Analysis of the implementation of an optimal detector in a computational auditory model

02458 - Cognitive Modelling

Paolo Attilio Mesiano (s162585),
Tanmayee Uday Pathre (s171931)

Contents

1 Introduction

Computational auditory models play a crucial role in understanding the functional signal processing of the human auditory system and in simulating human behaviour in response to acoustic stimuli.

In particular, reference is done to the computational auditory model by Dau *et al.* (1996). This model simulates various auditory processing stages of the human hearing system and implements an optimal detector as a decision device. In the original paper, the description of how the decision device is implemented is very dry and condensed for the reader that lacks of a very strong background in statistics and signal detection theory. The purpose of this report is to provide a more exhaustive description about the mathematical framework of such decision device. This is carried out in terms of an m-intervals, m-alternative forced-choice (AFC) framework in relation to a tone-in-noise detection experiment based on the principles of signal detection theory.

2 Structure of the model

The computational model implemented in Dau *et al.* (1996) is mainly divided into two main stages as shown in Figure 1.

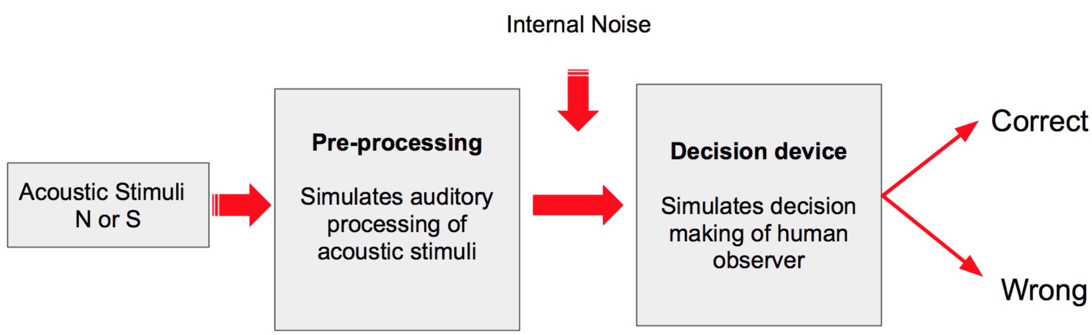

Figure 1: Schematic diagram of the model structure

The input is an acoustic stimuli (noise or tone) which is processed through various signal processing stages, intended to simulate the auditory processing pathway.

The representation of the acoustic stimulus in the human auditory pathway is limited in fidelity. This uncertainty can be described in terms of the so-called *internal noise*. Within the model, this limitation is reproduced by adding Gaussian noise with a given variance. The second stage of the model structure accounts for the decision device which simulates the observer's behaviour in decision making.

3 The implementation of the optimal detector

This section provides a detailed account of the realization of the decision device. The statistical framework of the optimal detector is reviewed with the purpose to provide a more thorough explanation of its implementation. The mathematical steps described in the Appendix of Dau *et al.* (1996) are reviewed and referenced to the relevant topics in signal detection theory and to the related literature. Throughout the following sections, reference is done to a tone-in-noise detection experiment run within an m-intervals, m-AFC experimental paradigm.

3.1 Tone-in-noise detection experiment and alternative-forced-choice paradigms

In psychoacoustics, the purpose of tone-in-noise detection experiments is to measure the detectability of an acoustic target signal (e.g., a pure tone) in adverse listening conditions such as in presence of background noise. For a given noise condition (with given level, spectral characteristics, time onset and duration) the minimum detectable level of the target signal (i.e., the minimum signal level necessary for detection) is measured. The realization of such experiments can be done within an m-interval, m-AFC framework, where at each trial of the experiment the observer is presented with m intervals of sound, one of which contains the tone embedded in noise and all the other $m-1$ intervals contain noise only. The observer's task is to determine which interval contains the tone. The experiment consists of several trials. In the first trial the level is well above the minimum detection level, i.e., at a supra-threshold level. During this first presentation the signal is very salient and it is assumed that the observer creates an auditory image, i.e. a *template* of the signal, by comparing the easily identifiable target interval with all the other noise-only intervals. This template can be used as a term of comparison during the rest of the experiment. In the following trials, depending on the correct or wrong responses of the observer, the level of the signal is varied according to a *tracking procedure* as described in Levitt (1971). When the signal level is below the detection threshold, the observer will not be able to detect the target signal and will therefore choose a random interval. The tone level is tracked as a function of trial number and the threshold level is estimated as the average of a certain number of last presented levels in the procedure.

