# Acoustic source localization in sensor networks with low communication bandwidth

Gyula Simon[1,2], László Sujbert[2]

[1]Department of Computer Science,
Pannon University, Veszprém, Hungary
simon@dcs.vein.hu

[2]Department of Measurement and Information Systems,
Budapest University of Technology and Economics, Budapest, Hungary
and
Embedded Information Technology Research Group,
Hungarian Academy of Sciences, Budapest, Hungary
{sujbert,simon}@mit.bme.hu

***Abstract —*** *A new acoustic localization scheme is proposed, which can be applied in sensor networks with low communication bandwidth. The first step of the two-stage algorithm is distributed and is evaluated at the sensors having possibly low communication capabilities. The sensors detect various events in the acoustic signal and transmit only the event descriptors to the base station where the sensor fusion is calculated in the second step using a novel generalized consistency function. Test results validate the performance of the proposed system in a noisy and reverberant environment. According to experiments, the proposed system decreases the necessary communication bandwidth with multiple orders of magnitude and still provides high accuracy comparable with that of the sophisticated beam-forming methods.*

## 1 Introduction

The localization of various acoustic sources is a hot research topic with many possible application areas, e.g. voice enhancement, intruder detection, sniper localization, or automatic tracking of speakers in an e-conferencing environment, just to name a few. The idea behind source localization systems is the use of multiple sensors (microphone arrays) placed at different locations. Since sound travels with a constant speed from the source to the sensors, the recorded signals can be used to calculate the possible location of the source, if the sensor locations are known. The techniques utilizing the phase shifts between the different recordings are theoretically simple (assuming line-of-sight measurements), but the disturbing reverberations present in most environments make the practical solution difficult. Many successful source localization solutions can be found in the literature, where cross-correlation techniques are used. Near-field beamforming has many variations, providing good localization accuracy even in noisy and reverberant environments [1],[2],[3]. These algorithms utilize the sound recordings of the sensors to calculate the most probable source location. However, the transfer of data from the sensors to the processing unit requires relatively high bandwidth, which can be a serious limitation. Another possible bottleneck in such systems is the huge amount of calculation to be performed, limiting the response time and thus the application field of the system.

Sensor networks provide a new and dramatically evolving research area. Small intelligent sensor units are used to form a sensor network, where the sensors can autonomously create an ad-hoc communication network, perform distributed sensing (and possibly actuating) activities, and also distributed processing. Due to the possible large number of the utilized sensors these units must be inexpensive. The low prize combined with the required small size results in sensors with rather limited resources, in terms of available power, processing capabilities and communication bandwidth [4].

Inexpensive sensors can be used to form an acoustic source localization system. The greatest advantage of wireless sensors vs. wired microphone arrays is flexibility, i.e. the system can be deployed quickly and practically everywhere, without the need of manual installation. The sensor network can perform self-localization, build the ad-hoc communication network and perform the required sensing. However, straightforward application of current localization techniques is not possible. The bandwidth required to transmit data records with satisfactory quality and length prohibits the use of such devices.

The proposed localization system utilizes a distributed two-stage approach. The first step of the operation is performed at the sensors where certain points of the acoustic signals are marked and their time of arrival is recorded. In the second step the central processing unit uses only this dramatically reduced amount of information to calculate the position of the signal source using consistency functions.

Consistency functions were first proposed for sensor fusion in [8] as part of a counter-sniper system. In this paper we generalize the original consistency function in two directions so that it can handle (1) *multiple occurrences* of (2) *different types* of events. The generalized consistency function enables the robust and accurate localization of practically all types of acoustic signals (e.g. speech), also in demanding reverberant environments.

The rest of the paper is organized as follows. Section 2 describes the proposed approach. In Section 3 the generalized consistency function is introduced. The test results of the proposed localization system are presented in Section 4. Finally, we present our future plans and conclusions.

## 2 Sensor Network-Based Acoustic Source Localization

In acoustic source localization systems multiple sensors (microphones or microphone arrays) placed at known positions are used to detect signals emitted from the source. The sound travels with constant speed to the sensors and thus the phase differences between the sensed signals depend on the distances between the sensor nodes and the source (provided line-of-sight measurements exist, as opposed to echoes or no detections at all).

One of the most successful acoustic source localization techniques is near-field beam-forming that can be used to detect multiple sources in noisy reverberant areas [1],[2],[3].

