ABSTRACT
The estimation of the time delay of arrival (TDOA) or the
direction of arrival (DOA) has been an important research
topic within the last few decades. In several applications, e.g.,
acoustic source localization or source separation, the estima-
tion of the delay or the direction is a significant processing
block. Most of the available algorithms estimate the delay in
the time domain. Therefore, the estimate is an integral mul-
tiple of the sampling period. Depending on the sampling rate
and other parameters, the resolution is restricted and possibly
not adequate for specific applications. To overcome the res-
olution problem, a new method based on the phase delay be-
tween two sensors and a line detection approach commonly
used in image processing is proposed. Experimental results
demonstrate the advantage of the introduced approach.

Index Terms— time delay estimation, direction of ar-
ival, Radon transform, resolution problem, generalized
cross-correlation

1. INTRODUCTION
The estimation of the delay or the direction of arrival is used
in several signal processing applications, wherein the require-
ment on the resolution can be high under certain conditions.
A prime example is the suppression of noise in hearing de-
vices, which is typically based on a direction-sensitive filter.
After estimating the incident angle from the time delay, the
source of interference can be suppressed. To calculate the de-
lay, well-known methods for TDOA estimation like the gen-
eralized cross-correlation (GCC) or the adaptive eigenvalue
decomposition (AEVD) are used [1]. Since the resolution for
methods in the time domain depends on the sampling rate and
the distance between the two sensors, which is shown in (3),
the accuracy of especially compact sensor setups is low. To
overcome this problem, a new method is proposed in this pa-
ter to estimate the delay or the direction of arrival.

The paper is structured as follows. At first, some funda-
mentals are presented and a short overview of the generalized
cross-correlation is given. In Sec. 3, the new approach is in-
troduced and compared to the GCC. Experiments with real
world data show the advantages of our method compared to
the GCC for direction of arrival estimation. At last, a conclu-
sion is drawn.

2. FUNDAMENTALS
2.1. Signal model
Sound propagation in a reverberant environment can be de-
scribed by multipath propagation. The signal at the j-th mi-
crophone is described as

\[ x_j(t) = \sum_{l=1}^{L} a_j^l s(t - t_j^l), \]

where \( s(t) \) corresponds to the source signal and \( L \) to the num-
ber of the considered propagation paths. The coefficient \( a_j^l \)
represents the attenuation for the \( l \)-th path between the source
and the \( j \)-th sensor. The coefficients of the direct paths (\( l = 1 \))
have the largest values. To estimate the DOA, a setup with
two microphones (\( j = 1, 2 \)) is chosen. The distance between
the microphones \( d \) is clearly shorter than the distances be-
tween source and sensors (\( d_1, d_2 \)). Therefore, the far-field
assumption for sound propagation is valid, where an incident
wave is modeled as a plane wavefront (Fig. 1). For this rea-

Fig. 1. Sound propagation (\( d \ll d_1, d_2 \)).

\[ \Delta t = \frac{d_1 - d_2}{c_{\text{air}}} = \frac{\Delta d}{c_{\text{air}}}, \]
where \( c_{\text{air}} \approx 343 \text{ m/s} \) is the speed of sound in air. Knowing \( \Delta t \), the DOA depends furthermore on the distance between the sensors:

\[
\theta = \arccos \left( \frac{\Delta t \cdot c_{\text{air}}}{d} \right). \tag{3}
\]

The angular resolution can be computed by considering \( \Delta t \) as integral multiple of the sampling time \( T_S = \frac{1}{f_S} \), where \( f_S \) is the sampling rate.

Due to the fact that the phase difference/phase delay between the sensors is used for estimation, the relation between the time delay and the frequency-dependent phase difference is shown:

\[
\Delta \varphi_D(f) = 2\pi f \Delta t, \tag{4}
\]

where the index \( D \) indicates the propagation along the direct path.

### 2.2. Generalized cross-correlation

The generalized cross-correlation is a widely used approach for time delay estimation (TDE). The GCC is based on the standard cross-correlation, that can be calculated either in the time domain

\[
R_{x_1,x_2}(\tau) = \int_{-\infty}^{\infty} x_1(t) x_2(t + \tau) \, dt \tag{5}
\]

or in the frequency domain

\[
R_{x_1,x_2}(\tau) = \int_{-\infty}^{\infty} S_{x_1,x_2}(f) \, e^{j2\pi f \tau} \, df, \tag{6}
\]

where \( S_{x_1,x_2}(f) = X_1(f) \cdot X_2^*(f) \) is the cross-power spectral density. In the frequency domain, the GCC is nearly equal to the cross correlation, except for a prefactor \( W(f) \) in the integrand, which leads to

\[
R_{x_1,x_2}^{\text{GCC}}(\tau) = \int_{-\infty}^{\infty} W(f) S_{x_1,x_2}(f) \, e^{j2\pi f \tau} \, df. \tag{7}
\]

