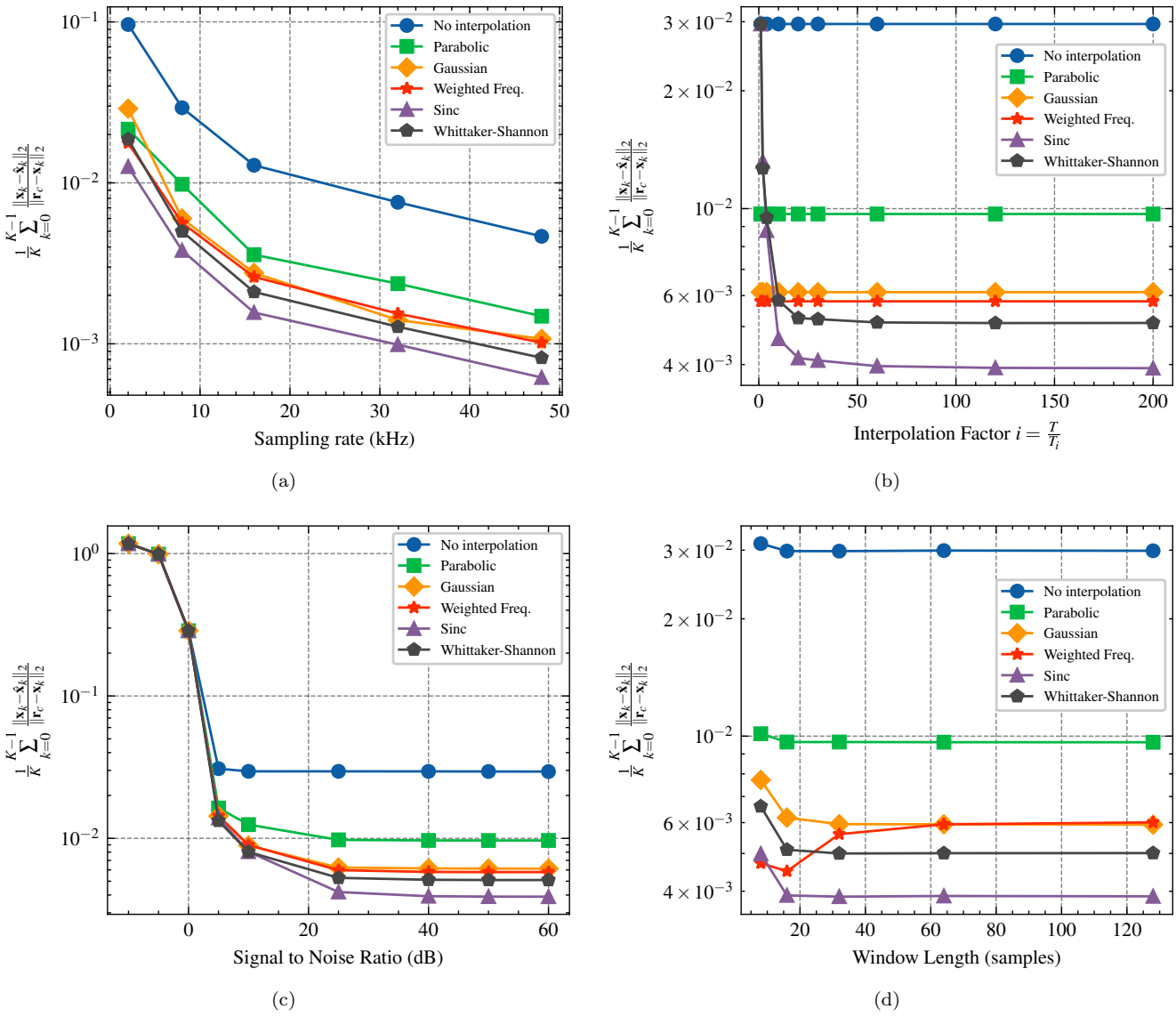


FIG. 9. The effect of varying (a) the sampling frequency f_s, (b) the interpolation factor i, (c) the SNR, and (d) the window length L on interpolation performance for determining the position in the XY plane of critically sampled acoustic reflections. The non-varying parameters were set to $f_s = 8$ kHz, $i = 200$, SNR $= 40$ dB, $L = 32$ samples, $S_{\text{sinc}} = 1$ sample, and $S_{\text{ws}} = 9$ samples.

timation. Sinc interpolation achieves the best result for critically sampled reflections, while Whittaker-Shannon interpolation outperforms the other interpolation methods for band-limited acoustic reflections.

