# Hyperacuity and Coarse Coding

David J. C. MacKay
Department of Computation and Neural Systems
California Institute of Technology 139–74
Pasadena CA 91125
mackay@hope.caltech.edu

24 January 1990

**Abstract**

In discussions of coarse coding, previous results have suggested that the larger the overlap between receptive fields is made, the sharper the achievable hyperacuity becomes. This note moderates this view by giving for an array of noisy analog receptors the scaling laws for the best achievable hyperacuity as a function of increasing receptive field size.

## 1 Introduction

The benefits and defects of coarse coding have been described in [1, Chapter 3]. The use of an array of receptors with large overlapping receptive fields makes it possible to achieve hyperacuity, that is, to achieve resolution sharper than the inter–receptor distance. However this work and other publications, *e.g.* [2][1], have suggested that it is generally true that the larger the size of the receptive fields is made, the sharper the achievable hyperacuity becomes. This statement is correct for the binary noise–free neurons considered in [1], but this note will show that it is not generally true for *analog neurons with intrinsic noise.*

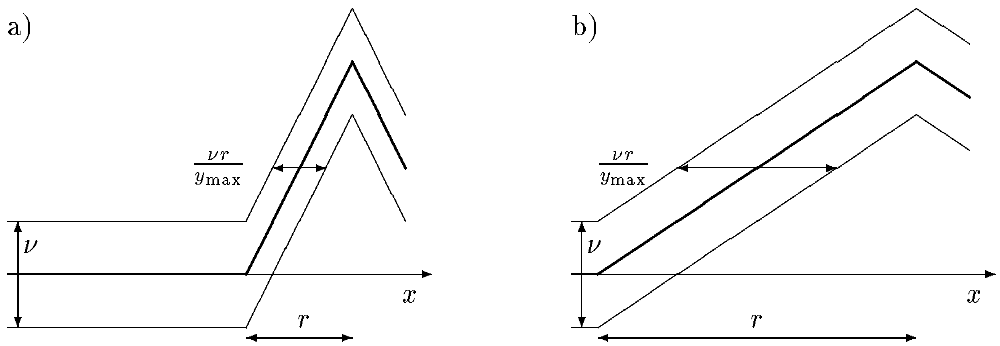
## 2 Assumptions

The same starting point and notation will be used as in [1]. Consider an array of receptors having identical spherically symmetric receptive fields of characteristic radius $r$; the centres of these receptive fields are uniformly distributed with density $n$ in a $k$–dimensional input space. A single stimulus will be considered, at one point in the input space.[2] No temporal aspect will be introduced. The arguments that follow all assume that $r$ is sufficiently large that a sufficient number of receptors are stimulated for hyperacute localisation to be performed. Scaling laws are derived for the behaviour of the *best achievable hyperacuity* for increasing $r$. No limits on the computational power available in subsequent processing are considered here. Such limits would set additional constraints on how large $r$ can be made without loss of hyperacuity, since the complexity of hyperacute localisation is expected to increase with the number of receptor responses that have to be integrated. This note concentrates on the limit imposed by the noise in the receptors themselves.

[1] In this reference, an array analog neurons without additive noise is considered, and a specific computational form of post–processing is considered.

[2] Note that this prior class of stimuli is very restricted. It is different from the band–limited class of functions considered by Nyquist sampling theory, and this is why hyperacuity can be achieved in 'violation' of the Nyquist limit.

Figure 1: **Scaling of spatial uncertainty:**
The thick line represents the average response $\bar{y}_l$ of a receptor as a function of stimulus location $\mathbf{x}_S$. The thin lines represent the range of $y_l$ that results when noise of standard deviation $\nu$ is added to $\bar{y}_l$. The spatial uncertainty given the response of a single receptor is proportional to receptive field width $r$.