3.2 Internal representation of the stimuli

In a single trial of the experiment, the sequence of stimuli in the m-AFC task presented to the observer can be represented as a vector $X = (x_1, x_2, x_3, ..., x_m)$ where x_i is the i^{th}-interval waveform. At the output of the pre-processing stage of the model, the acoustic stimulus contained in each interval is described by the internal representation of the acoustic waveform, indicated as $e(t)$. The perceptual evidence of the stimulus vector X can therefore be expressed in terms of an observation vector $E = (e_1, e_2, e_3, ..., e_m)$.

At the beginning of the AFC task, the noise interval and the target interval are clearly distinguishable. The observer creates the internal representation of the *expected signal* s (i.e., the *template*) by computing the difference between the internal representation of the stimulus containing the supra-threshold signal e_S and the internal representation of the stimulus containing noise only e_N:

$$s = e_S - e_N \tag{1}$$

In each of the following trials, the internal representation of the noise-only stimulus e_N is subtracted to the internal representation of the stimuli contained in all the intervals, giving the so-called *difference signal*:

$$\delta_i = e_i - e_N \qquad i = 1...m\,. \tag{2}$$

If the level of the tone is sufficiently high, the difference signal of the target interval will resemble the expected signal, i.e., the template s. If the level is too low, the presence of the signal will be hidden into the internal noise. By comparing the difference signal with the template the observer can determine the evidence about the presence of the tone in each interval.

3.3 The likelihood ratio and optimal decision criteria

For each interval there are two possible states (or hypothesis): N (noise only) and S (noise plus tone signal). It is of interest to introduce a quantity that describes the strength of the evidence for the "tone event" in each observation interval with a single value. Such quantity is the likelihood ratio that expresses the evidence of the tone rather than noise only in each interval:

$$l(\delta_i) = \frac{f(\delta_i|S)}{f(\delta_i|N)}\,. \tag{3}$$

The conditional probability density functions (pdfs) for each hypothesis $f(\delta|N)$ and $f(\delta|S)$ express the probability of a particular internal representation to occur given that the presented stimulus was actually N or S.

Due to the addition of internal noise to the stimuli, it is assumed that the internal representation of the i^{th} stimulus can be seen as a collection of independent and identically-distributed Gaussian random variables $\delta_i(n)$, one for each sample at time

$t = n/fs$ (with f_s sampling frequency and n sample index). Such Gaussian distributions are assumed to have mean 0 and variance σ^2 for the N stimulus, and mean $s(n)$ and same variance σ^2 for the S stimulus, where σ is the variance of the internal noise. Therefore the conditional densities for the n^{th} sample of the i$^{\text{th}}$ stimulus' internal representation are:

$$f(\delta_i(n)|N) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[-\frac{\delta_i(n)^2}{2\sigma^2}\Big] \tag{4a}$$

$$f(\delta_i(n)|S) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[-\frac{(\delta_i(n)-s(n))^2}{2\sigma^2}\Big]. \tag{4b}$$

Since all samples in a stimulus are assumed to be independent random variables, the conditional probability densities of the entire i^{th} stimulus are:

$$\begin{aligned} f(\delta_i|S) &= \prod_n f(\delta_i(n)|S) \\ f(\delta_i|N) &= \prod_n f(\delta_i(n)|N). \end{aligned} \tag{5}$$

