Scientific
Research

In-Solid Acoustic Source Localization Using Likelihood Mapping Algorithm

Ming Yang, Mostafa Al-Kutubi, Duc Truong Pham
School of Engineering, Cardiff University, Cardiff, UK
E-mail: *yangm@hotmail.co.uk*

Abstract

The significant challenge in human computer interaction is to create tangible interfaces that will make digital world accessible through augmented physical surfaces like walls and windows. In this paper, various acoustic source localization methods are proposed which have the potential to covert a physical object into a tracking sensitive interface. The Spatial Likelihood method has been used to locate acoustic source in real time by summing the spatial likelihood from all sensors. The source location is obtained from searching the maximum in the likelihood map. The data collected from the sensors is pre-processed and filtered for improvement of the accuracy of source localization. Finally a sensor fusion algorithm based on least squared error is presented to minimize the error while positioning the source. Promising results have been achieved experimentally for the application of acoustic tangible interfaces.

1. Introduction

With ordinary interface devices (e.g. keyboard, mouse and touch screen), the interaction of humans with computers is restricted to a particular device at a certain location within a small movement area. A challenge in human computer interaction research is to create tangible interfaces that will make the interaction possible via augmented physical surfaces, graspable objects and ambient media. In this paper acoustics-based remote sensing technology is presented since vibrations are the natural outcome of an interaction and propagate well in most solid materials. This means that the information pertaining to an interaction can be conveyed to a remote location [1-3] using the structure of the object itself as a transmission channel and therefore suppressing the need for an overlay or any other intrusive device over the area one wishes to make sensitive.

Time Difference of Arrival (TDOA) is commonly used for source localization without using a universal timing mechanism as shown in **Figure 1**. The principle of the TDOA localisation is to measure the time delays between the arrivals of the signals to the various sensors. These delays result from the distance differences from the acoustic impact sources to the sensors at known locations. To determine time delay of real signals, cross cor-

relation is commonly used. But in this application its result usually contains multiple peaks and therefore there is no guarantee that the peak will occur at the correct time difference. This problem of noisy cross correlation can be handled by accumulating the entire cross correlation vector of each sensor pair rather than selecting the single peak of each one. The summation of vectors from multiple sensor pairs yields a likelihood map. The highest likelihood in the map is the estimate of the source location.

The likelihood mapping method has been developed for room acoustics based on two different mathematical formulations. In [3] the processing is performed in the time domain and known by accumulated correlation, and in [4] the processing is performed in the frequency domain and known by spatial likelihood function.

In this paper the likelihood mapping method is enhanced for Tangible Acoustic Interface and referred to as the Enhanced Likelihood Mapping (ELM). The ELM is more robust in searching for the most likely source position and less prone to be affected by data fluctuation comparing to the TDOA method. Finally a sensor fusion algorithm based on least squared error is used to minimize the error of the estimated time differences while positioning the source.
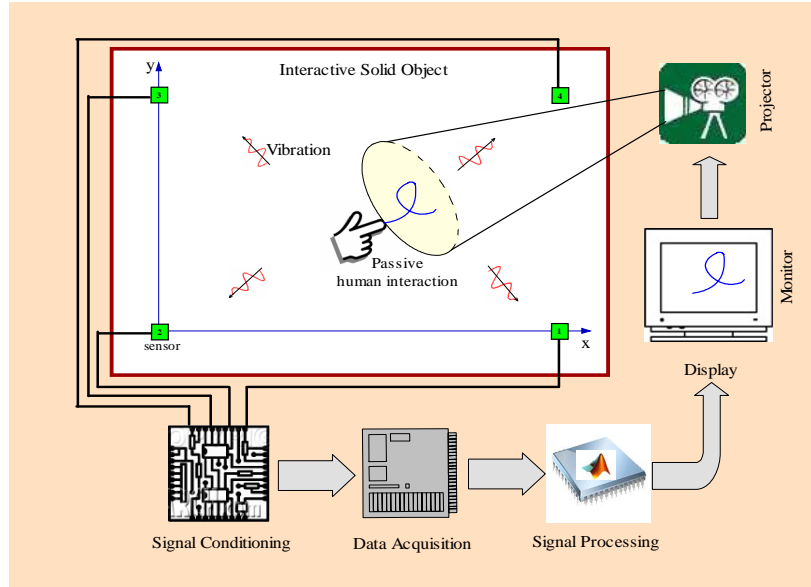
Figure 1. Tangible Acoustic Interface (TAI) model diagram.

