2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo) | 979-8-3503-9882-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/CINTI-MACRo57952.2022.10029405

Acoustic Moving Source Localization using Sparse Time Difference of Arrival Measurements

Gyula Simon
Óbuda University
Alba Regia Technical Faculty
Székesfehérvár, Hungary
simon.gyula@amk.uni-obuda.hu

***Abstract*—Time difference of arrival (TDoA) measurements are common in distributed localization systems, where events are emitted by the tracked source and the detected arrival times at the sensors, deployed at known locations, are used to compute the source location. Since erroneous measurements are common, e.g. due to non-line-of-sight conditions, the sensor fusion must be robust against such outlier measurements. In this paper the consensus-based TDoA fusion, which was proven very successful in various acoustic localization systems, is extended for the case when the source is moving and the event detections at the sensors are not reliable, thus sparse. The derivation of the extended consensus function will be introduced, and the performance is illustrated through simulation examples, showing accuracy around 10cm in realistic scenarios.**

Keywords—localization, sensor fusion, consistency, consensus, fault tolerance

I. Introduction

Localization is a key enabling service in many applications, e.g. indoor navigation [1][2], structural monitoring [3][4], or field biology [5], just to name a few. Specifically, acoustic source location has widespread application areas, too, e.g. speaker tracking [6], sound enhancement [7], structural health monitoring [3], and sniper [8] or animal [5] localization.

Beamforming techniques utilize the correlation between detected audio streams at the sensors [9], thus wide bandwidth is required to transmit data from the sensors to the sensor fusion unit. On the other hand, event-based solutions operate on specific, well defined events [10], where the detection of the events is performed at the sensor and only low amount of information (typically time of detection and possibly a short event descriptor) is transmitted between sensors and the fusion unit. The latter approach allows the creation of massively distributed localization systems containing inexpensive sensor elements, e.g. [8][11].

In event-based localization systems the sensors, deployed at known locations, detect the arrival time of the event. The event may have various forms, e.g. the sound of crack in a material [3], a muzzle blast of a gun [8], or specific patterns in human speech [12]. Since the exact emission time is not known in most practical applications, the times of flight of the signal from the source to the sensors are not known. Instead, the detection times of the events (or they differences, which are equivalent to differences of the times of flight) maybe used in the TDoA method, along with the positions of the sensors, to estimate the source position [13]. The solution is geometrically straightforward: each sensor pairs provides a time difference, or using the speed of sound, a distance, which defines a hyperboloid surface, the intersections of which is the source position. In 2 dimensions (2D), theoretically three (non-colinear) sensors are enough to determine the source position, while in 3 dimensions (3D) four (non-coplanar) sensors are required. The problem can also be formulated as a linear equation, but in this case four and five sensors are required in 2D and 3D, respectively [13]. When more sensor measurements are available, least squares solutions can provide the estimate [14].

Both the geometric and analytic solutions assume accurate, but possibly noisy measurements. Thus, these approaches tolerate small measurement inaccuracies, but outliers may cause significant estimation errors. However, in practical cases outliers are very common: either sensor errors, bad detections, or non-line-of-sight conditions can cause large measurement errors. Various approaches have been proposed to fight outliers: the most common solution is the RANSAC algorithm, which selects good measurements and eliminates outliers through numerous random trials [15],[16]. A completely different approach was proposed in [11], using consistency functions, where the estimation process automatically makes a distinction between good and bad measurements, providing extremely robust estimates.

The consistency function-based approach was successful in applications where the source was static and each emitted event was detected by enough sensors to provide an estimate. In this paper consistency functions will be extended in two directions:

- It is not required that each emitted event is detected by enough sensors to provide an estimate. Rather, multiple events are emitted by the source and the much weaker requirement is the following: (a) each event is detected by at least a predetermined number of sensors (not necessarily enough to compute the estimate), and (b) the total number of detections allows the localization.
- The source may move between event emissions.

The outline of the paper is the following: Section II reviews the consistency function and the corresponding TDoA source localization method. In Section III the extended consistency function will be introduced. Section IV illustrates the operation of the novel localization scheme through simulated examples. Section V concludes the paper.

II. Event-based TDoA Localization

In this section the consistency-based TDoA localization will be reviewed based on [8] and [11], focusing on 2D localization only.