For the decision process, the values of likelihood ratios need to be calculated for each interval. The likelihood ratio of the "event" interval with difference signal δ_i for the hypothesis S relative to hypothesis N is

$$\begin{aligned} l(\delta_i) &= \frac{f(\delta_i|S)}{f(\delta_i|N)} = \\ &= \frac{\prod_n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(\delta_i(n)-s(n))^2}{2\sigma^2}\Big)}{\prod_n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{\delta_i(n)^2}{2\sigma^2}\Big)} = \\ &= \frac{\prod_n \exp\Big(-\frac{(\delta_i(n)-s(n))^2}{2\sigma^2}\Big)}{\prod_n \exp\Big(-\frac{\delta_i(n)^2}{2\sigma^2}\Big)} \end{aligned} \tag{6}$$

This can be simplified by taking the natural logarithm obtaining:

$$\ln l(\delta_i) = \frac{1}{\sigma^2}\Big\{\sum_n \delta_i(n)s(n) - \frac{1}{2}\sum_n s(n)^2\Big\}. \tag{7}$$

The first term in the parenthesis of last equation is the correlation between the template and the i^{th} difference signal, $\rho_i = \sum_n \delta_i(n)s(n)$, while the second term is simply half the energy of the template signal. It is clear from Equation 7 that the correlation

ρ_i is a quantity that is monotonically related to the likelihood ratio. It is known from signal detection theory (D.M. and J.A, 1966) that any decision criterion based on the likelihood ratio (or any other quantity monotonically related to it) can be considered optimal in performing the observer's objective (for example in maximizing the probability of correct responses). "Optimal" means that any other criterion would not have a better performance on average. Hence, a decision rule based on the correlation value can also be considered optimal.

3.4 The objective: maximize the probability of correct responses

In the model implementation, the observer's objective is to maximize the number of correct responses. Depending on the particular stimuli conditions presented in a given trial of the m-AFC experiment, it is possible to calculate the probability of correct responses as follows.

Firstly, it is assumed that all intervals have equal prior probability to contain the tone. The situation of the tone in the first interval ($\langle S\,N\,N\,...\,N\rangle$) is presented here as it is equivalent to any other permutation of the S and $m-1$ N intervals. Second, it is assumed that the decision criterion is based on the selection of the interval with the largest correlation between the difference signal and the template, that is $\arg\max \rho_i$.

If the first interval has a correlation $\rho_1 = k$, and the decision rule is to select the interval with the largest correlation, the observer will be correct if and only if $\rho_i < k$ for $i = 2, ..., m$. Therefore a correct response corresponds to the joint occurrence of these events, which probability is

$$P_m(C) = P(\rho_1 = k, \rho_2 < k, \rho_3 < k, ..., \rho_m < k). \tag{8}$$

Since these are all independent events, their joint probability is the product of all individual probabilities

$$P_m(C) = P(\rho_1 = k) \prod_{i=2}^{m-1} P(\rho_i < k)\,. \tag{9}$$

In the original paper it is reported that if the template energy is normalized to one, than the likelihood ratios in Equations 4a and 4b have the same pdfs as the difference signals δ_i, and therefore also the correlation will have the same distributions. At the moment of writing this passage is not fully clear and will be investigated further. However, indicating the standard normal distribution as $\phi(x)$,

$$f(\rho_i|N) = \phi\left(\frac{\rho_i}{\sigma}\right) \tag{10a}$$

$$f(\rho_i|S) = \phi\left(\frac{\rho_i - d}{\sigma}\right). \tag{10b}$$

The probabilities of the single events in Equation 9 are

$$P(\rho_1 = k) = \phi\left(\frac{k-d}{\sigma}\right) \tag{11a}$$

$$P(\rho_i < k) = \int_{-\infty}^{k} \phi\left(\frac{t}{\sigma}\right)\, dt = \Phi(\sigma k)\,, \qquad \text{for} \quad i = 2...m \tag{11b}$$

