Applying a Bayesian Approach to Speech Sound Identification in Context and Exploring Possible Neural Network Mechanisms

Jay McClelland
SymSys 100, Spring 2009

This document builds on the document "*Using conditional relationships between events to make inferences*". First we re-express Bayes rule in terms of two derived quantities called the likelihood ratio and the prior odds. Then we go on to use the concepts of likelihood ratio and prior odds to offer a Bayesian account of the Ganong effect, in which word context affects speech perception. Your second homework assignment is to answer the questions posed as we proceed through these ideas, and also to answer a final question that attempts to link the Bayesian account for the Ganong effect with a mechanistic account based on the interactive activation model of Rumelhart & McClelland.

Re-expressing Bayes rule in terms of the likelihood ratio and the prior odds

Because it will be useful later, we define a function called the 'likelihood function', which is equal to the probability of the observed evidence given some hypothesis, $P(E|H)$. The use of the word 'likelihood' seems strange at first, but you can thing of it as expressing 'how likely would this evidence be if Hypothesis H were true?' The likelihood leads us to a useful quantity called the 'likelihood ratio', represented here by Λ (called 'lambda')

$$\Lambda = P(E|H)/P(E/{\sim}H)$$

The likelihood ratio can be seen as representing what the evidence tells us about the relative likelihood of the hypothesis and its negation. If the evidence is far more likely when the hypothesis is true than it is when it is false (Λ greater than 1), then the evidence tends to support the possibility that the hypothesis is true.

Another useful quantity is the 'prior odds' or 'odds ratio' represented by Ω ('omega'):

$$\Omega = P(H)/P({\sim}H)$$

When the hypothesis is very probable, the prior odds are large (much greater than 1), and when the hypothesis is improbable, the prior odds are very small (less than 1).

If we take Bayes Rule and divide the numerator and the denominator by P(E|~H)P(~H) we obtain:

$$P(H|E) = \frac{[P(E|H)/P(E|{\sim}H)][P(H)/P({\sim}H)]}{[P(E|H)/P(E|{\sim}H)][P(H)/P({\sim}H)]+1}$$

Or

$$P(H|E) = \Lambda\Omega/(\Lambda\Omega+1) \quad (1)$$

Q1: Compute Λ and Ω for the example in Q4 from the previous homework; show the results, then re-compute the answer to Q4 using (1).

Context Effects in Speech Perception

We now consider experiments investigating the Ganong effect. The Ganong effect is the finding that context influences the identification of speech sounds. For example, following the context 'fi…', an ambiguous sound somewhere between 's' and a 'sh' is more likely to be identified as 'sh'; following 'ki…', the same sound is more likely to be identified as 's'. The experiment you participated in addressed the Ganong effect.

One way to explain this effect is to see the context as affecting the odds that the sound will be 's' or 'sh'. For example, 'fi…', increases the odds that the final sound will be 'sh'. On the other hand 'ki…' reduces the odds that the final sound will be 'sh'. The longer contexts 'establi…' and 'malpracti…' should have a similar effect, possibly stronger than the effect of the shorter contexts. Here is a table with some possible values for the odds ratios in question. This numbers are chosen for illustrative purposes only.

Odds Table	Hypothetical Prior Odds for "sh" : "s" in Four Contexts			
	Malpracti…	Ki…	Fi…	Establi..
Ω	.333	.667	1.5	3

Now, let's consider the sounds 's' and 'sh'. Both are called 'fricatives' (they are made by the passage of air through a narrow constriction in the mouth) and they differ primarily in frequency ('sh' has lower frequencies than 's'). We can actually create a continuum of sounds by varying the frequency, and this creates a series of continuous gradations between 's' and 'sh'. From a Bayesian point of view, we can treat these sounds as varying in the probability that each one would occur when a speaker intended to say's' or 'sh'. For example, the most 's'-like sound would be most likely to be produced if we were trying to say 's', and least likely to be produced if we were trying to say 'sh'. I have made up some specific numbers for illustrative purposes in the following table. From that information we can then compute the corresponding likelihood ratios. In the table, the numbers are symmetrically arranged around a completely neutral midpoint – experimenters usually try to make their stimuli balanced in this way, but they do not necessarily succeed perfectly.