2. ELM Theory

If the source signal is $s(t)$, then the received signal can be defined as

$$g_i(t) = h_i(t) * s(t - \tau_i) + n_i(t), \quad (1)$$

where n_i is independent zero mean white Gaussian noise with variance σ_i, theoretically $g_i(t)$ can be regarded as an estimator for τ_i where $\tau_i = \|q - u_i\| / v$ given u_i the ith sensor location. Based on Bayes' Rule, the posterior probability that the source is located at q is

$$p = P(q, s \mid g_1, \cdots, g_N) = \frac{P(g_1, \cdots, g_N \mid q, s) P(q, s)}{P(g_1, \cdots, g_N)} \quad (2)$$

Since the denominator is a normalization constant, if assuming the prior $P(q, s)$ is uniform, then maximizing (2) becomes maximizing the likelihood $P(g_1, \cdots, g_N \mid q, s)$. With g_i is considered as independent random variable, it can be shown that

$$p' = \prod_{i=1}^{N} P(g_i \mid q, s) = \prod_{i=1}^{N} e^{-\int_{-T}^{T} \frac{[g_i(t + \tau_i) - s(t)]^2}{2\sigma_i^2} dt} \quad (3)$$

Substituting s with its maximum likelihood estimate given by

$$\hat{s} = \frac{1}{N} \sum_{i=1}^{N} g_i(t + \tau_i)$$

in (3), and assuming equal σ_i for all sensors, then taking the logarithm yields,

$$\log p' = -\sum_{i=1}^{N} \int_{-T}^{T} [g_i(t + \tau_i) - \hat{s}(t)]^2 dt = \frac{2}{N} V_C - \frac{N-1}{N} V_E \quad (4)$$

where

$$V_C = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \int_{-T}^{T} g_i(t + \tau_i) g_j(t + \tau_j) dt \quad (5)$$

is the accumulated correlation and

$$V_E = \sum_{i=1}^{N} \int_{-T}^{T} g_i^2(t + \tau_i) dt \quad (6)$$

is a constant representing the combined energy of the signals. Accordingly the estimated location can be found from the maximum of (5).

The theory of the algorithm is based on treating the cross correlation as an observational estimate of $P(g_1, \cdots, g_N \mid q, s)$ which is related to the posterior estimate as given in (1) under the same assumption for the prior probability $P(q, s)$ [4]. Thus by substituting the time difference $\tau_{ij}(q)$ between sensors i and j given as a function of source location q in General Cross Correlation (GCC) formula [5], the spatial likelihood function in the frequency domain for a pair of sensors is obtained as

$$SLF_{ij}(q) = \int_{-\infty}^{\infty} \Psi(f) G_i G_j^* e^{j2\pi f \tau_{ij}(q)} df \quad (7)$$

The advantage of (7) is that it allows for the filtering processes $\Psi(f)$ to be performed inclusively in the frequency domain.

3. ELM Algorithm

Given N sensors, the usable number of time differences is given by M none-repeated combinations of sensor pairs given by

$$M = C_2^N = \frac{N!}{2!(n-2)!} \tag{8}$$

With three sensors, three time differences are available. By adding a fourth sensor, the time delay difference is doubled to six. The ELM algorithm allows for sensors fusion by utilizing M time difference information to improve accuracy and robustness of the estimated location.