A notable observation is that the localization performance is improved for band-limited reflections over critically sampled reflections. This is attributed to the increased number of samples that are available for interpolation around the main lobe of the TDE function in band-limited conditions. As a result, the interpolation methods can achieve more accurate TOA and TDOA estimates, resulting in improved localization performance.

VII. EVALUATION WITH MEASURED DATA

This section assesses the performance of the different interpolation methods discussed in Sec. III in achieving acoustic reflector localization with subsample accuracy using real-world experimental measurements. The performance of the interpolation methods is evaluated in terms of the TDOA error between the estimated TDOA and a ground-truth TDOA, and in terms of the positional error of the acoustic reflections.

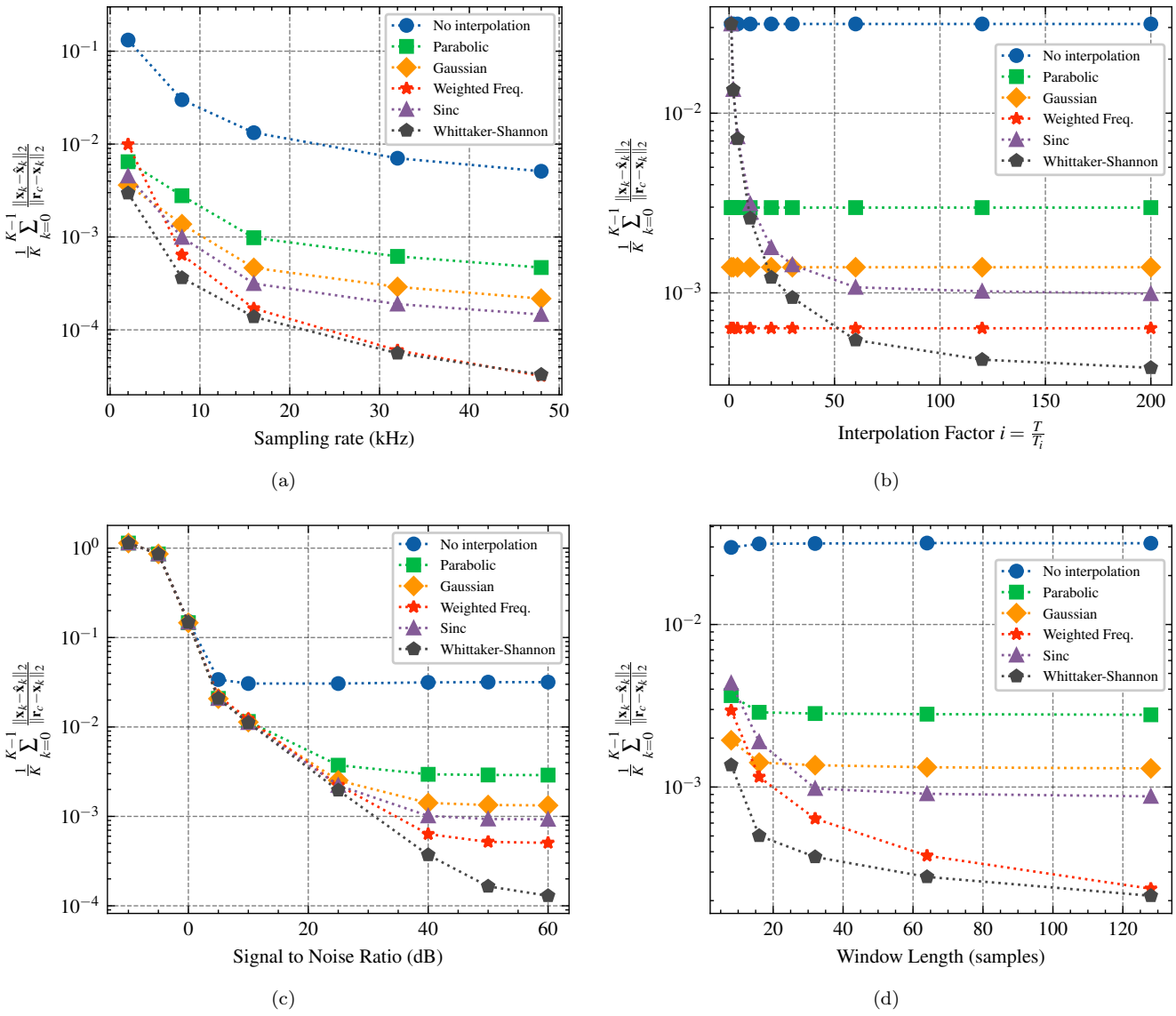


FIG. 10. The effect of varying (a) the sampling frequency f_s, (b) the interpolation factor i, (c) the SNR, and (d) the window length L on interpolation performance for determining the position in the XY plane of band-limited acoustic reflections. The bandlimit was set to $B = \frac{2f_s}{5}$. The non-varying parameters were set to $f_s = 8$ kHz, $i = 200$, SNR $= 40$ dB, $L = 32$ samples, $S_{\text{sinc}} = 32$ samples, and $S_{\text{ws}} = 9$ samples.

A. Experimental Setup

