Improved Acoustic Source Localization by Time Delay Estimation with Subsample Accuracy

Hannes Rosseel
Dept. of Electrical Engineering (ESAT-STADIUS)
KU Leuven
Leuven, Belgium
hannes.rosseel@esat.kuleuven.be

Toon van Waterschoot
Dept. of Electrical Engineering (ESAT-STADIUS)
KU Leuven
Leuven, Belgium
toon.vanwaterschoot@esat.kuleuven.be

***Abstract*—Discrete-time signal processing algorithms for Time Delay Estimation (TDE) generally yield a delay estimate that is an integer multiple of the sampling period. In applications that operate at a relatively low sampling rate or that require a highly accurate delay estimate, the TDE resolution obtained in this way may not be sufficient. One such application is 6DOF audio acquisition, in which accurate time delays are needed to estimate directions of arrival and sound source positions relative to microphone positions. Depending on the TDE algorithm and the envisaged application, several solutions have been proposed to increase the TDE accuracy, including parabolic interpolation, increasing the sampling rate, and increasing the distance between the microphones in the array. In this paper, we propose a novel method for solving the TDE resolution problem, which is directly rooted in the Nyquist-Shannon sampling theory. By fitting a continuous-time sinc function to the cross-correlation function of two measured acoustic impulses, a delay estimate can be obtained with a time resolution that is only a fraction of the sampling period. When applying this approach to a set of acoustic impulse responses measured between a single sound source and a microphone array, e.g., in a 6DOF audio acquisition scenario, the increase in TDE accuracy yields a more accurate estimate of the time differences of arrival of the source relative to the different microphones, which can eventually lead to improved source localization. A comparison of the proposed method with existing methods will be presented.**

***Index Terms*—source localization, time delay estimation, Nyquist-Shannon sampling theorem**

I. Introduction

Time Delay Estimation (TDE) is an active research topic that builds the foundation for numerous practical applications in various fields [1]–[3].

6DOF audio acquisition is one such application that depends on an accurate TDE resolution. In 6DOF audio acquisition, an acoustic source localization algorithm estimates the Direction of Arrival (DOA) of a propagating wavefront relative to a microphone array. When the distance of an acoustic sound source is relatively large compared to the distance between the microphones in the array, the propagating wavefront can be estimated using the plane wave propagation model [3]. The wavefront arrives in each sensor of the microphone array at different time instances. These time differences can be used to determine the Direction of Arrival (DOA) of the wavefront. Techniques that use time-differences to determine the DOA are called Time Difference of Arrival (TDOA) based DOA techniques.

One shortcoming of TDE on discrete-time signals is that the obtained delay estimate is generally an integer multiple of the sampling period. For specific applications that operate at a relatively low sampling rate or that require a highly accurate delay estimate, the TDE resolution obtained in this way may not be sufficient.

In general, TDE methods utilize the cross-correlation between measured signals. More specifically, generalized cross-correlation (GCC) methods [4], [5] are popular methods for TDE, considering their improved robustness against noise. In these methods, the TDE corresponds to the time instance where the cross-correlation function reaches its maximum absolute value. As stated above, one shortcoming of TDE, when working with discrete-time signals, is that the obtained TDE will be an integer multiple of the sampling period. A common technique that aims to improve the accuracy of TDE methods to merely a fraction of the sampling period, consists of fitting a parabolic [6] or Gaussian [7] function to the peak of the GCC. The improved TDE is found at the vertex of the fitted parabolic or Gaussian function.

This paper proposes a novel interpolation technique that is directly rooted in the Nyquist-Shannon sampling theorem [8], [9]. The proposed interpolation method relies on the fact that an acoustic impulse, measured by a microphone array at sampling rate f_s, is by definition bandlimited. This paper shows that the GCC between two bandlimited acoustic impulses results in a sinc function with a maximum value at the TDE. By fitting a critically sampled and truncated sinc-function to the GCC function, a TDE is obtained with a time resolution that is only a fraction of the sampling period. Sinc interpolation of the GCC was used earlier in [10], but several aspects proposed in our work go beyond the work in [10], i.e.,

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven internal fund VES/19/004, FWO Large-scale research infrastructure "The Library of Voices - Unlocking the Alamire Foundation's Music Heritage Resources Collection through Visual and Sound Technology" (I013218N), and FWO SBO Project "The sound of music - Innovative research and valorization of plainchant through digital technology" (S005319N). 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.

the use of a truncated sinc interpolant, the consideration of the resulting impact on DOA estimation, and the experimental evaluation with real measurements.

The remainder of this paper is organized as follows. Section II discusses the acoustic propagation model that is used in this paper and derives the GCC between acoustic impulses in the continuous-time domain and the discrete-time domain. Section III presents two existing interpolation techniques for obtaining a TDE with subsample accuracy. The proposed interpolation method for improving the TDE is also discussed in this section. Section IV explains a TDOA-based DOA estimation technique that is based on the slowness vector propagation model. Sections V and VI compare the performance of the proposed interpolation technique with existing interpolation techniques in simulations and anechoic measurements. Section VII concludes the paper.

II. Time Delay estimation of an ideal acoustic impulse

A. Continuous-time representation

In this section, the TDE of a propagating ideal acoustic impulse, representing a point on a spherical wavefront, is derived. The acoustic impulse is emitted from a point source located at a distance r from a microphone array. The microphone array consists of N spatially separated microphones. When free-field conditions are assumed, the signal $x_n(t)$ measured by each microphone n is modeled as

$$x_n(t) = \frac{\delta\left(t - \frac{r_n}{c}\right)}{4\pi r_n}, \tag{1}$$

where $\delta(\cdot)$ denotes the Dirac delta function, r_n is the distance between the point source and the microphone n, and c is the propagation speed of air (about $343\,\mathrm{m\,s^{-1}}$, at a temperature of $20\,^{\circ}\mathrm{C}$). A common method for estimating the time delay between two spatially separated microphone signals is by calculating the GCC between the microphone signals [4]. The GCC between two signals of microphones m and n is defined as

$$R_{m,n}(\tau) = \int_{-\infty}^{\infty} \Psi(\omega) \cdot G_{m,n}(\omega) \cdot e^{j\omega\tau} d\omega, \tag{2}$$

where $\Psi(\omega)$ denotes a general frequency weighting [4], and $G_{m,n}(\omega)$ is the cross power spectrum between the two signals.

