Adaptive decision fusion for unequiprobable sources

N.Ansari
J.-G.Chen
Y.-Z.Zhang

Indexing terms: Distributed detection, Probability of false alarm, Reinforcement learning

Abstract: An optimal decision rule has been derived by Chair and Varshney (1986) for fusing decisions based on the Bayesian criterion. However, to implement such a rule, the miss probability P_M and the probability of false alarm P_F for each local detector must be known, and these are not readily available in practice. To circumvent this situation, an adaptive fusion system for equiprobable sources has been developed. The system is extended to unequiprobable sources; thus its practicality is enhanced. An adaptive fusion model using the fusion result as a supervisor to estimate the P_M and P_F is introduced. The fusion results are classified as 'reliable' and 'unreliable'. Reliable results are used as a reference to update the weights in the fusion centre. Unreliable results are discarded. The convergence and error analysis of the system are demonstrated theoretically and by simulations. The paper concludes with simulation results that conform to the analysis.

1 Introduction

There is a growing interest in developing efficient and reliable distributed detection systems (multiple sensor systems) for target recognition and communications. Tenney and Sandell [1] were among the first to study the problem of detection with distributed sensors. They applied the classical single-sensor detection theory to a two-sensor, two-hypothesis testing. An optimum local decision rule was established to minimise a global cost. Sadjadi [2] generalised the work of [1] to n detectors and m hypotheses and reached similar conclusions. Chair and Varshney [3] assumed that each local detector had a predetermined decision rule and that each local decision was independent. With these assumptions, an optimum fusion model was generated. Optimal techniques have also been developed for other criteria. When *a priori* probabilities were unknown, Thomopoulos [4] used the Neyman–Pearson (NP) test both at the local detector level and at the decision fusion level. An optimal decision scheme was derived. Demirbas [5] applied the maximum *a posteriori* (MAP) concept for object recognition in a multi-sensor environment and showed that the MAP estimation approach minimised mean square error estimation.

IEE Proceedings online no. 19971176

Paper first received 18th March 1996 and in revised form 7th February 1997

The authors are with the Center for Communications and Signal Processing, Electrical and Computer Engineering Department, New Jersey Institute of Technology, University Heights, Newark, New Jersey 07102, USA

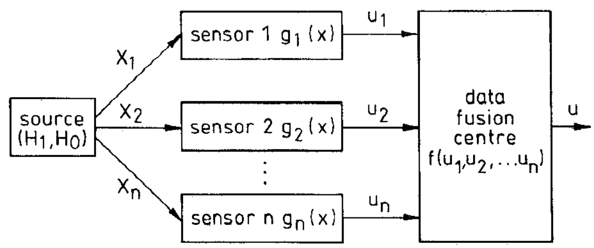

Fig. 1 *Distributed detection system with data fusion centre*

In the distributed system with data fusion shown in Fig. 1, some data processing is done at each sensor, and partial results are transmitted to the data fusion centre for further processing. The final results are then available at the data fusion centre. When the detection rule is fixed at each sensor, the optimal fusion rule developed for independent local decisions [3] is a weighted sum of local decisions. The weight is a function of the probability of detection P_D and the probability of false alarm P_F of the detector. However, in practice, neither P_D nor P_F is known. Furthermore, as the sensors are usually exposed to a changing environment, the performance of each individual detector may not always be the same, i.e. P_D and P_F may vary with time. To circumvent this situation, some learning rules for adaptive distributed detection have been proposed [6–9]. A learning algorithm for the tandem neural network structure was proposed for the binary hypothesis testing [6] in which only one parameter needed to be estimated. In [7], an on-line learning algorithm was adopted to estimate the thresholds for each local sensor and the fusion centre. Both approaches [6, 7] were based on the NP performance criterion instead of on estimation of the prior probabilities, which is our method. In [8], a set of stochastic approximation rules was proposed for the probability estimation in Bayesian hypothesis testing that was similar, in concept, to our method [9]. However, our estimation is much simpler and can be achieved by a simple counting process. Our method is based on reinforcement learning, which does not require a training sequence.

This paper is an extension of our previous work [9] and uses the fusion results as a supervisor to estimate P_M and P_F. The model for equiprobable sources has been reported [9]. This paper considers the performance and analysis for unequiprobable sources. Various components of the model will be covered: the fusion rule, classification of fusion results, an updating algorithm for the fusion centre, convergence, error analysis and computer simulations.

2 Data fusion rule and its properties

Let us consider a binary hypothesis testing problem with the following two hypotheses:

$$H_0: \quad \text{signal is absent}$$
$$H_1: \quad \text{signal is present}$$

The *a priori* probabilities of the two hypotheses are denoted by $P(H_0) = P_0$ and $P(H_1) = P_1$. As shown in Fig. 1, we assume that there are n detectors, and the observations at each detector are denoted by x_i, $i = 1, \ldots, n$. We further assume that the observations at the individual detectors are statistically independent and that the conditional probability is denoted by $P(x_i|H_j)$, $i = 1, \ldots, n$, $j = 0, 1$. Each detector employs a decision rule $g_i(x_i)$ to make a decision u_i, $i = 1, \ldots, n$, where

$$u_i = \begin{cases} -1 & \text{if } H_0 \text{ is declared} \\ +1 & \text{if } H_1 \text{ is declared} \end{cases}$$

The probabilities of false alarm and miss for each detector are denoted by P_{Fi} and P_{Mi}, respectively.

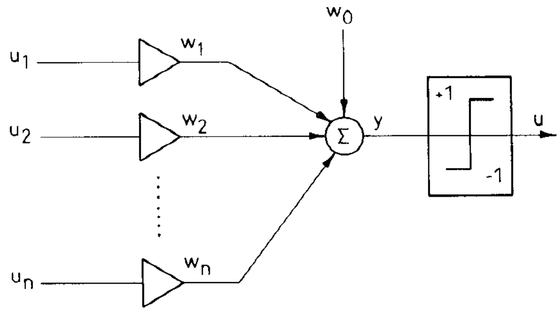

Fig.2 *Fusion centre structure*

After processing the observations locally, the decisions u_i are transmitted to the data fusion centre. The data fusion centre determines the overall decision u for the system based on the individual decisions, i.e.