The difference from [1] is that each receptor has an analog response with added noise. Thus the response of a receptor $l$ with field centred on $\mathbf{x}_l$ to a stimulus at $\mathbf{x}_S$ is

$$y_l(\mathbf{x}_S) = y_{\max} f\left(\frac{|\mathbf{x}_S - \mathbf{x}_l|}{r}\right) + \mathcal{N}$$

where $\mathcal{N}$ is additive noise, for example, gaussian with standard deviation $\nu$, and $f$ is a monotonic function with range $[0,1]$. I will assume that the amplitude of the noise is small compared with the dynamic range $y_{\max}$ of the neuron, *i.e.* $y_{\max} \gg \nu$. Various scaling laws for the magnitude of the noise can be considered: first I assume that $\nu$ is independent of the receptive field radius $r$; the results for $\nu \sim r^{k/2}$ are also stated; any other scaling law for the noise can easily be substituted.

## 3 Scaling laws: Informal argument

The best achievable hyperacuity might be expected to improve with increasing $r$ because an increasing number of receptors are stimulated by any point stimulus, each contributing information about the stimulus location. This number scales as $N = nr^k$. But the noise imposes a spatial uncertainty that also changes with $r$. For example, in one dimension, the lateral uncertainty in the location of a stimulus, given the reponse of a single receptor, is proportional to $\nu$ divided by the local gain of the reponse, $\frac{\partial}{\partial \mathbf{x}} y_l$. As the receptive field width $r$ is increased, the gain decreases as $y_{\max}/r$, so the spatial uncertainty scales as $s = r\nu/y_{\max}$ (Figure 1). The same holds in $k$ dimensions too: the spatial uncertainty, given the reponse of a single receptor, scales as $r\nu/y_{\max}$.

Now when we integrate together $N$ receptor responses each giving information about $\mathbf{x}_S$ with spatial uncertainty $s$, the resulting uncertainty in the location of $\mathbf{x}_S$ is $\sigma \sim s/\sqrt{N}$.[3] Thus the hyperacuity $\sigma$ scales as:

$$\sigma_{\text{Analog}} \sim \frac{r\nu/y_{\max}}{(nr^k)^{1/2}} \sim \frac{\nu}{y_{\max}} \frac{r^{1-k/2}}{n^{1/2}} \tag{1}$$

[3] This is a standard statistical result in the one–dimensional case; in $k$ dimensions some further justification is required: Each receptor contributes a spherically symmetric likelihood to the posterior probability distribution of the stimulus, indicating that the stimulus probably lies in a spherical shell of thickness $s$. Subject to the initial assumptions, each shell can be treated locally as a one–dimensional gaussian, indicating that the stimulus probably lies in a planar slice. The $N$ planar slices can be collected into orthogonal groups of $k$, each of which makes a spherical gaussian likelihood with radius $s$. These combine to give $\sigma = s/\sqrt{N/k}$.

This $\sigma_{\text{Analog}}$ scales differently with increasing $r$ and $n$ from the $\sigma_{\text{Binary}}$ for binary noise–free neurons [1]:

$$\sigma_{\text{Binary}} \sim \frac{r^{1-k}}{n}. \tag{2}$$

### Why binary neurons give a different result from analog neurons with noise.

The achievement of hyperacuity $\sim r^{1-k}/n$ with binary neurons relies on the assumption that the receptive field boundary of the binary neuron is exactly defined and reliable regardless of receptive field size; this means that the activity of a binary neuron can provide extremely precise locational information if the stimulus is known to be near the receptive field boundary. Given a noisy analog neuron, on the other hand, the locational information provided by the neuron's activity level is always fuzzy with a width that is proportional to the receptive field radius.

### Scaling law for alternative analog receptive field type.

The above scaling law is for a monotonic response function that scales uniformly as the radius increases, for example a conical or gaussian response profile. Bartlett Mel[4] suggested that another receptive field type should be examined, known locally as the inverted dog bowl. Here, the receptive field has a central plateau of variable radius $r$; at the edge of this plateau the flanks have a fixed profile, so that the gain of a single receptor is independent of $r$ in a shell of fixed thickness $w$ in the input space (rather than scaling as $1/r$). For $w \ll r$, the number of receptors that are put onto high gain operating points by a typical stimulus is $\sim nr^{k-1}w$, and the lateral uncertainty in the information provided by each receptor will be $\sim w\nu/y_{\max}$ (note that there is no $r$–dependence). So

$$\sigma_{\text{Dog Bowl}} \sim \frac{w\nu/y_{\max}}{(nr^{k-1}w)^{1/2}} \sim w^{1/2}\frac{\nu}{y_{\max}}\frac{r^{-(k-1)/2}}{n^{1/2}}. \tag{3}$$

So if the inverted dog bowl is implementable, it gives a better scaling law than a profile of the form $f\left(\frac{|\mathbf{x}_S - \mathbf{x}_l|}{r}\right)$, notably in two dimensions. However the engineerability of a dog bowl response for large radius $r$ is questionable.

