LOCATION AND POWER ESTIMATION OF MULTIPLE SOURCES BY NONLINEAR REGRESSION

U. Sandkühler and J.F. Böhme

Lehrstuhl für Signaltheorie, Ruhr-Universität Bochum
D-4630 Bochum, Fed. Rep. Germany

ABSTRACT

The problem of estimating certain parameters of sources generating signals by suitable array processing methods is investigated. The unknown source parameters are estimated by maximizing conditional likelihood functions which require to solve a nonlinear regression problem. Numerical experiments are used to determine the statistical properties of those estimates and to compare their variances with the theoretical Cramer Rao lower bound.

I. INTRODUCTION

The paper addresses the multiple source location problem which has been a research topic in radar, sonar and seismology for many years. Frequently, certain parameters of sources radiating signals are unknown and have to be estimated by appropriate array processing methods. We shall examine the performance of the nonlinear regression method which was described in {1} and {2} for estimating bearing, range and the spectral density matrix of unknown signals. The asymptotic properties of those estimates are known for a large number K of observations and are indicated in {2}. However, of most interest is the behavior for moderate K which is investigated by extensive numerical experiments. We shall show the results of such experiments and compare them with the predicted asymptotic properties of maximum likelihood estimates.
First of all, the data model is presented. In section III, we describe the nonlinear regression method and outline the numerical procedure. It follows the computation of the Cramer Rao lower bound for the unknown parameters in section IV. In section V, a description of the statistical tools for analysing the numerical experiments is given, and we finish with a presentation of the results and some concluding remarks.

II. MODEL AND DATA

In Fig. 1, the geometric relations between sources and antenna elements in the plane are shown. L sources are located in the farfield generating stationary signals. Some of them may be correlated, e.g. by multipath propagation which is a general problem in underwater acoustics. The correlation of source signals is described by the (LxL) spectral density matrix C_S. The distance from the origin 0 of the antenna to the l'th source (l=1,...,L) is indicated by ρ_l, while β_l is the bearing of this source with respect to the broadside direction of the antenna. An array of N sensors at known positions (a_n, α_n) (n=1,...,N) in polar coordinates samples the wavefield. Then, the propagation-reception conditions can be described by a (NxL) matrix H with elements $H_{nl} = N^{-1/2}\exp(-j\omega\tau_{nl})$. If the sources generate spherical and coherent wavefronts which are not disturbed by the medium of propagation, the

time delays
$$\tau_{nl} = \rho_l/c_\circ\left(1-(1+(a_n/\rho_l)^2-2(a_n/\rho_l)\cos(\alpha_n-\beta_l)\,)^{1/2}\right) \qquad (1)$$

result from the geometric relations in figure 1. The constant $c_\circ$ is the velocity of propagation. At the sensors, we measure the received signals and independent sensor noise. From these measurements, we want to determine the bearing β_l and range ρ_l of every source, the spectral density matrix C_S and the power of sensor noise ν.
In a first step, the data of the sensor outputs are Fourier-transformed with a window of length T to get successive data pieces $\underline{X}^k(\omega)=(\underline{X}^1(\omega),\ldots,\underline{X}^K(\omega))'$ $(k=1,\ldots,K)$. In the following, we omit the notation of ω and consider only the narrowband case for a single frequency. For large T, we can write $\underline{X}^k=H\underline{S}^k+\underline{U}^k$, where the $\underline{S}^k$'s are the vectors of the unknown Fourier-transformed signals and the $\underline{U}^k$'s describe the sensor noise. From the statistical properties of the Fourier-transformed data, we know that, for a large window length, $\underline{X}^1,\ldots,\underline{X}^K$ are independent random vectors with zero mean and covariance matrix C_U. In the following, we restrict ourselves to the simple case of $C_U=\nu I$, where I is the unit matrix. Any other structure of C_U can be reduced to this form by prewhitening the data if C_U is known except for a scaling. Then, we can write, considering conditionally distributed data given $\underline{S}^1,\ldots,\underline{S}^K$, the $\underline{X}^1,\ldots,\underline{X}^K$ are independent normally distributed random vectors with mean $H\underline{S}^k$ and covariance matrix νI.