$$u = f(u_1, \ldots, u_n) \tag{1}$$

Based on the above specification, Chair and Varshney developed the optimum fusion rule as

$$u = f(u_1, \ldots, u_n) = \begin{cases} +1 & \text{if } w_0 + \sum_{i=1}^{n} w_i u_i > 0 \\ -1 & \text{otherwise} \end{cases} \tag{2}$$

where

$$w_0 = \log \frac{P_1}{P_0} \tag{3}$$

$$w_i = \begin{cases} \log \frac{1-P_{M_i}}{P_{F_i}} & \text{if } u_i = +1 \\ \log \frac{1-P_{F_i}}{P_{M_i}} & \text{if } u_i = -1 \end{cases} \tag{4}$$

The optimum data fusion rule can be implemented as shown in Fig. 2, where

$$y = w_0 + \sum_{i=1}^{n} w_i u_i$$

Note that the above fusion rule has the following property:

Lemma 1: When each weight in the fusion centre is optimum, the conditional probability mass functions $P(y - w_0 = \zeta|H_1)$ and $P(y - w_0 = \zeta|H_0)$ satisfy the following equation:

$$e^{\zeta} = \frac{P(y - w_0 = \zeta|H_1)}{P(y - w_0 = \zeta|H_0)} \tag{5}$$

where ζ is a possible value of $y - w_0$.

Proof: see the Appendix (Section 7.1).

Eqn. 5 is a very interesting result. The ratio of the conditional probabilities under H_1 and H_0 only depends on the value $y - w_0$, as illustrated in Fig. 3.

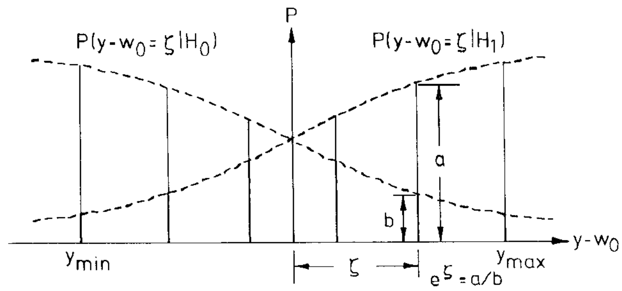

Fig.3 *Relationship between* $P(y - w_0 = \zeta|H_1)$ *and* $P(y - w_0 = \zeta|H_0)$

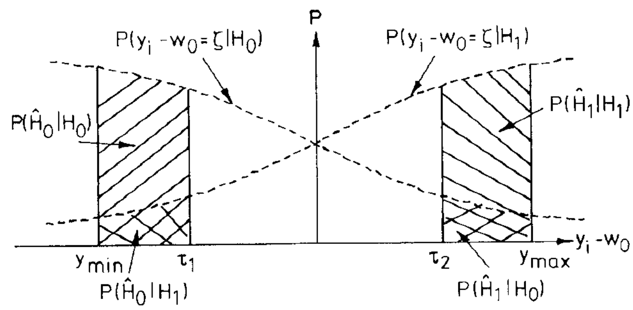

Fig.4 *Classification of fusion results*

Recall the data fusion centre structure shown in Fig. 2. If reference signals are given, they can be used as a 'reference' to train the system, so that weights will converge to the optimal values defined by eqns. 3 and 4. However, in practice, such a reference is not readily available, and, at the same time, the P_M and P_F of a detector may vary with time. As the fused decisions are usually better than local decisions, they can be considered as the reference. When the ith local decision u_i is equal to the fused decision u, then u_i is considered to be correct; otherwise, u_i is considered to be incorrect. As $y = w_0 + \Sigma_{i=1}^{n} w_i u_i$, the fused decision u has already taken into account the decision of the ith detector u_i. If u is used as a reference for u_i, a bias is established for u_i. Thus, in the proposed system, the decision of the ith local detector u_i is arbitrated by the fused decision of all the other $(n - 1)$ local detectors. Denote this fused decision as $\bar{u}_i$, and define

$$y_i = y - w_0 - w_i u_i = \sum_{j=1, j\neq i}^{n} w_j u_j \tag{6}$$

The decision $\bar{u}_i$ in the fusion centre for updating $\hat{w}_i$ depends on the value y_i. Here, $\hat{w}_i$ is the estimated weight. Using the same procedure, it can be shown that y_i has a similar property to y in Lemma 1. That is,

$$\frac{P(y_i = \zeta|H_1)}{P(y_i = \zeta|H_0)} = e^{\zeta} \tag{7}$$

where ζ is a possible value that y_i takes on. The range of y_i is divided into reliable and unreliable ranges. We denote the lower and upper limit of the unreliable range as τ_1 and τ_2, as shown in Fig. 4. Usually $\tau_2 \geq 0$, $\tau_1 \leq 0$. We call τ_1 and τ_2 the reliability thresholds. Only the fused decisions $\bar{u}_i$ that satisfy $y_i < \tau_1$ or $y_i > \tau_2$ are chosen to adapt each weight, denoted by $\hat{w}_i$. These decisions are considered reliable decisions, defined by $\hat{H}_1$ when $y_i > \tau_2$, and $\hat{H}_0$ when $y_i < \tau_1$. The decision is considered unreliable when $\tau_1 < y_i < \tau_2$, denoted by $\hat{H}_x$. Obviously,

$$P(\hat{H}_1) + P(\hat{H}_0) + P(\hat{H}_x) = 1 \tag{8}$$

This type of learning belongs to the class of reinforcement learning [10].

Based on the proposed fusion rule described in eqn. 6, we obtain the following two properties related to the steady state error:

Lemma 2: If $\alpha = P(\hat{H}_1|\hat{H}_0)/P(\hat{H}_1|H_1)$, $\beta = P(\hat{H}_0|H_1)/P(\hat{H}_0|H_0)$ and $\gamma = P(\hat{H}_1|H_1)/P(\hat{H}_0|H_0)$, then

(*a*) α is monotonically decreasing when τ_2 increases, β is monotonically decreasing when τ_1 decreases, and

$$\alpha_{min} = \prod_{j=1,j\neq i}^{n} \frac{P_{F_j}}{1 - P_{M_j}} \tag{9}$$

$$\beta_{min} = \prod_{j=1,j\neq i}^{n} \frac{P_{M_j}}{1 - P_{F_j}} \tag{10}$$

(*b*) when $\tau_2 = (y_i)_{max} = \Sigma_{j=1,j\neq i}^{n} \log((1 - P_{Mj})/P_{Fj})$ and $\tau_1 = (y_i)_{min} = -\Sigma_{j=1,j\neq i}^{n} \log((1 - P_{Fj})/P_{Mj})$

$$\gamma = \frac{P(\widehat{H_1}|H_1)}{P(\widehat{H_0}|H_0)} = \prod_{j=1,j\neq i}^{n} \left(\frac{1 - P_{M_j}}{1 - P_{F_j}}\right) \tag{11}$$

Proof: see the Appendix (Section 7.2).