However, most of these methods require the transmission of the sampled data records between sensor nodes and/or the base station, a prohibitive feature in typical wireless sensor networks.

There exist similar two-step techniques where in the first step the time differences between the signals are measured, and in the second step the location is calculated [5],[7]. The communication burden of the transmission of time data is acceptable. Note that the time delay measurement/calculation requires either special signals (e.g. where the start of the signal is clearly detectable, see [8]) or the usage of various correlation techniques requires high communication burden again.

Once the time delay data is available, a central fusion algorithm can calculate the source location. In theory four non-coplanar measurements are enough to determine a 3D location, provided the speed of sound is known. In practice, however, more measurements and thus an over-determined equation system is used to compensate errors due to inaccuracies in sensor localization and time delay measurement [3][5][7].

From correct time delay data the conventional analytical solutions using least-squares or maximum likelihood criteria can calculate the source location [7]. These methods tolerate noise and even a small amount of reverberant data, but the accuracy decreases dramatically if a large amount of reverberant data is present, as shown in the analysis presented in [6]. The main problem with the analytical solutions is that they use all of the available data, regardless of their quality. Iterative solutions containing classification of the data would solve this problem, but no computationally feasible solution is known.

In the localization system presented in [8] a two-stage approach was used. In the first step the time of arrival (TOA) of the signals were measured at the sensors (the special signal type allowed this solution). At the second step a new robust fusion algorithm was proposed that was able to find the maximum set of consistent time measurements, using a so-called consistency function, and thus find the source location with high accuracy even in the presence of high amount (up to 50%) of erroneous data.

The solution proposed in [8] works fine if the TOA of the signals can be measured with high accuracy. In [8] the special signal type allowed the detection and high precision TOA measurement of the starting point of the signal. However, for conventional acoustic signals (e.g. speech) the definition of such points is very hard (e.g. where is the exact starting point of a word?), especially in the presence of noise and the different acoustic transfer functions between the source and the microphones. The proposed approach overcomes this limitation.

Our proposed solution is also a two-stage approach: In the first step, not having one easily detectable point to measure, each sensor $S_i$, $i = 1,2,...,Q$ marks several ($M_i$) events $E_j$ $(j = 1,2,...,M_i)$ that match one of the characteristic properties from the predefined global set $\{\Pi_k\}$, $k = 1,2,...,K$. Such simple event can be, for example, the sudden increase of energy in a certain frequency band. Note that sensor $i$ detects $M_i$ different events, each of which can correspond to any one of the $K$ characteristic properties. Each event is thus characterized by the property $\Pi_k$ it matches, and the $t_n$ time of arrival of the event $(n = 1,2,...,M_i)$. The *event descriptor* tuple $\langle i,k,t_n \rangle$ is then routed to the central processing node using the wireless ad-hoc network. Note that the size of the data is very small, containing only a few bytes, allowing in-network aggregation to further reduce the communication cost.

In the second step the central processing node calculates the source location from the TOA data obtained from the $\sum_i M_i$ received event descriptors and the known sensor positions. Since the first preprocessing step completely removes the fine detail from the recorded data, the data fusion must use sophisticated methods to provide high accuracy. The fusion is based on the generalized consistency function, defined in the next section.

## 3 The Generalized Consistency Function

The consistency function first proposed in [8] used one event (namely the start of the detected muzzle blast), i.e. $K=1$, $M_j=1$. For the sake of clarity, we shortly review the original consistency function here.

If the *i*th sensor detects the TOA $t_i$ of an acoustic event at sensor position $(x_i, y_i, z_i)$ then the estimated time of emission of that event can be calculated as

$$t^i(x,y,z)=t_i-\frac{\sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}}{v}, \tag{1}$$

where *v* is the speed of sound and $t^i(x,y,z)$ is the emission time estimated by sensor *i*, *provided* the unknown source is located at $(x,y,z)$. If $(x,y,z)$ corresponds to the true source location and the measurements are accurate then every sensor estimates the very same $t^i(x,y,z)$ value. If the source location is not $(x,y,z)$ then the estimated emission times are not the same. If the measurements are not accurate but $(x,y,z)$ is the true location then the sensor estimates $t^i(x,y,z)$ are not exactly the same, but are in a close proximity of the true source time, the 'closeness' depending on the accuracy of the sensors. Evidently measurements with large errors become outliers.