Likelihood Table	Hypothetical Likelihoods and Likelihood Ratios for Different Final Segments								
	1	2	3	4	5	6	7	8	9
P(E|'s')	.095	.09	.075	.0667	.05	.0333	.025	.01	.005
P(E|'sh')	.005	.01	.025	.0333	.05	.0667	.075	.09	.095
Λ('sh':'s')	0.0526	.111	.333	.5	1	2	3	9	19

We can now ask, what will we get if we calculate the posterior probability P('sh'/stimulus), separately for each stimulus in each context, using the hypothetical prior odds in each context from the Odds Table and the Likelihood Ratio for each stimulus from the Likelihood Table. The results are shown in Figure 1. For each context, I have connected the points representing the posterior probability for the nine levels of the stimulus. The curves that are generated by this procedure have a 'sigmoid' or S-shape – the effect of the context is to shift the S along the x axis. Because the S-shape levels off at 0 on the left and 1 on the right, the curves tend to converge at the two ends and are separated in the middle. We can think of these curves as capturing the *pattern* we would expect to see if participants' responses corresponded to the posterior probability that the stimulus ends in 'sh', as calculated by using Bayes Rule. You can verify that, for example the left-most point on the curve labeled 'establi…' is what you get when you plug the Odds given for 'establish' in the odds table and the Likelihood ratio for stimulus 1 from the Likelihood table into Equation (1).

Q2: Compare the curves in Figure 1 to the qualitative features of the group data from the experiment in which you participated, included in Figure 2. Only the 'establi…' and 'malpracti' contexts are included. (a) What qualitative features of the curves in Figure 1 are reproduced in the actual data in Figure 2? Focus on the shapes and relationships between the curves, not their slopes or exact degree of separation.

Now let's get a bit more quantitative. Looking at the data, we can see that participants gave slightly more 'sh' than 's' responses overall. Perhaps the stimuli are not perfectly symmetrically arranged around the mid-point of the subjective 's'-'sh' continuum. Considering stimulus 5, from a Bayesian perspective we could say that participants treat this stimulus as being more likely given 'sh' than 's'. Let's adopt the value of 1.6 for the likelihood ratio, Λ('sh':'s'). (b) Given this value of Λ, what is the expected probability of the 'sh' response, if $\Omega = 1$ (prior probabilities of 'sh':'s' are equal)? (c) What would be the probability of the 'sh' response if $\Omega = 2$? Use Equation 1 to compute your answers. (d) The observed probabilities of 'sh' responses to stimulus 5 for the 'establi…' and 'malpracti…' contexts are approximately .72 and .50. Find a good estimate of Ω for 'establi…' and a good estimate for Ω for 'malpracti…' to explain these findings, keeping the value of Λ fixed at 1.6. Your value should fit the data to within about 1%. Describe in words what these numbers correspond to in terms of the prior odds. (e) Now consider the experimental data in Figure 2 from stimulus 7, for which the probability of responding 'sh' in the 'establi…' context is .95 and in the 'malpracti…' context the probability of responding 'sh' is .87. Keeping the values of Ω for 'establi…' and 'malpracti…' that you already calculated, what value of Λ is needed to approximate the observed results, again to within about 1%?

Q3: Describe in your own words how the interactive activation model of Rumelhart and McClelland might explain the results of the experiment, translating the model to apply to speech perception. Use the 'fi…' context as your example, and don't worry about the 'feature' level – consider only a word level and a 'speech sound' level, where units correspond to distinct speech sounds like 'f', 'k', 'i', 'sh', 's', etc. Discuss specifically what would happen when a final fricative sound that activates 's' and 'sh' equally is presented in the ‘fi…’ context. Explain how between-level excitatory influences and within-level inhibitory influences would increase the tendency to choose ‘sh’ instead of ‘s’ in this context. Finally, consider stimulus inputs ranging from very ‘s’-like to very ‘sh’-like, and discuss whether you think the model would approximate the way in which the effect of context varies across the stimulus continuum, as seen in the data and in the Bayesian account of context effects considered in the previous question (300 words maximum).