where $\Phi(x)$ indicates the cumulative density function (cdf) of the standard normal distribution. The total probability of being correct is obtained by integrating on all possible values of k:

$$\begin{aligned} P_m(C) &= \int_{-\infty}^{\infty} P(\rho_s = k)\ P(\rho_n < k)^{(m-1)}\, dk \\ &= \int_{-\infty}^{\infty} \phi\left(k - \frac{d}{\sigma}\right)\ \Phi(\sigma k)^{m-1}\, dk\,. \end{aligned} \tag{12}$$

As pointed out before, a decision criterion based on the correlation ρ between the difference signals and the template can be considered optimal in the sense that it maximizes the probability in Equation 12.

3.5 Computing the probability of correct responses

The solution of the integral in Equation 12 is non-trivial but can be computed numerically for given values of m, d and σ^2 (Green and Dai, 1991). Another way of deriving the probability of a correct answer in an m-AFC task is the following.

In the derivation of Equation 12 the joint probability of m independent events was considered to satisfy the decision rule. However, if $\rho_1 = k = \max_{i=1...m}$ is the maximum correlation then it is larger than the maximum correlation of all $m-1$ noise intervals, that is

$$\rho_1 > \max_{2...m} \rho_i\,, \tag{13}$$

Hence

$$\max_{2...m} \rho_i - \rho_1 < 0\,. \tag{14}$$

So the observer will be correct if this last condition is satisfied. In this alternative approach, instead of satisfying m events jointly, a correct response will correspond to one single event, described by Equation 14, and the probability of a correct response with this alternative approach is computed as

$$P_m(C) = P\left(\max_{2...m} \rho_i - \rho_1 < 0\right)\,. \tag{15}$$

3.5.1 On the distribution of extreme members of a sample

To compute the probability in Equation 15 it is needed to derive the pdf of the event $\max_{2...m} \rho_i - \rho_1 < 0$.

Consider a sample $X = (x_1, x_2, x_3, ..., x_n)$ of values of n independent and identically distributed random variables, with pdf $\phi(x)$. Suppose to sort these values in ascending order and consider the extreme values $\min_i x_i$ and $\max_i x_i$. It is possible to calculate the sampling distribution of the extreme values of such sample. More precisely, it is possible to estimate the sampling distribution of any ranked element of the sample, that is the pdf of the j^{th} larger value of the sample of n. For the purpose of this report, it will be considered the case of the sample's maximum value only. The minimum and all other ranking cases can be treated similarly.

If $t = \max_{i=1,...n} x_i$, it follows that $x_i \leqslant t$ for all $i = 1, ..., n$. The probability of these independent events to occur is equal to the cdf evaluated at the maximum value t of the sample

$$\begin{aligned}
\Psi(t) &= P(\max_{i=1,...n} < t) \\
&= P(x_1 < t, x_2 < t, ..., x_n < t) = \\
&= \prod_{i=1}^{n} P(x_i < t) = \\
&= \prod_{i=1}^{n} \int_{-\infty}^{t} \phi(x)dx = \\
&= \Phi(t)^n,
\end{aligned} \tag{16}$$

where it has made use of the fact that all $x_i < t$ are independent events with the same cdf $\Phi(t)$. The pdf of $\max_i x_i$ is

$$\psi(t) = \frac{d}{dt}\Psi(t) = n\,\Phi(t)^{n-1}\,\phi(t)\,. \tag{17}$$

The expected value and the variance of the distribution can be calculated respectively as

$$E\left[\max_{i=1,\dots n} x_i\right] = \int_{-\infty}^{\infty} t\,n\,\Phi(t)^{n-1}\,\phi(t)\,dt = \mu^* \tag{18}$$

$$E\left[(\max_{i=1,\dots n} x_i)^2\right] = \int_{-\infty}^{\infty} (t-\mu^*)^2\,n\,\Phi(t)^{n-1}\,\phi(t)\,dt = \sigma^{*2}. \tag{19}$$

It has been shown by Cramer (1946) that when the parent distribution $\phi(x)$ is Gaussian, the distribution of the maximum is approximately normal. His calculations are rather extended and still unclear at the moment of writing this report. They will be not demonstrated for the scope of this project and the pleasure of a step-by-step demonstration of such asymptotic behavior is left to the most curious reader.