Since our signal model consists of wide-band signals, direct weighting can be used for the GCC, i.e., $\Psi(\omega) \equiv 1$. Using direct weighting, (2) becomes the regular cross-correlation. We can calculate the time-domain cross-correlation between two microphone signals as defined in (1), as

$$R_{m,n}(\tau) = \int_{-\infty}^{\infty} \frac{\delta\left(t - \frac{r_m}{c}\right)}{4\pi r_m} \cdot \frac{\delta\left(t + \tau - \frac{r_n}{c}\right)}{4\pi r_n} dt \tag{3}$$

$$= \frac{1}{16\pi^2 r_m r_n} \cdot \delta\left(\frac{r_m - r_n}{c} + \tau\right). \tag{4}$$

It is clear that the cross-correlation function $R_{m,n}(\tau)$ is a scaled and time-shifted Dirac delta function which is non-zero only when τ is equal to the true TDOA, i.e., when

$$\tau = \frac{r_n - r_m}{c}. \tag{5}$$

B. Discrete-time representation

In practice, however, the signals $x_n(t)$ measured by the microphone array will be discretized by sampling the signals at a fixed sampling interval $T = \frac{1}{f_s}$, where f_s is the sampling rate. According to the Nyquist-Shannon sampling theorem [8], this discrete-time acquisition requires the introduction of a bandlimit on the measured signal less than or equal to the Nyquist frequency $\frac{f_s}{2}$. As a result, in the case of critical sampling, i.e., when a bandlimit equal to $\frac{f_s}{2}$ is imposed, the signal $x_n(t)$ measured in the free-field by each microphone n in the microphone array becomes

$$x_n(t) = \frac{\operatorname{sinc}\left(\pi f_s(t - \frac{r_n}{c})\right)}{4\pi r_n}, \text{ for } t = kT, k \in \mathbb{Z}, \tag{6}$$

where $\operatorname{sinc}(x) = \frac{\sin(x)}{x}$. We can estimate the TDOA between two spatially separated microphone signals by performing GCC with direct weighting as done in (3), which results in

$$R_{m,n}(\tau) = \int_{-\infty}^{\infty} \frac{\operatorname{sinc}\left(\pi f_s\left(t - \frac{r_m}{c}\right)\right)}{4\pi r_m} \cdot \frac{\operatorname{sinc}\left(\pi f_s\left(t + \tau - \frac{r_n}{c}\right)\right)}{4\pi r_n} dt \tag{7}$$

$$= \frac{1}{16\pi^2 f_s r_m r_n} \cdot \operatorname{sinc}\left(\pi f_s\left(\frac{r_m - r_n}{c} + \tau\right)\right). \tag{8}$$

The cross-correlation between two band-limited acoustic impulses is thus a scaled sinc function which is maximized when τ is equal to the true TDOA Δ, i.e., when

$$\tau = \frac{r_n - r_m}{c} \triangleq \Delta. \tag{9}$$

III. Improving Time Delay Estimation with subsample accuracy

When $R_{m,n}(\tau)$ is sampled at a sampling period $T = \frac{1}{f_s}$, $R_{m,n}[k] = R_{m,n}(kT)$ reaches its maximum absolute sample value when $k = k_0$. However, since the true TDOA Δ is not necessarily an integral multiple of T, i.e., it generally holds that $k_0 T \neq \Delta$. Instead, Δ lies between the discretized time instants $(k_0{-}1)T$ and $k_0 T$, or $k_0 T$ and $(k_0{+}1)T$. In the worst-case scenario, a maximum discretization error of $\frac{T}{2}$ seconds will be present in the TDE.

Previous research proposes several interpolation schemes which aim to estimate the lag where the cross-correlation function $R_{m,n}[k]$ reaches its maximum, with subsample accuracy. Popular interpolation schemes are parabolic interpolation [6] and Gaussian interpolation [7]. In this section, these interpolation techniques are discussed, after which our proposed interpolation scheme based on sinc interpolation is presented.

A. Parabolic interpolation

The parabolic interpolation scheme aims to improve the TDE when Δ is a fractional multiple of the sampling period T. This interpolation scheme fits a parabolic function of the form $y_p(\tau) = a\tau^2 + b\tau + c$ around lag k_0 at the cross-correlation function peak [6]. The parameters of $y_p(\tau)$ are found by fitting the parabolic function through points $R_{m,n}[k_0-1]$, $R_{m,n}[k_0]$, and $R_{m,n}[k_0+1]$. Since the true TDOA Δ is not an integer multiple of the sampling period T, we can find the fractional TDE at the vertex of the fitted parabolic function $\Delta_p = \frac{-b}{2a}$.

B. Gaussian interpolation

An alternative interpolation scheme can be used, which models the peak value of the cross-correlation function as a Gaussian function [7] of the form

$$y_g(\tau) = \alpha \cdot e^{-b(\tau-c)^2}. \tag{10}$$

The vertex location of a Gaussian function depends only on parameter c. By fitting a Gaussian function through points $R_{m,n}[k_0-1]$, $R_{m,n}[k_0]$, and $R_{m,n}[k_0+1]$, as done similarly in parabolic interpolation, we can find the value for c as

$$c = \frac{\ln(R_{m,n}[k_0+1]) - \ln(R_{m,n}[k_0-1])}{4\ln(R_{m,n}[k_0]) - 2(\ln(R_{m,n}[k_0+1]) + \ln(R_{m,n}[k_0-1]))}. \tag{11}$$

It is required that $R_{m,n}[k] > 0$, for $k \in [k_0-1, k_0, k_0+1]$. If this is not the case, for example when $R_{m,n}[k]$ shows a very sharp peak in the cross-correlation function, we propose to estimate parameter c after increasing $R_{m,n}[k]$ with an offset equal to $2 \cdot \min(R_{m,n}[k])$, for $k \in [k_0-1, k_0, k_0+1]$.