Let the source location be the unknown position $\boldsymbol{p_s} = (x_s, y_s)$, where the source emits an event at unknown time t_s. The sensors are $S_i,\ i = 1, 2, \dots, N$, where N is the number of sensors and the $\boldsymbol{p_i} = (x_i, y_i)$ location of sensor S_i is known. The sensors measure the time of incoming event at $t_i, i = 1, 2, \dots, N$. (Notice: the clocks of the sensors are synchronized, thus a common time-base is used.) The measurement scenario

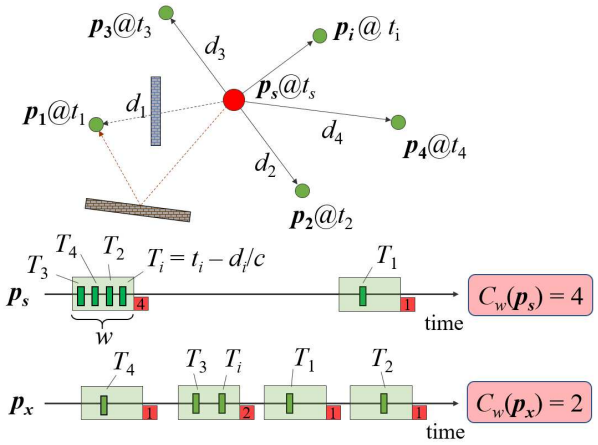

Fig. 1. The consistency function, evaluated in two points. For true source position p_s the values T_i are close to each other, except for the outlier T_1, resulting $c_w(p_s) = 4$. In another point p_x the values T_i are scattered, resulting lower $c_w(p_x) = 2$.

is shown in Fig. 1 with five sensors, one of which (S_1) has no line of sight and thus provides false (inconsistent) measurement.

Let us assume that the source is at position $\boldsymbol{p} = (x, y)$. In this case the distance between the source and sensor S_i can be calculated as follows:

$$d_i(x,y) = \sqrt{(x_i - x)^2 + (y_i - y)^2}. \tag{1}$$

The time of flight of the signal between the source and S_i, given the speed c of the sound, is

$$\tau_i = \frac{d_i(x,y)}{c}. \tag{2}$$

Let us define function $T_i(x,y)$ as follows:

$$T_i(x,y) = t_i - \tau_i = t_i - \frac{\sqrt{(x_i-x)^2+(y_i-y)^2}}{c}. \tag{3}$$

The physical meaning of $T_i(x,y)$ is the following: based on measurement t_i and sensor position p_i, sensor S_i estimates the emission time of the event as $T_i(x,y)$, *if* the source is at position $\boldsymbol{p} = (x, y)$. Now let us suppose that the measurements are accurate (there is no measurement error in t_i and p_i). In this case

$$T_i(x_s, y_s) = t_s, \ \forall i, \tag{4}$$

thus all sensors estimate the same emission time. If the measurements contain noise, the estimated values of $T_i(x_s, y_s)$ will not be exactly the same but will be close to each other in one large group (the "closeness" depending on the measurement noise, and the size of the group is N), as shown on the upper timeline of Fig. 1. On the other hand, if the T_i functions are evaluated at a random position $\boldsymbol{p_x}$ (apart from the true source location), each sensor will possibly provide a different emission time estimate, thus the estimates will be scattered in several smaller groups, as illustrated on the lower timeline of Fig. 1. Based on this observation the consistency function can be defined. First, let us define window width w and the corresponding function $K_w(x,y,t)$, as follows:

$$K_w(x,y,t) = \underset{i}{count}\left(|T_i(x,y) - t| < \frac{w}{2}\right) \tag{5}$$

The width w determines the required 'closeness' of the estimates: estimates closer than w are treated as one group. Function $K_w(x,y,t)$ gives the number of emission time measurements that are between $t - w/2$ and $t + w/2$, i.e. the size of a group around t. Using K_w, the consistency function C_w is defined as the size of the largest group:

$$C_w(x,y) = \max_{t \in \mathbb{R}} K_w(x,y,t), \tag{6}$$

Thus, consistency function C_w gives the maximum number of sensor measurements that support the hypothesis (with maximum error of $w/2$) that the source is in position (x, y). The position estimate is at the maximum of the consistency function:

$$(\hat{x}_s, \hat{y}_s) = \underset{(x,y)}{\operatorname{argmax}}\, C_w(x,y). \tag{7}$$

Notice that the value of t, for which the expression in (6) is maximal, is the estimate of the emission time.