If the inequity

$$|\, t^i(x,y,z)-t\,|\le\tau \tag{2}$$

holds for a time value *t* and for measurements $i=i_1,i_2,\ldots,i_{K(x,y,z,t,\tau)}$, where $K(x,y,z,t,\tau)$ is the number of consistent measurements for source position $(x,y,z)$ and emission time *t* with uncertainty $\tau$, then the consistency function is defined as

$$C_\tau(x,y,z)=\max_t K(x,y,z,t,\tau). \tag{3}$$

The optimal source location estimate $(x_s, y_s, z_s)$ is at the maximum of the consistency function:

$$(x_s,y_s,z_s)=\arg\max_{(x,y,z)} C_\tau(x,y,z). \tag{4}$$

The concept is illustrated in Figure 1, where the sensors and the source are placed on a grid of a plane. For simplicity, assume that the grid size is 1 and the speed of sound is 1. Sensors $S_1$, $S_2$, …, $S_6$ measure TOA values $t_1$, $t_2$, …, $t_6$, as shown in the figure. Note that measurement noise is present for all sensors, and $S_6$ reported an erroneous value (e.g. echo because of no line-of-sight). For positions $P_1$ and $P_2$ the estimated emission times for each sensor are indicated on the timelines. For $P_1$, which is the true source location, five sensors estimate emission times around 10, thus $K(P_1,10,0.6)=5$ (in the example $\tau=0.6$, the light boxes on the timelines indicate intervals of length $2\tau$). The timeline clearly

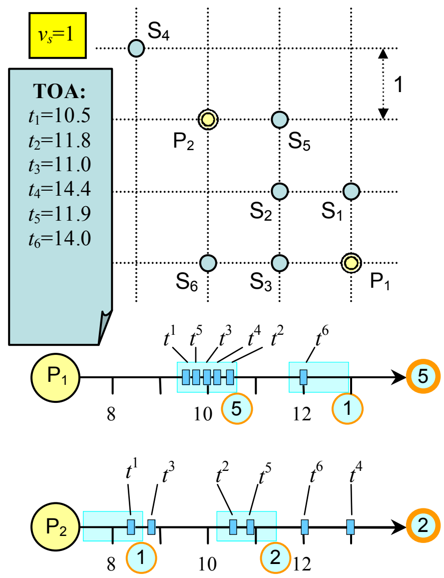


Figure 1. Illustration of the consistency function concept in 2D for a single event. The values of the consistency function are calculated at positions $P_1$ and $P_2$ from sensor data $t_1$, $t_2$, …, $t_6$. The true position is $P_1$ and the emission time is10.

shows that the faulty sensor $S_6$ estimated a completely different value, far from the dense group around 10. $K(P_1, 10, 0.6)$ is also the maximum defined by (3) for $P_1$, thus the consistency function value for position $P_1$ is 5. For $P_2$ the estimated emission times are scattered through the timeline, no definite cluster can be seen and the value of the consistency function is only 2. In this example the largest consistency function value points to the estimated source position $P_1$. Note that the true source position was indeed $P_1$ and the true emission time was 10.

The consistency function can be generalized to handle more event types and multiple occurrences of each event type. For each measurement $\langle i, k, t_n \rangle$ the estimated time of emission can be calculated for a hypothetical source location $(x, y, z)$ as

$$t^n(x, y, z) = t_n - \frac{\sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}}{v}. \tag{5}$$

If the inequity

$$| t^n(x, y, z) - t | \leq \tau, \tag{6}$$

holds for a time value $t$ and for measurements $< i_1, k, t_1 >$, $< i_2, k, t_2 >$, …, $< i_{K(x,y,z,t,\tau,k)}, k, t_{K(x,y,z,t,\tau,k)} >$, where $K(x, y, z, t, \tau, k)$ is the number of consistent measurements for events type $\Pi_k$ for position $(x, y, z)$ and time $t$ with uncertainty $\tau$, then

$$K^*(x, y, z, t, \tau, k) = \begin{cases} K(x, y, z, t, \tau, k) & \text{if } K(x, y, z, t, \tau, k) \geq \Lambda \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where $\Lambda$ is a constant defining the minimum amount of consistent measurements to be considered valid (in practice $\Lambda \geq 3$). The consistency function for event type $\Pi_k$ is defined as

