AN EXPERIMENTAL COMPARISON OF TIME DELAY WEIGHTS FOR DIRECTION OF ARRIVAL ESTIMATION

Tuomo W. Pirinen

Dept. of Signal Processing
Tampere University of Technology
Tampere, Finland
tuomo.pirinen@tut.fi

ABSTRACT

When direction of arrival is estimated using time differences of arrival, the estimation accuracy is determined by the accuracy of time delay estimates. Probability of large errors increases in poor signal conditions and reverberant conditions pose a significant challenge. To overcome the problems, reliability criteria for time delays and weighted least squares direction estimation have been proposed. This work combines these approaches, and compares several weight criteria for single-frame estimation experimentally. Testing is conducted on different types of audio signals in a loudspeaker experiment. As a result, an optimum combination of weights is found, whose performance exceeds earlier proposals and iterated weighting. Furthermore, the optimum weighting is not dependent on the source signal type, and the best weights are the ones that do not require information about the underlying time delay estimator.

1. INTRODUCTION

Direction of arrival (DOA) estimation is an essential part of signal processing in many array systems. Recently, use of acoustic signals and microphone arrays for surveillance [1] and multimedia applications [2] has received attention. For example, tracking of speakers is subject to special interest, due to the large number of potential applications. Applications within speech domain include, for example, automated camera steering, speech enhancement, and various diarization functions, such as segmentation [3].

In time difference of arrival (TDOA) based DOA estimation, time delays between signal pairs are first estimated, and DOA is then computed from these estimates. In comparison to other methods, e.g., steered-response [4] and parametric methods [5, 6], TDOA-based methods are best suited for single source scenarios involving large time-bandwidth products. TDOA based estimation does not restrict the array geometry and even small or arbitrarily shaped arrays are feasible. In addition, TDOA based estimation can be conducted on limited resources, e.g., one-bit sampling [7], or within a customized integrated circuit [8].

Accuracy of TDOA based DOA estimation is directly dependent on the accuracy of delay estimation. This is because errors in TDOA estimates propagate to the DOA estimation stage. If DOA estimator treats all delay estimates equally, e.g., to form a least squares (LS) solution, even a single outlier can cause a large error to the DOA estimate. In poor signal conditions, the probability of large errors in time delay estimation increases. Theoretical studies clearly demonstrate a threshold effect related to SNR decrease [9], and experiments have confirmed the drastic effects of reverberation [10].

The problems involved in TDOA estimation have motivated the use of reliability measures. These methods aim to evaluate the quality of delay estimates, and reject or weight the estimates accordingly. In [11], TDOA was estimated by performing regression in cross-spectral phase domain, and the residual error was used as criterion for the estimate quality. Dependency of multiple TDOA estimates was utilized in [12, 13] as a confidence and selection measure. Two reliability criteria computed from the cross-correlation of sensor signals were proposed in [14]. A similar criterion, and a constant weighting based value of TDOA, were utilized as weights in LS solution in [15]. The latter were found to enhance the performance compared to an unweighted solution.

This article examines weights for the LS solution and compares them experimentally. The research expands the approach of [15] by using more candidate weights and a more comprehensive experiment set-up. Weights are evaluated using different types of audio signals in a reverberant room. Results show that a proper combination of weights increases the performance of single-frame DOA estimation. Furthermore, optimal weighting is not dependent on the signal content.

The remainder of the paper is organized as follows. Section 2 explains the weighted DOA estimation method and weight candidates are introduced in Section 3. Experiment set-up for weight comparison is described in Section 4, followed by a discussion of results in 5. Concluding remarks are given in Section 6.

2. DIRECTION OF ARRIVAL ESTIMATION

Given a planar wave passing a pair of sensors located at $\mathbf{p}_1$ and $\mathbf{p}_2$, the resulting TDOA is given by

$$\tau = \frac{\|\mathbf{p}_2 - \mathbf{p}_1\|}{c}\cos(\theta) \tag{1}$$

where c is the wave propagation speed. Angle θ is between vector $\mathbf{p}_2 - \mathbf{p}_1$ and direction of wave front propagation. Letting $\mathbf{x}_{1,2} = \mathbf{p}_2 - \mathbf{p}_1$, Eqn. (1) simplifies to [16]

$$\tau = (\mathbf{p}_2 - \mathbf{p}_1)^{\mathrm{T}}\mathbf{k} = \mathbf{x}_{1,2}^{\mathrm{T}}\mathbf{k} \tag{2}$$

where $\mathbf{k}$ is the propagation vector of the planar wave. This vector has the direction of wave front propagation and magnitude $1/c$. Using an array of sensors, and thus several sensor pairs, all TDOA are given by

$$\boldsymbol{\tau} = \mathbf{X}\mathbf{k} \tag{3}$$

where $\boldsymbol{\tau}$ is a vector of TDOAs and $\mathbf{X}$ is a matrix of row vectors $\mathbf{x}_{i,j}^{\mathrm{T}}$. Consequently, direction of arrival can be estimated from (3) by LS inversion [7]. Similar estimation methodologies have also been proposed in [17, 15, 18]. Advantage of the inversion is that knowledge on the signal propagation speed is not required.