The fractional TDE is estimated as $\Delta_g = c$.

C. Sinc interpolation

This paper proposes an alternative interpolation scheme that exploits the fact that the cross-correlation function between two bandlimited acoustic impulses can be modeled as a sinc function, as shown in (8). Because of this, we argue that sinc interpolation is the most suitable interpolation scheme for obtaining a TDE with subsample accuracy.

We propose a sinc interpolation scheme that fits a critically sampled and truncated sinc function around the maximum value of the cross-correlation function. This fitting, which consists in solving the following minimization problem, with cost function

$$J(\tau) = \int_{-\infty}^{\infty} \left| \operatorname{sinc}(\pi f_s(t-\tau)) - \frac{R_{m,n}(t)}{\max(R_{m,n}(t))} \right|^2 dt, \tag{12}$$

yields a correct TDE if the model in (8) holds, i.e.,

$$\tau_{\text{sinc}} = \arg\min_\tau J(\tau) = \frac{r_n - r_m}{c}. \tag{13}$$

When $R_{m,n}(\tau)$ is sampled at a sampling period $T = \frac{1}{f_s}$, we can write the optimization problem for sinc fitting in the discrete-time domain as

$$\kappa_{\text{sinc}} = \arg\min_\kappa \sum_{k=k_0-S}^{k_0+S} \left| \operatorname{sinc}(\pi f_s(kT - \kappa T_i) - \frac{R_{m,n}[k]}{R_{m,n}[k_0]} \right|^2, \tag{14}$$

where S is a design parameter specifying the number of samples around the lag value k_0 before truncating the sinc function. For TDOA estimation purposes, it is recommended that S is greater than the maximum propagation delay $\Delta_{\max}$ between the microphones. For any given microphone array with a maximum distance $d_{\max}$ between the microphones, $\Delta_{\max} = \frac{d_{\max}}{c} f_s$. T_i represents the interpolation sampling period, where $T_i < T$. This design parameter determines the subsample accuracy of the interpolation. A detailed discussion of the design parameters can be found in Section V. The fractional TDE with sinc interpolation is estimated as $\Delta_{\text{sinc}} = k_0 T + \kappa T_i$.

IV. Time Difference of Arrival based Direction of Arrival Estimation

A propagating acoustic wave traveling towards a microphone array will arrive in each microphone at different time instants. Correctly estimating these relative time delays between microphones is of crucial importance to TDOA-based DOA estimation. These DOA estimation methods work in two steps. First, the TDOAs between all microphone pairs in an array are estimated. From this TDOA information, a source direction relative to the microphone array is estimated. TDOA-based DOA estimation is challenging since any inaccuracy in the TDE will result in an error in the DOA estimation [11].

Assuming plane wave propagation, as is common in TDOA-based DOA estimation, the normal vector of a wavefront can be defined, describing the speed and direction of the propagating plane wave. This vector is called the Slowness Vector (SV) and can be used to find the DOA of a plane wave [11]. The SV $\boldsymbol{k}$ is defined as

$$\boldsymbol{k} = c^{-1}\boldsymbol{n}, \tag{15}$$

where $\boldsymbol{n}$ is the normal vector of the wavefront, defining the propagation direction in space. The SV can be estimated for a microphone array containing N microphones by first calculating the TDOAs between all $M = \binom{N}{2}$ microphone pairs. The vector of TDOA estimates is denoted by

$$\hat{\boldsymbol{\tau}}_{\boldsymbol{M}} = [\hat{\tau}_{1,2}, \hat{\tau}_{1,3}, \ldots, \hat{\tau}_{N-1,N}]^{\mathrm{T}}. \tag{16}$$

We can then estimate the SV using the least squares approach given by [11]

$$\hat{\boldsymbol{k}} = \boldsymbol{V}^{+}\hat{\boldsymbol{\tau}}_{\boldsymbol{M}}, \tag{17}$$

where $(\cdot)^+$ is the Moore-Penrose pseudo-inverse, and $\boldsymbol{V}$ is the matrix of microphone position difference vectors. $\boldsymbol{V}$ is dependent on the relative microphone position vectors $\boldsymbol{p}_n \in \mathbb{R}^3$ in the array and is calculated as

$$\boldsymbol{V} = [\boldsymbol{p}_1 - \boldsymbol{p}_2, \boldsymbol{p}_1 - \boldsymbol{p}_3, \ldots, \boldsymbol{p}_{N-1} - \boldsymbol{p}_N]^{\mathrm{T}}. \tag{18}$$

The direction of the SV in a three-dimensional space can be expressed as an azimuth angle $\theta_{\boldsymbol{k}}$ and elevation angle $\phi_{\boldsymbol{k}}$.

$$\theta_{\boldsymbol{k}} = \operatorname{atan}\left(\frac{k_y}{k_x}\right), \tag{19}$$

$$\phi_{\boldsymbol{k}} = \operatorname{atan}\left(\frac{k_z}{\sqrt{\left(k_x^2 + k_y^2\right)}}\right), \tag{20}$$

where k_x, k_y, and k_z represent the x, y, and z coordinate values of the SV $\boldsymbol{k}$, respectively.

V. Simulations

This section describes the simulation setup and simulation results. A comparison is made between the simulation performance of TDOA-based DOA estimation, using the proposed sinc interpolation scheme, and using the interpolation schemes introduced in Section III.

A. Setup

The simulation setup consists of a 3-dimensional microphone array containing three coincident microphone pairs, where each microphone pair is positioned along a different Cartesian axis. The distance between two microphones in a pair is set to $d = 50\,\mathrm{mm}$. The acoustic propagation delay between a microphone and a sound source located at distance r is modeled by a truncated Lagrange fractional delay FIR filter of prototype filter order $L = 40$, and filter order $O = 20$ [12], [13]. This fractional delay filter has a magnitude response of which the transition band depends on the fractional delay Δ_f, as shown in Fig. 1. The propagation speed of sound is set to $c = 343\,\mathrm{m\,s^{-1}}$.