The measurements were obtained from the publicly available *MYRiAD* dataset (Dietzen *et al.*, 2023), which consists of room acoustic measurements acquired using diverse microphone-array and loudspeaker configurations. The measurements considered in this study were recorded in the Alamire Interactive Lab (AIL), employing a circular microphone array with a radius of 20 cm, comprising eight microphones. A total of 20 distinct source positions were utilized, resulting in 160 unique microphone-source pairs. The source positions that were used are denoted as *SL1-SL8* and *SU1-SU12* in the *MYRiAD* dataset. The AIL has a volume of approximately 208 m^3 and exhibits a reverberation time of $T_{20} = 0.5$ s (Dietzen *et al.*, 2023).

The measurements are originally sampled at a rate of $f_s = 44.1$ kHz. To allow down-sampling to 8 kHz by an integer factor, the measurements were resampled to a sampling rate of $f_s = 48$ kHz.

To compensate for the frequency response of the loudspeakers and microphones, a matched filter $m_n[\kappa]$ was applied to each measured RIR (Antonacci *et al.*, 2012; Carlo *et al.*, 2021). This matched filter was estimated for each loudspeaker-microphone pair by isolating the direct-path component. The matched filter $m_n[\kappa]$

was then correlated with the measured RIR $\tilde{h}_n[\kappa]$ to obtain the compensated RIR $h_n[\kappa]$, as,

$$h_n[\kappa] = \sum_{i=0}^{M-1} \tilde{h}_n[\kappa] m_n[\kappa + i], \tag{38}$$

where M is the number of samples in the measured RIR $h_n[\kappa]$. The compensated RIRs were used to estimate the TOA of each acoustic reflection and TDOAs between each microphone-source pair.