More specifically, in [15], a weighted least squares (WLS) solution was used to enhance robustness against delay estimation errors. The WLS solution to (3) is

$$\hat{\mathbf{k}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\hat{\tau} \tag{4}$$

where $\mathbf{W}$ is the weight matrix. If errors in TDOA estimates are zero mean, (4) is the best linear unbiased estimator when $\mathbf{W}$ is the inverse of the error covariance matrix [19]. In practice, the statistics of TDOA errors are unknown, nonstationary, and thus difficult to utilize in (4). This has motivated the use of single-frame confidence measures, which are measured from the time delay estimates or the (correlation) function used in estimation.

3. TIME DELAY WEIGHT CANDIDATES

An estimate of time delay between two windowed sensor signals is obtained by computing a similarity measure between the signals, and locating the delay that maximizes (or minimizes) the measure. There are several possible methods of estimation [11, 20, 21, 22, 23], and the generalized cross correlation (GCC) family of methods [24] is especially well known. The phase transform variant (GCC-PHAT) has been popular in recent works, and found to have some robustness against reverberation [4, 15, 25].

This work compares the weights listed below. More details on the weights and their development can be found from the given references. In weight definitions, T_i is the maximum possible TDOA in the i-th sensor pair. All weight matrices are diagonal, and notation $q_{i,i}^{(k)}$ refers to the i-th diagonal element of weight matrix $\mathbf{Q}_k$.
$\mathbf{Q}_1$: Quadratic weighting of TDOA [15]

$$q_{i,i}^{(1)} = 1 - \frac{\min(T_i^2, \tau_i^2)}{T_i^2}. \tag{5}$$

$\mathbf{Q}_2$: Cosine weighting of TDOA, as an alternative to $\mathbf{Q}_1$

$$q_{i,i}^{(2)} = \frac{1}{2\pi}\cos^{-1}\left(\frac{\min(T_i, |\tau_i|)}{T_i}\right). \tag{6}$$

$\mathbf{Q}_3$: Ratio of two largest peaks in GCC-function, denoted by c_{i1} and c_{i2} [15, 26]

$$q_{i,i}^{(3)} = 1 - \max\left(\frac{c_{i2}}{c_{i1}}, 0\right). \tag{7}$$

$\mathbf{Q}_4$: Value of the largest peak in GCC-function [26]

$$q_{i,i}^{(4)} = \max\left(c_{i1}, 0\right). \tag{8}$$

$\mathbf{Q}_5$: Confidence factor based on TDOA dependency [13]

$$q_{i,i}^{(5)} = \frac{1}{N-2}\sum_{\substack{n=1\\ n\neq a,b}}^{N} \frac{|\hat{\tau}_{a,b} + \hat{\tau}_{b,n} + \hat{\tau}_{n,a}|}{2/c\left(\|\mathbf{x}_{a,b}\| + \|\mathbf{x}_{b,n}\| + \|\mathbf{x}_{n,a}\|\right)}. \tag{9}$$

In (9), $\hat{\tau}_{a,b}$ and $\mathbf{x}_{a,b}$ denote the TDOA estimate and sensor vector, respectively, between sensors a and b, which are the sensors corresponding to the i-th sensor pair. See [13] for more information.