III. ESTIMATION OF PARAMETERS

For the estimation of the unknown parameters, we use the nonlinear regression method developed in {1} and {2}. The conditional likelihood function for data $\underline{X}^k$ can be written
$$\Lambda = \prod_{k=1}^{K} (2\pi)^{-N}(\det C_U)^{-1}\exp(-(\underline{X}^k-H\underline{S}^k)^* C_U^{-1}(\underline{X}^k-H\underline{S}^k)) \ . \qquad (2)$$
The asterisk means the hermitian operation. This equation can be reduced to a least square problem for $C_U=\nu I$:
$$L = \ln\Lambda = -NK\ln(2\pi\nu)-\frac{1}{\nu}\sum_{k=1}^{K}|\underline{X}^k-H\underline{S}^k|^2 \ . \qquad (3)$$
Now we have to minimize the criterion $\sum_{k=1}^{K}|\underline{X}^k-H\underline{S}^k|^2$ over all parameters summarized to the vectors $\underline{\beta}$, $\underline{\rho}$ and $\underline{S}$. Fixing the wave parameters $\underline{\beta}$ and $\underline{\rho}$ and solving for $\underline{S}$, we get
$$\underline{S}^k = (H^*H)^{-1}H^*\underline{X}^k \qquad (4)$$
if H^*H is nonsingular. Using the result for $\underline{S}^k$ in (3), we can find the parameters $\underline{\beta}$ and $\underline{\rho}$ by minimizing the new criterion
$$q(\underline{\beta},\underline{\rho}) = \mathrm{tr}(P(\underline{\beta},\underline{\rho})\hat{C}_X) \qquad (5)$$

over the unknown parameters, where $\hat{C}_X = {}^1/_K \sum_{k=1}^{K} \underline{X}^k \underline{X}^{k*}$ is an estimate of the spectral density matrix of the sensor outputs and $P=I-H(H^*H)^{-1}H^*$ is the projection matrix into the noise space orthogonal to the signal subspace spanned by the columns of H.

$$\bar{\nu} = q(\bar{\underline{\beta}},\bar{\underline{\rho}})/(N-L) \tag{6}$$

is a reasonable estimate of the noise power ν for optimum estimates $\bar{\underline{\beta}}$ and $\bar{\underline{\rho}}$ which is the maximum likelihood estimate for ν except for a scale factor.

The wave parameters are determined by the Gauss-Newton iterative procedure for minimizing the criterion q. We can write $\underline{\theta}=(\underline{\beta}',\underline{\rho}')'$ and

$$\underline{\theta}^{n+1} = \underline{\theta}^n - (\nabla\nabla' q)^{-1}\nabla q \,\big|_{\underline{\theta}^n} \tag{7}$$

where ∇ is the operator for the gradient and ' indicates a transposed vector. Instead of calculating the Hessian matrix of second derivatives $(\nabla\nabla' q)$ which is an extensive computational task, we use a much simpler approximation replacing the estimate $\hat{C}_X$ by the covariance matrix C_X inside the Hessian matrix:

$$(\nabla\nabla' q)_{ik} = \frac{\partial^2 q}{\partial\theta_i \partial\theta_k} = tr\left(\hat{C}_X \frac{\partial^2 P}{\partial\theta_i \partial\theta_k}\right) \simeq tr\left(C_X \frac{\partial^2 P}{\partial\theta_i \partial\theta_k}\right)$$

$$= 2Re\left\{tr\left(C_S \frac{\partial H^*}{\partial\theta_i} P \frac{\partial H}{\partial\theta_k}\right)\right\} \qquad (i,k=1,\ldots,2L) \tag{8}$$

where $C_X = HC_SH^* + \nu I$.

For C_S we use the estimate
$$\hat{C}_S = \frac{1}{K}\sum_{k=1}^{K} \underline{S}^k\underline{S}^{k*} - \bar{\nu}(H^*H)^{-1} \,\big|_{\underline{\theta}^n}$$
$$= (H^*H)^{-1}H^*\hat{C}_X H(H^*H)^{-1} - \bar{\nu}(H^*H)^{-1} \,\big|_{\underline{\theta}^n} \tag{9}$$

that was given by Wax in {3}.