As an aside it should be noted that the scaling law for the binary neuron cannot be obtained from (3) by letting $w \to 0$, because in that limit the number of receptors on high gain operating points vanishes, violating the initial assumptions. The two scaling laws do join up at the point where the assumptions break down, as can be seen by substituting for the $w$ such that the number of receptors on high gain operating points is $O(1)$. Substituting $N \sim nr^{k-1}w = 1$ in (3), $\sigma_{\text{Dog Bowl}} \sim \frac{r^{1-k}}{n}\frac{\nu}{y_{\max}}$. So at the point at which the assumptions break down, $\sigma_{\text{Dog Bowl}}$ scales in the same way as $\sigma_{\text{Binary}}$.

### Scaling law for noise increasing with $r$.

The preceding derivations assumed $\nu$ independent of $r$. It is also reasonable to consider the case $\nu \sim r^{k/2}$. In this case there is *no improvement* in hyperacuity with increasing $r$ in any number of dimensions for either type of receptive field profile.

[4] Personal communication

## 4 Summary

| Receptor type | Binary | Analog, $f\left(\frac{\lvert \mathbf{x}_S - \mathbf{x}_l \rvert}{r}\right)$ | | Analog, 'inverted dog bowl' | |
|---|---|---|---|---|---|
| | | $\nu$ constant | $\nu \sim r^{k/2}$ | $\nu$ constant | $\nu \sim r^{k/2}$ |
| Scaling of $\sigma$ with increasing $r$ | $\frac{r^{1-k}}{n}$ | $\frac{\nu}{y_{\max}} \frac{r^{1-k/2}}{n^{1/2}}$ | $\frac{1}{y_{\max}} \frac{r}{n^{1/2}}$ | $w^{1/2} \frac{\nu}{y_{\max}} \frac{r^{-(k-1)/2}}{n^{1/2}}$ | $\frac{w^{1/2}}{y_{\max}} \frac{r^{1/2}}{n^{1/2}}$ |
| for $k = 1$ | no change | worse | worse | no change | worse |
| for $k = 2$ | better | no change | worse | better | worse |
| for $k = 3$ | better | better | worse | better | worse |

## References


[1] Rumelhart et al. *Parallel Distributed Processing,* MIT Press. 1986.

[2] Baldi, Heiligenberg. How sensory maps could enhance resolution through ordered arrangement of broadly tuned receivers, *Biol. Cyb.* **59**, pp. 313–318, 1988.

## Appendix

The scaling argument for the hyperacuity can be formalised by examining how the width of the Bayesian posterior scales with $r$. This is illustrated for the one–dimensional case.

### Optimal hyperacuity in one dimension

Consider a one–dimensional spatial array of receptors labelled by an index $l$. Let the average response of the $l$th receptor to the stimulus at $\mathbf{x}_S$ be a continuous bell–shaped curve with a peak value of $y_{\max}$ at $x = l$, and characteristic width $r$:

$$\text{for all } l,\ \bar{y}_l(x_S) = y_{\max} f(\tfrac{x_S - l}{r})$$

The actual response at any moment is subject to additive noise, $\mathcal{N}$.

$$y_l = \bar{y}_l(x_S) + \mathcal{N}_l$$

If the $\mathcal{N}_l$ are independently gaussian distributed with standard deviation $\nu$, then we can write the log–likelihood function:

$$\ln P(y_l|x_S) = -\frac{(y_l - \bar{y}_l(x_S))^2}{2\nu^2} + c \tag{4}$$

We wish to assess how accurately the input $x$ can be inferred given the outputs $\{y_l\}$. How this hyperacuity depends on the width of the receptive field $r$ will then be examined.

