ICSV13 - Vienna
The Thirteenth International Congress on Sound and Vibration
Vienna, Austria, July 2-6, 2006

TRACKING NOISE SOURCES USING MULTIPLE MOBILE MICROPHONE ARRAYS

Dan Mennitt[1], Philip Gillett[1], James Carneal[1] and Marty Johnson[1]

[1]Department of Mechanical Engineering, Virginia Tech
147 Durham Hall, Blacksburg, VA 24060, USA
jcarneal@vt.edu (e-mail address of lead author)

Abstract

A network of mobile microphone arrays (nodes) has been created to identify and track noise sources in an outdoor environment. Collaborative signal processing is exploited: individual nodes with limited processing power and access to the local environment relay information to a central-level composite tracking system (fusion center). The nodes are able to detect noise sources, calculate the relative direction, and perform gross classification. The fusion center is responsible for associating detections to form noise events, associating events to objects, and then tracking those objects. Association is accomplished by a series of classification algorithms and decision logic. Information is stored on the objects' location, speed, heading, track, and noise character: impulsive, tonal, or broadband. Results from experiments run in an outdoor suburban environment demonstrate performance and the complexity of the problem. Challenges and directions for future research are discussed.

INTRODUCTION

The study of distributed sensor networks is a rapidly growing field with a wide array of applications. One important application is acoustic surveillance, or the localization and tracking of noise sources in an environment. Such as system could localize snipers, and/or track military vehicles. If the system is mounted on autonomous vehicles, it could eventually replace humans on dangerous scouting missions.

To ensure a robust system (i.e. capable of handling the loss of sensors and/or sensor nodes), a collaborative signal processing architecture has been adopted for the current research of tracking noise sources in a realistic environment. The system architecture will now be discussed referring to Figure 1, also introducing terminology in *italics*. Multiple acoustic *sensors* are attached to a *node*, which has local computational processing ability, GPS for position, a magnetic compass for orientation, and wireless networking capability. The node processes all the information from the acoustic sensors and sends data on *detections*, which are currently defined as high acoustic levels, or changes in one or several acoustic metrics (relative angle, power spectra levels, etc). This reduced information is sent to the central command *fusion center*, which will fuse the node information to events (source *localization*), merge events to objects (object *classification*), and merging objects to paths (*tracking*).

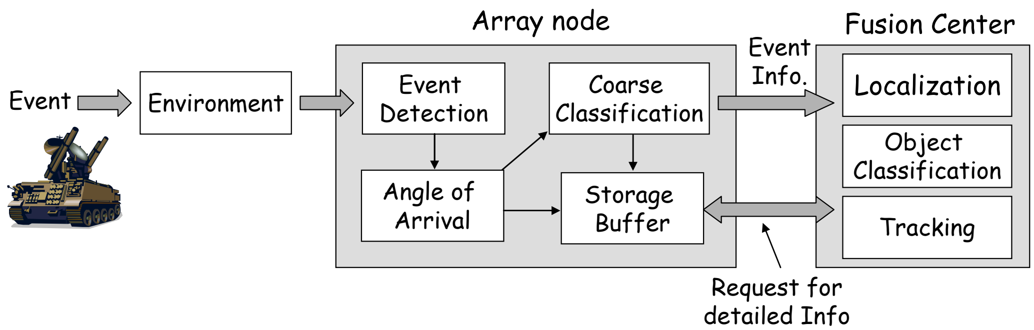


Figure 1. Schematic of system architecture.

The goal of this system is to develop a reliable estimate of the acoustic environment. Since this system is statistically based, events that are detected by a subset of nodes will be incorporated into the current model of the environment by the fusion center. Persistent events that are detected by one node only will also be modeled, and the fusion center may also re-direct a mobile node to investigate further to triangulate, or to acquire information for a more reliable estimate.

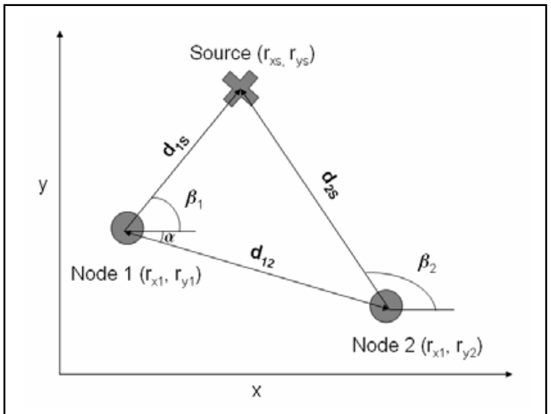
This paper will review the developed node processing and fusion center algorithms, and present the experimental investigation and results, followed by conclusions and future work.