For a successful application of the Gauss-Newton procedure it is very important to have a good initial estimate of the wave parameters close to the optimal values maximizing the likelihood function. A simple and computationally efficient method satisfying these requirements was introduced in {1}. This method yields good estimates if the columns of H are approximately orthogonal.

IV. CRAMER RAO LOWER BOUNDS OF PARAMETER ESTIMATES

The asymptotic properties of the estimates are only known for a large window length T and a large number of observations K. However, these conditions are not satisfied in practice. To have an objective measure of the performance of the nonlinear regression method, we can compare the theoretical lower bounds for the estimates which are given by Cramer and Rao with the results of numerical experiments. This is done for one special configuration of signals and different signal to noise ratios in this paper. For the theoretical bounds, we use the nonconditional likelihood function which is given by

$$\bar{\Lambda} = \prod_{k=1}^{K} (2\pi)^{-N}(\det C_X)^{-1}\exp(-tr(\hat{C}_X C_X^{-1})) \ . \tag{10}$$

Here, $\hat{C}_X = \frac{1}{K}\sum_{k=1}^{K}\underline{X}^k\underline{X}^{k*}$ is the estimate of the spectral density matrix of the sensor outputs $\underline{X}^k$ which are complex normally distributed with mean zero and covariance matrix $C_X=H(\underline{\beta},\underline{\rho})C_S H(\underline{\beta},\underline{\rho})^*+\nu I$. The lower bound of the covariance matrix of unbiased parameter estimates is given by

$$\mathrm{cov}(\hat{\underline{\beta}},\hat{\underline{\rho}},\hat{\underline{S}}) \geq J^{-1}(\underline{\beta},\underline{\rho},\underline{S}) \quad . \tag{11}$$

J is the Fisher information matrix defined by

$$J_{ik} = -E\left(\frac{\partial^2 \ln\bar{\Lambda}}{\partial\theta_i\partial\theta_k}\right) \qquad \underline{\theta} = (\underline{\beta}',\underline{\rho}',\underline{S}')' \quad .$$

We get

$$J_{ik} = \mathrm{tr}\left(C_X^{-1}\frac{\partial C_X}{\partial\theta_i}C_X^{-1}\frac{\partial C_X}{\partial\theta_k}\right) \qquad (i,k=1,\ldots,3L) \quad . \tag{12}$$

The explicite calculation of J_{ik} is a difficult task and only possible for a few simple signal configurations, e.g. uncorrelated and clearly separated signals:$C_S=\mathrm{diag}(S_1^2)$ and $H^*H=I$. Under these conditions all parameter estimates are uncorrelated. The numerical calculation of $J^{-1}(\underline{\beta},\underline{\rho},\underline{S})$ shows that bearing and power estimates for closely spaced (separated less than about 2/3 of a beamwidth) but independent signals are correlated. The correlation increases with decreasing signal to noise ratio or decreasing separation. For an equispaced line array, the range estimates are uncorrelated from all other parameters. Details of this investigation and other results are part of a forthcoming PhD dissertation of the first author. A comparison of theoretical bounds and experimental results is given in VI.

V. STATISTICAL ANALYSIS

In this section, we outline the statistical methods which were used for the interpretation of the numerical experiments. Using statistical tools for the analysis of data, we have to pay attention to 4 points. First, the number of numerical experiments has to be large enough to be of statistical evidence. This is only a problem of computer power. Secondly, we have to take in mind that the parameter estimates are not normally distributed in general. However, many statistical tools work well only for normally distributed data. It is known that the median can be a better estimate for the expectation value than the mean, if the distribution of the parameter is not gaussian. Therefore, we use the median $\bar{\mu}$ instead of the mean in the following and define the median deviation as the length $\bar{s}$ of an interval for which 68.27 % of the data are inside the interval $(\bar{\mu}-\bar{s},\mu+\bar{s})$. This is identical to the standard deviation in the case of a normal distribution. For range and power, we use the logarithms of the estimates related to the array length or the power of noise, resp. because there distributions are distinctly nonnormal and skew. Thirdly, we have to deal with outliers in a useful manner. Especially, estimates of higher moments using powers of the data values for their calculation are falsified by outliers. To avoid this, the rank correlation coefficient of Spearman {4} is used as measure for correlation between parameter estimates. It is also well appropriate for nonnormal data. The definition of the median deviation given above simplifies