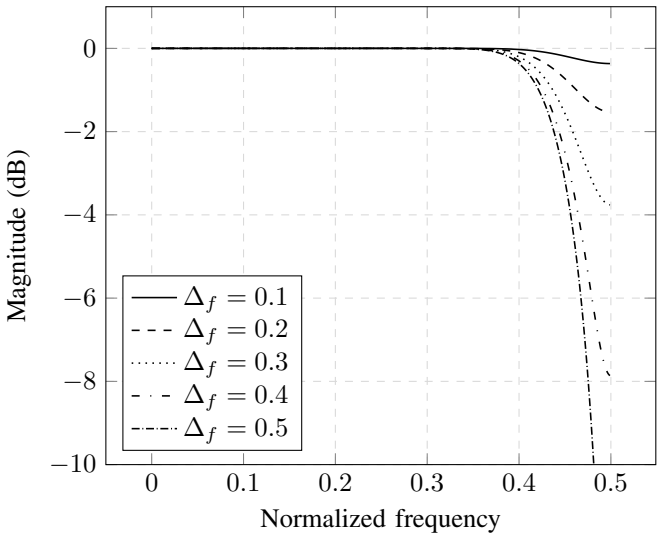


Fig. 1. Magnitude response of a truncated Lagrange fractional delay FIR filter of order $O = 20$ and prototype filter order $L = 40$, designed for varying fractional delays Δ_f.

White noise was added to each microphone signal to account for measurement noise in the simulations. The signal-to-noise ratios (SNRs) of each microphone signal in the performed simulations were set to $40\,\mathrm{dB}$. The ideal acoustic impulses were sampled at a sampling rate of $f_s = 4\,\mathrm{kHz}$. The interpolation interval half width of the proposed method was set to $S = 0.05 f_s$ and the interpolation factor $i = \frac{T}{T_i}$ was set to 100. The distance r between the microphone and sound source was $2.2\,\mathrm{m}$.

The TDOA-based DOA estimate is calculated at different sound source positions. At each position, the azimuth angle θ, relative to the microphone array, is taken from the set $\{0°, 30°, 60°, 90°, 120°, 150°, 180°\}$. The average DOA estimation error over all azimuth angles θ is compared for different cross-correlation interpolation schemes, i.e., no interpolation, parabolic interpolation, Gaussian interpolation, and the proposed sinc interpolation.

This comparison is repeated for different varying simulation parameters, such as the distance r between the microphone array and the sound source, the sampling frequency of the simulation, the interpolation factor T_i of our proposed sinc interpolation method, and the SNR of the microphone measurements.

B. Results

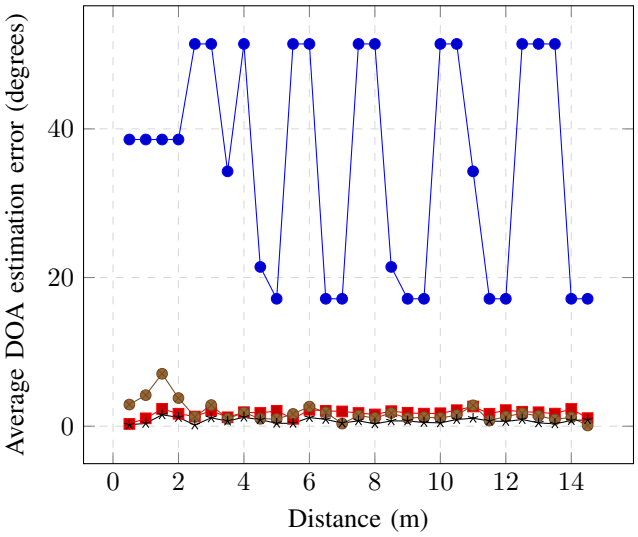
In this section, four different simulation comparisons are discussed to validate the performance of the proposed interpolation method relative to the performance of existing interpolation methods for TDOA-based DOA estimation.

In the first comparison, the distance r between the sound source and microphone array is varied. It can be seen in Fig. 2 that, for most distances r, parabolic and Gaussian interpolation methods produce a higher average DOA estimation error across all directions θ than the proposed sinc interpolation.

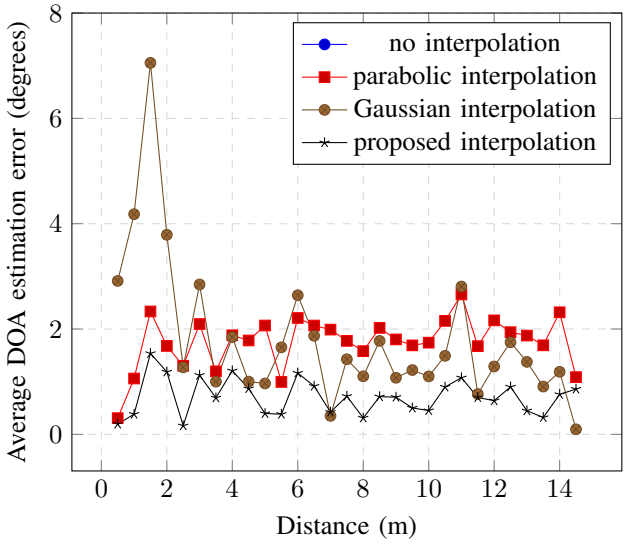
In the second comparison, as shown in Fig. 3, the sampling rate f_s is varied over all divisions of $f_s = 192\,\mathrm{kHz}$. The average DOA estimation error decreases for all interpolation methods when the sampling rate f_s increases. The proposed sinc interpolation performs similarly to parabolic and Gaussian interpolation.

A third comparison is presented, where the interpolation factor i is varied. In Fig. 4, we observe that the average DOA estimation error for the proposed interpolation decreases as the interpolation factor i increases. This decreasing DOA estimation error can be explained by the increase in TDE resolution as $T_i \ll T$. It should be noted that a high interpolation factor in the proposed method requires an increase in computational power.