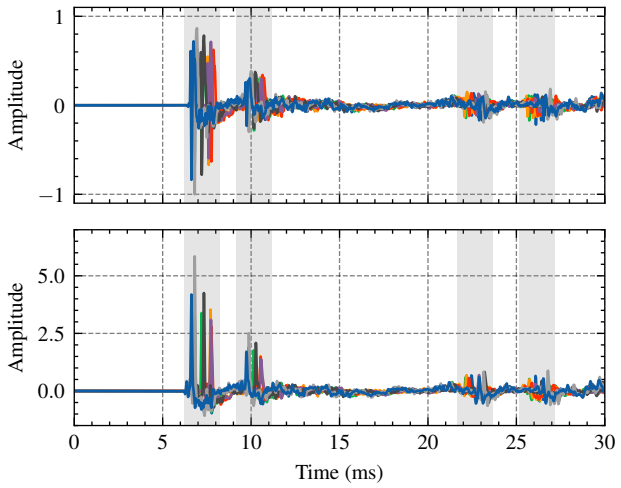


FIG. 11. Normalized RIRs from the *MYRiAD* dataset at loudspeaker position *SL1*, captured using the eight-microphone circular array. The top panel plots the initial 30 ms of measured RIRs including the direct path and first three reflections. The bottom panel plots the compensated RIRs after match-filtering. The gray-shaded regions indicate the time windows used to isolate the direct path and reflections, identified using a peak-picking algorithm.

The direct-path component and first three reflections were subsequently identified by applying a peak-picking algorithm to the match-filtered RIRs. The acoustic events were isolated using a 2 ms long window, corresponding to a window length of $L = 96$ samples. In Fig. 11, the first 30 ms of the RIRs measured using an eight-microphone circular array at loudspeaker position *SL1* are shown in the top panel, while the first 30 ms of the match-filtered RIRs are shown in the bottom panel. The gray-shaded regions indicate the time windows used to isolate the direct path and reflections, which were identified using the peak-picking algorithm.

The ground-truth TDOA, TOA, and position estimates were obtained at 48 kHz without applying interpolation methods. These ground truths serve as a reference for evaluating the performance of the different interpolation methods. Limiting the measurements to the direct-path and the first three reflections was done to ensure that the reflections were accurately determined in time and to avoid the influence of multiple reflections in a single time window.

TABLE I. Median, mean, and standard deviation of the TDOA error for the different interpolation methods on *MYRiAD* dataset measurements. The best performing interpolation method is highlighted in bold.

Interpolation method	Median (µs)	Mean (µs)	Std (µs)
No interpolation	41.67	31.96	20.67
Parabolic	7.56	7.55	4.82
Gaussian	3.97	5.40	4.95
Weighted Frequency	2.65	4.34	4.85
Sinc	2.92	3.46	2.86
Whittaker-Shannon	**1.67**	**2.30**	**2.00**

It is worth noting that, due to idealized assumptions made in the derivation of the interpolation methods, model mismatches will inevitably be introduced under real-world measurement conditions. For instance, the specular-reflection assumption made in Sec. II does not hold in practice, since acoustic reflectors impose frequency-dependent filtering on reflected signals. Moreover, the image-source model is based on point-like source and receiver approximations that do not hold in real environments.

The performance of the interpolation methods was evaluated by first down-sampling the isolated early reflections to 8 kHz, resulting in a window length of 16 samples. Subsequently, the TOA and TDOA were estimated for all microphone-pairs and loudspeaker positions from the downsampled windows while applying the different interpolation methods discussed in Sec. III. The performance of all interpolation methods was assessed in terms of the TDOA error and positional error, defined as the absolute difference between the estimated TDOA and the ground-truth TDOA, and the distance between the estimated reflection position and the ground-truth reflection position, respectively. It is important to mention that the results presented in this section are based on the assumption that ground-truth TDOA and TOA estimates are correct, which, due to sampling rate limitations and the presence of noise in the ground-truth measurements, may not be the case in practice. Therefore, the results presented in this section should be interpreted with caution.

The interpolation parameters for sinc interpolation and Whittaker-Shannon interpolation were set to $S_{\text{sinc}} = 3$ samples and $S_{\text{ws}} = 13$ samples, respectively, based on the results obtained in the simulation study presented in Sec. VI. For sinc and Whittaker-Shannon interpolation, the interpolation factor was set to i = 500.

B. Results

TDOA Error

The TDOA error for the different interpolation methods is presented in Table I. The best performing interpolation method is highlighted in bold. The results

show that Whittaker-Shannon interpolation outperforms the other interpolation methods in terms of the median, mean, and standard deviation of the TDOA error on the direct-path and first three reflections in the *MYRiAD* dataset measurements.