Lemma 3: Let $\varepsilon_i = |w_i - \hat{w}_i|$, $i = 0, 1, \ldots, n$, represent the estimation error. The minimum ε_i that can be achieved is

$$\begin{cases} \epsilon_0 = \log\left(\dfrac{1 + \frac{\alpha_{min}}{r_0}}{1 + r_0\beta_{min}}\right) + \log(\gamma) \\ \epsilon_i = \begin{cases} \log \frac{1+\beta_{min} r_0}{1+\frac{\alpha_{min}}{r_0}} + \log \frac{1+\frac{\alpha_{min}}{r_0 r_i}}{1+\beta_{min} r_0 r_i} & \text{if } u_i = +1 \\ \log \frac{1+\frac{\alpha_{min}}{r_0}}{1+\beta_{min} r_0} + \log \frac{1+\frac{\beta_{min} r_0}{r_i}}{1+r_i \frac{\alpha_{min}}{r_0}} & \text{if } u_i = -1 \end{cases} \end{cases} \tag{12}$$

Proof: If $w_i = \log r_i$ and $\hat{w}_i = \log \hat{r}_i$, we have

$$\begin{cases} r_0 = \dfrac{P_1}{P_0} \\ r_i = \begin{cases} \frac{P(u_i=1|H_1)}{P(u_i=1|H_0)} & \text{if } u_i = +1 \\ \frac{P(u_i=-1|H_0)}{P(u_i=-1|H_1)} & \text{if } u_i = -1 \end{cases} \end{cases} \tag{13}$$

and

$$\begin{cases} \hat{r}_0 = \dfrac{P(\hat{H}_1)}{P(\hat{H}_0)} \\ \hat{r}_i = \begin{cases} \frac{P(u_i=1|\hat{H}_1)}{P(u_i=1|\hat{H}_0)} & \text{if } u_i = +1 \\ \frac{P(u_i=-1|\hat{H}_0)}{P(u_i=-1|\hat{H}_1)} & \text{if } u_i = -1 \end{cases} \end{cases} \tag{14}$$

Using the total probability theorem $P(BA) = P(B|A)P(A)$,

$$P(u_i = 1|\hat{H}_1) = \frac{P(u_i = 1|H_0)P(\hat{H}_1|u_i = 1, H_0)P(H_0)}{P(\hat{H}_1)} + \frac{P(u_i = 1|H_1)P(\hat{H}_1|u_i = 1, H_1)P(H_1)}{P(\hat{H}_1)}$$

Similarly,

$$P(u_i = 1|\hat{H}_0) = \frac{P(u_i = 1|H_0)P(\hat{H}_0|u_i = 1, H_0)P(H_0)}{P(\hat{H}_0)} + \frac{P(u_i = 1|H_1)P(\hat{H}_0|u_i = 1, H_1)P(H_1)}{P(\hat{H}_0)}$$

$$P(u_i = -1|\hat{H}_0) = \frac{P(u_i = -1|H_0)P(\hat{H}_0|u_i = -1, H_0)P(H_0)}{P(\hat{H}_0)} + \frac{P(u_i = -1|H_1)P(\hat{H}_0|u_i = -1, H_1)P(H_1)}{P(\hat{H}_0)}$$

$$P(u_i = -1|\hat{H}_1) = \frac{P(u_i = -1|H_0)P(\hat{H}_1|u_i = -1, H_0)P(H_0)}{P(\hat{H}_1)} + \frac{P(u_i = -1|H_1)P(\hat{H}_1|u_i = -1, H_1)P(H_1)}{P(\hat{H}_1)}$$

Using eqn. 14 and the above formulas, if $u_i = +1$,

$$\hat{r}_i = r_i \cdot \frac{P(\hat{H}_1|H_1)}{P(\hat{H}_0|H_0)} \cdot \frac{P(H_1)}{P(H_0)} \cdot \frac{P(\hat{H}_0)}{P(\hat{H}_1)} \cdot \frac{\frac{1}{r_i}\frac{P(\hat{H}_1|H_0)}{P(\hat{H}_1|H_1)}\frac{1}{r_0} + 1}{1 + r_i \frac{P(\hat{H}_0|H_1)}{P(\hat{H}_0|H_0)} r_0} \tag{15}$$

Here,

$$\frac{P(\hat{H}_0)}{P(\hat{H}_1)} = \frac{P(\hat{H}_0|H_1)P(H_1) + P(\hat{H}_0|H_0)P(H_0)}{P(\hat{H}_1|H_1)P(H_1) + P(\hat{H}_1|H_0)P(H_0)}$$

Thus,

$$\frac{P(\hat{H}_1|H_1)}{P(\hat{H}_0|H_0)} \cdot \frac{P(H_1)}{P(H_0)} \cdot \frac{P(\hat{H}_0)}{P(\hat{H}_1)} = \frac{\frac{P(\hat{H}_0|H_1)}{P(\hat{H}_0|H_0)} \cdot \frac{P(H_1)}{P(H_0)} + 1}{1 + \frac{P(\hat{H}_1|H_0)}{P(\hat{H}_1|H_1)} \cdot \frac{P(H_0)}{P(H_1)}} \tag{16}$$

Substituting α, β and r_0 into eqn. 15, after simplification, we obtain

$$\hat{r}_i = r_i \cdot \frac{1 + \beta r_0}{1 + \frac{\alpha}{r_0}} \cdot \frac{1 + \frac{\alpha}{r_0 r_i}}{1 + \beta r_0 r_i} \tag{17}$$

Similarly, if $u_i = -1$,

$$\hat{r}_i = r_i \cdot \frac{1 + \frac{\alpha}{r_0}}{1 + \beta r_0} \cdot \frac{1 + \frac{\beta r_0}{r_i}}{1 + r_i \frac{\alpha}{r_0}} \tag{18}$$

From eqn. 14,

$$\hat{r}_0 = \frac{P_1 + P_0\alpha}{P_1\beta + P_0}\gamma = r_0 \frac{1 + \frac{\alpha}{r_0}}{1 + r_0\beta}\gamma \tag{19}$$