If the maximum of C_w is not one point but an area, the location estimate is the center of the area.

If all of the sensors provide correct measurements (with small noise only, the allowed noise level determined by w) then the maximum value of C_w is N at the true position (x_s, y_s). If some (but not too many) outlier measurements are also present, the outliers will not contribute to the maximum value. In the illustration of Fig. 1, sensor S_1 provided an outlier measurement, thus the emission time estimate T_1 is not correct, being far from the group of correct estimates. In the example of Fig. 1, the value of the consistency function at $\boldsymbol{p_s}$ is 4, instead of the maximal possible value of 5. Nevertheless, the maximum of C_w will still be at the correct position (x_s, y_s). The robustness of the consistency function is shown by the fact that the number of outliers can be even more than the number of correct measurements, provided that the number of correct measurements are still high enough [11].

Notice that the consistency function of (6) assumes that one event is emitted from a source, and this event is detected by enough sensors. In case the event detection is not reliable (sparse), multiple events may be utilized to gather enough number of measurements, as proposed in [12]. This approach, however, assumes that the source is static. In the following section an extension will be proposed, which allows moving sources.

III. Consistency Function for Moving Sparse Detections

A. General Concept

Let the emission model be the following: The source emits events $e^1, e^2, e^3, \ldots$ at discrete time instants $t_s^1, t_s^2, t_s^3, \ldots$, where the emissions are not closer than w, i.e. $t_s^k + \mathrm{w} \leq t_s^{k+1}, \forall k$. The measurement is performed in rounds, the length of each round being T_R. The processing is performed at the end of each round, thus now we will focus on a single round.

Each sensor S_i starts listening at the beginning of the round and may detect some of the events during the round, providing measurement times t_i^j. (Notice that measurement t_i^j

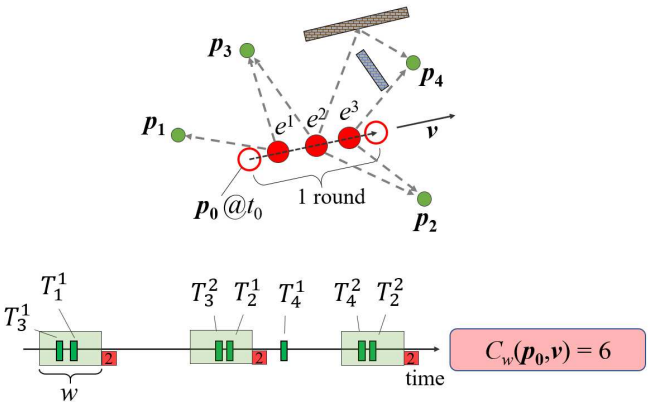

Fig. 2. The computation of the consistency function in case of a moving source. The source is at $\boldsymbol{p_0}$ at the beginning of the round and its speed vector is $\boldsymbol{v}$. The events e_i, emitted along the track. Sensors that detect an event are connected to the event by a dashed arrow. The estimated emission times of each event are gathered in groups, when calculated in the true source position.

does not provide information on which of the events was detected.) Each measurement can be correct (with small measurement noise) or an outlier. At the end of each round, measurements of each sensor are collected by the fusion unit and the estimation is performed.

The concept is illustrated in Fig. 2. In the example the source is at position $\boldsymbol{p_0}$ at the beginning of the measurement round and emits three events (e^1, e^2, e^3) during the round. Event e^1 is sensed by S_1 and S_3, which produce emission time estimates T_1^1 and T_3^1, respectively. Event e^2 is sensed by S_2, S_3 and S_4, the corresponding estimated emission times being T_2^1, T_3^2, and T_4^1. Notice that the measurement of S_4 is an outlier, thus emission time estimate T_4^1 is incorrect. Finally, event e^3 is sensed by S_2 and S_4, with corresponding emission time estimates being T_2^2, and T_4^2. The emission time estimates are shown on the timeline of Fig. 2, where the groups, corresponding to each event, are clearly visible. Notice, that the estimate T_4^1, corresponding to the outlier measurement, does not belong to either group. Also notice that the event detections were sparse (i.e. each event was sensed by only a few sensors), which would have not necessary been enough to make a reliable estimate. The collection of the detections during the measurement round, however, will allow the calculation of the consistency function.