### 3. TDE IN THE TIME-FREQUENCY DOMAIN

#### 3.1. Statistical consideration of the phase difference

As motivation for the new approach, the distributions of the phase differences \( \Delta \varphi(f) \) are considered in detail. An adequate description of the sensor signals is achieved by performing an \( N_T \)-point short-time Fourier transform (STFT). On the one hand, this allows a frequency-dependent representation of the signals. On the other hand, the time-varying structure of the speech signals is also considered. The time-frequency coefficients \( X_i(n, f_k) \) describe the \( i \)-th signal at a specific time \( n \) and a frequency \( f_k \), where \( f_k \in \left\{ 0, \frac{1}{N_T}, \ldots, \frac{N_T-1}{N_T} f_S \right\} \).

Therefore, the phase difference is calculated between the phase values of the related time-frequency coefficients:

\[
\Delta \varphi(n, f_k) = \arg \left[ X_1(n, f_k) \cdot X_2^*(n, f_k) \right]. \tag{8}
\]

To gain a statistical interpretation of the phase difference, a normalized histogram \( h(\Delta \varphi_H(f_k)) \) for every frequency \( f_k \) with \( N_H \) sections is calculated based on all available measures \( \Delta \varphi(n, f_k) \). As shown in Fig. 2(a), the phase differences are distributed around a maximum associated with the direct path. The variations are caused by multipath propagation, since the signal of the direct path is superimposed by reflections with an arbitrary phase difference and a lower magnitude. The histograms are calculated for the \( N_F \) possi-

![Fig. 2. Statistical interpretation of the phase difference for one source in reverberant environment (RT_{60} = 250 \text{ ms}).](image)

The obvious linear structure in the image \( H(f_k, \Delta \varphi_H) \) offers a new possibility to estimate the TDOA or the DOA. Based on a line detection approach used in image processing [2], a new method to calculate the direction of arrival is derived. Due to the fact that the phase difference of the direct path \( \Delta \varphi_D(f_k) \) mainly defines the underlying structure, the DOA can be estimated with a line detection approach applied to \( H(f_k, \Delta \varphi_H) \) followed by some postprocessing steps to consider special properties of the data.

The ordinary Radon transform \( R_{s,\theta}\{\rho\} \) is defined as a two-dimensional integration of the density distribution \( \rho(x, y) \):

\[
R_{s,\theta}\{\rho\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x, y) \delta(s - x \cos \theta - y \sin \theta) \, dz \, dy, \tag{9}
\]

where the second multiplicant describes a straight Dirac line defined by the distance \( s \) to the origin and the angle \( \theta \) between the straight line and the negative \( y \)-axis [3]. Based on this general description, a modified transform can be derived from the following items.

---

2717
1. The indefiniteness of $2\pi$ causes a discontinuity in $H(f_k, \Delta \varphi_H)$ (see Fig. 2 (b) at $f = 0.35f_S$). Nevertheless, both lines depend on the same source.

2. Lines describing TDOA pass either the point $(f_k = 0, \Delta \varphi_H = 0)$ due to (4) or points at $f_k = 0$ and $\Delta \varphi_H = m \cdot 2\pi$ due to the indefiniteness of $2\pi$.

At first, (9) must be replaced by a discrete implementation. The discrete Radon transform is designed to work on images. To easily apply existing realizations of the transform to our problem, $H(f_k, \Delta \varphi_H)$ is considered as an image with $N_H$ rows and $N_F$ columns. To calculate the discrete Radon transform of $H(f_k, \Delta \varphi_H)$, the equation

$$R_{s, \theta} \{H\} = \sum_{k=1}^{N_F} \sum_{n_H=1}^{N_H} H(f_k, \Delta \varphi_H) \delta(s - n_H \cos \theta - k \sin \theta)$$

(10)

is computed for an arbitrary set of angles. $R_{s, \theta} \{H\}$ represents the intensity of the projection of $H(f_k, \Delta \varphi_H)$ on a straight line with the parameters $s$ and $\theta$, so high intensities indicate the presence of a line. Below, the origin is chosen according to Fig. 3 and $s$ is the distance in pixels. The Radon transform has the form (with an angular resolution of $1^\circ$) for the previous histogram image is illustrated in Fig. 4. The distinctive peak in the Radon space is due to the upper line structure, whereas the second line is hardly identifiable. Regarding the items 1

Thus, the related distances

$$s(\theta, n_\delta) = \sin(\theta - 90^\circ) \cdot \left( \frac{N_F}{2} - n_\delta \cdot \frac{N_H}{\tan(\theta - 90^\circ)} \right)$$

(12)

with $n_\delta = 0, \ldots, N_\delta(\theta)$ are computed. The values of $s$ with respect to $\theta$ are shown in Fig. 4 (b). The resulting intensity is calculated according to

$$I(\theta) = \sum_{n_\delta=0}^{N_\delta(\theta)} R_{s(\theta, n_\delta), \theta} \{H\}$$

(13)

An efficient implementation of (13) can be achieved with the definition of masks describing the straight lines of interest. An example of such a mask is depicted in Fig. 5 (b). The masks for all desired $\theta$ must be calculated in advance. The intensity for a specific angle can be computed according to

$$I(\theta) = \sum_{k=1}^{N_F} \sum_{n_H=1}^{N_H} H(f_k, \Delta \varphi_H) \circ M_{\theta}(f_k, \Delta \varphi_H),$$

(14)

where ‘$\circ$’ indicates a pointwise multiplication. This approach is less flexible, but computationally more efficient.

