USE YOUR POWERS WISELY: RESOURCE ALLOCATION IN PARALLEL CHANNELS

Santiago Jaramillo *Barak A. Pearlmutter*

Hamilton Institute
National University of Ireland Maynooth, Co. Kildare, Ireland

ABSTRACT

This study evaluates various resource allocation strategies for simultaneous estimation of two independent signals from noisy observations. We focus on strategies that make use of the underlying dynamics of each signal, exploiting the difference in estimation uncertainty between them. This evaluation is done empirically, by exploring the parameter space through computer simulations. Two cases are studied: one in which an initial allocation is maintained during estimation of the variables, and one in which allocation can be dynamically changed at each time step according to the uncertainty of the estimate from each channel. The results suggest that there are conditions in which it is advantageous to assign a high signal-to-noise ratio (SNR) to only one of the signals and *guess* the other one. Furthermore, comparison between the two allocation strategies shows that the dynamic strategy significantly improves estimation performance in low SNR scenarios when the signals have similar dynamics.

1. INTRODUCTION

Knowledge about the dynamics of a random process and the ability to assign more resources to represent features of interest can be used to improve statistical inference of the underlying process from noisy observations. This paper evaluates different resource allocation strategies for simultaneous estimation of two independent signals contaminated with noise. The dynamics of the signals to be estimated are assumed to be known, but the signals themselves are *hidden* and only noisy observations are available. Resources in our case refer to the quality of the observation, where a constraint on the available resources sets a limit on the average quality across channels.

This problem is motivated by studies of bottom-up attention, saliency and active perception [1–3], in which it is hypothesized that we *orient* our senses towards features of interest in order to maximize the gathering of information relevant for a particular task. An analogy of the problem described in this paper is that of a basketball player who has to keep track of other players in the court. To do this, he/she must decide where to look, how often to change gaze, and how to track more than one player without having to look directly at any. Good gazing strategies will presumably depend on the predictability and relevance of each of the targets being tracked.

These ideas are also related to the concept of bit allocation for signal coding and compression, in which a limited number of bits is distributed between different features (*e.g.* subbands) according to their perceptual relevance and statistical characteristics [4].

After presenting the general problem of allocation in parallel channels, we describe the simple setup that will be the core of this paper. We then evaluate the performance of the allocator/estimator over different sets of parameters. Optimal strategies are found by exploring the parameter space through computer simulations. We compare two different conditions for allocation: one that requires fixing an SNR for each channel and maintaining it during estimation, and a second one in which SNR is dynamically reassigned at each time step according to the changing uncertainties of the estimation for each channel.

2. TWO-CHANNEL MODEL

In our context, the problem of resource allocation can be stated as follows. Consider multiple independent noisy channels in parallel, with a common power constraint. The goal is to distribute the total power among the channels so as to minimize some error function between the original signals and their estimates. If the signals are known to have predictable dynamics, an optimal strategy should presumably combine prediction and power allocation. For the remainder of this paper we will keep the power of the signals fixed and instead vary the noise level for each channel.

For our particular setup the goal is to estimate, at each time step, two independent binary signals each one contaminated by Gaussian noise. The signals are generated by a Markov process with known transition probabilities, and the only link between the two channels is a constraint on the total noise of the system. The allocation problem consists of deciding how much of the total noise should be assigned to each channel so that the average number of errors in either channel is minimal.

The system is illustrated in Fig 1. States variables s_i are binary and change according to transition probability matrices

Supported by Science Foundation Ireland grant 00/PI.1/C067 and the Higher Education Authority of Ireland.

