Speech perception, phonetics and phonology:

Native language experience and the perception of voice contrast[1]

DUE DATES

7 October: Email your personal dataset, as per instructions below, by 5pm.

13 October: Hand in, or e-mail me, the write-up by the beginning of class on this date.

NOTE ON COLLABORATION

You are permitted to work in groups to analyze your data. In fact, I encourage it. Students will invariably have different levels of experience with manipulating numbers, using Excel, understanding the phonetics or psychology background, etc. Working together is a good opportunity to share strengths and shore up weaknesses. It will make following the instructions below easier and help you think through the ideas necessary to write the lab report.

However, you *must* do your own analysis. You are not allowed to simply copy a group member's completed spreadsheet/chart/etc, and insert your own data. Likewise, the report must be written individually.

Introduction

In this lab exercise you will test your own speech perception. There are four short experiments, all of which you will run on yourself.

The experiments examine your perception of sounds from English and Russian. The English experiments focus on the contrast between voiced /d/ and voiceless, aspirated /t/. The Russian experiments focus on the contrast between voiced /d/ and voiceless, unaspirated /t/. For greater background on this topic, see the excerpt of Ashby & Maidment (2005), *Introduction to Phonetic Science,* or the Introduction and Materials & Methods section of Kazanina et al. (2006), both of which are available on the readings page. Pay attention to the how the voicing contrast is realized differently in Russian and English.

For each language, there is a **labeling** or **identification** task. In this task, you will hear single sound and you must classify it as either a /d/ or /t/. For each language, there is also **discrimination** task. In this task, you hear two sounds in series, and must decide whether they are the same or different. These sounds are separated by 1200 ms, a period referred to as the *inter-stimulus interval (ISI).* Since the sounds are 300ms long, the time between sound onset to sound onset – the *stimulus onset asynchrony (SOA)* – is 1500 ms.

[1] This lab is based on the lab originally created by Colin Phillips (UMD). Phillips generated the English sounds with a speech synthesizer. The Russian sounds are computer-edited natural speech samples, created by Nina Kazanina (U. Bristol, UK). The experiments in the lab are run using PsyScope scripts originally created by Phillips and improved by Henny Yeung (U. British Columbia, CA) and Brian Dillon (UMD).

Running the experiments

Stimulus presentation and data collection is handled by *PsyScope*. PsyScope is free software for running psychology experiments[2]. To complete the assignment, you will need to gain access to a Mac. There are two options. You can install PsyScope on your own computer (or a friends', or in a computer lab … if any of them are Macs), then download and run the experiment scripts. Or you can sign up for a time to use one of the computers in the Linguistics labs.

Use a pair of headphones. This will ensure higher quality, but more importantly, will keep you from driving your roommate/neighbor crazy. We have some you can use in the Linguistics lab, if you run the experiment there. Also, if you're not using your own machine, have a USB key handy (or be able to log in to webmail), so you can send yourself the data.

Self-install

1. The PsyScope home page is located here: http://psy.ck.sissa.it. Take a second to look around. The latest version can be downloaded from this link: http://psy.ck.sissa.it/Files to Download/currentversion/PsyScopeXCurrentVersion.zip

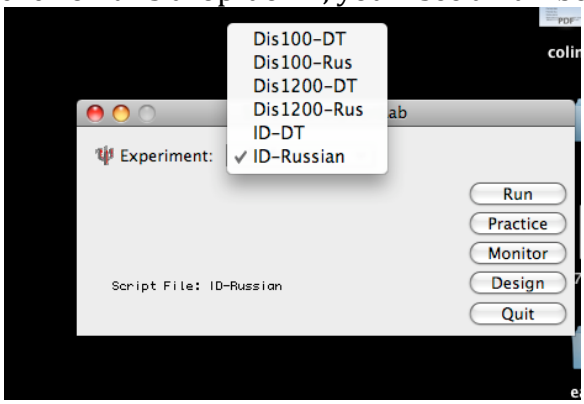
2. Unzip the downloaded file. There will be two versions of the software 'PsyScope X B53' and 'PsyScope X B53D' (next to icons that look like the following ). You will want to use 'B53', not 'B53D'. Drag this version to your Applications folder or to the Desktop.
3. Download the Speech Perception lab scripts, located on the class website. The file is downloaded as a .zip file. When you unzip it, it will create its own folder.

Running the experiment.

1. Open PsyScope B53.
2. Open the script entitled 'SpeechPerceptionLab', contained within the directory 'SpeechLab' (File >> Open).
3. You will probably initially see two windows, like the ones in the screenshot below. The large window to the left is a graphical representation of how the experiment is designed, while the smaller window to the right is the control panel for the experiment. You can close or minimize the larger window.