$$C_{k,\tau}(x,y,z) = \max_T \sum_{t \in T} K^*(x,y,z,t,\tau,k), \tag{8}$$

where $T$ is a set of time values such that for every two values $t_1, t_2 \in T$ the inequity $|t_1 - t_2| > \tau$ holds. Loosely speaking, $C_{k,\tau}(x,y,z)$ is the total number of consistent measurements for position $(x,y,z)$ supporting event emission times $t \in T$ of events type $\Pi_k$, with uncertainty $\tau$. Because of uncertainty $\tau$ in the measurements the minimum distance between emission times must be at least $\tau$. Now the consistency function is defined as

$$C_\tau(x,y,z) = \sum_k C_{k,\tau}(x,y,z). \tag{9}$$

The concept is illustrated in Figure 2, where the original and the generalized consistency functions are shown. The left-hand side shows the single-event case with the original consistency function, while on the right-hand side the generalized case is illustrated with three events: $E_1$ and $E_3$ match property $\Pi_1$, while $E_2$ match property $\Pi_2$. Sensors $S_1$, $S_2$, $S_3$ sense all of the events, as shown in Figure 2(a). Figure 2(b) and (c) show the timelines at the correct ($P_1$) and at an incorrect ($P_2$) source position, respectively. The corresponding consistency function values are also shown. In the generalized example $C_{1,\tau}(P_1) = 6$ and $C_{2,\tau}(P_1) = 3$, thus $C_\tau(P_1) = 9$, while $C_{1,\tau}(P_2) = 4$, $C_{2,\tau}(P_2) = 2$ and $C_\tau(P_2) = 6$. In the example for the sake of simplicity $\Lambda$=2 was used.

Note that the consistency function must be evaluated at every point of the space where the source may be present. A fast search method presented in [6] makes the search possible for the original consistency function in large areas, even in real time. The application of the fast search for the general case needs further investigations.

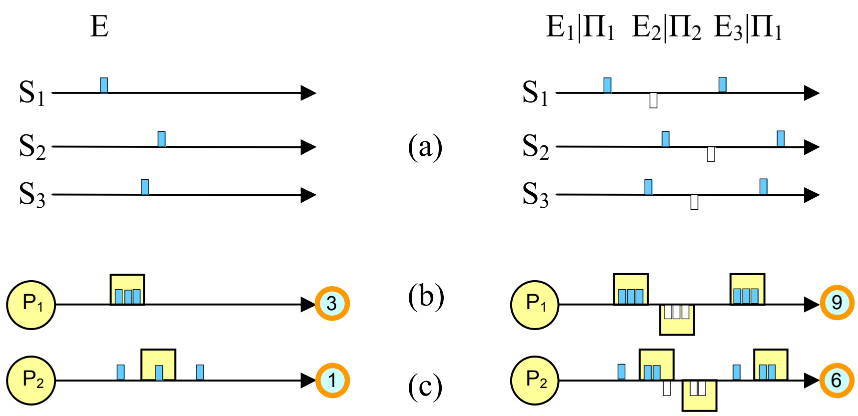


Figure 2. Comparison of the original (left) and the generalized (right) consistency functions: (a) events shown at each sensor's timeline, (b) consistency calculated at the correct source position (for sake of visibility the events are not exactly aligned) and (c) consistency calculated at a wrong position. In the generalized case $\Lambda$=2 was used.

## 4 Experimental Results

The original consistency function was tested in a large-scale counter-sniper application with approximately 60 sensors. As was reported in [8] the accuracy of the localization system was around 1m in the test area with the size of a small village. The proposed generalized consistency function was tested in a small-scale indoor experiment in a medium-size auditorium at the Budapest University of Technology and Economics. The auditorium provides relatively large reflections. The setup contained only 5 sensors placed in a plane. The speaker was also close to the plane, as shown in Figure 3. The microphone signals were recorded and later processed in Matlab.

The test signal was a Hungarian children's rhyme, approximately 7 seconds long. The signal recorded by one of the microphones is shown in Figure 4. As the plot shows, the signal-to-noise ratio was quite low (see the high background noise from 0 to 2s).