Table 1: *RMS errors (degrees) of WLS DOA estimation.*

Weight ($\mathbf{W}$)	Noise	Music	Speech
$\mathbf{Q}_1\mathbf{Q}_3\mathbf{Q}_4\mathbf{Q}_5$	2.96	11.65	20.71
$\mathbf{Q}_1\mathbf{Q}_3\mathbf{Q}_4$	3.08	11.93	21.11
$\mathbf{Q}_2\mathbf{Q}_3\mathbf{Q}_4\mathbf{Q}_5$	3.09	12.12	21.34
$\mathbf{Q}_2\mathbf{Q}_3$	3.19	12.25	21.34
$\mathbf{Q}_2\mathbf{Q}_3\mathbf{Q}_5$	3.21	12.26	21.39
$\mathbf{Q}_1\mathbf{Q}_3$	3.23	12.34	21.58
$\mathbf{Q}_2\mathbf{Q}_3\mathbf{Q}_4$	3.23	12.41	21.68
$\mathbf{Q}_1\mathbf{Q}_3\mathbf{Q}_5$	3.25	12.41	21.68
$\mathbf{Q}_1\mathbf{Q}_2\mathbf{Q}_3\mathbf{Q}_4\mathbf{Q}_5$	3.47	12.45	21.70
$\mathbf{Q}_1\mathbf{Q}_2\mathbf{Q}_3\mathbf{Q}_4$	3.48	12.57	21.74
$\mathbf{Q}_1\mathbf{Q}_2\mathbf{Q}_3\mathbf{Q}_5$	3.50	12.60	21.77
$\mathbf{Q}_1\mathbf{Q}_2\mathbf{Q}_3$	3.51	12.66	21.77
$\mathbf{Q}_2\mathbf{Q}_3\mathbf{Q}_5$	3.53	12.71	21.77
$\mathbf{Q}_1$	3.55	12.78	21.77
$\mathbf{Q}_2$	3.59	13.03	21.77
$\mathbf{Q}_5$	3.63	13.14	21.77
IRLS	3.65	13.14	21.77
LS ($\mathbf{I}$)	3.66	13.14	21.77
$\mathbf{Q}_4$	3.68	13.14	21.77
$\mathbf{Q}_3$	3.92	14.15	25.32

Weights $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are based on the cosine relation (1). Assuming that error in TDOA is relatively small (considerably less than $T_i/2$), its effects are more severe when the error occurs near the maximum delay values [15]. Thus, the solution favors TDOA estimates that are closer to zero.

4. EXPERIMENT SET-UP

The proposed weights were tested using data recorded in a hall-like room with a 0.36 m four microphone tetrahedron array. Eight loudspeaker locations, in approximate elliptic fashion around the array, were used as sources. Reverberation time varied between 0.46–0.56 s, depending on the locations of loudspeakers and microphones.

Twelve test signals used consisted of white noise, music, and speech from the TIMIT database. Each test signal was 15 seconds in duration and was separately played through each of the loudspeakers. This provided a total of approximately 29 minutes of test audio. Depending on the signal content, SNR varied between 0–25 dB. More details on the recording setup can be found in [27].

TDOAs were estimated using GCC-PHAT on 8192 sample windows with 50% overlap, and interpolated beyond the resolution allowed by the sampling rate using parabolic interpolation [28]. Secondary peaks needed in $\mathbf{Q}_3$ weighting were extracted using the three-point-peak method described in [15], and also interpolated. DOA was estimated from DOA using (4). As a comparison point to iterative methods, DOA was also estimated using iterative reweighted least squares (IRLS) [29].

DOA estimation accuracy is measured as the angular RMS error. Table 1 lists the accuracies achieved using weights and their combinations, separately for each signal type. For brevity, performances of weight combinations are listed only for combinations which provided better performance than any of the weights alone.

5. RESULTS AND DISCUSSION

From the results in Table 1, it is clear that DOA estimation accuracy is dependent on the source signal type. Noise is an ideal signal, because its autocorrelation function is an impulse. The signal is also continuous, and thus estimation frames are filled by it. Consequently, the best performance is achieved on noise, and absolute improvement is at most only $0.69°$ RMS.

Music and speech signals have a more complicated autocorrelation and spectra. In addition, speech signals are not continuous, but contain intermittent pauses, and thus the frames are not completely filled by the signal. This is observed as a lesser estimation accuracy, because the signals are disturbed more by reverberation effects. However, larger absolute improvements can be achieved by weighting: $1.49°$ RMS for music and $1.08°$ RMS for speech. The best combined performance is achieved with combination $\mathbf{Q}_1\mathbf{Q}_3\mathbf{Q}_4\mathbf{Q}_5$. Relative improvements in comparison to least squares are 19%, 11%, and 5%, for noise, music, and speech signals, respectively. The results are conclusive and valid for comparing the weights, because the weight ranking is the same for all signal types. Thus, the performance of weights relative to each other is independent of signal type.

Of the individual features, $\mathbf{Q}_1$ is the optimal. This is interesting, because $\mathbf{Q}_1$ does not depend on the apparent error in the delay estimate, but only assigns a weight according to the estimated delay value. However, performance improvement from $\mathbf{Q}_1$ alone is small. It does not give any improvement for speech signals, and the improvements are minor for noise ($0.10°$ RMS) and music ($0.36°$ RMS).