the handling of outliers, too. The last point is only important if we have more than one source. In this case, we must find the right association between the true sources and their estimates. The identification is done in the order to their closeness with the condition that for each true source, we have exactly one estimated source.
To make the accuracy of the estimates, their distribution and the relationship between parameters visible in only one picture, we use 2-dim. estimated confidence regions which are calculated as follows. For every interesting pair of the 3L+1 parameters ((bearing + range + power) x L + noise), we compute a 2-dim. histogram in polar coordinates centered at the median of these parameters. Radially departing from the median, we find a boundary for which a given percentage of the data is located inside this boundary. This is done for levels of 20, 40, 60, 80, 90, 95 and 98 per cent of the estimates. For the statistical analysis, the results of all experiments including outliers were used.

VI. RESULTS

In our numerical experiments, we used a line array of 15 sensors spaced by half a wavelength. We analysed several different signal configurations. The results of one interesting configuration are presented in the following. A weak source of -3 dB signal to noise ratio ($10 \log C_{Sii}/\nu$) located in a distance of 40 times the array length was surrounded by 2 strong sources of 4 dB power separated in bearing by 2/3 of a beamwidth in a distance of 32 and 20 array lengths, respectively. All signals were independent. The matrices $\hat{C}_X$ were simulated by complex Whishart matrices with 20 degrees of freedom. The results of 2048 independent experiments are shown in Fig. 2a and 2b. The positions of the true sources are marked by big cross s.
Confidence regions of all signals numbered from left to right are plotted in Fig. 3a-c for bearing and power and in Fig. 4a-c for bearing and range. Small crosses indicate the mean and small circles the median of the estimates. The rank correlation coefficient is noted by RS. There are a lot of outliers of the weak signal scattered over all bearings which are not visible in Fig. 2 but obvious by the deformations of the confidence regions.
The experiments show that estimates of closely spaced sources are influenced by each other. Bearing and power estimates are correlated even for independent sources. However, range estimates are uncorrelated with all other parameters which corresponds to the theoretical results. The estimates of well separated sources are stable and uncorrelated. The bearing estimates are approximately normally distributed in contrast with power and range estimates. If sources are less separated than ≃1/3 of a beamwidth, the estimates of bearing can deviate from a normal distribution, too. The median is a better estimate for the unknown parameter than the mean.
In Fig. 5a-c, we compare the Cramer Rao bounds (continous lines) and the experimental results (dashed lines) for the parameters of signal 1-3. The lines were calculated in 1 dB steps equally lifting up or setting down the power of all signals relative to the configuration described above. Here, we used 256 experiments for each step.
For strong and independent signals, the Cramer Rao bound is a suitable prediction for the variance of the estimates. But for weak and correlated estimates, there are increasing discrepancies, e.g. the theoretical bound for the

range deviation of the weak signal is much higher (by about 14 dB) than the experimental result.

VII. CONCLUDING REMARKS

We showed that the lower bounds for the variances of the parameter estimates can be approached by the nonlinear regression method for strong signals. Parameter estimates are correlated if the sources are not well separated even if the signals are independent. The accuracy and the distribution of the estimates were represented for one special configuration of signals.
In the examples we assumed the number of signals, but in the general case L has to be estimated. Space limitations do not permit us to investigate this problem more detailed. The results of other experiments show that Rissanen's criteria, as discussed in {5} , give a good estimate for L, while Akaike's criteria overestimate the number of signals.

REFERENCES

{1} Böhme J.F., "Estimation of Source Parameters by Maximum Likelihood", Proc. IEEE-ICASSP San Diego 1984, 7.3.1 - 7.3.4

{2} Böhme J.F., "Source Parameter Estimation by Approximate Maximum Likelihood and Nonlinear Regression", Special Issue on Beamforming, IEEE J. Oceanic Engineering, July 1985

{3} Wax M., Shan T.J., Kailath T., "Location and the Spectral Density Estimation of Multiple Sources", Proc. 16 th Asimolar Conference 1982, 322 - 326

{4} Sachs L., "Angewandte Statistik", Springer Verlag 1984, 309 - 311, (also available in English)