ALGORITHMS

Node Processing

There are three algorithms applied to reduce the broadband acoustic data to event features. The first is an event detection algorithm, which compares the current power at the sensors to the previous power. If the power has increased, or is above a certain level (approximately 50 dB) and is tonal, then the algorithm assumes there is an event to process. The second algorithm is the localization algorithm, based on previous work by Carneal, et. al[i] and is presented in a companion paper[ii]. This work efficiently determines the relative angle of arrival (AOA) of the event to the node (sensor array). The third algorithm is a characterization algorithm, where the beamformed signal is characterized as impulsive, tonal, or broadband, and the third octave power is calculated. These algorithms will not be discussed in detail, but can be found in the aforementioned references.

Fusion Center



Without loss of generality, consider two nodes coplanar with one far-field noise event

The Fusion Center employs composite tracker system architecture, incorporating inputs from several sensor nodes to arrive at a single classified event. After detection and local processing, a node transmits a feature vector (FV) to the Fusion Center containing the node name, node location, and concentrated features corresponding to the detection: time of detection, angle of arrival, impulsiveness, tonality, and 1/3 octave band power levels. The feature vectors are combined through a series of data association algorithms which yields a classified event. Tracking objects is accomplished using a Kalman filter to process events. This section outlines the procedure of three data association algorithms (associating node detections to observations, fusing observations to an event, associating events to objects) and tracking detected objects.

Detection Association

Event localization is performed using triangulation of AOA from individual nodes. Without loss of generality, consider two nodes coplanar with one far-field noise event as described by the geometry of Figure 2. The sound from an impulsive source *S* reaches *Node 1* at time t_{d1} at AOA β_1 and *Node 2* at time t_{d2} at AOA β_2.

For an event to be fused, the lines of bearings (LOBs) must intersect, which is dependant on the angle (α), between the two nodes in question. If this condition is met, the LOBS are triangulated to obtain a position fix of the source in cartesian coordinates according to equations:

$$r_{xs} = \frac{\delta y - \delta x[\tan(\beta_2)]}{\tan(\beta_1) - \tan(\beta_2)} + r_{x1} \tag{1}$$

$$r_{ys} = (r_{xs} - r_{x1})\tan(\beta_1) + r_{y1} \tag{2}$$

where δ_x and δ_y are the distance between the nodes in the *x* and *y* directions respectively. One consequence of transforming the two degrees of freedom from angular measurements to cartesian coordinates is geometric dilution of precision[iii]. This error is dependant on the node source spatial configuration. The individual variances of the angle measurements can be propagated to a coupled cartesian covariance matrix using uncertainty analysis[iv].

Using the proposed position fix calculated using equations 1,2, the time of event (*TOE*) is calculated by back propagating the acoustic wavefront.

$$TOE_n = \left(t_{dn} - \frac{d_{nS}}{c}\right) \tag{3}$$

A range test is then made by comparing the *TOEs* of the following form:

$$TOE_1 - TOE_2 < T_d \tag{4}$$

where the event detection window, T_d, is some threshold dependant on the measurement noise covariance and detection time quality.

After an event detection window is closed, associated detections are sent to the observation fuser. Each pair of detections is localized separately to yield an independent observation and covariance matrix. If an event detection window containing a single detection closes, the noise is assumed to be a disturbance local to that node (windblown debris for example) and the detection is ignored.

Observation Fuser

Due to geometric dilution of precision, the noise source location estimate in one dimension could be very poor while the other dimension is acceptable, as reflected in the measurement cross covariance matrix *R*. Therefore, the *x* and *y* components of the observations are combined independently. Consider the fusion of observations corresponding to the *x* coordinate of the source for simplicity, the *y* coordinate is evaluated identically. The error of the i^{th} observation is assumed to follow a normal distribution with mean μ_i and variance σ_i^2. It can be shown that the mean and variance of the fused observation is[v]:

$$\mu_\zeta = \sum_{i=1}^{N} a_i \mu_i \tag{5}$$

$$\sigma_\zeta^2 = \sum_{i=1}^{N} a_i^2 \sigma_i^2 \tag{6}$$

After the fused source location is calculated, the *TOE* is recalculated using equation 3. A Kalman filter tracker requires measurements in temporal order. By knowing the *TOE*, the common sensor network drawback of out of sequence measurements due to communication delays and acoustic wave propagation is circumvented. The track can be rolled back, updated and re-propagated.