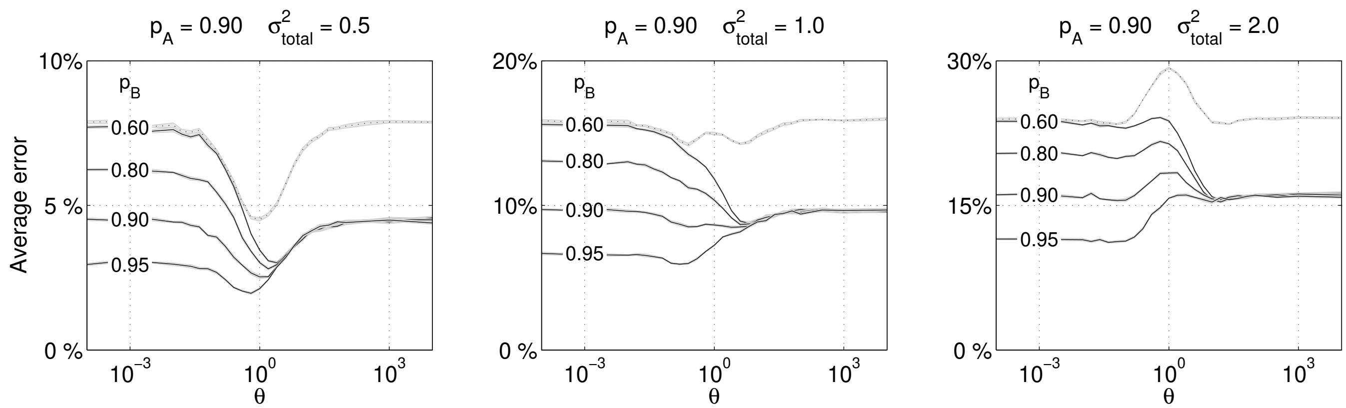

Fig. 2. Fixed allocation. Average error as a function of noise ratio θ. Each panel presents results for a given SNR using one p_{A} and many p_{B}. The dotted line corresponds to the performance of an estimator that ignores the dynamics of the signals. Gray areas around the curves represent the standard error of the mean of the measurements.

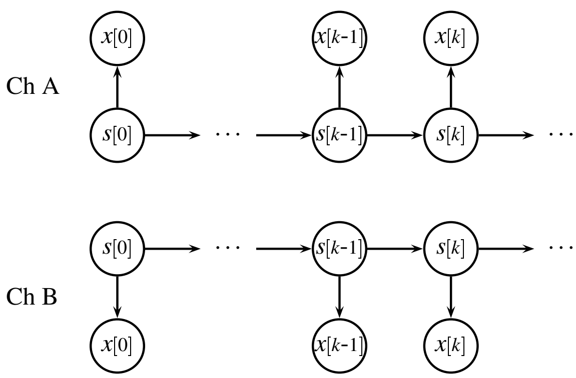

Fig. 1. Two channel Markov model. States $s[k]$ are binary and the probability distribution of the observations given the states $p(x[k]|s[k])$ is Gaussian. The sum of the observation noise across channels at each time k is constant.

T_i (a different one for each channel). To simplify the system we make the transition matrices symmetric. The dynamics can then be described by one variable p_i per channel:

$$s_i[k] \in \{-1, 1\}, \quad i = \mathrm{A}, \mathrm{B} \tag{1}$$

$$T_i = \begin{bmatrix} p_i & 1 - p_i \\ 1 - p_i & p_i \end{bmatrix} \tag{2}$$

$$p_i = P(s_i[k] = s_i[k-1]) \tag{3}$$

The observations are drawn from a Gaussian distribution with mean equal to the current state for each channel:

$$x_i[k] \sim \mathcal{N}\left(s_i[k], \sigma_i^2\right) \tag{4}$$

The constraint on the noise level, which can be interpreted as having a common power source or a constant total signal-to-noise ratio, is given by

$$\sigma^2_{\text{total}} = \sigma^2_{\mathrm{A}} + \sigma^2_{\mathrm{B}} \tag{5}$$

Throughout the paper we refer to the ratio between noise variances as θ:

$$\theta \equiv \frac{\sigma^2_{\mathrm{A}}}{\sigma^2_{\mathrm{B}}} \tag{6}$$

Inference of the state (for each channel) at each time k is done by calculating the maximum a-posteriori (MAP) estimate for the individual state, given the observation sequence:

$$\hat{s}_i[k] = \underset{q \in \{-1,1\}}{\arg\max}\, P(s_i[k] = q \mid x_i[0], \ldots, x_i[k]) \tag{7}$$

$$= \underset{q \in \{-1,1\}}{\arg\max}\, \frac{P(s_i[k] = q, \tilde{x}_i[0], \ldots, \tilde{x}_i[k])}{P(\tilde{x}_i[0], \ldots, \tilde{x}_i[k])} \tag{8}$$