According to the definitions of ε_i, $\hat{r}_i$ and r_i, the following weight error is obtained:

$$\begin{cases} \epsilon_0 = \log\left(\dfrac{1 + \frac{\alpha}{r_0}}{1 + r_0\beta}\right) + \log(\gamma) \\ \epsilon_i = \begin{cases} \log \frac{1+\beta r_0}{1+\frac{\alpha}{r_0}} + \log \frac{1+\frac{\alpha}{r_0 r_i}}{1+\beta r_0 r_i} & \text{if } u_i = +1 \\ \log \frac{1+\frac{\alpha}{r_0}}{1+\beta r_0} + \log \frac{1+\frac{\beta r_0}{r_i}}{1+r_i \frac{\alpha}{r_0}} & \text{if } u_i = -1 \end{cases} \end{cases} \tag{20}$$

From eqn. 20, we know that, when $\alpha = 0$, $\beta = 0$ and $\gamma = 1$, ε_i (for $i = 0, 1, \ldots, n$) would achieve its minimum. In Lemma 2, we have proved that α and β are monotonically decreasing with τ_1 and τ_2. Thus, when α and β

achieve their minimum, ε_i (for i = 0, 1, ..., n) also acquires its minimum, and thus

$$\begin{cases} \epsilon_0 = \log\left(\dfrac{1+\frac{\alpha_{min}}{r_0}}{1+r_0\beta_{min}}\right) + \log(\gamma) \\ \epsilon_i = \begin{cases} \log\frac{1+\beta_{min}r_0}{1+\frac{\alpha_{min}}{r_0}} + \log\frac{1+\frac{\alpha_{min}}{r_0 r_i}}{1+\beta_{min}r_0 r_i} & \text{if } u_i = +1 \\ \log\frac{1+\frac{\alpha_{min}}{r_0}}{1+\beta_{min}r_0} + \log\frac{1+\frac{\beta_{min}r_0}{r_i}}{1+r_i\frac{\alpha_{min}}{r_0}} & \text{if } u_i = -1 \end{cases} \end{cases} \tag{21}$$

where α_{min}, β_{min} and γ are defined in Lemma 2, and r_i and r_0 are defined in eqn. 13. Note that the minimum error is uniquely determined by P_1, P_0 and the parameters of sensors (P_{Fi} and P_{Mi}).

Note that $(y_i)_{max}$ and $(y_i)_{min}$ vary from sensor to sensor. To enable every sensor to adjust its weight and achieve the least error, the maximum value of τ_2 is chosen to be the minimum of all $(y_i)_{max}$, and the minimum value of τ_1 is chosen to be the maximum of all $(y_i)_{min}$. That is

$$(\tau_2)_{max} = \min\{(y_1)_{max}, (y_2)_{max}, \ldots, (y_n)_{max}\} \tag{22}$$

$$(\tau_1)_{min} = \max\{(y_1)_{min}, (y_2)_{min}, \ldots, (y_n)_{min}\} \tag{23}$$

Lemmas 2 and 3 illustrate how close the estimated weights are to optimal weights.

3 Reinforcement updating rule

The distributed decision system is assumed to have no knowledge of the probability mass functions of the observations. Thus, the estimated probability of detection and false alarm for the ith detector $\hat{P}_{Di}$ and $\hat{P}_{Fi}$ can be approximated by relative frequencies. Let m be the number of $\hat{H}_1$, n be the number of $\hat{H}_0$, and

m_{1i} be the number of $u_i = +1$ and $\bar{u}_i = +1$

m_{0i} be the number of $u_i = -1$ and $\bar{u}_i = -1$

n_{1i} be the number of $u_i = +1$ and $\bar{u}_i = -1$

n_{0i} be the number of $u_i = -1$ and $\bar{u}_i = +1$

Hence, m, n, m_{1i}, m_{0i}, n_{1i} and n_{0i} can simply be obtained by counting in the simulations. That is,

$$\frac{m}{n} \approx \frac{P(\hat{H}_1)}{P(\hat{H}_0)}$$
$$\frac{m_{1i}}{n_{1i}} \approx \frac{P(u_i = +1, \hat{H}_1)}{P(u_i = +1, \hat{H}_0)}$$
$$\frac{m_{0i}}{n_{0i}} \approx \frac{P(u_i = -1, \hat{H}_0)}{P(u_i = -1, \hat{H}_1)} \tag{24}$$

We shall next develop the updating rule for the fusion centre. From eqn. 14,

$$\hat{w}_0 = \log\frac{P(\hat{H}_1)}{P(\hat{H}_0)}$$
$$\hat{w}_i = \begin{cases} \log\frac{P(u_i=+1|\hat{H}_1)}{P(u_i=+1|\hat{H}_0)} & \text{if } u_i = +1 \\ \log\frac{P(u_i=-1|\hat{H}_0)}{P(u_i=-1|\hat{H}_1)} & \text{if } u_i = -1 \end{cases}$$

Using the Bayes rule, $P(x, y) = P(x|y)P(y)$,

$$\hat{w}_i = \begin{cases} \log\frac{P(u_i=+1,\hat{H}_1)P(\hat{H}_0)}{P(u_i=+1,\hat{H}_0)P(\hat{H}_1)} & \text{if } u_i = +1 \\ \log\frac{P(u_i=-1,\hat{H}_0)P(\hat{H}_1)}{P(u_i=-1,\hat{H}_1)P(\hat{H}_0)} & \text{if } u_i = -1 \end{cases} \tag{25}$$

Applying eqns. 24 and 14 to eqn. 25 yields

$$\hat{w}_0 \approx \log\frac{m}{n}$$
$$\hat{w}_i \approx \begin{cases} \log\frac{m_{1i}}{n_{1i}} - \hat{w}_0 & \text{if } u_i = +1 \\ \log\frac{m_{0i}}{n_{0i}} + \hat{w}_0 & \text{if } u_i = -1 \end{cases} \tag{26}$$

and

$$m \approx e^{\hat{w}_0} n$$
$$m_{1i} \approx n_{1i}\exp(\hat{w}_i + \hat{w}_0) \quad \text{if } u_i = +1$$
$$m_{0i} \approx n_{0i}\exp(\hat{w}_i - \hat{w}_0) \quad \text{if } u_i = -1 \tag{27}$$

Taking the partial derivative of eqn. 26 with respect to m, n, m_{1i}, m_{0i}, n_{1i} and n_{0i}, respectively,

$$\frac{\partial\hat{w}_0}{\partial m} \approx \frac{1}{m}$$
$$\frac{\partial\hat{w}_0}{\partial n} \approx -\frac{1}{n} = -\frac{1}{m}e^{\hat{w}_0} \tag{28}$$

and

$$\frac{\partial\hat{w}_i}{\partial m_{1i}} \approx \frac{1}{m_{1i}}, \quad \frac{\partial\hat{w}_i}{\partial n_{1i}} \approx -\frac{1}{m_{1i}}e^{\hat{w}_i+\hat{w}_0} \quad \text{if } u_i = +1 \tag{29}$$

$$\frac{\partial\hat{w}_i}{\partial m_{0i}} \approx \frac{1}{m_{0i}}, \quad \frac{\partial\hat{w}_i}{\partial n_{0i}} \approx -\frac{1}{m_{0i}}e^{\hat{w}_i-\hat{w}_0} \quad \text{if } u_i = -1 \tag{30}$$