From the above theory and by introducing Hilbert envelop detection operator Θ, the proposed ELM algorithm for TAI can be formulated in compact form for both of time domain and frequency domain as follows and the algorithm architectures are shown in **Figure 2**.

$$ELM_t(x,y) = \sum_{\substack{i=1,j=2 \\ i \neq j \\ i,j \neq j,i}}^{N-1,N} \Theta \left[\int_{-\infty}^{\infty} g_i^{BPF}(t) g_j^{BPF}\left(t + \tau_{ij}(x,y)\right) dt \right] \tag{9}$$

$$ELM_f(x,y) = \sum_{\substack{i=1,j=2 \\ i \neq j \\ i,j \neq j,i}}^{N-1,N} \int_{-\infty}^{\infty} \Psi(f) G_i G_j^* e^{j2\pi f \tau_{ij}(x,y)} df \tag{10}$$

given,

$$\tau_{ij}(x,y) = \sqrt{(x-x_i)^2 + (y-y_i)^2} - \sqrt{(x-x_j)^2 + (y-y_j)^2} / v \tag{11}$$

where $g^{BPF} = h_{BPF}(t) \otimes g(t)$ is the band pass filtered signal resulting from the convolution of the signal g and the impulse response of the band pass filter $h_{BBF}(t)$. $\Psi(f)$ is the weighting filter. The estimated locations $\hat{x}, \hat{y}$ can then be found by locating the maximum of (9) or (10).

Considering the example of having four sensors located on the surface of a tangible object as in **Figure 1**, the theoretical time difference for all sensor pairs can be computed numerically from the hyperbola defined by (11). However, by summing the spatial likelihood from all pairs, the source location can be more reliably obtained from the maximum in the likelihood map as shown in **Figure 3**.

4. Filtering Process in ELM

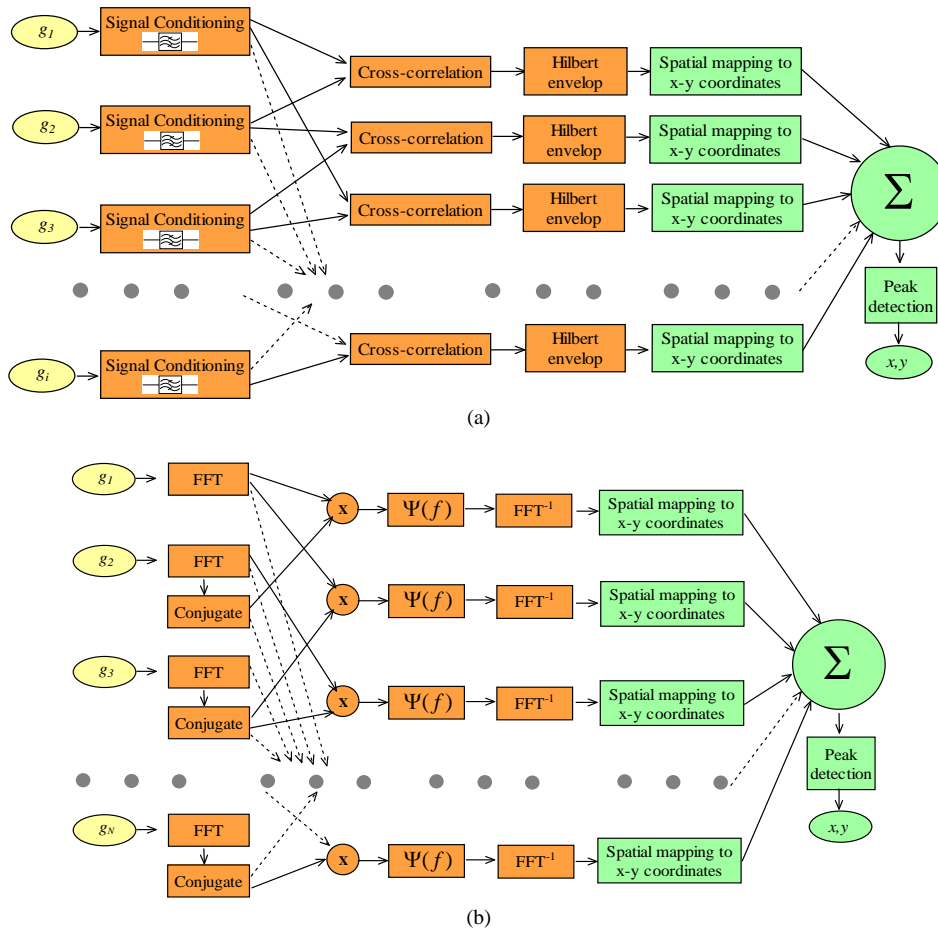In the previous section the ELM algorithm is verified for

