Cognition, Language & Communication 2014

Assignment IV

due: 16-10-2013 (before class)

In this final assignment we will look at some of the data that we might use to answer the big questions we addressed in the class: do humans have a unique skill in detecting particular patterns in a stream of sounds, and is that skill one of the crucial abilities underlying language? One proposal for such a unique skill is the ability to detect abstract rules underlying the data, such as the AXC pattern that we discussed.

Existing data from artificial language learning experiments (cf. Pena, Toro) suggests human can pick up such patterns under the right circumstances, whereas animals (rats in this case) cannot. Before jumping to conclusions about qualitative differences between the learning abilities of these species, we need to investigate whether a quantitative difference in motivation / attention / etc. can already explain the experimental results, as the model of Alhama et al. suggests. In the experiment that you all participated in we addressed one prediction of that model: very skewed distributions of responses for both words and partwords, such that many individual partwords are preferred over many individual words even if, on the average, words are preferred over partwords.

A completely different way to address questions about the human uniqueness of complex sequential patterns is to look at learned vocalizations produced by non-human species. A strong contender for complex patterns in vocalizations are the sea mammals. In particular, Suzuki et al. (2006) claim that humpback whale vocalizations show hierarchical structure. In the second part of this assignment you will briefly look at some of the humpback whale data that they used for that study.

1 Artificial Language Learning experiment

In the experiment you were exposed to an AXC pattern. The data from the experiments are in the files `cond3_w.csv`, `cond3_pw.csv`, `cond4_w.csv` and `cond4_pw.csv` on Blackboard. (Conditions 3 and 4 differed in the actual syllables that instantiated a AXC pattern, but were otherwise identical; `w` contains the responses to words at test time; `pw` contains the responses to partwords at test time. All test items had been present in the familiarization stream. Responses are recorded as judgment (+1 or -1) times confidence (scale from 1 tot 7)). Download these files, and load them into R by using a command like:

```
cond3_w <- read.csv("cond3_w.csv",row.names=1)
```

Now, it's always a good idea to look at the data per individual and per stimulus first. Look at the the 4 datasets by typing their names, and plot all values as a scatter plot with a command like:

```
plot(1:length(unlist(cond4_w)),unlist(cond4_w))
```

It's difficult to see a pattern in there. So let's plot the averages per test stimulus (you can glue together words and partwords with the `cbind()` command and compute the means of every column with `colMeans()`):

```
colmeans3 = colMeans(cbind(cond3_w,cond3_pw))
barplot(colmeans3,names.arg=names(colmeans3),las=2)
```

Question 1 *Give the plot for condition 4, and report for both conditions what the 3 stimuli with the highest response are. Are these words or partwords? Do the stimuli with high responses have anything in common (e.g., sound similar)?*

The prediction we would like to evaluate is that the response distribution is skewed, and different in every individual, such that some partwords receive a higher response than many of the words, even though they are 3 times less frequent in the familiarization stream. To investigate the skew, it is useful to order the response per individual:

```
sorted3w = matrix(0,dim(cond3_w)[1],dim(cond3_w)[2])
for (i in 1:dim(sorted3w)[1]) sorted3w[i,]<-sort(as.matrix(cond3_w)[i,],decreasing=TRUE)
```

We can now plot the mean response per 1st, 2nd, 3d etc ranked stimulus (ranked per person) as follows:

```
barplot(colMeans(sorted3w),names.arg=1:18,cex.names=.8)
```

Question 2 *Show the mean responses per rank for the 2 conditions and words and partwords separately. Informally, does the pattern of results seem to support the prediction? (We won't deal with the appropriate statistical tests here).*

2 Humpback whale song

On the course website you can find a small sample from the transcribed recordings of Humpback whale songs (from 1977!) in the file `humpbacksong-March1977.txt`. In the transcription, a different letter is used for each new element in the songs. Read in the data, and have a look at the sequence of elements:

```
library(tau)
song = readLines("humpbacksong-March1977.txt")
song
```

The proper way to analyse the structure is with the model selection techniques we have seen in previous weeks. In the interest of time, we will now only visualize the data. With the following commands, we can draw for the first element we encounter a dot at the bottomleft, and for each *new* element we encounter after that a dot one step higher. Dots for elements we have seen before are drawn at the same height as the first time we see them, and all dots are then connected.

```
mono = textcnt(song, n = 1,  method="string")
number = mono
number[1:26]=1:26
lst = strsplit(song," ")
plot(1:978,number[lst[[1]]],type="l")
```

We can also zoom in on the first elements:

```
plot(1:100,number[lst[[1]]][1:100],type="b",yaxt="n")
text(-1,1:25,names(mono))
```

Question 3 *Give the plot, and describe its structure. What does this suggest about the structure of humpback song? Can we in few or many cases predict the next element in a song from the current?*

Finally, we can make a similar plot of the utterances of a child of 1 year and 8 months old, now going up one step in the plot for every new word we encounter. This shows a dramatic difference.

```
speech = readLines("eve-file5-first500.txt")
lst = strsplit(speech, " ")
mon = textcnt(speech,n=1,method="string")
number = mon
number[1:213]=1:213
plot(1:1298,number[lst[[1]]],type="l")
```

Question 4 *Show the plot and explain briefly how you interpret the difference you see here between humpback whale song and child speech.*