Alternatively, to show that the maximum of n samples is Gaussian distributed, computational simulations have been performed with MATLAB. These simulations consisted in the realization of N sets of n samples drawn from the standard normal distribution $\mathcal{N}(0,1)$ (any other Gaussian distribution with mean μ and variance σ^2 can be referred to this by a simple rescaling $\frac{x-\mu}{\sigma}$). Of each of the N sets the maximum was selected and stored in an array. At the end of the N iterations, the mean and variance of the N maxima was calculated. Different values of n ranging from 2 to 8 have been considered. The resulting distributions of maxima have been compared to a Gaussian distribution by means of an Anderson-Darling test, which tests the hypothesis that a sample of values is distributed accordingly to a given pdf. The test accepted the hypothesis of the maxima being Gaussian distributed if the simulations were run for values of N iterations below a certain limit. To show this behavior, simulation results for $N = 1000$ and $N = 10000$ are reported in Tables 1 and 2 and in Figures 2 and 3.

n	h	p	μ	σ
2	0	0.20	0.56	0.85
3	0	0.57	0.89	0.77
4	0	0.35	1.03	0.71
5	0	0.24	1.17	0.67
6	0	0.32	1.25	0.64
7	0	0.10	1.32	0.64
8	0	0.06	1.44	0.61

Table 1: Results for the simulations N = 1000. n is the number of elements in each samples, h is 1 if the hypothesis is rejected, 0 if is accepted, p is the probability of observation, μ and σ are the mean and standard deviation of the distribution.

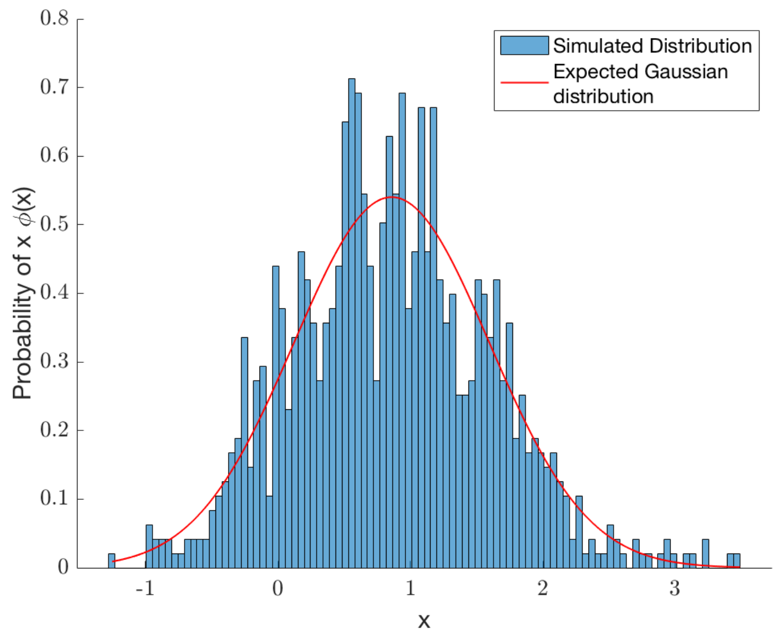

Figure 2: Distribution of the N = 1000 maxima of sample of Gaussian distributed random variables with $\mu = 0$ and unit variance.

n	h	p	μ	σ
2	1	0.20	0.57	0.83
3	1	0.02	0.86	0.75
4	1	0.00	1.02	0.69
5	1	0.00	1.15	0.67
6	1	0.00	1.26	0.64
7	1	0.00	1.34	0.62
8	1	0.00	1.41	0.61

Table 2: Results for the simulations N = 10000. Symbols are as in 1.