(a)

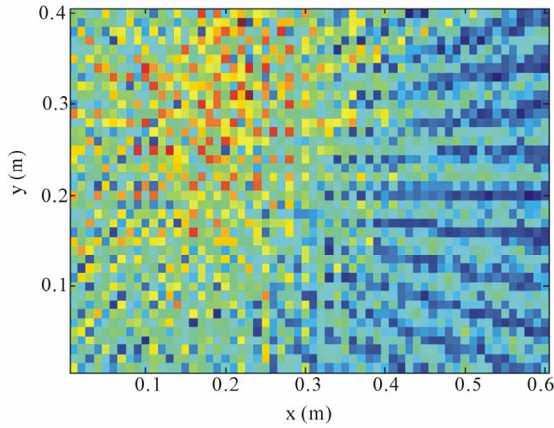(b)

Figure 2. Algorithm diagram for (a) ELM_t **and (b)** ELM_f.

Figure 3. Spatial Likelihood of the source at (0.2 m, 0.3 m).

Figure 4. Spatial Likelihood of the source at (0.2 m, 0.3 m) using filtered signals.

Figure 5. Spatial likelihood map using PHAT process.

TAI application using raw signals. However, the localization accuracy and robustness can be further improved by use of various filters. The purpose of the pre-filtering is to remove the noise from the signal as a result of low frequency components and the high frequency components from the non linear response of the sensors. The use of the popular IIR filter found to be adequately successful. The designed digital filter is a 10th order band-pass Elliptic filter with lower cut off frequency of 500 HZ and upper cut off frequency of 8 KHz. It is clear that pre-filtering has significantly improved the reliability of the estimation as can be seen from the smoothness achieved in the likelihood map in **Figure 4** where the local maximum becomes more distinctive compared to the multiple peaks in **Figure 3** using just raw signals.

The second filtering type employed here is the Phase Transform (PHAT) given by

$$\Psi_{PHAT} = \frac{1}{|X_1(f)||X_2(f)|} \tag{12}$$

This PHAT processor performs well in a moderately reverberant room. It has been used extensively to localize acoustic source in a room [6] and in robotics applications [7]. This is achieved by substituting the filtering process $\Psi_{PHAT}(f)$ given in (12) into (10). The resulting map of ELM_f produced for the same signals used for generating **Figure 3** is shown in **Figure 5**. It is apparent that sharper peak is obtained compared to the pre-filtering method in the time domain. A significant advantage of using ELM_f over ELM_t is that PHAT process doesn't require any design parameters, while the pre-filtering in ELM_t requires knowledge of the dominant signal components and noise which is normally obtained by analyzing the signals. That means if these parameters have been considerably changed as a result in changing the object material for example, the filter of ELM_t (IIR, FIR or wavelet) has to be redesigned but it doesn't for PHAT.
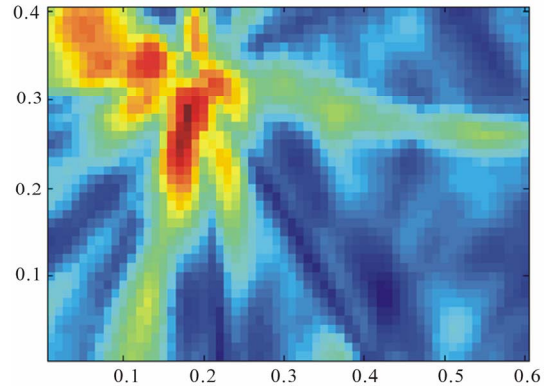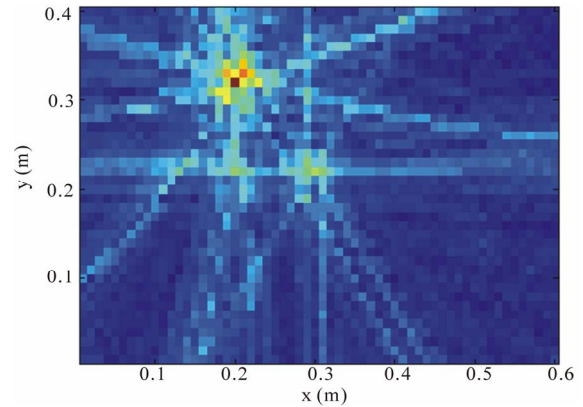
5. Temporal Smoothing