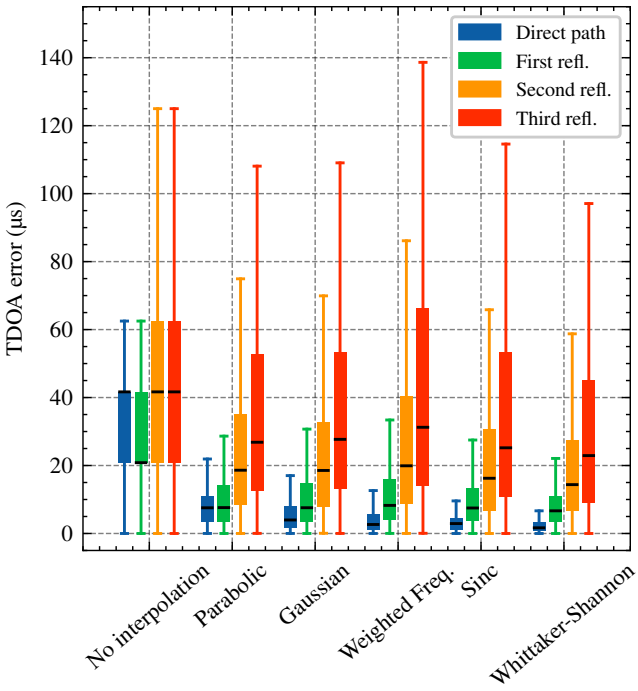


FIG. 12. TDOA error for the different interpolation methods after isolating the direct-path and first three acoustic reflections in the *MYRiAD* dataset measurements. The results are presented as box plots, where the central mark indicates the median TDOA error, the edges of the box represent the 25th and 75th percentiles, and the whiskers extend to the most extreme data points excluding outliers.

In Fig. 12, the TDOA error for the different interpolation methods is visualized as box plots for the direct-path and first three reflections in the *MYRiAD* dataset measurements. The results show that Whittaker-Shannon interpolation demonstrates the lowest median TDOA error and the smallest interquartile range for the direct-path and first three reflections, indicating that Whittaker-Shannon interpolation provides the most consistent performance across the different microphone-source pairs and loudspeaker positions. Sinc interpolation demonstrates the second-best performance in terms of the median TDOA error and the interquartile range. It can also be observed that weighted frequency interpolation performs well for the direct-path and first reflection, but its performance deteriorates for the second and third reflections, as indicated by the larger interquartile range.

TABLE II. Median, mean, and standard deviation of the positional error for the different interpolation methods on *MYRiAD* dataset measurements. The best performing interpolation method is highlighted in bold.

Interpolation method	Median (cm)	Mean (cm)	Std (cm)
No interpolation	2.49	4.14	3.64
Parabolic	0.39	0.93	1.09
Gaussian	0.84	0.95	0.70
Weighted Frequency	0.57	0.78	0.61
Sinc	**0.27**	0.56	0.61
Whittaker-Shannon	0.32	**0.46**	**0.42**

Positional Error

The positional error for the different interpolation methods is presented in Table II. The results show that sinc interpolation demonstrates the lowest median positional error, followed by Whittaker-Shannon interpolation. The results also show that the latter method demonstrates the lowest mean and standard deviation of the positional error.