If the current local detector's decision conforms to the reliable fusion, its weight $\hat{w}_i$ should be reinforced. In this case,

$$\Delta\hat{w}_i \approx \begin{cases} \frac{1}{m_{1i}}\Delta m_{1i} = \frac{1}{m_{1i}} & \text{if } u_i = +1 \text{ and } \hat{H}_1 \\ \frac{1}{m_{0i}}\Delta m_{0i} = \frac{1}{m_{0i}} & \text{if } u_i = -1 \text{ and } \hat{H}_0 \end{cases} \tag{31}$$

On the other hand, if the current local decision contradicts the reliable decision, its weight $\hat{w}_i$ should be reduced. That is,

$$\Delta\hat{w}_i \approx \begin{cases} -\frac{1}{n_{1i}}\Delta n_{1i} = -\frac{1}{m_{1i}}e^{\hat{w}_i+\hat{w}_0}\Delta n_{1i} \\ \qquad = -\frac{1}{m_{1i}}e^{\hat{w}_i+\hat{w}_0} & \text{if } u_i = +1 \text{ and } \hat{H}_0 \\ -\frac{1}{n_{0i}}\Delta n_{0i} = -\frac{1}{m_{0i}}e^{\hat{w}_i-\hat{w}_0}\Delta n_{0i} \\ \qquad = -\frac{1}{m_{0i}}e^{\hat{w}_i-\hat{w}_0} & \text{if } u_i = -1 \text{ and } \hat{H}_1 \end{cases} \tag{32}$$

and

$$\Delta\hat{w}_0 \approx \begin{cases} \frac{1}{m}\Delta m = \frac{1}{m} & \text{when } \hat{H}_1 \text{ occurs} \\ -\frac{1}{n}\Delta n = -\frac{1}{m}e^{\hat{w}_0}\Delta n \\ \qquad = -\frac{1}{m}e^{\hat{w}_0} & \text{when } \hat{H}_0 \text{ occurs} \end{cases} \tag{33}$$

Thus, we obtain the following updating rule:

$$\hat{w}_i^+ = \hat{w}_i^- + \Delta\hat{w}_i, \quad i = 0, 1, 2, \ldots \tag{34}$$

where $\hat{w}_i^+$ and $\hat{w}_i^-$ represent the weight after and before each update. As the steady state $\hat{w}_i$s are what we are trying to compute, for actual implementation, we use the current estimated weight $\hat{w}_i^-$ to compute $\Delta\hat{w}_i$. That is, to update the weights according to eqn. 34, $\Delta\hat{w}_i$ is computed according to Table 1.

Table 1: Computation of $\Delta\hat{w}_i$

	$\hat{H}_1$		$\hat{H}_0$	
	$u_i = +1$	$u_i = -1$	$u_i = +1$	$u_i = -1$
$\Delta\hat{w}_0$	$1/m$		$-(1/m)e^{\hat{w}_0^-}$	
$\Delta\hat{w}_i$	$1/m_{1i}$	$-(1/m_{0i})e^{(\hat{w}_i^- - \hat{w}_0^-)}$	$-(1/m_{1i})e^{(\hat{w}_i^- + \hat{w}_0^-)}$	$1/m_{0i}$

Lemma 4: Using the updating rule according to eqn. 34 and Table 1, $\hat{w}_i^-$ will converge to the desired steady state estimated weight $\hat{w}_i$.

Proof: At steady state,

$$E[\hat{w}_i^+ - \hat{w}_i^-] = 0 \tag{35}$$

Using the definition $E[X] = \Sigma\, x_i\, P(x_i)$, the updating rule according to eqn. 34 and Table 1, with $u_i = +1$, eqn. 35 becomes

$$\frac{1}{m_{1i}} P(u=+1, \hat{H}_1) - \frac{1}{m_{1i}} e^{\hat{w}_i^- + \hat{w}_0^-} P(u=+1, \hat{H}_0) = 0$$

Using eqn. 25 for further simplification yields

$$\hat{w}_i^- + \hat{w}_0^- = \hat{w}_i + \hat{w}_0 \qquad (36)$$

Similarly, if $u_i = -1$, we have

$$\hat{w}_i^- - \hat{w}_0^- = \hat{w}_i - \hat{w}_0 \qquad (37)$$

For $i = 0$, the following condition can similarly be obtained at steady state:

$$\frac{1}{m} P(\hat{H}_1) - \frac{1}{m} e^{\hat{w}_0^-} P(\hat{H}_0) = 0$$

Thus,

$$\hat{w}_0^- = \hat{w}_0 \qquad (38)$$

Hence, $\hat{w}_i^- \to \hat{w}_i$, for $i = 0, 1, \ldots, n$.

Convergence is a crucial criterion in determining whether an adaptive algorithm is workable. Lemma 4, which proves the convergence of our proposed algorithm, analytically justifies the validity of the algorithm.

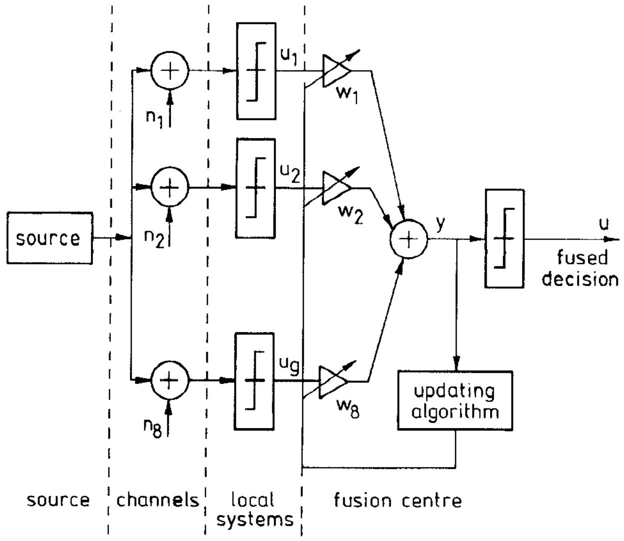

Fig. 5 *Computer simulation diagram*

4 Simulations

Fig. 5 shows the simulation set up to validate the proposed adaptive fusion model. In the simulation presented here, the source produces a binary signal with $P(H_1) = 0.3$ and $P(H_0) = 0.7$, where H_1: +1 and H_0: −1. Eight sensors are used. The probabilities of false alarm and miss P_F and P_M of each sensor are fixed, but not known to the system. The channel is additive Gaussian noise. The Gaussian random variables are generated according to the following transformation:

$$\begin{cases} x = (-2\ln r_1)^{1/2} \cos 2\pi r_2 \\ y = (-2\ln r_1)^{1/2} \sin 2\pi r_2 \end{cases}$$

where r_1 and r_2 are uniformly distributed on (0, 1), and (x, y) becomes a pair of orthogonally normalised Gaussian random variables. The additive Gaussian variable for each sensor is zero-mean with a standard deviation σ ranging from 0.5 to 1.27.