Further enhancement in the ELM algorithm is achievable by treating the dispersion effect in solids. Theoretically, in non dispersive multiple input system the output of the cross correlation reaches the maximum at time lag equal to the time difference between the arrival of the input signals. On the other hand, in dispersive system, where the wave propagation velocity is a function of frequency, the output peak of the cross correlation envelop occurs at the time lag equals to the group delay of the wave [8]. This fact can be interpreted in practice using Hilbert transform.

The analytical signal of a given function $z(t)$ is defined by

$$Z(t) = z(t) + j\tilde{z}(t) \tag{13}$$

where the imaginary part in (13) is the Hilbert transform of $z(t)$ given by

$$\hat{z}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{z(\zeta)}{t - \zeta} d\zeta \tag{14}$$

the envelope function of $z(t)$ is defined as

$$\ell(t) = \left[z^2(t) + \hat{z}^2(t) \right]^{1/2} \qquad (15)$$

The Equation (15) can be used in (9) or (10) to reduce the error in cross correlation caused by dispersion.

To visualize the difference between algorithms and the effect of dispersion treatment, the ELM algorithm is applied to impact signals shown in **Figure 6** and scratch signals shown in **Figure 7**. With raw signals the algorithm produces multiple peaks in the likelihood map with several sharp local maxima comparable to the global maximum as shown in **Figures 6(a)-7(a)**. By conditioning the input signals, the local maxima are compressed as shown in **Figures 6(b)-7(b)**. When Hilbert envelop is applied, it is observable that the peak is enhanced by shifted local maxima towards the global peak and the overall ELM surface is smoothed as shown in **Figures 6(c)-7(c)**. It is clear from **Figures 6(d)-7(d)** that PHAT process produces sharper peak with lower side lobes. The Temporal Smoothing has significantly improved the ELM results as seen from the enhanced global maximum and reduced local maxima.

Although scratch signals produce more local maxima than the impact signals, the proposed ELM algorithm has significantly improved the results as seen from the enhanced global maximum and reduced local maxima. The result shows that Hilbert envelope can be regarded as an effective temporal smoothing filter, and has considerably better improvement on revealing the global peak when used with pre-filtering.

6. Time Difference Based Localisation

The accuracy of the time difference based localization vastly depends on the level of error in the time difference values. Therefore it becomes crucial to develop a reliable algorithm to estimate time differences with less error as possible. An efficient algorithm is developed for estimating time differences based on spectral estimation.