Let us assume that all the receptor responses $y_l$ are at our disposal, and we have no limits on computational power or precision for our estimation of the true location $x_S$. The most complete inference we can make given the responses $\{y_l\}$ is the posterior probability distribution for $x_S$:

$$P(x_S = x|\{y_l\}) \sim P(\{y_l\}|x)P(x) \sim \prod_l P(y_l|x)P(x)$$

The likelihood function appearing here, $P(y_l|x)$, is the same as in (4), but Bayes' rule changes its role from a predictive function of $y$ to an inductive function of $x$. As a function of $x$, each of the functions $P(y_l|x)$ is two–peaked, since an output $y_l$ arises with equal probability if the input $x$ is to the right or to the left of $x = l$. One of the peaks is spurious and the other is typically close to the true stimulus location $\mathbf{x}_S$. So when the evidence is combined from all the receptors $l$, most of the functions $P(y_l|x)$ will have one of their peaks in a common location. If the prior $P(x)$ is flat, $P(x|\{y_l\})$ will typically have a single peak near $x_S$. As long as several receptors are contributing peaks to the posterior, all the other spurious peaks will be vanishingly small compared to this 'probably' correct peak. The width of this peak is the measure of the best achievable hyperacuity. To evaluate the hyperacuity we first evaluate the width of the peaks in the individual functions $P(y_l|x)$.

We assume that the standard deviation $\nu$ of the noise $n_l$ that is added to $f_l(x)$ is small compared with $y_{\max}$. This allows us to approximate quadratically the log likelihood function (4) for $x$ close to the maximum likelihood (best fit) $x = x_l^*$ for the datum $y_l$:

$$\ln P(y_l|x) \quad \simeq \quad -\frac{(\bar{y}_l'(x_l^*)\delta x_l))^2}{2\nu^2} + c$$

where $\delta x_l = x - x_l^*$. Now we add together these log likelihoods and obtain, if all the $\mathbf{x}_l^*$ are close enough to each other for the expansions to be simultaneously valid:

$$\ln P(x|\{y_l\}) \quad = \quad \sum_l \ln P(y_l|x) + c$$

$$\simeq \quad -\sum_l \frac{(\bar{y}_l'(x_l^*)\delta x_l))^2}{2\nu^2} + c$$

$$\simeq \quad -\frac{(x-x^*)^2}{2\sigma^2} + c$$

where

$$1/\sigma^2 = \sum_l \bar{y}_l'(x_l^*)^2/\nu^2 \tag{5}$$

$$x^* = \sigma^2 \sum_l x_l^* \bar{y}_l'(x_l^*)^2/\nu^2 \tag{6}$$

So the posterior is a gaussian with mean $x^*$ which is the maximum likelihood $x$ for the entire data set,[5] and width $\sigma$, which is the best achievable hyperacuity.

The magnitude of each of the terms in (5) goes down as $1/r^2$ because $|\bar{y}_l'(x_l^*)|$ is proportional to $1/r$:

$$\bar{y}_l'(x^*) = \tfrac{y_{\max}}{r} f'\left(\tfrac{x^*-l}{r}\right) \tag{7}$$

so

$$1/\sigma^2 = \sum_l \frac{y_{\max}^2}{r^2\nu^2} f'\left(\tfrac{x_l^*-l}{r}\right)^2 \tag{8}$$

If the receptors are distributed with density $n$, the number of terms contributing to the sum (8) increases as $nr$ for $r \gg 1/n$. In the limit of large increasing $r$ the sum can be approximated by an integral:

$$\sum_l f'\left(\tfrac{x_l^*-l}{r}\right)^2 \sim nr \int f'(m)^2 dm$$

So the hyperacuity $\sigma$ scales in accordance with

$$1/\sigma^2 \sim \frac{nry_{\max}^2}{r^2\nu^2} \sim \frac{ny_{\max}^2}{r\nu^2} \tag{9}$$

So $\sigma$ scales as $\frac{\nu}{y_{\max}}\sqrt{r/n}$.

[5] NB this $x^*$ is not a weighted sum of the receptor locations $l$. It's a weighted sum of the individual maximum likelihood $x_l^*$'s.