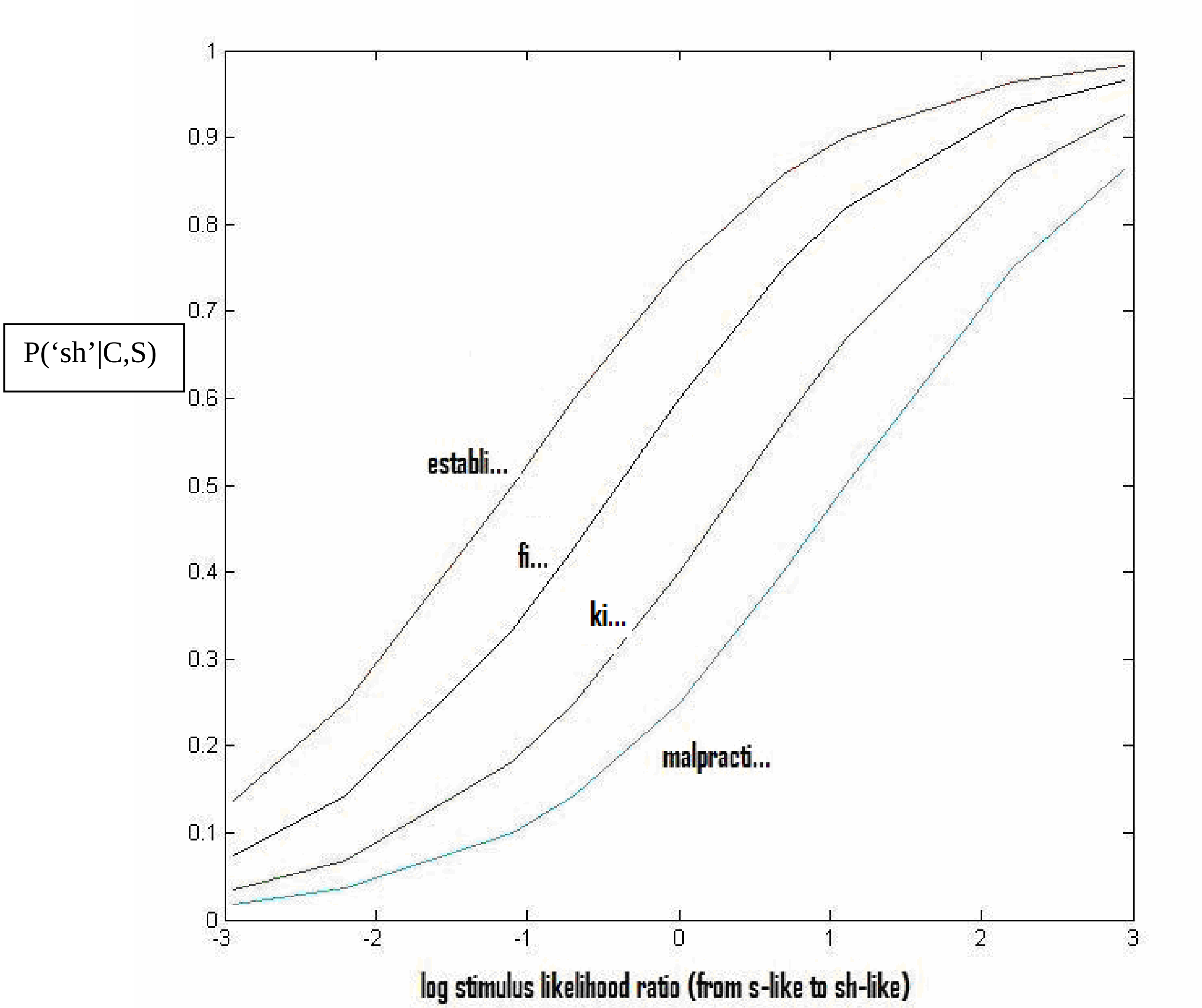

Figure 1. Calculated posterior probability that the stimulus is a ‘sh’ for each context (C) and stimulus (S), using the values of Λ and Ω for each stimulus and context from the tables in the text. We have used the logarithm of the likelihood ratio on the x-axis of this figure because it spaces the likelihood