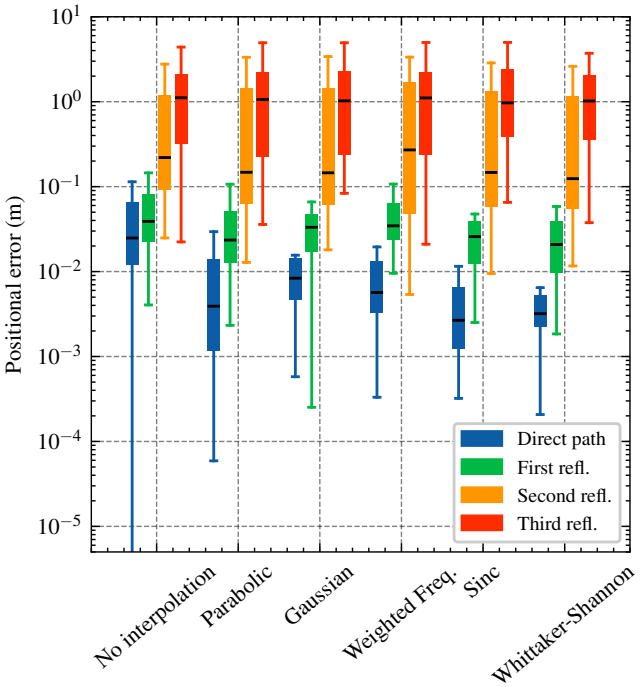


FIG. 13. Positional error for the different interpolation methods after isolating the direct-path and first three acoustic reflections in the *MYRiAD* dataset measurements. The results are presented as box plots, where the central mark indicates the median TDOA error, the edges of the box represent the 25th and 75th percentiles, and the whiskers extend to the most extreme data points excluding outliers.

In Fig. 13, the positional error is visualized for the different interpolation methods as box plots for the direct-path and first three reflections in the *MYRiAD* dataset measurements. The results are plotted on a logarithmic scale and show that sinc interpolation demonstrates the lowest median positional error, followed closely by Whittaker-Shannon interpolation.

The findings of this study indicate that the interpolation methods discussed in this paper can enhance the accuracy of acoustic reflector localization in real-world environments. However, given that the sampling rate utilized in the *MYRiAD* dataset measurements was 44.1kHz, the ground truth TDOA and reflection position exhibited limited time-resolution. Consequently, the outcomes presented in this section may not fully reflect the performance of the interpolation methods. Future research will focus on evaluating the performance of the interpolation methods in achieving subsample precision in acoustic reflector localization, using real-world measurements with more accurate ground truth estimates.

VIII. CONCLUSION

In this paper, we have presented a comprehensive study on time delay interpolation with subsample accuracy for acoustic reflector localization. We have derived the Whittaker-Shannon interpolation formula from the previously proposed sinc interpolation by assuming infinite support of the sinc function over discrete time. This reformulation allows for the application of Whittaker-Shannon interpolation to windowed TDE functions, which is essential to achieve accurate estimation in practical scenarios where acoustic reflections are often windowed in reverberant conditions. The Whittaker-Shannon interpolation method has not been previously applied to the subsample TDE of acoustic reflections. It was shown through simulation that sinc interpolation and Whittaker-Shannon interpolation are able to outperform parabolic, Gaussian, and weighted frequency interpolation in terms of the TDOA error and positional error for critically sampled and band-limited signals, respectively. The simulations also show that sinc and Whittaker-Shannon interpolation is robust for small window lengths, low sampling rates, and low SNRs compared to existing interpolation methods. The performance of the interpolation methods was further evaluated on real-world measurements from the *MYRiAD* dataset, showing that sinc and Whittaker-Shannon interpolation are able to outperform existing interpolation methods in terms of the TDOA error and positional error for the direct-path and first three reflections. The results indicate that Whittaker-Shannon interpolation provides the most consistent performance across different sensor-source pairs and loudspeaker positions. The results of this study can be used to improve the accuracy of acoustic reflector localization methods, which are essential for various applications, such as room acoustics analysis, sound source localization, and acoustic scene analysis.

ACKNOWLEDGMENTS

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of FWO Large-scale research infrastructure "The Library of Voices - Unlocking the Alamire Foundation's Music Heritage Resources Collection through Visual and Sound Technology" (I013218N), FWO SBO Project "The sound of music - Innovative research and valorization of plainchant through digital technology" (S005319N), and SBO Project "New Perspectives on Medieval and Renaissance Courtly Song" (S005525N). This research received funding from the Flemish Government (AI Research Program). This research received funding from internal KU Leuven funds: C14/21/075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices", C3/23/056 "HELIXON: Hybride, efficiënte en vloeiende interpolatie van geluid in uitgebreide realiteit". The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

AUTHOR DECLARATIONS

The authors have no conflicts to disclose.