Finally, a comparison is presented where the SNR of the microphone measurements is varied, as shown in Fig. 5. The SNR is controlled by adding white noise to the measured acoustic impulses. We can see that, at very low SNR values, the average DOA estimation error across all interpolation schemes shows a high variance. This indicates that current TDOA-based DOA estimation techniques are not accurate in these conditions. The proposed interpolation technique performs similarly to

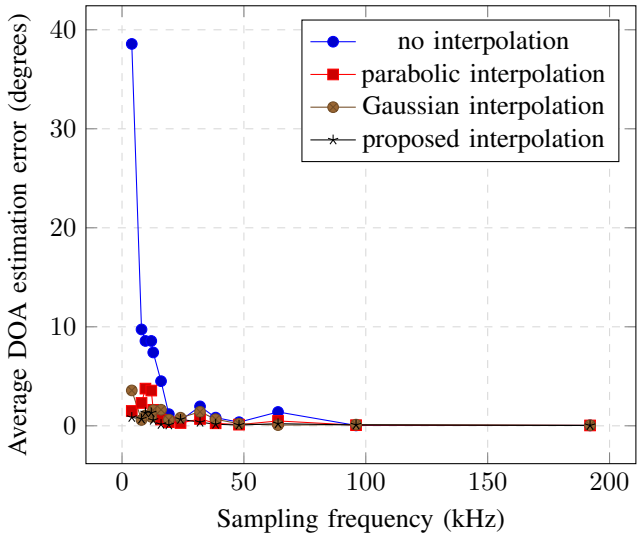


(a)

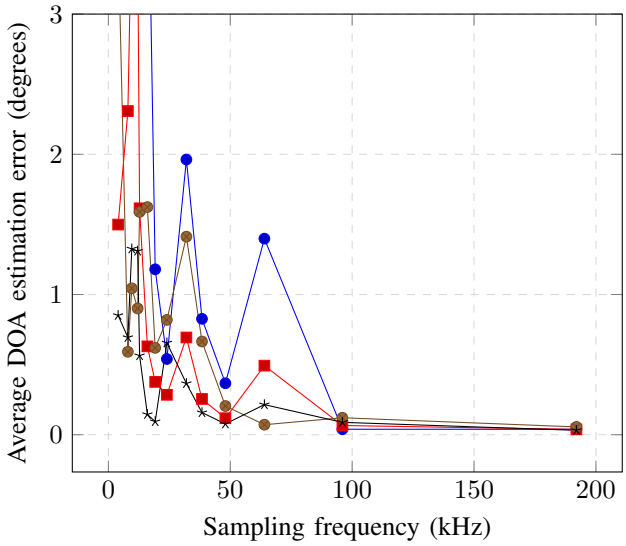


(b)

Fig. 2. Simulation results: (a) average DOA estimation error in degrees versus the distance r between the sound source and the microphone array in meters. The simulation parameters are: sampling frequency $f_s = 4\,\mathrm{kHz}$, interpolation factor $\frac{T}{T_i} = 100$, interpolation interval half width $S = 0.05 f_s$, and SNR $= 40\,\mathrm{dB}$. (b) Average DOA estimation error plotted from 0 to 8 degrees.



(a)



(b)

Fig. 3. Simulation results: (a) average DOA estimation error in degrees versus the sampling frequency f_s. The simulation parameters are: distance $r = 2.2\,\mathrm{m}$, interpolation factor $\frac{T}{T_i} = 100$, interpolation half width $S = 0.05 f_s$, and SNR $= 40\,\mathrm{dB}$. (b) Average DOA estimation error plotted from 0 to 3 degrees.

parabolic and Gaussian interpolation schemes at low SNR values. At higher SNR values, starting at 35 dB, the average DOA estimation error stabilizes. The proposed interpolation technique outperforms parabolic and Gaussian interpolation at high SNR values.

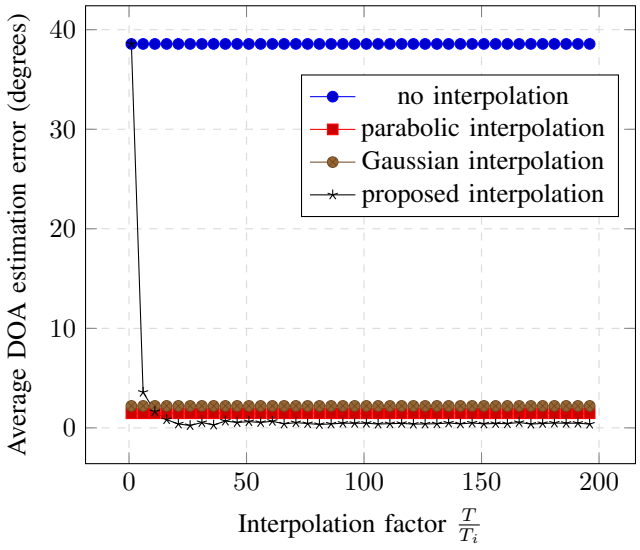
VI. Anechoic Measurements

This section describes the measurement setup and measurement results, which further validate the proposed interpolation scheme. These acoustic measurements took place in the semi-anechoic chamber at the KU Leuven Department of Physics.

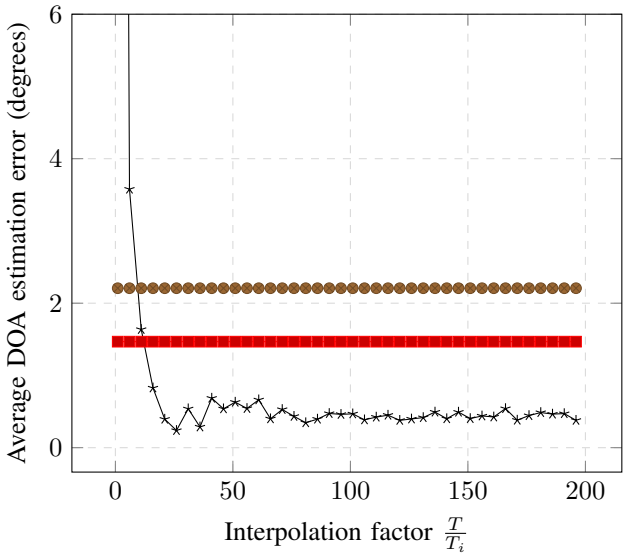
A. Setup

A microphone array of type G.R.A.S. vector intensity probe 50VI-1, consisting of six omnidirectional microphones, was placed in a semi-anechoic chamber. This microphone array contains three coincident, phase-matched microphone pairs on each axis, separated by a spacer of $50\,\mathrm{mm}$.