The calculation of the emission time estimates will be detailed in Section III.B, while the consistency function for moving source will be introduced in Section III.C.

B. *Emission Time Estimation*

The source may move while emitting events. Let us assume that the movement can be reasonably approximated by piecewise linear approximation in such a way that in each round of length T_R the movement is approximated with constant speed. Let the state of the source described by vector $(\boldsymbol{p_0}, \boldsymbol{v})$, where $\boldsymbol{p_0} = (x_0, y_0)$ is the position of the source at the beginning of the measurement round at time instant t_0, while $\boldsymbol{v} = (v_x, v_y)$ is the speed vector of the source during the measurement round.

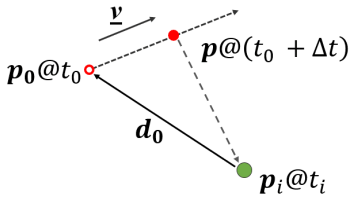

Fig. 3. Calculation of the emission time of the moving source. The initial position at t_0 is $\boldsymbol{p_0}$, while at time instant $t_0 + \Delta t$ the position is $\boldsymbol{p}$.

Now let us calculate the estimated emission time for a source state (x_0, y_0, v_x, v_y) if the sensor's position is $\boldsymbol{p_i} = (x_i, y_i)$, as shown in Fig.3. Let the source emit an event at time instant $t_0 + \Delta t$, which is detected by the sensor at time instant t_i. The source's position at $t_0 + \Delta t$ is $\boldsymbol{p_0} + \boldsymbol{v}\Delta t$, thus the distance d between the source and sensor at the time of emission is

$$d = |\boldsymbol{p_0} + \boldsymbol{v}\Delta t - \boldsymbol{p_i}| = |\boldsymbol{d_0} + \boldsymbol{v}\Delta t|, \tag{8}$$

where $\boldsymbol{d_0} = (d_x, d_y) = \boldsymbol{p_0} - \boldsymbol{p_i}$ is the vector between the sensor and the source at the beginning of the round, as shown in Fig. 3. The emission time estimate T_i can be expressed, as follows:

$$T_i = t_i - d/c, \tag{9}$$

where c is the speed of sound. Using (8), (9), and the fact, that the emission time T_i is $t_0 + \Delta t$, it follows that

$$t_i - (t_0 + \Delta t) = \frac{|\boldsymbol{d_0} + \boldsymbol{v}\Delta t|}{c}. \tag{10}$$

From (10), the following quadratic equation of Δt follows:

$$\begin{aligned} &\Delta t^2\left(c^2 - v_x^2 - v_y^2\right) \\ &+2\Delta t\left((t_i - t_0)^2 c^2 - v_x d_x - v_y d_y\right) \\ &+c^2(t_i - t_0)^2 - d_x^2 - d_y^2 = 0. \end{aligned} \tag{11}$$

The solution of (11) gives the estimate $\widehat{\Delta t}$ of Δt, from which the emission time estimate T_i is the following:

$$T_i = t_0 + \widehat{\Delta t}. \tag{12}$$

Since (11) is quadratic, there can be zero, one, or two solutions. In any case, the solutions are checked against trivial physical constraints, and only those solutions are accepted that satisfy the following inequities:

$$t_0 - w < T_i < t_i + w\,. \tag{13}$$

The left-hand side inequity assures that the event was emitted after the start of the measurement round, while the right-hand side enforces that the event was emitted before it was detected. In both constraints the term w is used to allow measurement errors.

C. *Consistency Function*

The consistency function can be calculated for a given $(\boldsymbol{p_0}, \boldsymbol{v})$ hypothesis, as follows:

Step 1. At the end of the measurement round the measurements t_i^j are collected.

Step 2. For each measurement the corresponding emission time T_i^j is estimated, using the method described in Section III.B, and the emission time estimates are

added to the solution pool. A sorted solution pool is illustrated on the timeline of Fig. 2.