The first step of the processing contained the event detection. In real applications sensor devices with limited computing capabilities are used, thus the event detection must be computationally simple. In the experiment 6 rather simplistic event classes were defined. The raw signal was filtered by a bank of bandpass filters containing 6 banks (implemented by a block FFT). For band $i$ at each (downsampled) time instant $k$ we calculated the signal's maximum amplitude before ( $A_i^{before}$ ) and after ( $A_i^{after}$ ) the given time $k$ within a time window of 1ms. If and when the ratio $A_i^{after}(k)/A_i^{before}(k)$ at sensor $j$ increased above a limit, an event descriptor $\langle j,i,k\rangle$ was generated.

The search space was set to $x$ = [-5…5], $y$ = [-10…10]. Figure 5 shows the results: the surface of the consistency function and the contour of the consistency function zoomed into the central region. It's clearly visible that the maximum is around the true location, which can be detected with 0.2m accuracy. The power of the consistency function approach is shown by the fact that less than five percent of the detected events were correct.

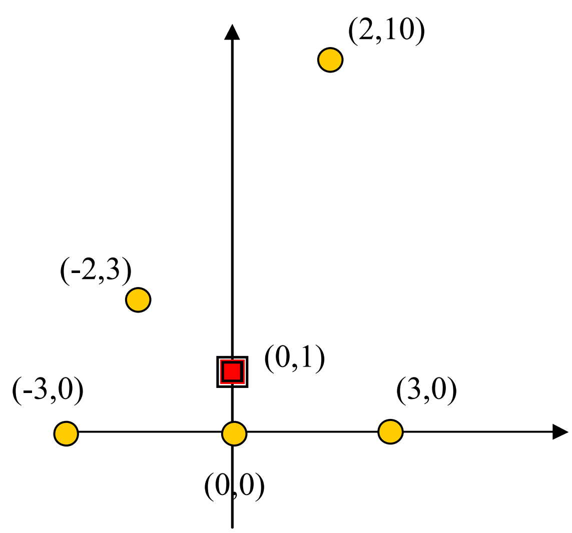


Figure 3. Experimental layout for testing the acoustic localization system containing 5 microphones (circles) and a source (square).

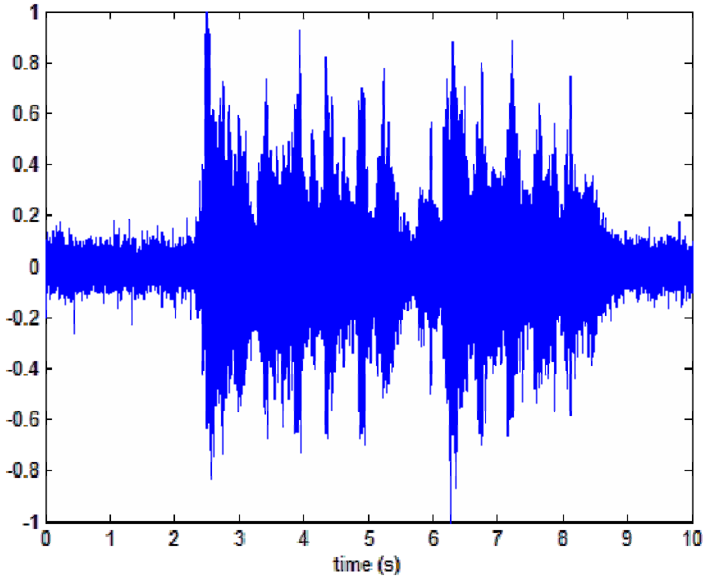


Figure 4. The test signal (a Hungarian rhyme) recorded by the microphone placed at location (-2,3)