6.1. Linear Cross Spectral Phase

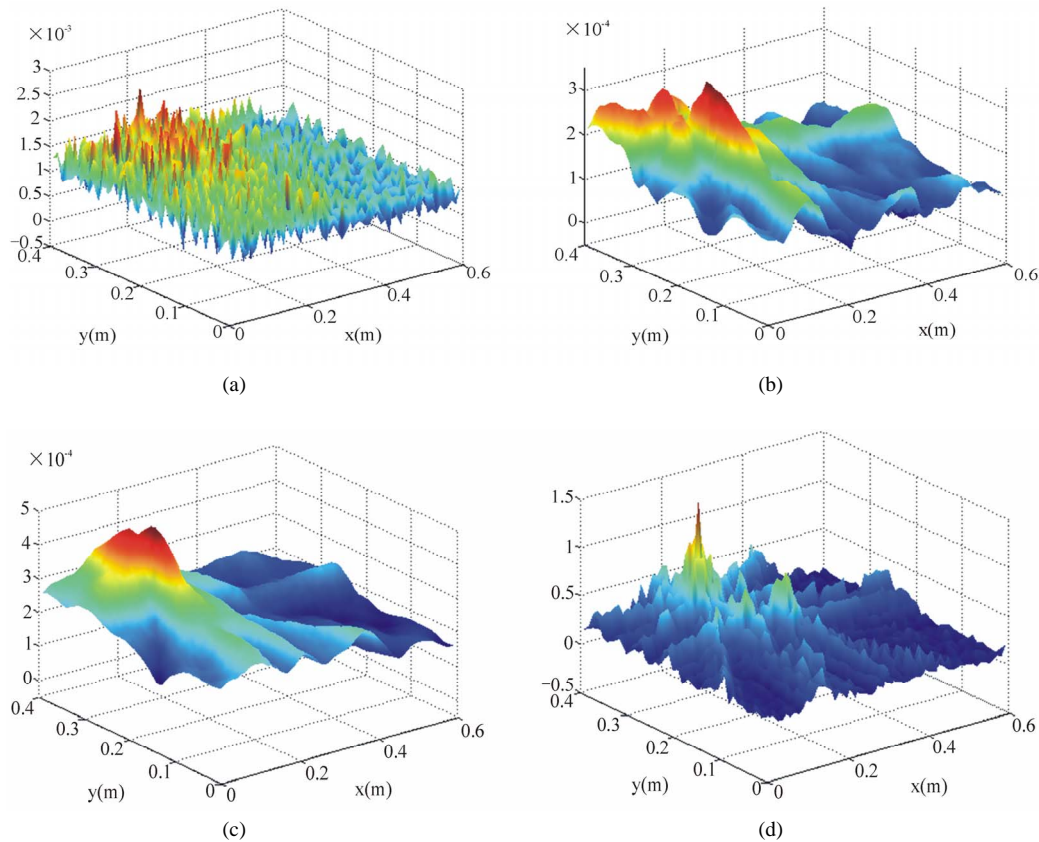The classical time difference estimation can be improved by pre filtering the signals or applying the most popular

Figure 6. ELM of impact signals using (a) raw signals, (b) conditioned signals, (c) as in (b) with Hilbert envelop and (d) PHAT.
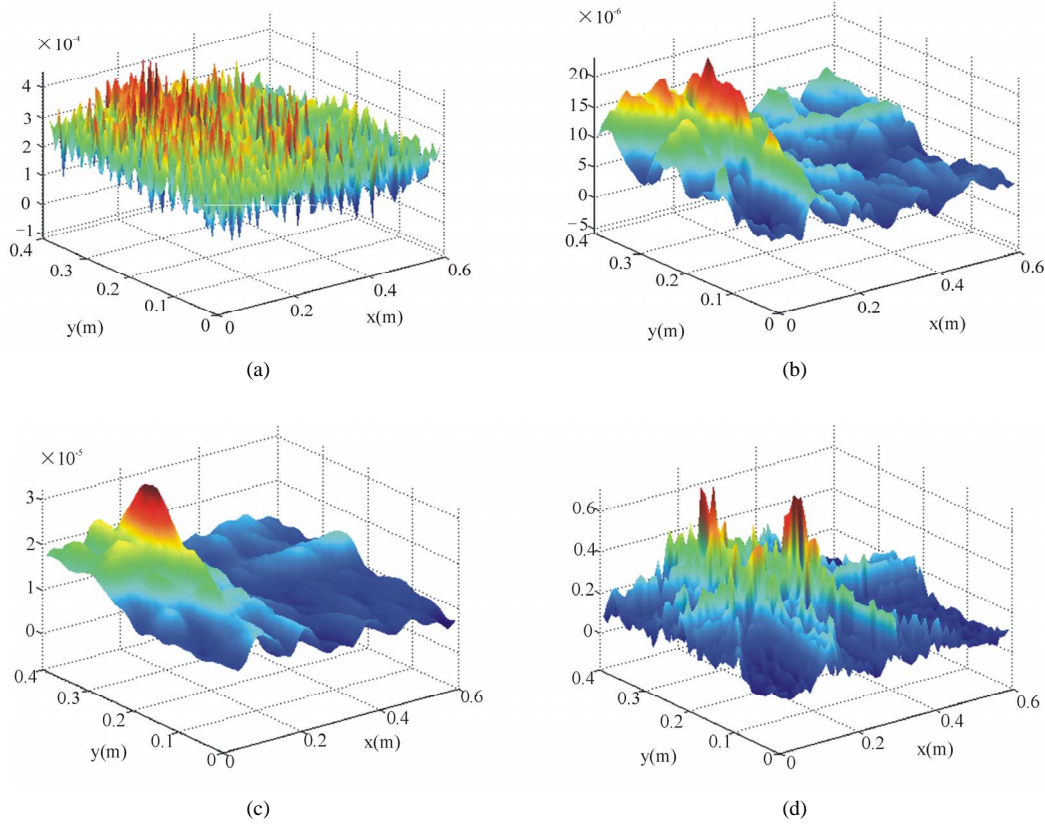
Figure 7. ELM of scratch signals using (a) raw signals, (b) conditioned signals, (c) as in (b) with Hilbert envelop and (d) PHAT.