DATA AVAILABILITY

The code used to generate the simulation results presented in this paper is available at https://github.com/hrosseel/TDE_INTERP.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Amer. **65**(4), 943–950.

Amengual Garí, S. V., Arend, J. M., Calamia, P. T., and Robinson, P. W. (**2021**). "Optimizations of the spatial decomposition method for binaural reproduction," J. Audio Eng. Soc. **68**(12), 959–976.

Annibale, P., Filos, J., Naylor, P. A., and Rabenstein, R. (**2012**). "TDOA-based speed of sound estimation for air temperature and room geometry inference," IEEE Trans. Audio Speech Lang. Process. **21**(2), 234–246.

Antonacci, F., Filos, J., Thomas, M. R., Habets, E. A., Sarti, A., Naylor, P. A., and Tubaro, S. (**2012**). "Inference of room geometry from acoustic impulse responses," IEEE Trans. Audio Speech Lang. Process. **20**(10), 2683–2695.

Antonacci, F., Sarti, A., and Tubaro, S. (**2010**). "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *Proc. 2010 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '10)*, pp. 2822–2825.

Camilleri, J., Kaplanis, N., and De Sena, E. (**2019**). "Evaluation of car cabin acoustics using auralisation over headphones," in *Proc. 2019 AES Int. Conf. Immersive and Interactive Audio*.

Carlo, D. D., Tandeitnik, P., Foy, C., Bertin, N., Deleforge, A., and Gannot, S. (**2021**). "dEchorate: a calibrated room impulse response dataset for echo-aware signal processing," EURASIP J. Audio, Speech, Music Process. **2021**(1), 39.

Céspedes, I., Huang, Y., Ophir, J., and Spratt, S. (**1995**). "Methods for estimation of subsample time delays of digitized echo signals," Ultrasonic Imaging **17**(2), 142–171.

DeFrance, G., Daudet, L., and Polack, J.-D. (**2008**). "Detecting arrivals within room impulse responses using matching pursuit," in *Proc. 11th Int. Conf. Digital Audio Effects (DAFx '08)*, pp. 1–4.

DeFrance, G., Daudet, L., and Polack, J.-D. (**2009**). "Using matching pursuit for estimating mixing time within room impulse responses," Acta Acustica United with Acustica **95**, 1082–1092.

Dietzen, T., Ali, R., Taseska, M., and van Waterschoot, T. (**2023**). "MYRiAD: a multi-array room acoustic database," EURASIP J. Audio, Speech, Music Process. **2023**, 17.

Ehrenberg, J. E., Ewart, T. E., and Morris, R. D. (**1978**). "Signal-processing techniques for resolving individual pulses in a multipath signal," J. Acoust. Soc. Amer. **63**(6), 1861–1865.

El Baba, Y., Walther, A., and Habets, E. A. (**2017**). "3D room geometry inference based on room impulse response stacks," IEEE/ACM Trans. Audio Speech Lang. Process. **26**(5), 857–872.

Farina, A. (**2000**). "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Preprints Audio Eng. Soc. Convention 108*.

Kaplanis, N., Bech, S., Tervo, S., Pätynen, J., Lokki, T., van Waterschoot, T., and Jensen, S. H. (**2017**). "A rapid sensory analysis method for perceptual assessment of automotive audio," J. Audio Eng. Soc. **65**(1/2), 130–146.

Knapp, C., and Carter, G. (**1976**). "The generalized correlation method for estimation of time delay," IEEE Trans. Audio Speech Lang. Process. **24**(4), 320–327.

Kuttruff, H. (**2017**). *Room Acoustics*, 6th ed. ed. (CRC Press).

Laakso, T., Välimäki, V., Karjalainen, M., and Laine, U. (**1996**). "Splitting the unit delay [FIR/all pass filters design]," IEEE Signal Process. Mag. **13**(1), 30–60.