Step 3. On the sorted solution pool such eligible groups are searched, which have size of minimum Λ, starting from the lowest value in the solution pool, until the highest value. In the illustration of Fig. 2, for sake of simplicity, $\Lambda = 2$, but in real cases higher values (at least 3) should be used. In the illustration three satisfactory groups were found, each of size 2.

Step 4. The value of the consistency function is calculated as the sum of the eligible group sizes. In the illustration of Fig. 2, the value of the consistency function is $C_w = 2 + 2 + 2 = 6$.

D. *Computing considerations*

The calculation method, described by Steps 1-4 of Section III.C, provides the value of the consistency function for one particular input parameter $(\boldsymbol{p_0}, \boldsymbol{v}) = (x_0, y_0, v_x, v_y)$. For localization purposes the point $(\boldsymbol{p_0}, \boldsymbol{v})_{max}$ is searched where the consistency function has its maximal value. Since the parameter space is now four-dimensional (as opposed to the two-dimensional search space of the static consistency function), brute force computation is not feasible for the full parameter space. Rather, search may be performed only around the previously estimated position. The new search area can be determined from the maximum speed of the object, its maximum acceleration, and the length T_R of the measurement round.

To start the tracking, either a rough initial estimation of the object's location must be known, or it must be ensured that the source is in a static position. In the latter case the initial search space is only two-dimensional, and even a full search is a feasible solution.

IV. Simulation Examples

The performance of the proposed localization method is examined through simulation examples. The test setup with 6 sensors $(S_1, S_2, \dots, S_6)$ is shown in Fig. 4. The track of the source has for segments, as shown in Fig 4. In the first segment T0 the source is standing at position (5,5), then accelerates in a linear path during T1, and then moves with constant speed during T2. During track segment T3 both the speed and the direction of the movement is changing.

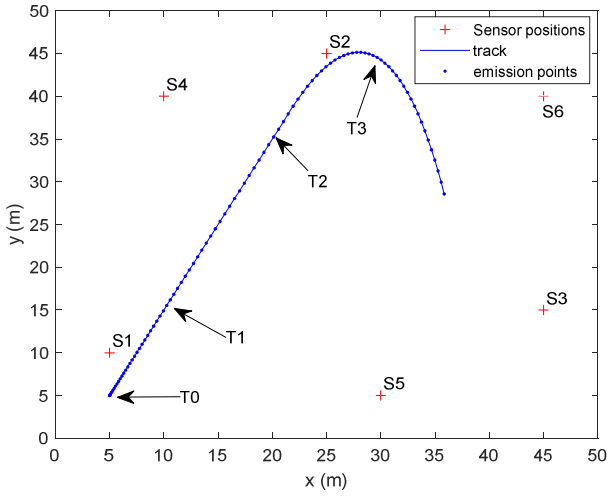

Fig. 4. Test setup with sensor positions, source track, and emission points

In the simulation the source emits events in every 300ms along its trajectory, thus $t_s^j = j \cdot 0.3s$; the positions of the emissions are also shown in Fig. 4. Notice that the temporarily equidistant emission is not necessary, it was chosen only to simplify the presentation.

In the simulation ideal measurement times were simulated for each sensor S_i, as follows:

$$t_{i,ideal}^j = t_s^j + d_i(x_s, y_s), \tag{14}$$

where $d_i(x, y)$ is defined in (1). For each ideal measurement value a noise component n_i^j, simulating measurement noise, with normal distribution of $n_i^j = N(0, \sigma_{noise})$ was added:

$$t_i^j = t_{i,ideal}^j + n_i^j. \tag{15}$$

In each measurement, out of the 6 sensors, $N_{outlier}$ sensors produced outlier measurements, simulated with noise component $n_i^j = N(0, \sigma_{outlier})$. The sensors, producing outliers, were chosen randomly in each measurement instant.