PHAT process using GCC which involve filtering in the frequency domain then returning to the time domain to extract the time difference (9) and (10). An alternative method for estimating the time difference is the Linear Cross Spectral Phase (LCSP). The LCSP algorithm estimates the time difference entirely in the frequency domain making the estimation process more efficient and robust than the time domain algorithms particularly for tracking a continuous source.

Let the received signal $g(t)$ assumed a broad sense stationary process. The cross spectral density of signals $g_i(t)$ and $g_j(t)$ can be found from

$$P_{ij}(f) = G_i(f)G_j^*(f) \qquad (16)$$

where $G(f)$ if the Fourier transform of $g(t)$. Since $g_j(t)$ is time delayed from $g_i(t)$ by τ, then in terms of the auto spectral density $A_{ii}(f)$ of $g_j(t)$, Equation (16) can be expressed by

$$P_{ij}(f) = A_{ii}(f)e^{-j2\pi f\tau} = \left|P_{ij}(f)\right| \angle \phi(f) \qquad (17)$$

The time difference τ appears only in the phase angle ϕ of (17) as linear function of the frequency f. Since the group velocity is used to compute the length difference, the group delay must be extracted from the phase

function in (17) as given by [9]

$$\tau = -\frac{\mathrm{d}\phi(f)}{\mathrm{d}f} \qquad (18)$$

The cross spectrum and auto spectrum functions can be effectively estimated based using Short Time Fourier Transform (STFT). Then Equation (18) can be computed numerically for the quantities given in samples using linear regression of the form

$$\sum (f_i - \bar{f})\phi_i / \sum (f_i - f) f_i \quad [10,11].$$

6.2. Maximum Likelihood Positioning

Given M pair of sensors, the Maximum Likelihood algorithm (ML) proposed here for TAI can handle the error by minimizing the error between the given time difference $\hat{\tau}_m$ of the mth pair and the ideal time difference $\tau_m(q)$ associated with the searched location q. If the estimated time differences is modeled by the random variable $\hat{\tau}_m = \tau_m + e_m$ where e_m is zero-mean additive white Gaussian noise with known standard deviation σ_m, then by assuming the time differences from each pair of sensors are statistically independent, the likeli-

hood function can be expressed by the conditional probability density function given by [12]

$$p\left(\hat{\tau}_1,\cdots,\hat{\tau}_M \mid q\right) = \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{\frac{-[\hat{\tau}_m - \tau_m(q)]^2}{2\sigma_m^2}} \qquad (19)$$

taking the log of both sides of (19) yield

$$\ln\left(p\left(\hat{\tau}_1,\cdots,\hat{\tau}_M \mid q\right)\right) =$$
$$-\frac{1}{2}\sum_{m=1}^{M}\ln\left(2\pi\sigma_m^2\right) - \sum_{m=1}^{M}\frac{\left(\hat{\tau}_m - \tau_m(q)\right)^2}{2\sigma_m^2} \qquad (20)$$