{5} Wax M.,Kailath T., "Determining the Number of Signals by Information Theoretic Criteria", Proc. IEEE-ASSP Spectrum Estimation Workshop II 1983, 192 - 196

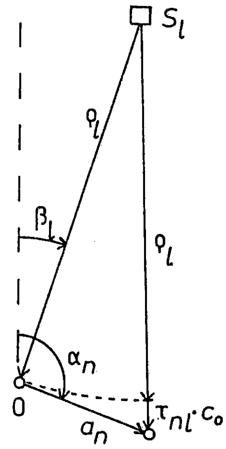

Fig. 1 Geometric relation between sources and array elements: S_l l th source with bearing β_l and range ρ_l related to the origin 0 of the antenna; (a_n, α_n) position of the n th sensor; τ_{nl} time delay; c_0 velocity of propagation

LOCATION AND POWER ESTIMATION OF MULTIPLE SOURCES BY NONLINEAR REGRESSION

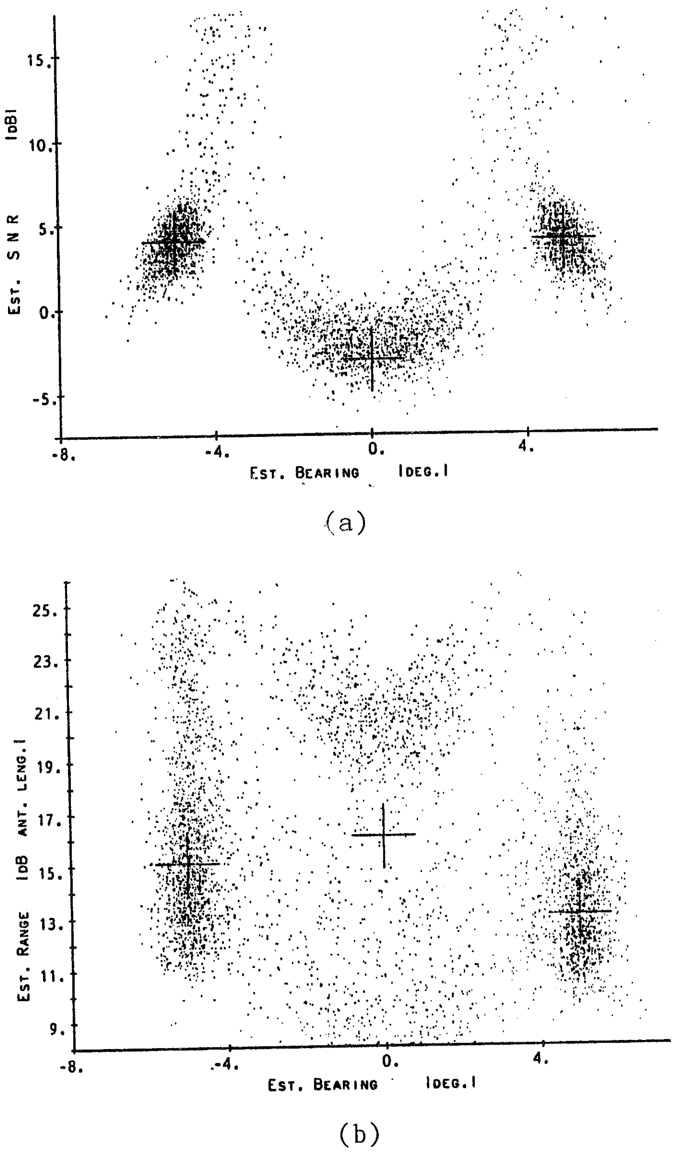

Fig. 2 Estimated SNR (a) and est. range (b) over est. bearing for 2048 experiments. The true sources are indicated by crosses.