Event to Object Association

Information (*TOE*, r_{xs}, r_{ys}, R, number of nodes detecting the event, sound type) is archived for each event. Similarly, time dependent cues are maintained for each object utilizing a track before detect approach: spectral character, current position and speed and the corresponding covariance. Similar to the detection association, each object-event combination is subject to a series of tests to discard unlikely matches. Association is based on spectral characteristics and location. All objects characterized with an identical noise type as the event are considered for association. For the location test, we use the all neighbors approach for simplicity[vi]. The correlation decision is based on a Chi-squared test of the following form:

$$\left[\hat{x}_p - z\right]^T \left[P_P + R\right]^{-1} \left[\hat{x}_p - z\right] < A \tag{7}$$

where x_p denotes the position components of the object at the *TOE*, P_p is the position components of the covariance matrix relating to the uncertainty in the current position. The decision criterion, or “gate” is denoted by *A*. If no existing objects are within the gate a new object and track is initiated. Widening gates and maneuvers are not considered.

As the number of targets increases, it may be useful to solicit more data from a node and classify an object and measurement with more detail, allowing thresholds to be tightened.

Tracking

A four state linear coupled Kalman filter[vii] is employed for tracking objects. The performance gains over a linear uncoupled filter are similar to that of the extended filter, with the additional benefit of robustness to initialization errors[viii]. The Kalman filter works in two stages, prediction and correction, to estimate the state of a dynamic system. Prior to incorporation of measurements, the state vector, $x \in \Re^n$, is projected ahead according to a discrete-time difference equation of the following form:

$$\hat{x}_k^- = \Phi \hat{x}_{k-1} + Bu_k + w_{k-1} \tag{8}$$

where the hat denotes an estimate, the subscript k refers to the time step, and the superscript minus refers to a predicted estimate before incorporation of measurements. Φ is the state transition matrix which relates $\hat{x}_{k-1}$ to $\hat{x}_k^-$ according to some deterministic motion model. The matrix B relates the control input u to the state x, which is unknown and set to zero for this application. w_k is a Gaussian white noise sequence that accounts for the uncertainty in the process model. A constant velocity motion model is used for simplicity. While a lower order model is subject to divergence during maneuvers, it allows for better position estimates with fewer measurements.

EXPERIMENTAL SETUP

To test the aforementioned algorithms, a field experiment was performed to localize and track impulsive and tonal sources. Three static diffracting arrays were placed in an outdoor environment. These arrays were all mounted on 8 inch diameter cylindrical tubes; two of the arrays had 12 microphones while the third array had 6 microphones. The arrays were placed near intersection of two roads to record a static impulsive source and two moving tonal sources (transportation buses). All diffracting arrays were connected to the same data acquisition system and computer to guarantee time alignment of the signals, and the positions of the arrays were measured with respect to landmarks such

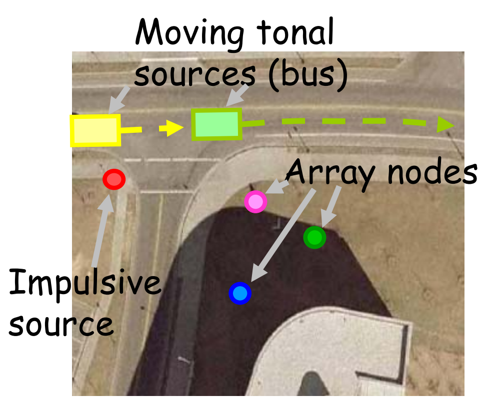


Figure 3. Aerial view of three static nodes, the impulse, and the moving tonal sources.

as stop signs. Once several tests were performed, the previously described node processing algorithms were tested and refined on the recorded data.