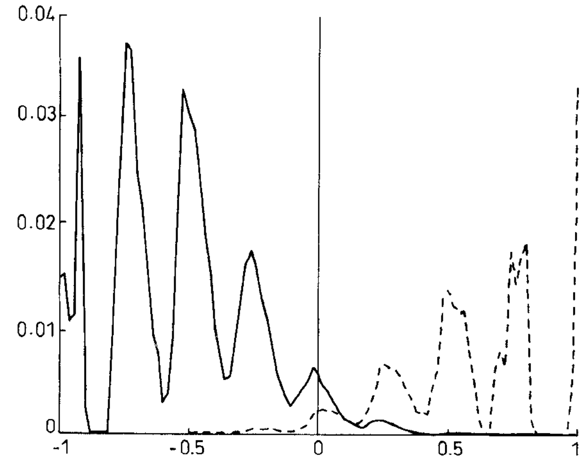

Fig. 6 *Probability mass functions $P(y|H_1)$ and $P(y|H_0)$*

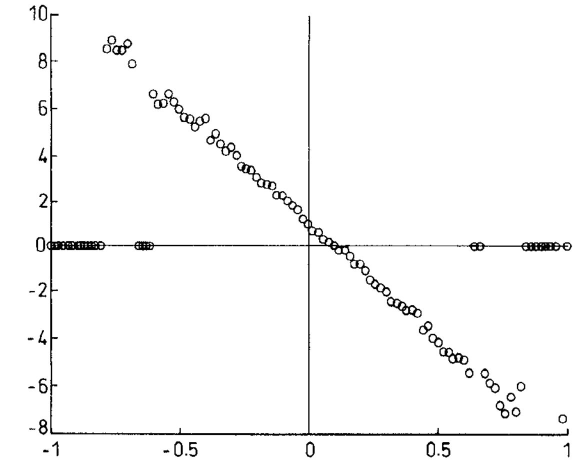

Fig. 7 *Log ratio of probability mass functions log $P(y|H_1)/P(y|H_0)$*

4.1 Conditional probability mass function of y

Fig. 6 shows the histograms of $P(y = \zeta|H_0)$ and $P(y = \zeta|H_1)$ for eight sensors and 250 000 samples. We can see that the waveforms are not monotonic. Fig. 7, which illustrates log $(P(y = \zeta|H_1)/P(y = \zeta|H_0))$, is almost a straight line, conforming to Lemma 1:

$$e^{\zeta} = \frac{P(y - w_0 = \zeta|H_1)}{P(y - w_0 = \zeta|H_0)}$$

4.2 Convergence of weights

Fig. 8 shows average errors of weights $|w_i - \hat{w}_i|$ for different τ, $\tau = 0$, $0.1y_{max}$, and $0.4y_{max}$. Here, $\tau = |\tau_1| = |\tau_2|$. As shown in the Figure, the larger the τ, the smaller the error, which agrees with Lemma 2. As the number of unreliable samples increases, the training time becomes longer. Fig. 9 shows how a weight adapts when the system is changed (P_{Di} and P_{Fi} are modified) in the middle of a simulation; that is, the variance of the noise of the system is modified at the 2500th iteration (steady state). Note that the weight can still track the changes but is not as responsive as when the system is reinitialised. Thus, to enable prompt response, a 'reset' procedure at which changes are made should be initiated. Another alternative is to use a sliding window (of a fixed size) to estimate the weight. The detection of changes at which the 'reset' is initiated and the concept of sliding window are, however, beyond the scope of this paper, and will be addressed in future research.

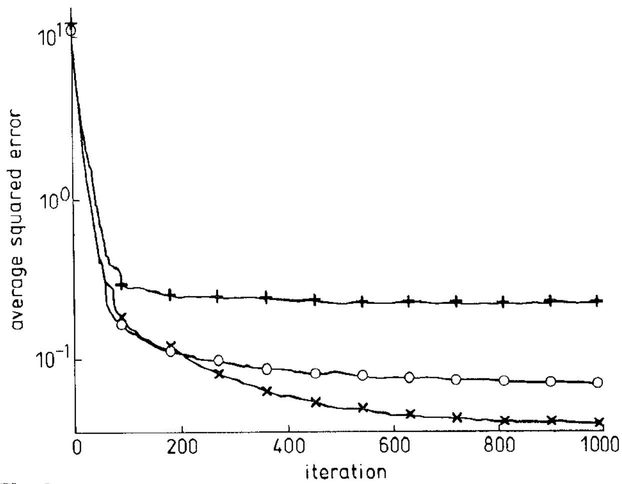

Fig. 8 *Error with various reliability thresholds*
$+\ \tau = 0.0$
$\bigcirc\ \tau = 0.1\ y_{max}$
$\times\ \tau = 0.4\ y_{max}$

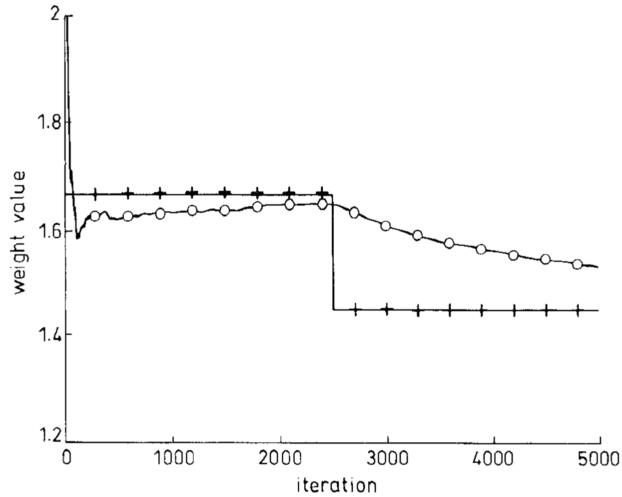

Fig. 9 *Adaptation of weight of second sensor when system environment is modified at 2500th iteration*
+ theoretical weights for $\sigma^2 = 1$ and $\sigma^2 = 1.3$
○ actual weights for $\sigma^2 = 1$ and $\sigma^2 = 1.3$

5 Conclusions

In the real-world environment, the probability mass functions of the observations at local detectors may not be known, and the performance of the local detectors may not be consistent. Under such circumstances, a system that can adapt itself during the decision-making process is needed. The major advantage is that the system can still have smaller error and does not need *a priori* knowledge of the probability mass functions of the observations. Simulation results conform to our theoretical analysis.