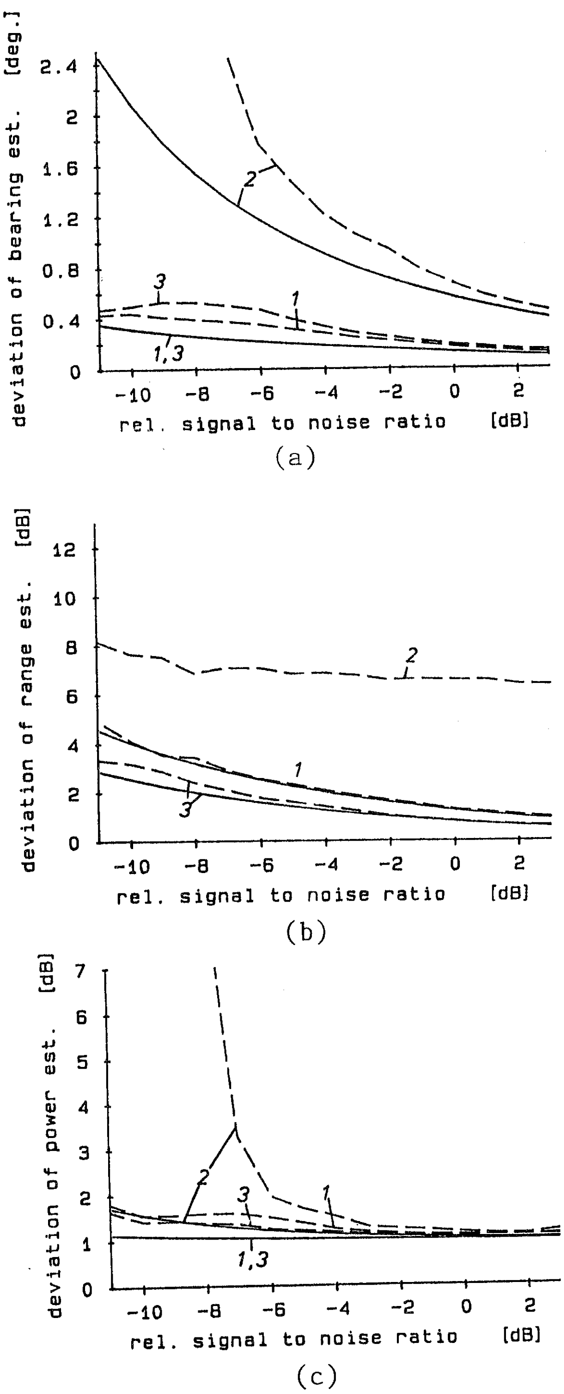

Fig. 5 Cramer Rao bound (——) and median deviation (- -) for bearing (a), range (b) and power (c). The sources are numbered from left to right.