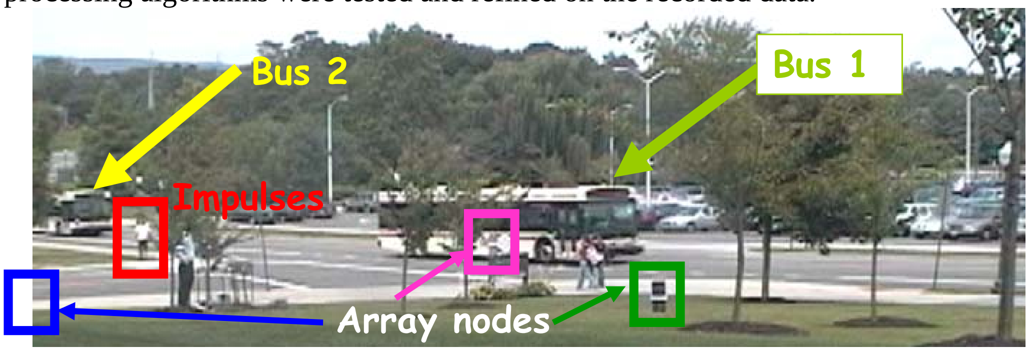


Figure 4. Perspective photograph of three static nodes, the impulse, and the moving tonal sources.

EXPERIMENTAL RESULTS

Over a 40 second time interval, 105 feature vectors were transmitted to the fusion center. Several feature vectors were rejected since they did not meet the aforementioned conditions. For example, a node may be listening to a more dominant source from another direction. Another situation occurred when the LOBs intersected but either the range was too broad or the impulsive power threshold was not exceeded.

Table 1.Percentage of feature vectors associated with known sources.

	Clap (Impulsive)	Bus 1 (Tonal)	Bus 2 (Tonal)	Total
Transmitted FVs	30	51	24	105
Associated FVs	27	45	20	92
% Associated	90	88.2	83.3	87.6

Both buses were strongly tonal and the tonal indicator consistently exceeded the threshold. As a bus approached the nodes, closer nodes measured the bus while nodes farther were registering other noise sources, such as another vehicle. Table 1 shows the number of feature vectors associated with the three objects. All of the events were associated to the correct objects by design of the gate threshold. The targets were well separated within the small scale studied such that this approach, coupled with the noise type match condition, was sufficient to classify the targets.

Tracking

Figure 5 shows an aerial view of the impulse track, bus 1 track, and bus ground truth in the experiment. As can be seen, the acoustic tracking system is able to correctly associate and track both the impulsive and tonal (bus) sources. Note that there is some erroneous wandering present in the impulse track due to the constant velocity model being applied to a stationary object.

For the bus track, the constant velocity model is able to track and predict the motion of the bus well. The process noise model is able to account for the slight acceleration through the turn. The measurements are biased in the *y* direction, possible due to reflections off the nearby building. Furthermore, although modeled as a point source, the bus is a large object with many noise sources: tires, engine, exhaust, etc. which makes precise localization difficult. Overall, the ability of the distributed node system to identify and track impulsive and tonal sources has been demonstrated with little error.

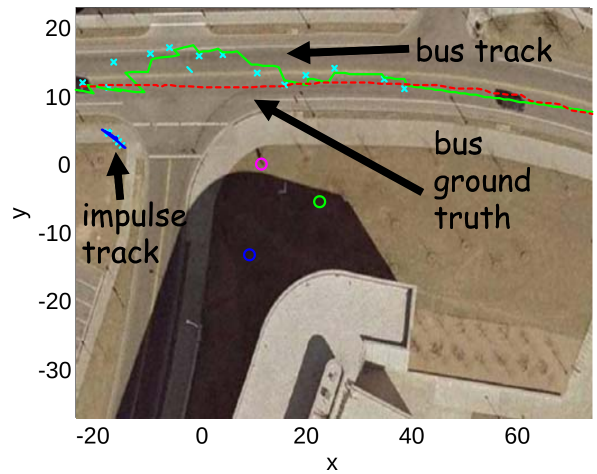
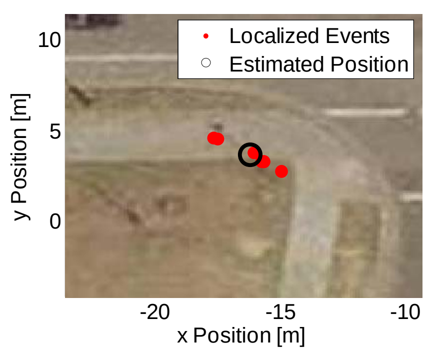