A loudspeaker of type Genelec 8030C was positioned at a distance of $2.2\,\mathrm{m}$ and an angle θ relative to the microphone array, as shown in Fig. 6. The loudspeaker position was changed between measurements so that the angle θ, relative to the microphone array, varied from $0°$ to $-180°$ in steps of $-30°$. During the measurements, the sampling frequency was set to $f_s = 192\,\mathrm{kHz}$.

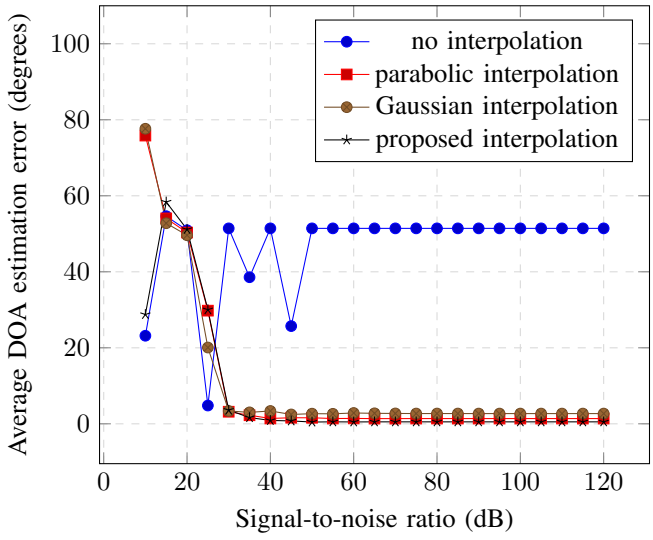


(a)

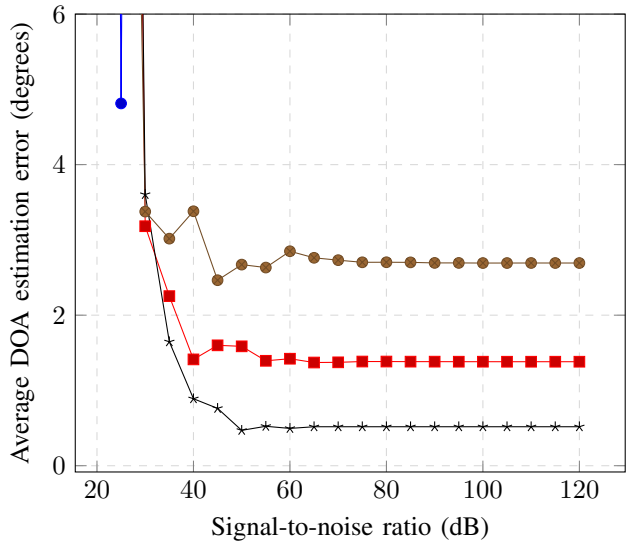


(b)

Fig. 4. Simulation results: (a) average DOA estimation error in degrees versus the interpolation factor $i = \frac{T}{T_i}$. The simulation parameters are: sampling frequency $f_s = 4\,\text{kHz}$, distance $r = 2.2\,\text{m}$, interpolation half width $S = 0.05 f_s$, and SNR $= 40\,\text{dB}$. (b) Average DOA estimation error plotted from 0 to 6 degrees.



(a)



(b)

Fig. 5. Simulation results: (a) average DOA estimation error in degrees versus the signal-to-noise ratio of the acoustic impulses. The simulation parameters are: sampling frequency $f_s = 4\,\text{kHz}$, distance $r = 2.2\,\text{m}$, interpolation factor $\frac{T}{T_i} = 100$, interpolation half width $S = 0.05 f_s$. (b) Average DOA estimation error plotted from 0 to 6 degrees.

In an anechoic environment, the acoustic impulse response between a loudspeaker and a microphone is a time-shifted impulse. At each angle θ, the acoustic impulse response at each microphone was measured using the Synchronized Swept-Sine method [14]. For this, an exponential sine sweep was sent through the loudspeaker with starting frequency $f_1 = 40\,\text{Hz}$, and final frequency $f_2 = 96\,\text{kHz}$, at a sampling rate of $f_s = 192\,\text{kHz}$. The duration of the sweep was 3 seconds. The acoustic impulse response between the loudspeaker and each of the microphones is obtained after applying an inverse filter [14] to the measured microphone signals.

By calculating the TDOA-based DOA estimate for every loudspeaker position, a comparison is made between the different interpolation techniques, as done similarly in Section V. The average DOA estimation error over all azimuth angles θ is compared for different cross-correlation interpolation methods, i.e., no interpolation, parabolic interpolation, Gaussian interpolation, and the proposed sinc interpolation.

This comparison is repeated for varying measurement parameters. First, the sampling frequency is varied by downsampling the original measurement between $f_s = 4\,\text{kHz} - 192\,\text{kHz}$. This downsampling is performed by first applying an order 30 FIR anti-aliasing filter with a Hamming window to the measured signals, after which the desired number of samples are decimated. The second comparison varies the