The ML estimation of location q is the position that maximizes the likelihood function (20) or equivalently that minimizes the second term since the first term is not a function of q which results in the following localization criterion

$$J_{ML}(q) = \arg\min_q \left(\sum_{m=1}^{M} \frac{\left[\hat{\tau}_m - \tau_m(q)\right]^2}{\sigma_m^2} \right) \qquad (21)$$

Here J_{ML} is a weighted least error estimator. If no statistics considered or σ_m is the same for all sensor pairs, then the denominator is constant and (21) is reduced to the following formula

$$\hat{J}_{ML}(q) = \arg\min_q \left(\sum_{m=1}^{M} \left[\hat{\tau}_m - \tau_m(q)\right]^2 \right) \qquad (22)$$

A significant difference between ML and ELM can be observed that ML doesn't suffer from side lobes but on the cost of sharpness which means that the ML algorithm provides more stability while ELM algorithm provides higher accuracy. The similarity between the ML maps of the impact and scratch signals is due to the dependence of the ML algorithm on the time differences already estimated not on the signal themselves.

7. Conclusions

In this paper, the in-solid acoustic source localization is developed based on measuring the time difference of arrivals between spatially separated sensors. For efficient operation of TAI with this approach, two methods are proposed for the source localization. In the one-step method the ELM performs the localization based on two algorithms, one encounters time domain processing with conventional post filtering and the other employs PHAT filtering in the frequency domain. In the two-steps method, TDOA values are found first using either GCC or LCSP then based on these values the source is localized using ML algorithm. The effect of dispersion is treated by introducing Hilbert envelop smoothing. A criterion is proposed to detect outlier estimations that can

happen from domestic noise as door shut.

9. References

[1] W. Rolshofen, D. T. Pham, M. Yang, Z. Wang, Z. Ji and M. Al-Kutubi, "New Approaches in Computer-Human Interaction with Tangible Acoustic Interfaces," *IPROMs* 2005 *Virtual Conference*, May 2005.

[2] X. Wang, Z. Wang and B. O'Dea, "A TOA-Based Location Algorithm Due to NLOS Propagation," *IEEE Transactions on Vehicular Technology*, Vol. 52, No.1, January 2003, pp. 112-116.

[3] S. Birchfield and D. Gillmor, "Fast Bayesian Acoustic Localization," *Proceedings of the IEEE International Conference on Speech and Signal Processing*, Florida, Vol. 2, May 2002, pp. 1793-1796.

[4] P. Aarabi and S. Zaky, "Robust Sound Localization Using Multi-Source Audiovisual Information Fusion," *Elsevier*, *Information Fusion*, Vol. 2, No. 3, 2001, pp. 209-223. doi:10.1016/S1566-2535(01)00035-5

[5] C. Knapp and G. Carter, "The Generalized Correlation Method for Estimation of Time Delay", *IEEE Transactions on Acoustics*, *Speech, & Signal Processing* Vol. 24, No. 4, 1976, pp. 320-327. doi:10.1109/TASSP.1976.1162830

[6] M. Omologo and P. Svaizer, "Acoustic Localization in Noisy and Reverberant Environment Using CSP Analysis," *Proceedings of the IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing Conference*, Atlanta, Vol. 2, 7-10 May 1996, pp. 921-924.

[7] J. Valin, F. Michaud, J. Rouat, D. LCtoumeau, "Robust Sound Source Localization Using a Microphone Array on a Mobile Robot," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, 27-31 October 2003, Vol. 2, pp. 1228- 1233.

[8] J. Bendat and A. Piersol, "Random Data," John Wiley & Sons, 1986.

[9] A. Mertins, "Signal Analysis," John Wiley & Sons, 1999.

[10] L. Danfeng and S. Levinson, "A Linear Phase Unwrapping Method for Binaural Sound Source Localization on a Robot," *Proceedings of the IEEE Conference on Robotics and Automation*, Washington, May 2002, pp. 19-23.

[11] H. Poor, "An introduction to Signal Detection and Estimation," 2nd Edition, Springer, New York, Berlin, Heidelberg, Hong Kong, London, Milan, Paris, Tokyo, 1994.

[12] N. Gershenfeld, "The Nature of Mathematical Modeling", Cambridge University Press, Cambridge, 1999.