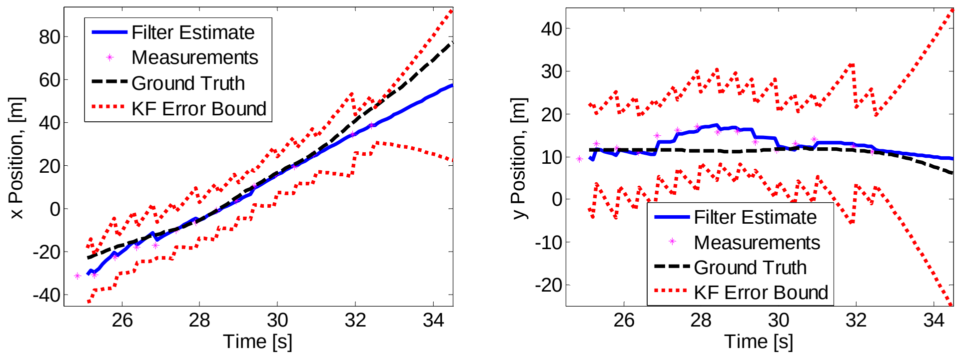
Figure 5. Overview of experimental environment showing the tracks of the claps (blue) and the first bus. Events are signified by an 'x'. The projected ground truth of the bus is also shown (dotted line).

Further analysis of the impulsive and tonal (bus) track data is now presented in **Error! Reference source not found.** and Figure 7, respectively. The plots display the Kalman filter estimate, the actual measurements, the ground truth and the Kalman filter error bounds for the x- and y- coordinates. Note that the error bound is a measure of uncertainty in the estimate. As can be seen in these plots, the system is able to track both the impulsive and tonal sources. Referring to **Error! Reference source not found.**, the localized events are centered around the true position of the source, and the system is able to accurately determine the position of the source within 0.3 meters with less than one meter standard deviation. Referring to Figure 7, the system is able to accurately characterize, associate, and track the moving source. As previously mentioned, the y-estimate of the tonal (bus) source exhibits some error in the 27 to 30 second timeframe. Since this error is position dependent, it can be attributed to strong building reflections.



Parameter	Value (m)
Ground Truth	(-16,4)
Average	(-16.3,3.7)
Standard Deviation	(0.65,0.96)

Figure 6. Kalman filter track for the impulsive source and comparison of track with ground truth.



(a) (b)

Figure 7. (a) Kalman filter track (x coordinate) for bus 1. (b) Kalman filter track (y coordinate) for bus 1.

CONCLUSIONS

A system for tracking noise sources using multiple mobile microphone arrays has been developed and demonstrated on real-world data. A distributed hierarchical processing methodology has been implemented and was able to track impulsive and tonal sources. The node processing algorithms have been developed to reduce the time-series microphone data to feature vectors. These feature vectors were then sent to a fusion center which correctly localized, associated, and tracked the noise source objects.

Future work will include refinement of all the algorithms, from node processing to data association and fusion. More advanced Kalman filter models to include acceleration and motion of the array itself will be investigated.

REFERENCES

[i] James Carneal, Marty Johnson, and Philip Gillett,"Comparison of a Diffracting and a Non-diffracting Circular Microphone Array," Proceeing of ICASSP 2006, Toulouse, France, May 14-19, 2006.

[ii] Marty Johnson, James Carneal and Philip Gillett, " Comparison of a Diffracting and a Non-diffracting Cylindrical Microphone Array," Proceedings of ICSV 13, Vienna, Austria, July 2-6, 2006.

[iii] R. A. Poisel, “Introduction to Communication Electronic Warfare Systems,” Artech House, 2002.

[iv] R. S. Figliola, and D.E. Beasley, “Theory and Design for Mechanical Measurements,” John Wiley & Sons, Inc. 2000.

[v] P. G. Hoel, S. C. Port, C. J. Stone, “Introduction to Probability Theory,” Houghton Mifflin Company, 1971.

[vi] M.P Dana, “Registration: A Prerequisite for Multiple Sensor Tracking,” article in “Multitarget-Multisensor Tracking: Advanced Applications," Artech House, 1990.

[vii] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” Journal of Basic Engineering, vol. 82, No.1, 1960, pp. 35-46.

[viii] P. Zarchan and H. Musoff, “Fundamentals of Kalman Filtering: A Practical Approach,” American Institute of Aeronautics and Astronautics, Inc. 2005.