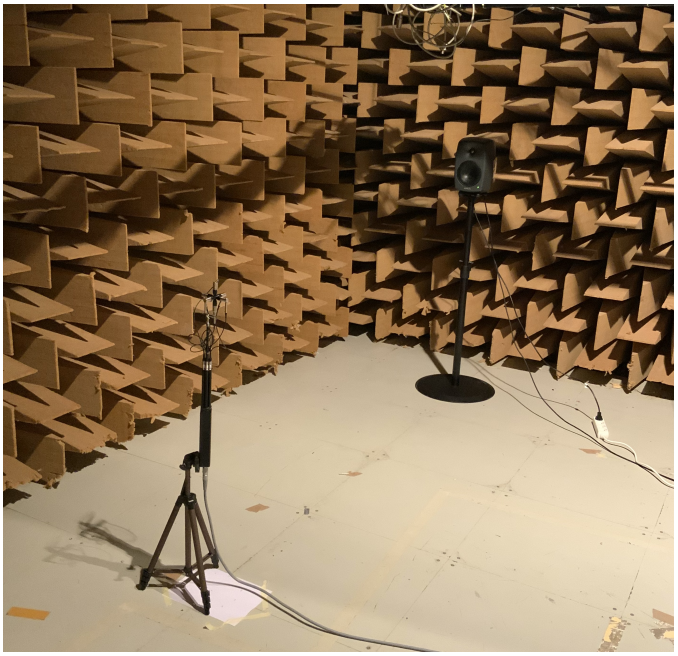


Fig. 6. The measurement setup consisting of the G.R.A.S. vector intensity probe type 50VI-1 microphone array and a Genelec 8030C loudspeaker positioned at 0° azimuth relative to the microphone array. The equipment is placed inside the semi-anechoic chamber at the KU Leuven Department of Physics.

interpolation period T_i at a measurement sampling rate of $f_s = 4\,\text{kHz}$.

B. Results

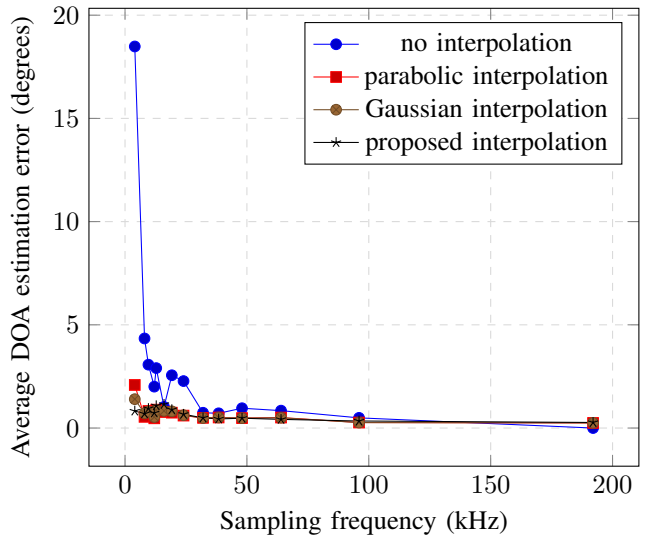
In this section, two comparisons are presented which validate the performance of the proposed interpolation method against the performance of existing interpolation methods for TDOA-based DOA estimation.

In the first comparison, shown in Fig. 7, the average DOA estimation error is plotted against the sampling frequency f_s for different interpolation methods. It can be seen that our proposed sinc interpolation method performs similarly to the parabolic and Gaussian interpolation for all sampling frequencies f_s.

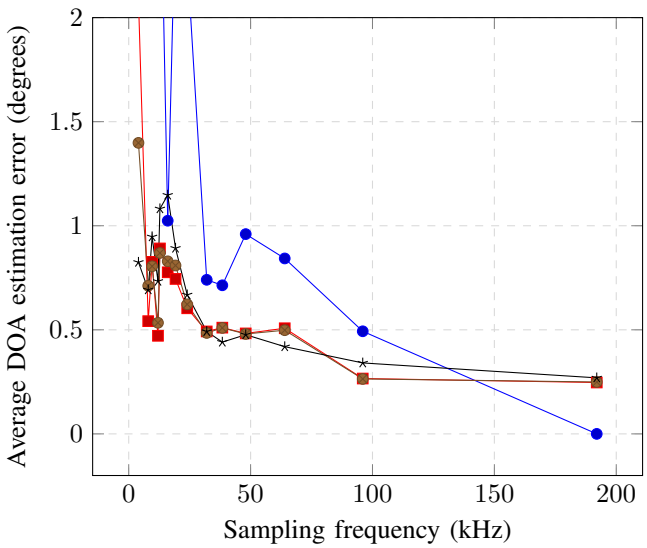
The second comparison is presented in Fig. 8. The average DOA estimation error is plotted against the interpolation factor i for all interpolation methods. We can observe that the measurements show a similar result as the simulations for this comparison. The average DOA estimation error decreases as the interpolation factor $\frac{T}{T_i}$ increases. We see that for high interpolation factors, the proposed interpolation method is able to outperform both parabolic and Gaussian interpolation methods.

VII. Conclusion

This paper proposed a novel method for solving the TDE resolution problem, which is directly rooted in the Nyquist-Shannon sampling theory. By fitting a continuous-time sinc function to the discrete-time cross-correlation function of two



(a)



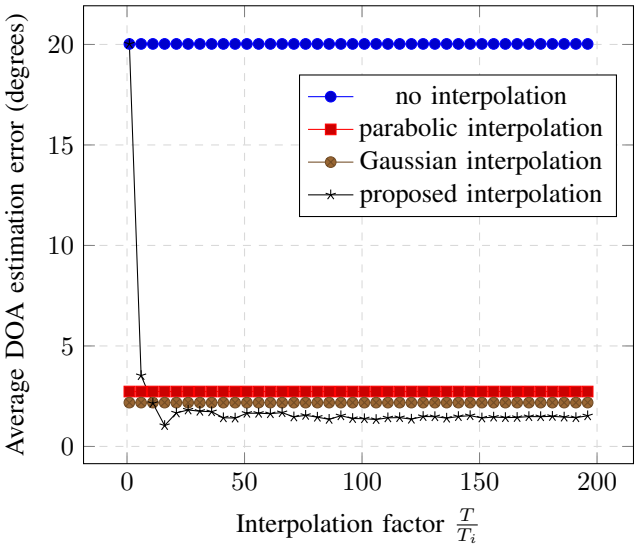
(b)

Fig. 7. Measurement results: (a) average DOA estimation error in degrees versus the sampling frequency f_s. The measurement parameters are distance $r = 2.2\,\text{m}$, interpolation factor $\frac{T}{T_i} = 100$, and interpolation half width $S = 0.05 f_s$. (b) Average DOA estimation error plotted from 0 to 2 degrees.