LOCATION AND POWER ESTIMATION OF MULTIPLE SOURCES BY NONLINEAR REGRESSION

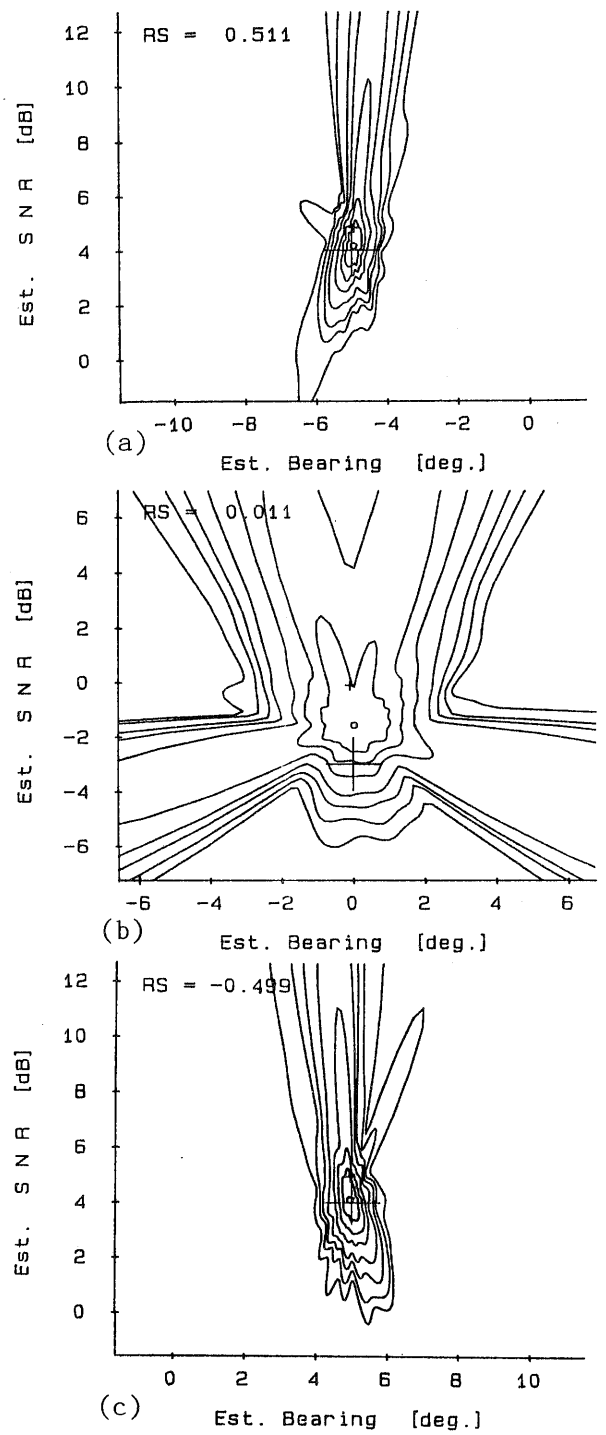

Fig. 3 Est. confidence regions for SNR over bearing. (a) 1st source, (b) 2nd source, (c) 3rd source; + mean; o median; + true source.

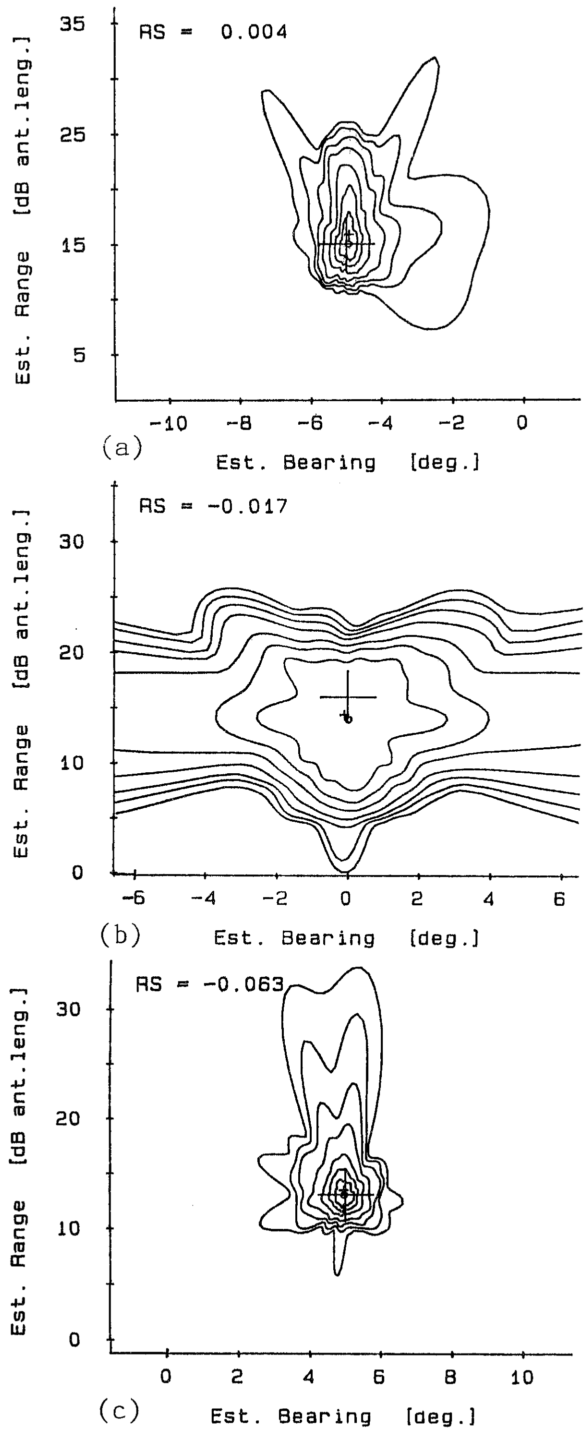

Fig. 4 Est. confidence regions for range. Notation as in Fig. 3.