Here we have written $\tilde{x}_i[k]$ instead of $x_i[k]$ so we can write expressions for the observations as probability mass functions. The variables $\tilde{x}_i[k]$ can be interpreted as discretized versions of the observations, and are only used for describing the algorithm, which is unaffected by this approximation. Since the marginal of the observations does not affect the maximum, the problem can be solved by finding the joint probability between the state at time k and the observation sequence. This value corresponds to the forward variable $\alpha[k]$ described in [5]. In vector form we have:

$$\alpha_i[k] = \begin{bmatrix} P(s_i[k] = -1, \tilde{x}_i[0], \ldots, \tilde{x}_i[k]) \\ P(s_i[k] = \ \ 1, \tilde{x}_i[0], \ldots, \tilde{x}_i[k]) \end{bmatrix} \tag{9}$$

$$= \frac{\breve{\alpha}_i[k]}{\sum_{q \in \{-1,1\}} \breve{\alpha}_i[k](q)} \tag{10}$$

$$\breve{\alpha}_i[k] = \left(T_i\ \alpha_i[k-1]\right) \odot f(x_i[k], s_i[k], \sigma_i^2) \tag{11}$$

$$\text{with} \quad \breve{\alpha}_i[0] = \pi_i \odot f(x_i[k], s_i[k], \sigma_i^2) \tag{12}$$

Here, $\odot$ denotes the element-by-element (or Hadamard) product, π_i is a 2-element vector that represents the initial state probabilities, and $f(x, s, \sigma^2)$ is a 2-element vector containing the likelihood (for each possible state) that the sample x came from the distribution defined in (4).

The estimation method described here maximizes the expected number of correct individual states by choosing the most likely state at each time step for each channel. Other optimality criteria imply different methods, *e.g.*, the Viterbi algorithm [6] finds the most likely state sequence for a given observation sequence. In this paper we will focus only on the individual state MAP estimate described above.

Our goal is to evaluate the performance of this estimation method for different allocation strategies. Since we are interested in simultaneous tracking of the two channels, we measure performance by calculating the average number of time steps in which errors occur in either channel:

$$\bar{\mathcal{E}} = E\left[\hat{s}_{\mathrm{A}}[k] \neq s_{\mathrm{A}}[k] \vee \hat{s}_{\mathrm{B}}[k] \neq s_{\mathrm{B}}[k]\right] \tag{13}$$

where $E[\cdot]$ represents the expectation over time. A lower $\bar{\mathcal{E}}$ indicates better performance.

3. ALLOCATION STRATEGIES

We explore two allocation methods: one in which the ratio θ is set initially and maintained during estimation, and one in which the ratio can be changed at each time step.

3.1. Fixed allocation

In the fixed case we want to find the constant ratio θ that minimizes the average error $\bar{\mathcal{E}}$, given the system parameters and total noise power:

$$\theta^* = \underset{\theta=\sigma_{\mathrm{A}}^2/\sigma_{\mathrm{B}}^2}{\arg\min}\, \bar{\mathcal{E}} \qquad \text{s.t.} \qquad \sigma_{\text{total}}^2 = \sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{B}}^2 \tag{14}$$

One way to solve this problem is to derive an expression for the probability of making an error when estimating the states $P(\hat{s}_i[k] \neq s_i[k])$ at each time step k for each channel i, calculate the expectation over time of errors made in either channel, and then solve the minimization problem either analytically or numerically. We take instead an empirical approach in which we simulate the system for fixed values of p_i and σ_{total}^2, and find the value of θ that gives minimal error.

Results for different noise ratio θ are compared to the performance achieved by an estimator that assumes a white process and ignores the dynamics of the system.

3.2. Dynamic allocation

Many strategies for dynamic allocation could be considered. Here we explore one in which the ratio is changed at each step according to the relative uncertainties of the two channels. In this case, we define the certainty of our estimate as:

$$c_i[k] = \left|0.5 - \alpha_i[k]^{(1)}\right| \tag{15}$$

where $\alpha_i[k]^{(1)}$ corresponds to the first element of the joint probability vector $\alpha_i[k]$ for channel i at time k. This quantity represents how close to 1 is the probability of being in one state, indicating a level of certainty about that estimate. From these values, we set the ratio at each time step as:

$$\theta[k] = \frac{\sigma_{\mathrm{A}}^2[k]}{\sigma_{\mathrm{B}}^2[k]} = \left(\frac{c_{\mathrm{A}}[k-1]}{c_{\mathrm{B}}[k-1]}\right)^{\phi} \tag{16}$$

And we want to find the optimal exponent:

$$\phi^* = \underset{\phi}{\arg\min}\, \bar{\mathcal{E}} \qquad \text{s.t.} \qquad \sigma_{\text{total}}^2 = \sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{B}}^2 \tag{17}$$

The exponent ϕ enables a non-linear relation between the certainty ratio and the noise ratio at the next time step. Note that this function includes the case in which the noise ratio is always 1 (ignoring the certainties) and the case in which resources are completely moved from one channel to the other at each time step.