6 References

1 TENNEY, R., and SANDELL, J.N.R.: 'Detection with distributed sensors', *IEEE Trans.*, 1981, **AES-17**, pp. 501–510
2 SADJADI, F.: 'Hypotheses testing in a distributed environment', *IEEE Trans.*, 1986, **AES-22**, pp. 134–137
3 CHAIR, Z., and VARSHNEY, P.K.: 'Optimal data fusion in multiple sensor detection systems', *IEEE Trans.*, 1986, **AES-22**, pp. 98–101
4 THOMOPOULOS, S., VISWANATHAN, R., and BOUGOOULIAS, D.: 'Optimal decision fusion in multiple sensor systems', *IEEE Trans.*, 1987, **AES-23**, pp. 644–653
5 DEMIRBAS, K.: 'Maximum *a posteriori* approach to object recognition with distributed sensors', *IEEE Trans.*, 1988, **AES-24**, pp. 309–313
6 BATALAMA, S.N., KOYIANTIS, A.G., PAPANTONI-KAZAKOS, P., and KAZAKOS, D.: 'Feedforward neural structures in binary hypothesis testing', *IEEE Trans. Commun.*, 1993, **41**, pp. 1047–1062
7 PADOS, D.A., PAPANTONI-KAZAKOS, P., KAZAKOS, D., and KOYIANTIS, A.G.: 'On-line threshold learning for Neyman-Pearson distributed detection', *IEEE Trans. Syst. Man. Cybern.*, 1994, **24**, pp. 1519–1531
8 NAIM, A., and KAM, M.: 'On-line estimation of probabilities for Bayesian distributed detection', *Automatica*, 1994, **30**, (4) pp. 633–642
9 ANSARI, N., HOU, E., ZHU, B., and CHEN, J.: 'An adaptive fusion model for distributed detection systems', *IEEE Trans. Aerosp. Electron. Syst.*, 1996, **32**, pp. 524–531
10 HASSOUN, M.H.: 'Fundamentals of artificial neural networks' (MIT Press, Cambridge, MA, 1995)

7 Appendix

7.1 Proof of Lemma 1

Consider the structure shown in Fig. 2. We have

$$y = w_0 + \sum_{j=1}^{n} w_j u_j \tag{39}$$

or

$$y = w_0 + \sum_{j\in S^+} w_j - \sum_{j\in S^-} w_j \tag{40}$$

where $S^+ = \{j{:}\ u_j = 1\}$, and $S^- = \{j{:}\ u_j = -1\}$. From eqns. 4 and 40,

$$y = w_0 + \log\left[\prod_{j\in S^+} \frac{P(u_j = 1|H_1)}{P(u_j = 1|H_0)} \Big/ \prod_{j\in S^-} \frac{P(u_j = -1|H_0)}{P(u_j = -1|H_1)}\right] \tag{41}$$

$$\exp(y - w_0) = \frac{\prod_{j\in S^+} P(u_j = 1|H_1) \prod_{j\in S^-} P(u_j = -1|H_1)}{\prod_{j\in S^+} P(u_j = 1|H_0) \prod_{j\in S^-} P(u_j = -1|H_0)} \tag{42}$$

Let ζ be a possible value of $y - w_0$, and let each local decision u_j be independent:

$$P(y - w_0 = \zeta|H_1) = \sum_{\mathbf{u}\in U} P(\mathbf{w}^T\mathbf{w} = \zeta|H_1)$$

where $\mathbf{u}$ is a vector with elements u_i, $i = 1, 2, ..., n$; $\mathbf{w}$ is a vector with elements w_i, $i = 1, 2, ..., n$, and

$$U = \{\mathbf{u} : \mathbf{w}^T\mathbf{u} = \zeta\}$$

By defining S as

$$\left\{ \{S^+, S^-\} : \text{a combination of } S^+ \text{ and } S^- \text{ such that } \sum_{j\in S^+} w_j - \sum_{j\in S^-} w_j = \zeta \right\}$$

$$P(y - w_0 = \zeta|H_1) = \sum_S \prod_{j\in S^+} P(u_j = 1|H_1) \prod_{j\in S^-} P(u_j = -1|H_1)$$

and

$$P(y - w_0 = \zeta|H_0) = \sum_S \prod_{j\in S^+} P(u_j = 1|H_0) \prod_{j\in S^-} P(u_j = -1|H_0)$$

Thus,

$$\frac{P(y - w_0 = \zeta|H_1)}{P(y - w_0 = \zeta|H_0)} = \frac{\sum_S \prod_{j\in S^+} P(u_j = 1|H_1) \prod_{j\in S^-} P(u_j = -1|H_1)}{\sum_S \prod_{j\in S^+} P(u_j = 1|H_0) \prod_{j\in S^-} P(u_j = -1|H_0)} \tag{43}$$

From eqn. 42 and the following equality:

$$\frac{a}{b} = \frac{c}{d} = k \Rightarrow \frac{a+c}{b+d} = \frac{bk+dk}{b+d} = k$$
$$(b \neq 0, d \neq 0, b+d \neq 0)$$

then

$$\frac{P(y-w_0=\zeta|H_1)}{P(y-w_0=\zeta|H_0)} = \frac{\prod_{j\in S^+} P(u_j=1|H_1) \prod_{j\in S^-} P(u_j=-1|H_1)}{\prod_{j\in S^+} P(u_j=1|H_0) \prod_{j\in S^-} P(u_j=-1|H_0)}$$

$$\frac{P(y-w_0=\zeta|H_1)}{P(y-w_0=\zeta|H_0)} = e^{y-w_0} = e^{\zeta} \tag{44}$$

7.2 Proof of Lemma 2

$$\alpha = \frac{\sum_{j=1}^{m} P(y_i=\zeta_j|H_0)}{\sum_{j=1}^{m} P(y_i=\zeta_j|H_1)}$$

Without loss of generality, assume that $\zeta_1 > \zeta_2 > \ldots > \zeta_m > \tau_2$ are all possible y_j. Note that, as τ_2 becomes larger, m becomes smaller.