The first experiments were performed with $\sigma_{noise} = 0.02m/c$, i.e. the time measurement error was equivalent to $2cm$ of distance error. The speed of sound was $c = 340m/s$. For outliers noise level $\sigma_{outlier} = 10m/c$ was utilized. The search window width was $w = 0.1m/c$, the count parameter was $\Lambda = 3$, and the length of the measurement round was $T_R = 1.2s$. The number of outliers changed between 0 and 3. An example result is shown in Fig. 5 for $N_{outlier} = 3$. The figure shows the estimated speed, orientation, and the position error. The effect of the piece-wise linear approximation is

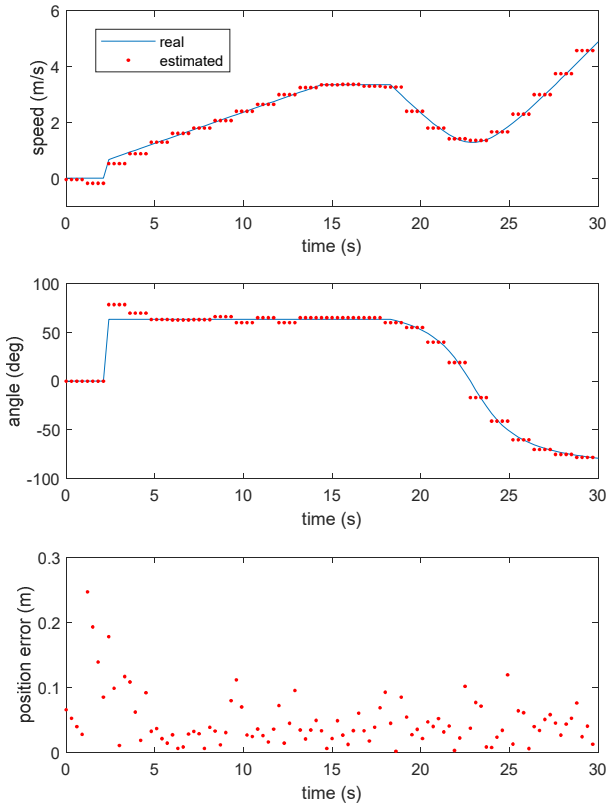

Fig. 5. Speed and orientition estimates, and the absolute position error for $N_{outlier} = 3$.

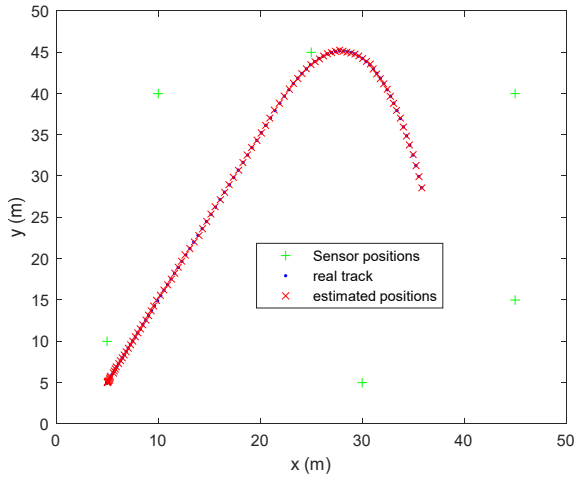

Fig. 4. Tracking result for $N_{outlier} = 3$.

clearly visible; since in each measurement round four measurements were present, each linear segment contains four points. (Notice that in real situations the number of measurements in a round would probably vary in time.)

The real and estimated positions at the times of event emissions are shown in Fig 6. Notice that despite the high number of outliers (half of the measurements were bad) the tracking results are quite good.

The RMS value of the position error, as a function of the number $N_{outlier}$ of outliers, is shown in Table I. For this noise level ($\sigma_{noise} = 0.02m/c$) and for small number of outliers the error is around 4.8cm, which is close the size of the search grid (5cm) in the experiments. As the number of outliers increase the estimation error also increases, but even for $N_{outlier} = 3$ the error level of 6.3 cm is remarkably good. The slight increase in error is due to the decrease of the number of good measurements.

TABLE I. RMS POSITION ERROR VS. NUMBER OF OUTLIERS

$N_{outlier}$	RMS position error
0	4.8 cm
1	4.7 cm
2	5.4 cm
3	6.3 cm

The RMS position error, as a function of the measurement noise, is shown in Table II. In the experiment $N_{outlier} = 1$ was constant and the noise level σ_{noise} was varied between $0.01m/c$ and $0.3m/c$. The Table shows the noise level in distance equivalent. For low noise levels the error of the discrete search grid dominates (around 5cm). For higher noise levels the position error grows almost proportionally with the measurement noise. The actual value, however, may greatly depend the geometric dilution of precision (GDOP) [17].