### 3.3. Relation between DOA and Radon angle

The angle $\theta$ is not equal to the DOA, but the corresponding angle can be calculated as

$$\delta = \arccos \left( \frac{\tan(\theta - 90^\circ) \cdot N_F \cdot c_{\text{air}}}{N_H \cdot f_S \cdot d} \right).$$

(15)

Depending on the requirements on the DOA estimation, the relevant Radon angles can be calculated in advance. These values can be extracted from the Radon space or equivalent masks can be prepared.

### 3.4. Comparison with the GCC

In this section, the GCC is compared to the proposed method in order to achieve a better understanding. The prefilter is chosen as $W_{\text{PHAT}} = 1/|S_{x_1, x_2}(f)|$. It has been observed that this way the algorithm is more immune to reverberation [1]. Under this condition, the GCC can be simplified to

$$R_{x_1, x_2}^{\text{GCC}}(\tau) = \int_{-\infty}^{\infty} \frac{1}{|S_{x_1, x_2}(f)|} S_{x_1, x_2}(f) e^{j 2\pi f \tau} df$$

$$= \int_{-\infty}^{\infty} e^{j [\arg(S_{x_1, x_2}(f))] \tau} e^{j 2\pi f \tau} df$$

The argument of the cross-power spectral density is equal to the phase difference between the two sensors as shown in (8). The discrete realization of the previous equation is equal to

$$R_{x_1, x_2}^{\text{GCC}}(n_T) = \sum_{k=0}^{N_F} e^{j \Delta \varphi(f_k)} e^{j 2\pi n_T \frac{\tau}{T}}.$$
Similar to Sec. 3.1, the signals are transformed to the frequency domain using an $N_T$-point Fourier transform. The time delay can only be estimated as a multiple of the sampling time. An interpretation akin to the Radon transform can be derived by calculating the values of a matrix

$$M_{n_T}(f_k, \Delta \varphi_H) = e^{j(\Delta \varphi(f_k) + 2 \pi n_T f_T)}$$

for $\Delta \varphi \in [-\pi, \pi]$ for all positive frequencies and a fixed delay $n_T$. This matrix (Fig. 5(a)) shows a similar line structure as the Radon mask for the same delay. These GCC masks $M_{n_T}(f_k, \Delta \varphi_H)$ can also be used to estimate the delay of arrival according to

$$I^*(n_T) = \sum_{k=1}^{N_F} \sum_{n_H=1}^{N_H} H(f_k, \Delta \varphi_H) \circ M_{n_T}(f_k, \Delta \varphi_H).$$

For an exact implementation of (16), the image $H(f_k, \Delta \varphi)$ must be calculated from the values of one realization of the Fourier transform.

### 4. RESULTS

The proposed approach has been tested with data from SiSEC [4]. Two live recordings ($f_S = 16$ kHz) in a reverberant room ($RT_{60} = 250$ ms) with two different microphone spacing was used. The parameters of the proposed algorithm were chosen to $N_H = 128$, $N_F = 128$, $N_T = 256$ and 180 different angles $\theta$ corresponding to an angular resolution of the incident angle of 1°. To simplify the comparison to the GCC, the mean values have been removed from the results and the maximum was scaled to 1.

A setup with a large sensor distance ($d = 1.00$ m) has been used to compare the method to the GCC. In the case of large sensor distances, the GCC has a high resolution. As can be seen in Fig. 6, the results agree approximately. For smaller sensor setups ($d = 0.05$ m), the advantage of the proposed algorithm can be seen clearly (Fig. 7). The resolution is not limited by the sampling time and the incident angle can be estimated nearly exactly.

### 5. CONCLUSIONS

In this paper, an alternative approach to DOA estimation has been presented and the relation to the GCC has been shown. Due to the fact that the resolution is not restricted to the sample rate, a higher accuracy can be achieved. Therefore, the proposed method could be used in application with compact sensor setups for direction/delay estimation. The applicability for more reverberant rooms must be shown, but due to the apparent resemblance to GCC, similar results as in [1] are expected. The usage of modified (less sharp) masks could be considered. In the context of GCC, a modified approach with non-integer values of $n_T$ should also be evaluated.

### 6. REFERENCES