From eqn. 7,

$$e^{-\zeta} = \frac{P(y_i=\zeta|H_0)}{P(y_i=\zeta|H_1)} \tag{45}$$

we have

$$\frac{P(y_i=\zeta_1|H_0)}{P(y_i=\zeta_1|H_1)} < \frac{P(y_i=\zeta_2|H_0)}{P(y_i=\zeta_2|H_1)} < \cdots < \frac{P(y_i=\zeta_m|H_0)}{P(y_i=\zeta_m|H_1)} \tag{46}$$

Denote $A_k = \Sigma_{j=1}^{k} P(y_i = \zeta_j|H_1)$, $B_k = \Sigma_{j=1}^{k} P(y_i = \zeta_j|H_0)$, and $\alpha_k = A_k/B_k$. The objective is to show that $\alpha_k > \alpha_{k-1}$ for $k = 1, 2, \ldots, n$. First, we need to show $\alpha_2 > \alpha_1$.

$$\alpha_1 = \frac{P(y_i=\zeta_1|H_0)}{P(y_i=\zeta_1|H_1)}$$
$$\alpha_2 = \frac{P(y_i=\zeta_1|H_0)+P(y_i=\zeta_2|H_0)}{P(y_i=\zeta_1|H_1)+P(y_i=\zeta_2|H_1)}$$

Using the following inequality and eqn. 46,

$$\frac{X}{Y} < \frac{a}{b} \Rightarrow \frac{X}{Y} < \frac{X+a}{Y+b} < \frac{a}{b} \quad (Y, b > 0) \tag{47}$$

we have

$$\alpha_2 > \alpha_1 \tag{48}$$

Next, we shall show that, if $\alpha_k > \alpha_{k-1}$ then $\alpha_{k+1} > \alpha_k$. As

$$\alpha_{k-1} < \alpha_k \tag{49}$$

$$\frac{A_{k-1}}{B_{k-1}} < \frac{A_k}{B_k} \Rightarrow \frac{A_{k-1}}{B_{k-1}} < \frac{A_{k-1}+P(y_i=\zeta_k|H_0)}{B_{k-1}+P(y_i=\zeta_k|H_1)} \tag{50}$$

Using inequality eqn. 47 again,

$$\frac{A_{k-1}}{B_{k-1}} < \frac{A_{k-1}+P(y_i=\zeta_k|H_0)}{B_{k-1}+P(y_i=\zeta_k|H_1)} < \frac{P(y_i=\zeta_k|H_0)}{P(y_i=\zeta_k|H_1)} \tag{51}$$

Applying eqns. 46 and 50 to eqn. 51 yields

$$\frac{A_{k-1}+P(y_i=\zeta_k|H_0)}{B_{k-1}+P(y_i=\zeta_k|H_1)} < \frac{P(y_i=\zeta_k|H_0)}{P(y_i=\zeta_k|H_1)} < \frac{P(y_i=\zeta_{k+1}|H_0)}{P(y_i=\zeta_{k+1}|H_1)}$$

Using eqn. 47,

$$\frac{A_{k-1}+P(y_i=\zeta_k|H_0)}{B_{k-1}+P(y_i=\zeta_k|H_1)} < \frac{A_{k-1}+P(y_i=\zeta_k|H_0)+P(y_i=\zeta_{k+1}|H_0)}{B_{k-1}+P(y_i=\zeta_k|H_1)+P(y_i=\zeta_{k+1}|H_1)}$$

$$\Rightarrow \frac{A_k}{B_k} < \frac{A_{k+1}}{B_{k+1}} \Rightarrow \alpha_k < \alpha_{k+1} \tag{52}$$

From eqns. 48, 51 and 52, α decreases monotonically with τ_2.

However, τ_2 cannot go to infinity; the maximum value of τ_2 is $(y_i)_{max}$. When τ_2 attains its maximum, α reaches its minimum value. According to the definition of α, the minimum of α is

$$\alpha_{min} = \frac{P(y_i=(y_i)_{max}/H_0)}{P(y_i=(y_i)_{max}/H_1)} = \exp(-(y_i)_{max}) \tag{53}$$

When $P_{Di} = 1 - P_{Mi}$ (the probability of detection) is greater than P_{Fi} for each sensor (which is the usual case), every weight in y_i (eqn. 6) is positive. Thus, the maximum value for y_i is

$$(y_i)_{max} = \sum_{j=1, j\neq i}^{n} \log \frac{1-P_{M_j}}{P_{F_j}} \tag{54}$$

Likewise, the minimum value of y_i is

$$(y_i)_{min} = -\sum_{j=1, j\neq i}^{n} \log \frac{1-P_{F_j}}{P_{M_j}} \tag{55}$$

Thus,

$$\alpha_{min} = \exp(-(y_i)_{max}) = \exp\left(-\sum_{j=1,j\neq i}^{n} \log \frac{1-P_{M_j}}{P_{F_j}}\right)$$
$$= \prod_{j=1,j\neq i}^{n} \frac{P_{F_j}}{1-P_{M_j}} \tag{56}$$

By the same reasoning, we can prove that β is decreasing when τ_1 decreases, and

$$\beta_{min} = \frac{P(y_i=(y_i)_{min}|H_1)}{P(y_i=(y_i)_{min}|H_0)} = \exp((y_i)_{min})$$
$$= \prod_{j=1,j\neq i}^{n} \frac{P_{M_j}}{1-P_{F_j}} \tag{57}$$

When $\tau_2 = (y_i)_{max}$ and $\tau_1 = (y_i)_{min}$,

$$P(y_i \geq \tau_2|H_1) = \prod_{j=1,j\neq i}^{n} P(u_j=1|H_1) = \prod_{j=1,j\neq i}^{n} (1-P_{M_i}) \tag{58}$$

$$P(y_i \leq \tau_1|H_0) = \prod_{j=1,j\neq i}^{n} P(u_j=-1|H_0) = \prod_{j=1,j\neq i}^{n} (1-P_{F_i}) \tag{59}$$

Thus,

$$\gamma = \frac{P(\widehat{H}_1|H_1)}{P(\widehat{H}_0|H_0)} = \prod_{j=1,j\neq i}^{n} \left(\frac{1-P_{M_i}}{1-P_{F_i}}\right) \tag{60}$$