4. PERFORMANCE MEASUREMENTS

4.1. Fixed allocation

The system was simulated using the fixed allocation method for different transition probabilities and three SNR scenarios ($\sigma_{\text{total}}^2 = 0.5, 1, 2$). The average error was measured following (13) as the noise ratio θ is varied. Results for $p_{\mathrm{A}} = 0.9$ are shown in Fig. 2. This figure includes the performance given by an estimator that assumes a white process, *i.e.* an estimator that ignores the dynamics of the system and simply sets a decision boundary at $x_i[k] = 0$.

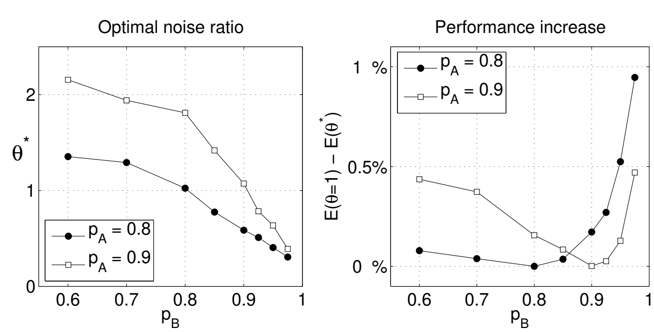

Fig. 3. Optimal fixed allocation in high SNR ($\sigma_{\text{total}}^2 = 0.5$). The left panel shows the value of θ that produces minimal average error. The right panel shows the difference between the optimal error and the error achieved when both channels are assigned the same amount of noise.

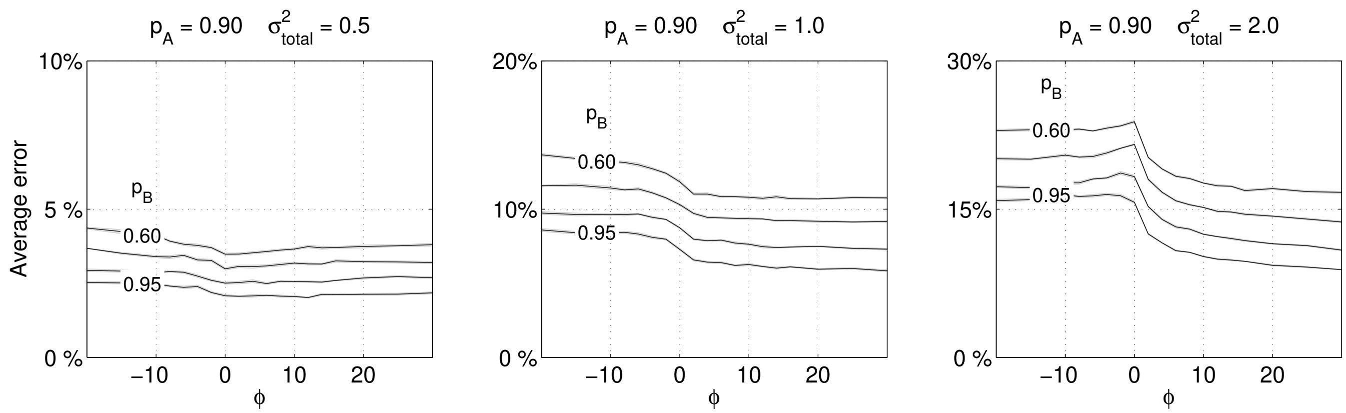

Fig. 4. Dynamic allocation. Average error as a function of exponent ϕ. Each panel presents results for a given SNR using one p_{A} and many p_{B}. Gray areas represent the standard error of the mean at each point simulated.

For each total noise level we obtained a different behavior. In low SNR (right column of Fig. 2) the best strategy seems to be assigning most resources to one of the channels and *guessing* (estimate with very high observation noise) the other one. The curves confirm the intuition that more resources (lower variance) should be assigned to the least predictable channel.

In contrast, when the SNR is high, there is a non-extreme ratio at which the error is minimum. This ratio depends on the relative values of the transition probabilities as shown in the left panel of Fig. 3. These results imply that when both channels have identical dynamics it is optimal to split the resources equally across the channels, which is not the case when the SNR is low. The right panel of Fig. 3 shows the difference between the error obtained with $\theta = 1$ and the minimal error. The improvement in performance when using the optimal ratio is lower than 1% for the transition probabilities shown here. This implies that, when the SNR is high, it may not be worth trying to find the optimal allocation ratio, but simply distributing resources equally across the channels.