ratios evenly and symmetrically around the midpoint of the scale.

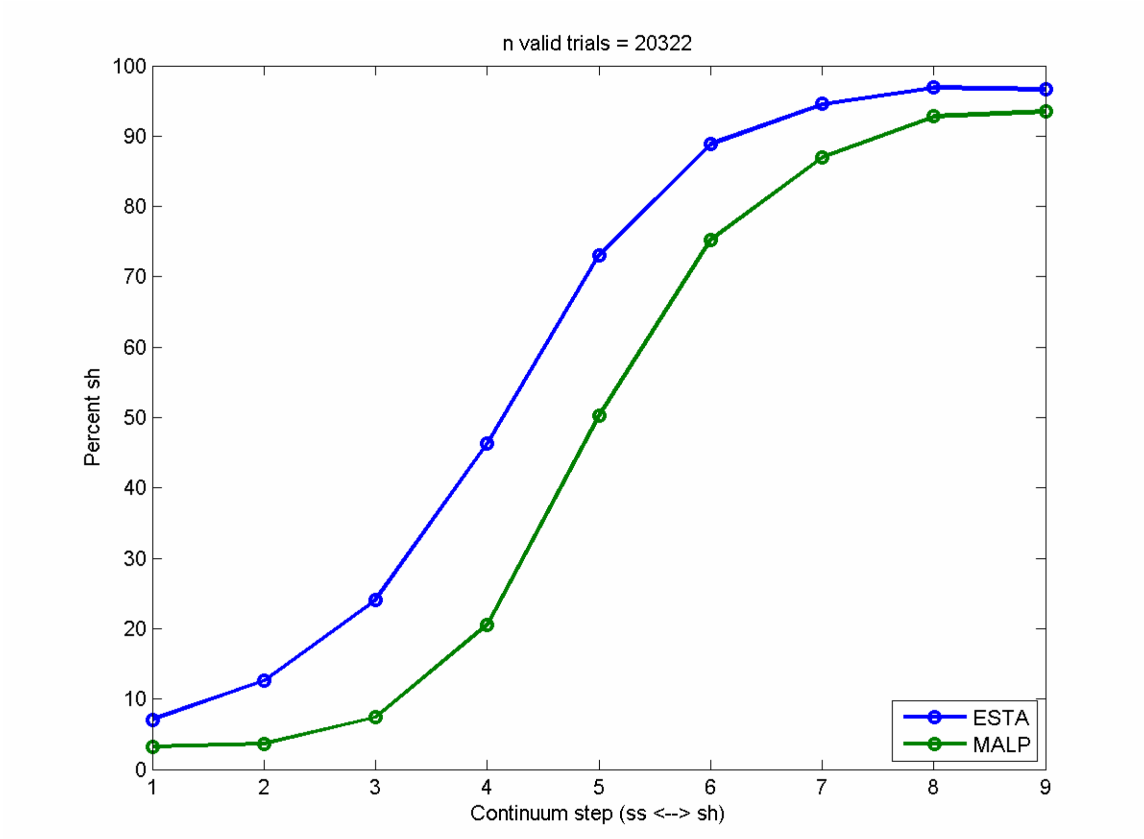

Figure 2. Data from the behavioral experiment for approximately 75 participants. For each of nine steps on a continuum ranging from ‘ss’ to ‘sh’, percent ‘sh’ responses is graphed, separately for the ‘establi’ context and the ‘malpracti’ context.