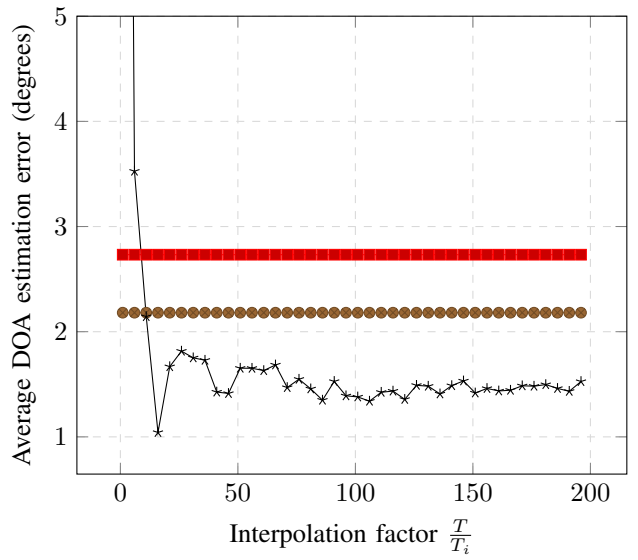
measured acoustic impulses in the free field, we can obtain a time resolution that is only a fraction of the sampling period.

The performance of the proposed interpolation scheme was compared with the performance of existing interpolation schemes in Time-Difference of Arrival based Direction of Arrival estimation, using simulations and measurements in a semi-anechoic chamber. These results show that the proposed interpolation scheme is a valid method for solving the TDE resolution problem and can outperform existing interpolation schemes.

One shortcoming of current interpolation schemes is that, at low signal-to-noise ratios, the interpolation schemes are unable to obtain an accurate Direction of Arrival estimate. This can be explained by the fact that current interpolation schemes are



(a)



(b)

Fig. 8. Measurement results: (a) average DOA estimation error in degrees versus the interpolation factor $\frac{T}{T_i}$. The measurement parameters are sampling frequency $f_s = 4\,\mathrm{kHz}$, distance $r = 2.2\,\mathrm{m}$, and interpolation half width $S = 0.05 f_s$. (b) Average DOA estimation error plotted from 0 to 5 degrees.

unable to model the noise of the measured signals.

In this paper, free-field conditions were assumed. In reverberant environments, the measured signals are typically split up in overlapping short-time frames before performing Time-Difference of Arrival based Direction of Arrival estimation on each frame. Future research involves investigating the impact of the proposed interpolation scheme on short-time frames.

References

[1] G. Carter, "Time delay estimation for passive sonar signal processing," IEEE Trans. Acoust., Speech, Signal Process., vol. 29, no. 3, pp. 463–470, Jun. 1981, doi: 10.1109/TASSP.1981.1163560.

[2] S. Tervo, J. Pätynen, and T. Lokki, "Acoustic Reflection Localization from Room Impulse Responses," Acta Acustica united with Acustica, vol. 98, no. 3, pp. 418–440, May 2012, doi: 10.3813/AAA.918527.

[3] J. Yli-Hietanen, K. Kalliojarvi, and J. Astola, "Low-complexity angle of arrival estimation of wideband signals using small arrays," in Proceedings of 8th Workshop on Statistical Signal and Array Processing, Corfu, Greece, 1996, pp. 109–112. doi: 10.1109/SSAP.1996.534832.

[4] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process., vol. 24, no. 4, pp. 320–327, Aug. 1976, doi: 10.1109/TASSP.1976.1162830.

[5] M. Azaria and D. Hertz, "Time delay estimation by generalized cross correlation methods," IEEE Trans. Acoust., Speech, Signal Process., vol. 32, no. 2, pp. 280–285, Apr. 1984, doi: 10.1109/TASSP.1984.1164314.

[6] Xiaoming Lai and H. Torp, "Interpolation methods for time-delay estimation using cross-correlation method for blood velocity measurement," IEEE Trans. Ultrason., Ferroelect., Freq. Contr., vol. 46, no. 2, pp. 277–290, Mar. 1999, doi: 10.1109/58.753016.

[7] Lei Zhang and Xiaolin Wu, "On Cross Correlation Based Discrete Time Delay Estimation," in Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., Philadelphia, Pennsylvania, USA, 2005, vol. 4, pp. 981–984. doi: 10.1109/ICASSP.2005.1416175.

[8] R. J. Marks, Introduction to Shannon sampling and interpolation theory. New York: Springer-Verlag, 1991.

[9] A. V. Oppenheim and R. W. Schafer, "Sampling of Continuous-Time signals" in Discrete-time signal processing," 3rd ed. Upper Saddle River: Pearson, 2010, pp. 153-273.

[10] Bo Qin, Heng Zhang, Qiang Fu and Yonghong Yan, "Subsample time delay estimation via improved GCC PHAT algorithm," 2008 9th International Conference on Signal Processing, 2008, pp. 2579-2582, doi: 10.1109/ICOSP.2008.4697676.

[11] T. Pirinen, "Confidence Scoring of Time Delay Based Direction of Arrival Estimates and a Generalization to Difference Quantities," Ph.D. thesis, Tampere University of Technology, 2009.

[12] T. I. Laakso, V. Valimaki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay [FIR/all pass filters design]," IEEE Signal Process. Mag., vol. 13, no. 1, pp. 30–60, Jan. 1996, doi: 10.1109/79.482137.

[13] V. Valimaki and A. Haghparast, "Fractional Delay Filter Design Based on Truncated Lagrange Interpolation," IEEE Signal Process. Lett., vol. 14, no. 11, pp. 816–819, Nov. 2007, doi: 10.1109/LSP.2007.898856.

[14] A. Novak, P. Lotton, and L. Simon, "Synchronized Swept-Sine: Theory, Application, and Implementation," J. Audio Eng. Soc., vol. 63, no. 10, pp. 786–798, Nov. 2015, doi: 10.17743/jaes.2015.0071.