4.2. Dynamic allocation

The dynamic allocation method was evaluated using similar parameters to those for the fixed case but using instead the exponent ϕ as the abscissa, and calculating the ratio $\theta[k]$ at each time step. Results for $p_{\text{A}} = 0.9$ and three SNR scenarios are shown in Fig. 4.

For a high SNR ($\sigma^2_{\text{total}} = 0.5$), performance remained almost constant as the exponent ϕ was varied, tending to be lower for high values of ϕ. For a low SNR, the curves indicate that better performance is achieved as the exponent is increased. This implies that even for very small differences between the certainties on state estimates, we should assign all resources to only one channel: that with lower certainty. Thus, according to (16), a higher variance is assigned to the channel with higher certainty.

The extreme case as the exponent increases gives rise to a strategy in which resources are completely shifted from one channel to the other at each time step. This is illustrated in Fig. 5, which shows an example of the dynamic estimation procedure for $\phi = 20$, $p_{\text{A}} = 0.9$, $p_{\text{B}} = 0.8$ and $\sigma^2_{\text{total}} = 2$. Note the alternating behavior of σ^2_{A} in the lower panel of this figure.

4.3. Comparison: fixed *vs.* dynamic

In addition, we want to compare the performance of fixed and dynamic strategies, and find those conditions in which one is more advantageous than the other. For $p_{\text{A}} = 0.9$, we can calculate the minimal error achieved with each method and plot them as a function of p_{B} (Fig. 6).

At high SNRs, performance for both methods is very high and relatively similar. In this case the error is already low enough, leaving little room for improvement. For low SNRs in contrast, there are significant differences in performance from both methods and a clear region in which dynamic allocation is better. This implies that in cases when the dynamics of both channels is very similar it is better to switch from one channel to the other at each time step (best dynamic strategy) than to give all resources to only one channel and guess the other (best fixed strategy). For this particular case, the average error was lowered by 5%.

5. DISCUSSION

The reader should note that the estimation procedure used in the simulations was not derived from the general optimization problem described initially since a slightly different error measure was used. We limited ourselves to finding the optimal noise ratio for a given estimator, but a more general goal would be to devise a theory that encompasses both estimation and allocation. The estimation method used here was selected

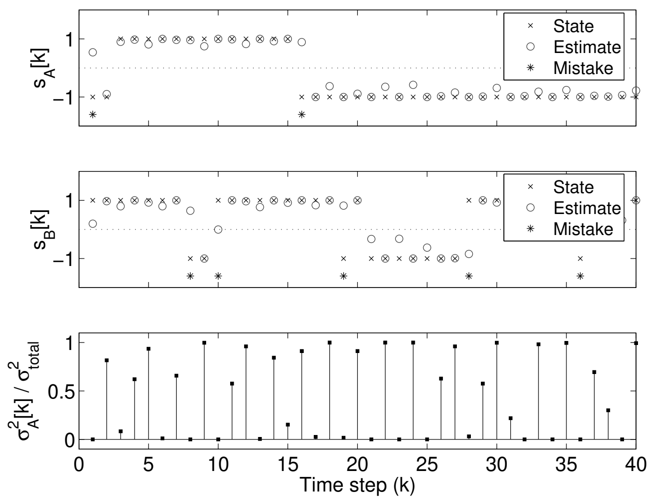

Fig. 5. Example of dynamic allocation. Top two panels show the states and soft-estimates of channels A and B. Soft-estimates represent the probability of being in a particular state scaled to the range $[-1, 1]$. Bottom panel displays the (normalized) amount of noise assigned to channel A at each time step.

for its simplicity and low storage requirements, and because the quantities it relies on relate directly to concepts of uncertainty and confidence of the estimates.