TABLE II. RMS POSITION ERROR VS. MEASUREMENT NOISE

σ_{noise}	RMS position error
1 cm	4.0 cm
2 cm	4.7 cm
5 cm	8.7 cm
10 cm	20.7 cm
20 cm	40.8 cm
30 cm	50.5 cm

V. CONCLUSIONS

A novel consistency function-base localization method was introduced, which can be used when the source is moving and the detections are sparse. For the calculation of the consistency function, a movement model is used: it is assumed that the movement can be reasonably modelled by a piecewise linear model with constant speed during the time of one (short) measurement round.

The performance of the method was illustrated by simulation experiments. The experiments showed that the method is indeed highly tolerant towards outliers, and the position estimation error is proportional to the measurement noise. In the realistic simulation experiments the average localization error was between 5cm and 50cm, depending on the measurement noise level.

REFERENCES

[1] M. Štancel, J. Hurtuk, M. Hulič, J. Červeňák, "Indoor Atlas Service as a Tool for Building an Interior Navigation System," Acta Polytechnica Hungarica vol. 18, no. 9, pp. 87-110, 2021.

[2] T. T. Mac, C. Copot, R. Cajo, "Towards an Autonomous Landing System in Presence of Uncertain Obstacles in Indoor Environments," Acta Polytechnica Hungarica vol. 18, no. 3, pp. 197-220, 2021.

[3] T. Kundu, "Acoustic source localization," Ultrasonics, vol. 54, pp. 25-38, Jan. 2014.

[4] L. Kou, "A Review of Research on Detection and Evaluation of the Rail Surface Defects," Acta Polytechnica Hungarica vol. 19, no. 3, pp. 167-186, 2022.

[5] A.M. Ali, et al, "An Empirical Study of Collaborative Acoustic Source Localization," Journal of Signal Processing Systems, vol. 57, pp. 415–436, 2009.

[6] M. F. Fallon and S. J. Godsill, "Acoustic Source Localization and Tracking of a Time-Varying Number of Speakers," IEEE Trans. Audio, Speech, and Language Processing, vol. 20, no. 4, pp. 1409-1415, May 2012.

[7] A. Xenakia, J.B. Boldt, "Sound source localization and speech enhancement with sparse Bayesian learning beamforming," Journal of the Acoustical Society of America vol. 143, paper 3912, 2018.

[8] A. Ledeczi, et al., "Countersniper System for Urban Warfare," ACM Trans. Sensor Networks, vol. 1, pp. 153-177, Nov. 2005.

[9] J. C. Chen, Kung Yao, R. E. Hudson, "Source localization and beamforming," IEEE Signal Processing Magazine, vol. 19, no. 2, pp. 30-39, March 2002.

[10] M. Cobos, et al., "A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks," Wireless Communications and Mobile Computing, vol. 2017, Article ID 3956282, 2017.

[11] G. Simon, et al., "Sensor Network-based Countersniper System," Proc. Second ACM Conference on Embedded Networked Sensor Systems (SenSys 04), pp. 1-12, Baltimore, MD, USA, November 2004.

[12] G. Simon, L. Sujbert, "Acoustic source localization in sensor networks with low communication bandwidth," 2006 International Workshop on Intelligent Solutions in Embedded Systems, 2006, pp. 1-9.

[13] A. Mahajan, M.Walworth, "3-D position sensing using the differences in the time-of-flights from a wave source to various receivers," IEEE Trans. Robot. Autom., vol. 17, no. 1, pp. 91–94, Feb. 2001.

[14] J. Smith, J. Abel, "Closed-form least-squares source location estimation from range-difference measurements," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 35, no. 12, pp. 1661-1669, Dec. 1987.

[15] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[16] G. Vakulya and G. Simon, "Fast Adaptive Acoustic Localization for Sensor Networks," IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 5, pp. 1820-1829, May 2011.

[17] I. Sharp, K. Yu and Y. J. Guo, "GDOP Analysis for Positioning System Design," IEEE Transactions on Vehicular Technology, vol. 58, no. 7, pp. 3371-3382, 2009.