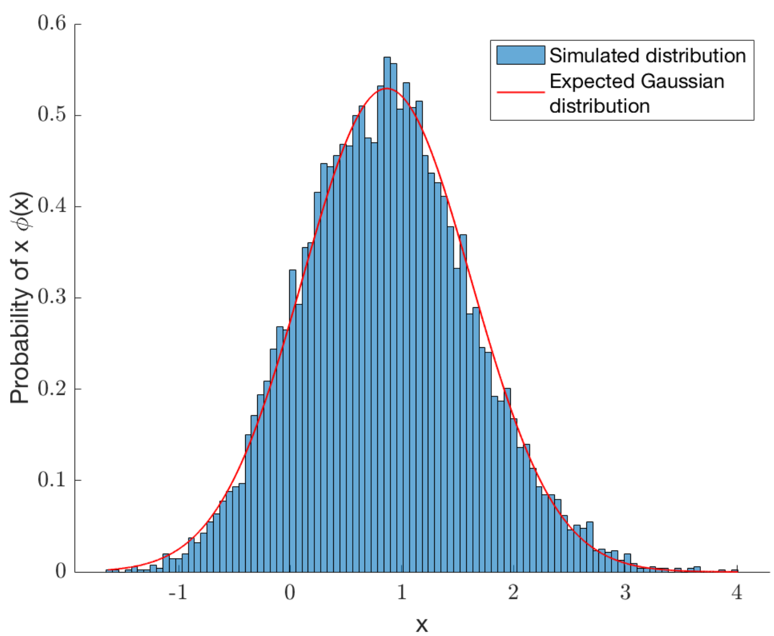

Figure 3: Distribution of the N = 10000 maxima of sample of Gaussian distributed random variables with $\mu = 0$ and unit variance.

3.5.2 The model decision making

In this section, it will be described how the decision is taken about the presence of the tone, based on the theory described so far.

In every experimental trial, the information about which interval actually contains the tone is available to the model. The difference between the correlation of such interval and the maximum correlation of all other intervals is computed together with the probability of being correct as in Equation 15. The decision making process about such event can be seen as a *yes/no* task about the fulfillment of the condition in Equation 14. The probability of correct responses $P_m(C)$ is mapped as a function of the difference $\max_{i=2\ldots m} \rho_i - \rho_1$ onto a psychoacoustic curve. The calculated $P_m(C)$ is compared with the probability of correct responses $\widehat{P_m(C)}$ calculated according to the tracking rule used in the AFC procedure (as reported in Levitt (1971)). If $P_m(C) > \widehat{P_m(C)}$ a "correct" response will be provided by the model, a "wrong" response otherwise.

4 Conclusions

In conclusion, this project helped gaining a deep insight in the implementation of the decision device in the computational auditory model. The mathematical framework explored in the report was useful in understanding the optimal decision rule in determining observer's response with respect to the given objective. The only weak point in the project was not being able to demonstrate a link between the likelihood ratio and correlation between the template and difference signal.

List of acronyms

AFC alternative forced-choice

cdf cumulative density function

pdf probability density function

References

Cramer, H. (**1946**). *Mathematical Methods of Statistics* (Princeton University Press).

Dau, T., Pueschel, D. C. M., and Kohlrausch, A. (**1996**). "A quantitative model of the "effective" signal processing in the auditory system. i. model structure.", The Journal of the Acoustical Society of America **99 6**, 3615–22.

D.M., G. and J.A, S. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Green, D. M. and Dai, H. (**1991**). "Probability of being correct with 1 of m orthogonal signals", Attention, Perception, & Psychophysics **49**, 100–101.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics", The Journal of the Acoustical society of America **49**, 467–477.