

## Results of the voicing perception experiment

The results are in the Excel spreadsheet. The full results are in the upper part of the sheet. The main point of the experiment is to assess the relevant strength of the voicing cues in the ‘closure’ and ‘release’ portions of the VCV, where the closure consists of the first vowel plus the consonant constriction and the release consists of the consonant release plus the second vowel. The relative strength of these cues is evaluated by the perception of the conflicting cue stimuli – does the perceived voicing depend on the closure portion or the release portion? Accordingly the results for the conflicting cue stimuli are presented in terms of how often they were identified according to the release portion in the lower part of the spreadsheet. Where a subject categorized a stimulus according to its release, this is marked with a ‘1’. The last two columns show the total number of times that a stimulus was categorized in accordance with its release as an absolute frequency and as a percentage.

It is clear that stops and fricatives patterned differently: Overall fricative stimuli were identified in accordance with the release 19% of the time, while categorization of stop stimuli followed the release 81% of the time. However it also looks like individual stimuli patterned differently, e.g. many of the conflicting fricative stimuli were always identified according to the closure, but perception of stimulus 4 followed the release for 10/12 subjects. There are also ‘exceptional’ stop stimuli, e.g. 40. It would be interesting to examine these stimuli acoustically to see if you can understand why they differ from the rest. Sound files can be accessed at: <http://web.mit.edu/~flemming/www/experiment/> (Different stimuli were constructed in different ways since you each did this individually – in a real experiment consistency would be desirable, but the systematic results of the variation may be very instructive in thinking about how to run a full version of this experiment.)

### Statistical analyses:

The main hypotheses we want to test concern whether a given set of conflicting cue stimuli were identified more often according to closure or release. A simple approach to testing these hypotheses is to use the chi-square ( $\chi^2$ ) test to test whether the responses were significantly unevenly divided between closure and release. Chi-square is used to assess the chance of getting a particular distribution of frequencies, given the null hypothesis that they should follow a particular expected pattern of frequencies.

For example, our null hypothesis is that fricative conflicting cue stimuli should be categorized according to the release as often as they are categorized according to the closure. We have 12 conflicting fricative stimuli and 12 subjects for a total of 144 categorizations, so according to the null hypothesis we would expect 72 to match the release and 72 to match the closure, but the observed frequencies are 27 for release and 117 for closure. The Chi-square statistic is calculated from the following formula:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad \text{where } O_i = \text{observed frequency, } E_i = \text{expected frequency}$$

$$\text{So in our example, } \chi^2 = (27-72)^2/72 + (117-72)^2/72 = 56.25$$

The number of degrees of freedom is 1. The probability of obtaining a  $\chi^2$  this large or larger by chance is much less than 0.01. We report this result as follows:  $\chi^2(1) = 56.25, p < 0.01$ .

The number of release responses is significantly greater than 50% for the stops:  $\chi^2(1) = 50.94, p < 0.01$ .

Looking at individual stimuli, we have 12 judgments on each stimulus. 2/12 (or less) and 10/12 (or more) are significantly different from even ( $p < 0.05$ ), according to the chi-square test.