Combination $\mathbf{Q}_1\mathbf{Q}_3$ proposed in [15] is good, but performance is further improved by including $\mathbf{Q}_4$ and $\mathbf{Q}_5$ to the weight combination. The results also confirm that combination $\mathbf{Q}_1\mathbf{Q}_3$ is better than either of the weights alone as suggested, but not verified in [15]. Interestingly, replacing $\mathbf{Q}_1$ in this pair with cosine weight $\mathbf{Q}_2$ yields a slightly better result. However, the best combination $\mathbf{Q}_1\mathbf{Q}_3\mathbf{Q}_4\mathbf{Q}_5$ uses $\mathbf{Q}_1$ instead of $\mathbf{Q}_2$.

IRLS estimation does not improve the results. This is due to the small number of TDOAs (six in a four sensor array), and the fact that all TDOA are prone to error. Linear regressors are limited by a 50% breakdown probability, and thus the regression fit may be bad if all delay estimates contain large errors. Therefore, assigning the weights by the residual distance from the fit is not helpful.

Features $\mathbf{Q}_3$ and $\mathbf{Q}_4$, when used alone, actually degrade the performance. But when combined with other weights, they improve the performance even further. For example, $\mathbf{Q}_1\mathbf{Q}_5$ is worse than $\mathbf{Q}_1$ alone (and thus not displayed in Table 1), but $\mathbf{Q}_1\mathbf{Q}_3\mathbf{Q}_4\mathbf{Q}_5$ is the best weighting.

Weights $\mathbf{Q}_1$, $\mathbf{Q}_2$, and $\mathbf{Q}_5$ utilize only the delay estimates but not the values of the underlying estimation function. They can be used with any delay estimator, not just GCC-PHAT. This is important, because there is not a single optimum delay estimator for all scenarios [30], but the estimator has to be selected by the application. Weights $\mathbf{Q}_3$ and $\mathbf{Q}_4$ rely on the peak values of the GCC-function, and thus their performance depends on the estimation function as well as the peak extraction method. Their usability in conjunction with other delay estimation methods should be studied further. $\mathbf{Q}_5$ -weight is only slightly worse than $\mathbf{Q}_1$ and $\mathbf{Q}_2$. It has a further advantage that delay values close to maximum values can also be utilized, whereas $\mathbf{Q}_1$ and $\mathbf{Q}_2$ always assign a small weight to estimates close to endpoints of the delay range.

The results demonstrate that the performance of single-frame DOA estimation can be improved by utilizing weights on delay estimates. The weights are computationally light, and the increase in computational load from the WLS solution is negligible in comparison to the demands of time delay estimation. Previous research works have shown that larger performance improvements can be achieved by utilizing multiple estimation frames and temporal tracking of delay estimates [15], DOA estimates, or confidence values [27]. However, a more accurate single-frame estimate can be achieved by weighting, and this is also helpful for further tracking.

As a concluding comment, we would like to point out the relation between the optimal individual weight $\mathbf{Q}_1$, and the array design results of [31]. In [31], it was derived that in an optimum array geometry for localization, the sensor vectors should be as independent as possible. This is equivalent to designing the array such that the number of TDOA having their values close to zero is maximized regardless of the source direction. As observed with $\mathbf{Q}_1$, such delays provide better DOA estimation accuracy.

6. CONCLUSIONS

This research compared different weightings of time delays in direction of arrival estimation. The weights were incorporated into a weighted least squares solution, and estimation accuracy was tested using an extensive set of audio signals. It was found that an optimum combination of weights exists and yields better single-frame estimation accuracies than previously proposed weights. Weighting is not dependent on the signal type, and the best individual weights are also independent of the time delay estimator.

7. ACKNOWLEDGMENTS

The author would like to thank Prof. David Wessel for providing the CNMAT facilities for recording the test data, and Dr. Konsta Koppinen and Dr. Michael Zbyszyński for their assistance during the recordings.

8. REFERENCES

[1] J. C. Chen, K. Yao, and R.E. Hudson, "Source localization and beamforming," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 30–39, 2002.

[2] Y. Huang, J. Benesty, and G. W. Elko, "Source localization," in *Audio Signal Processing for Next Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds. Kluwer Academic Publishers, 2004.

[3] G. Lathoud and I. McCowan, "Location based speaker segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, pp. 176–179.

[4] Joseph DiBiase, H. Silverman, and M. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds., pp. 157–180. Springer-Verlag, 2001.

[5] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996.

[6] A. Zeira and B. Friedlander, "Direction of arrival estimation using parametric signal models," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 339–350, 1996.