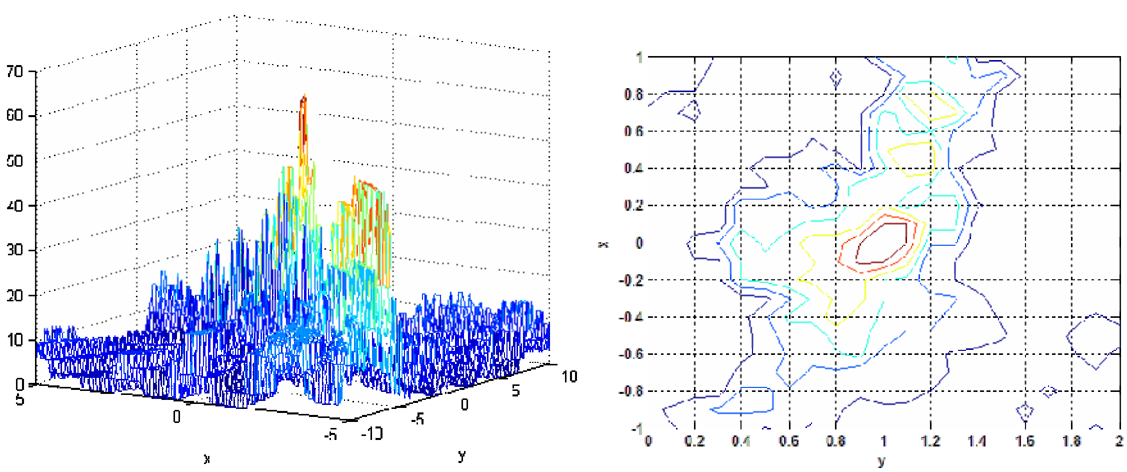

Figure 5. Result of the localization process. The maximum value of the consistency function is 62. The true location was at x=0, y=1, which is detected with 0.2m accuracy.

In the test the algorithm provided surprisingly high accuracy, even though rather simplistic event detection was used, the signal to noise ratio was low, and the auditorium provided lots of echoes. The size of the event descriptors, which had to be transmitted, was 1.1% of that of the raw signal. This means that compared to beam-forming, two orders of magnitude smaller communication bandwidth is enough to provide similar accuracy, even with such simple event detection algorithm. A more sophisticated event detection would probably further reduce the amount of detected false events and thus the necessary communication bandwidth as well.

## 5 Conclusions

A new acoustic localization system was proposed that can be used in sensor networks with low communication bandwidth. The proposed localization system requires multiple orders of magnitude lower bandwidth than the conventional beam-forming techniques. The system uses a two-stage approach, where in the first stage various events are detected in the acoustic signals by the sensors, and in the second stage a generalized consistency function is used to calculate the source location by the central computing unit.

The unique event detection concept and the associated event descriptors make it possible to use a very limited amount of data to reach high accuracy.

The choice of the events is important: the signal must be sufficiently rich in events, but the speed of the system degrades if too many events are used. Also, the identification of the events must be easy using the sensors with limited computational capabilities. The automatic data classification built in the consistency function provides resistance against reverberant data and also bad measurements.

The system showed promising results in the experimental tests, but further research is necessary to answer open questions: how to handle multiple and possibly moving sources.

## 6 Acknowledgement

This research was partially funded by the Hungarian Government under contract NKFP2-00018/2005.

## 7 References

[1] Chen, J. C., Hudson, R. E. and Yao, K. Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field. *IEEE Trans. Signal Processing*, 50(8):1843–1854, Aug. 2002.

[2] Chen, J. C., Yao, K., and Hudson, R. E. Source localization and beamforming. *IEEE Signal Processing Magazine*, 19(2):30-39, March 2002.

[3] Chen, J., Yip, L., Elson, J., Wang, H., Maniezzo, D., Hudson, R., Yao, K., and Estrin, D. Coherent acoustic array processing and localization on wireless sensor networks. *Proceedings of the IEEE*, 91(8):1154-1162, Aug. 2003

[4] Hill, J., Culler, D. Mica: A Wireless Platform for Deeply Embedded Networks. *IEEE Micro*, 22(6):12–24, Nov. 2002.

[5] Huang, Y., Benesty, J., and Elko, G. W. Passive acoustic source localization for video camera steering. Proc. IEEE ICASSP, vol. 2, June 2000, pp. 909–912.

[6] Ledeczi A., Nadas A., Volgyesi P., Balogh G., Kusy B., Sallai J., Pap G., Dora S., Molnar, K., Maroti M., Simon G. Countersniper System for Urban Warfare. *ACM Transactions on Sensor Networks*, 1(2):153-177, Nov. 2005.

[7] Mahajan, A. and Walworth, M. 3-D Position Sensing Using the Differences in the Time-of-Flights from a Wave Source to Various Receivers. *IEEE Trans. On Robotics and Automation*, 17(1):91-94, February 2001.

[8] Simon, G., Maróti, M., Lédeczi, A., Balogh, G., Kusy, B., Nádas, A., Pap, G., Sallai, J., Frampton, K. Sensor Network-based Countersniper System. In *Proceedings of the Second ACM Conference on Embedded Networked Sensor Systems (SenSys 04)*, pages 1-12, Baltimore, MD, USA, November 2004.