Lai, X., and Torp, H. (**1999**). "Interpolation methods for time-delay estimation using cross-correlation method for blood velocity measurement," IEEE Trans. Ultrason., Ferroelect., Freq. Control **46**(2), 277–290.

Lokki, T., Pätynen, J., Kuusinen, A., and Tervo, S. (**2016**). "Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences," J. Acoust. Soc. Amer. **140**(1), 551–562.

Marks, R. J. (**1991**). *The Cardinal Series*, 33–56 (Springer New York, New York, NY).

Merimaa, J., and Pulkki, V. (**2005**). "Spatial impulse response rendering I: Analysis and synthesis," J. Audio Eng. Soc. **53**, 1115–1127.

Meyer-Kahlen, N., Amengual Garí, S., and Lokki, T. (**2022**). "What the spatial decomposition method can and cannot do," in *Proc. 24th Int. Congr. Acoust. (ICA '22)*.

Müller, S., and Massarani, P. (**2001**). "Transfer-function measurement with sweeps," J. Audio Eng. Soc. **49**.

Novak, A., Lotton, P., and Simon, L. (**2015**). "Synchronized swept-sine: Theory, application, and implementation," J. Audio Eng. Soc. **63**(10), 786–798.

Pätynen, J., Tervo, S., and Lokki, T. (**2013**). "Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses," J. Acoust. Soc. Amer. **133**(2), 842–857.

Pirinen, T. (**2009**). "Confidence scoring of time delay based direction of arrival estimates and a generalization to difference quantities," Ph.D. thesis, Tampere University of Technology.

Pulkki, V., and Merimaa, J. (**2006**). "Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests," J. Audio Eng. Soc. **54**, 3–20.

Puomio, O. (**2021**). "Parametric analysis and reproduction of indoor acoustic environments," Ph.D. thesis, Aalto University, Finland.

Remaggi, L., Jackson, P. J. B., Coleman, P., and Wang, W. (**2017**). "Acoustic reflector localization: Novel image source reversion and direct localization methods," IEEE/ACM Trans. Audio Speech Lang. Process. **25**(2), 296–309.

Rosseel, H., and van Waterschoot, T. (**2021**). "Improved acoustic source localization by time delay estimation with subsample accuracy," in *Proc. 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–8.

Rosseel, H., and van Waterschoot, T. (**2025**). "Python Code for Time Delay Interpolation for Acoustic Reflector Localization" https://github.com/hrosseel/TDE_INTERP/.

Shannon, C. (**1949**). "Communication in the presence of noise," Proc. IRE **37**(1), 10–21.

Svilainis, L. (**2019**). "Review on time delay estimate subsample interpolation in frequency domain," IEEE Trans. Ultrason., Ferroelect., Freq. Control **66**(11), 1691–1698.

Svilainis, L., Lukoseviciute, K., Dumbrava, V., and Chaziachmetovas, A. (**2013**). "Subsample interpolation bias error in time of flight estimation by direct correlation in digital domain," Measurement **46**(10), 3950–3958.

Tervo, S., Laukkanen, P., Pätynen, J., and Lokki, T. (**2014**). "Preferences of critical listening environments among sound engineers," J. Audio Eng. Soc. **62**(5), 300–314.

Tervo, S., Pätynen, J., and Lokki, T. (**2012**). "Acoustic reflection localization from room impulse responses," Acta Acustica United with Acustica **98**(3), 418–440.

Tervo, S., Pätynen, J., Kuusinen, A., and Lokki, T. (**2013**). "Spatial decomposition method for room impulse responses," J. Audio Eng. Soc. **61**(1/2), 17–28.

Tervo, S., and Tossavainen, T. (**2012**). "3D room geometry estimation from measured impulse responses," in *Proc. 2012 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '12)*, pp. 513–516.

Välimäki, V. (**1995**). "Discrete-time modeling of acoustic tubes using fractional delay filters," Ph.D. thesis, Helsinki University of Technology.

Zhang, L., and Wu, X. (**2005**). "On cross correlation based-discrete time delay estimation," Proc. 2005 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '05) **4**, iv/981–iv/984.