Using different cost functions with problems similar to the one described in this paper may yield significantly different optimal strategies. For example, if our goal is to maximize the total channel capacity in a system of parallel channels, the solution corresponds to giving more resources to the channel with lowest noise, following a process called *water-filling* (see [7] for the Gaussian case). In contrast, if one wants to represent a Gaussian vector with a fixed number of bits and minimal squared error distortion, the best strategy (from Rate Distorsion Theory, see *inverse-water-filling* [7]) assigns more bits to variables with higher variance, *i.e.*, to those affected more by quantization error. In these two examples, improving representation in any channel reduces the total error function, balancing out errors in other channels. When using instead the cost function defined in this paper, improving estimation of one channel will not decrease the total cost when an error has occurred in another channel. The intuition for the fixed allocation strategies found here for the low SNR case is that making an error in either channel is expensive, and estimating the channel with higher p is easier than estimating other channels; therefore, resources should be used to improve estimation of the least predictable channel.

One motive for approaching the problem of resource allocation is to derive theories that predict strategies for active perception. In this context, the noise constraint of a common power source (total sum of variances) may not be appropriate. In the case of vision, for example, there may be a complex

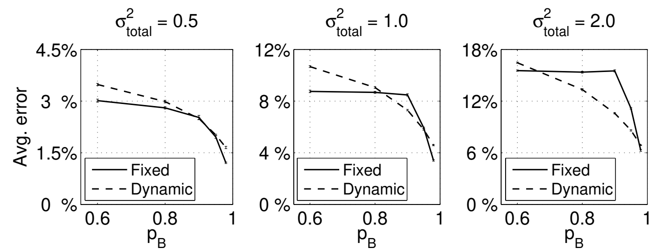

Fig. 6. Performance comparison between fixed and dynamic allocation methods. All curves are calculated for $p_A = 0.9$. Error bars correspond to the standard error of the mean.

function that describes how the quality of the observation depends on gaze.

This study proposes a dynamic allocation method which uses a measurement of certainty to derive the noise ratio at each time step. The method was based on the idea that it is advantageous to give more resources to uncertain targets since they are harder to predict. This is clearly not the only possible dynamic method and further work is necessary in order to find a more general solution by solving the optimization problem over the space of possible dynamic strategies. Furthermore, the dynamic solution described here does not take into account the cost of reassigning resources. It could be the case that switching back and forth between channels gives the lowest error, but is so expensive (in term of energy or other constraints) that becomes suboptimal.

Throughout the paper we assumed perfect knowledge of the dynamics of the signals to be estimated. Further work is necessary to evaluate the robustness of the fixed and dynamic allocation strategies when parameters are not known accurately. Furthermore, the allocation problem should be stated in a learning framework in which estimation of the parameters is done simultaneously (perhaps at a slower time-scale) than estimation of the signals. Some researchers have investigated similar ideas in the context of competitive allocation of learning between stimuli according to their relative uncertainties [8].

Finally, an analytical solution for the optimal noise ratio may provide further insight on the effects of the different system parameters (total system noise and relative predictability of the signals) on performance. An empirical approach like the one used here may miss some of these details, due to the limited sampling of the parameter space.

6. CONCLUSIONS

This paper presents an empirical evaluation of different strategies for distributing noise across parallel Gaussian channels in order to minimize signal estimation error. Noise levels were constrained by a total minimum system noise, and estimation of the transmitted signals was done by finding the most probable input given the history of noisy observations. The allocation strategies made use of the knowledge about the dynamic properties of the signals to be estimated.

Optimal strategies with respect to the error measure defined above depend on the total amount of system noise, giving qualitatively different results for low and high SNRs. With a fixed allocation and a low SNR, the best strategy is to give most resources to the least predictable channel. At a high SNR, performance can be slightly improved by choosing a noise ratio close to one. When allowing dynamic allocation at each time step, the optimal strategy is to rapidly switch between the channels. This strategy, compared to a fixed allocation, significantly increases performance in the low SNR case when both channels have similar dynamic properties.

7. ACKNOWLEDGEMENTS

The authors would like to thank R. Stanojevic for helpful discussions.

8. REFERENCES

[1] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):996–1005, 1988.

[2] L. Itti and C. Koch. Computational modelling of visual attention. *Nat Rev Neurosci*, 2(3):194–203, 2001.

[3] M. Hayhoe and D. Ballard. Eye movements in natural behavior. *Trends Cogn Sci*, 9(4):188–94, 2005.

[4] A. Gersho and R. M. Gray. *Vector quantization and signal compression*. Kluwer Academic Publishers, Boston, 1992.

[5] L. R. Rabiner. Tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–86, 1989.

[6] G. D. Forney. The Viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, 1973.

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

[8] P. Dayan, S. Kakade, and P. Montague. Learning and selective attention. *Nat Neurosci*, 3 Suppl:1218–23, 2000.