[7] J. Yli-Hietanen, K. Kalliojärvi, and J. Astola, "Low-complexity angle of arrival estimation of wideband signals using small arrays," in *Proc. IEEE Signal Process. Workshop Stat. Signal, Array Signal Process.*, 1996, pp. 109–112.

[8] D. Halupka, N. Mathai, P. Aarabi, and A. Sheikholeslami, "Robust sound localization in 0.18 μm CMOS," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2243–2250, 2005.

[9] Brian M. Sadler and Richard J. Kozick, "A survey of time delay estimation performance bounds," in *Proc. IEEE Sensor Array, Multichan. Signal Process. Workshop*, 2006, pp. 282–288.

[10] B. Champagne, S. Bédard, and A. Stéphenne, "Performance of time-delay estimation in the presence of room reverberation," *IEEE Trans. Speech Audio Process.*, vol. 4, no. 2, pp. 148–152, 1996.

[11] M. Brandstein and H. Silverman, "A practical methodology for speech source localization with microphone arrays," *Computer Speech & Language*, vol. 11, no. 2, pp. 91–126, 1997.

[12] J. Yli-Hietanen, K. Koppinen, and J. Astola, "Time-delay selection for robust angle of arrival estimation," in *Proc. IASTED Int. Conf. Signal Image Process.*, 1999, pp. 81–83.

[13] T. Pirinen, "Normalized confidence factors for robust direction of arrival estimation," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2005, vol. 2, pp. 1429–1432.

[14] D. Bechler and K. Kroschel, "Reliability criteria evaluation for TDOA estimates in a variety of real environments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, vol. IV, pp. 985–988.

[15] K. Varma, T. Ikuma, and A. A. Beex, "Robust TDE-based DOA-estimation for compact audio arrays," in *Proc. IEEE Sensor Array, Multichan. Signal Process. Workshop*, 2002, pp. 214–218.

[16] S. Haykin, J. Justice, N. Owsley, J. Yen, and A. Kak, *Array signal processing*, Prentice-Hall, 1985.

[17] R. Blumrich and J. Altmann, "Medium range localisation of aircraft via triangulation," *Applied Acoustics*, vol. 61, no. 1, pp. 65–82, 2000.

[18] G. Rosen and P. Hasler, "Chemical source localization in unknown turbulence using the cross-correlation method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, pp. 1116–1119.

[19] Jerry M. Mendel, *Lessons in estimation theory for signal processing, communications, and control*, Prentice-Hall, 1995.

[20] F. Viola and W. F. Walker, "Comparison of time delay estimators in medical ultrasound," in *IEEE Ultrason. Symposium*, 2001, pp. 1485–1488.

[21] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 525–533, 1993.

[22] J. Chen, Y. A. Huang, and J. Benesty, "Performance of GCC- and AMDF- based time delay estimation in practical reverberant environments," *EURASIP J. Applied Signal Process.*, vol. 2005, no. 1, pp. 25–36, 2005.

[23] J. Benesty, Y. Huang, and J. Chen, "Time delay estimation via minimum entropy," *IEEE Signal Process. Lett.*, vol. 14, no. 3, pp. 157–160, 2007.

[24] C. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.

[25] M. Omologo and P. Svaizer, "Use of the crosspower-spectrum phase in acoustic event location," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 3, pp. 288–292, 1997.

[26] D. Bechler and K. Kroschel, "Confidence scoring of time difference of arrival estimation for speaker localization with microphone arrays," in *13. Konferenz "Elektronische Sprachsignalver-arbeitung (ESSV)*, 2002.

[27] T. Pirinen and A. Visa, "Signal independent wideband activity detection features for microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, vol. IV, pp. 1109–1112.

[28] X. Lai and H. Torp, "Interpolation methods for time-delay estimation using cross-correlation method for blood velocity measurement," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 46, no. 2, pp. 277–290, 1999.

[29] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Commun. Statist. — Theor. Meth.*, vol. A6, no. 9, pp. 813–827, 1977.

[30] A. Hero, H. Messer, J. Goldberg, D. Thomson, M. Amin, G. Giannakis, A. Swami, J. Tugnait, A. Nehorai, A. Swindlehurst, J.-F. Cardoso, Lang Tong, and J. Krolik, "Highlights of statistical signal and array processing," *IEEE Signal Process. Mag.*, vol. 15, no. 5, pp. 21–64, 1998.

[31] B. Yang and J. Scheuing, "Cramer-Rao bound and optimum sensor array for source localization from time differences of arrival," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. 961–964.