[2] PsyScope was originally developed in the early 1990s at Carnegie Mellon University, by Jonathan Cohen, Matthew Flatt, Brian MacWhinney and Jefferson Provost. Its newest incarnation, for Mac OS X, is developed by SISSA Language, Cognition and Development Lab in Trieste, Italy; the head of the project is Luca Bonatti.

4. In the control panel there is a drop-down box next to 'Experiment.' If you click on this drop-down, you'll see a number of separate experiments.



For this assignment, you must run the following experiments:

a. ID-DT (English/Identification)
b. ID-Russian (Russian/Identification)
c. Dis-1200-DT (English/Discrimination)
d. Dis-1200-Rus (Russian/Discrimination)

You can run the Dis100-DT and Dis100-Rus experiments and include them in your write-up if you want to deepen your write-up. In these two experiments, the ISI is 100 ms [so, what's the SOA?]. How might different ISIs affect performance in the task?

Counterbalancing. When you run an experiment with a series of different tasks or items, you want to avoid a potential confound caused by the specific order of tasks. Depending upon particular experimental demands, the participants' mental state, attitude toward the task, etc., performance might improve over the session, as the participant gains experience through practice; or performance might worsen, as the participant gets fatigued. In advance you don't know what is going to happen. But If the sub-parts of the

experiment always occurred in the same order, performance patterns in any sub-part may reflect not just the experimental effect you're studying, but a bias from practice or fatigue. To avoid this problem, you change the order of tasks/items across the experiment for each new participant. In that way the same task/item does not always occur in the same position and there is thus no systematic bias in the dataset as a whole. This procedure is called *counterbalancing.* For the write-up you will only analyze your own data. However, we will still counterbalance, choosing 4 different orders (of course with 4 possible 'slots', there are many more orders, in total 4! or 24 – so we are only partially counterbalancing).

Flip a coin twice. On the first flip, if it turns up heads, run both discrimination experiments first. On the second flip, if it turns up heads, run the English version of each experiment first. Note the order, so you can report in it your write-up.

H/H: English discrim, Russian discrim, English ID, Russian ID
H/T: Russian discrim, English discrim, Russian ID, English ID
T/H: English ID, Russian ID, English discrim, Russian discrim
T/T: Russian ID, English ID, Russian discrim, English discrim

5. Practice. Once you've decided the order, select the first experiment. Clicking 'Run' will start the experiment, and the software will record your responses. Before doing that, you should first run an experiment in 'Practice' mode. While there is some practice built into the experiment, trying out an experiment in Practice mode will allow you not only to experiment using the software and listening to the speech sounds, but importantly to verify that your set-up is working.

Click the 'Practice' button. This runs through the experiment, but does not record data. When you press the "f" or the "j" key to indicate your response after each trial, you should immediately hear a "beep", and then the next trial will begin. If you do not give any response, the beep will automatically sound after 6 seconds, and then the next trial will begin [try this out!]. **Important:** if you there is no beep when you press the response key, and the experiment does not immediately proceed to the next trial, then your response is not being recorded properly. You should stop the experiment, *check the caps lock key,* and then restart it. If the caps lock is accidentally set to "on", your response will not be detected properly.

After a few trials in practice mode, you should break out of the experiment by pressing the Command key (⌘) and the period (".") simultaneously. A dialog box will then prompt you to end the experiment. Do so.

The other important thing to verify in your practice session is whether the audio is loud enough. If it isn't, adjust it using the Sound control pull-down menu at the upper-right of the screen; or the volume keys on the keyboard.

6. Running the experiment. Start with the first experiment that you picked by the coin-flipping procedure. Click 'Run'.

   Before the experiment begins you will be asked a couple of questions about yourself , which will be recorded in your data file. Enter your name in the format *Firstname Lastname* -- your name will identify your data file. If you encounter a message that "a person by this name has already been run" - ignore this and proceed. If you are prompted to enter a "Run Number" use "1". Once you have answered these questions a screen of instructions will appear. Read the instructions.

7. Switching to each subsequent experiment. Once you have completed the first experiment, you can use the Control console window to select the next condition that you are supposed to run. When the new condition opens you may be shown an alert dialog which asks you whether you want to save the changes to the previous condition - answer "no".

   The Identification experiment will present 100 sounds, without any breaks (in experimental terminology, in one block). The discrimination experiment contains about 200 pairs of sounds, interrupted occasionally so you can take a break (i.e., in multiple blocks).

   It should take you 30-40 minutes to complete all four experiments. Do not proceed immediately from one experiment to the next. Take a brief break (stretch, get a drink of water if you need it). You will do your best across the tasks if you are not tired.

Wrap-up.

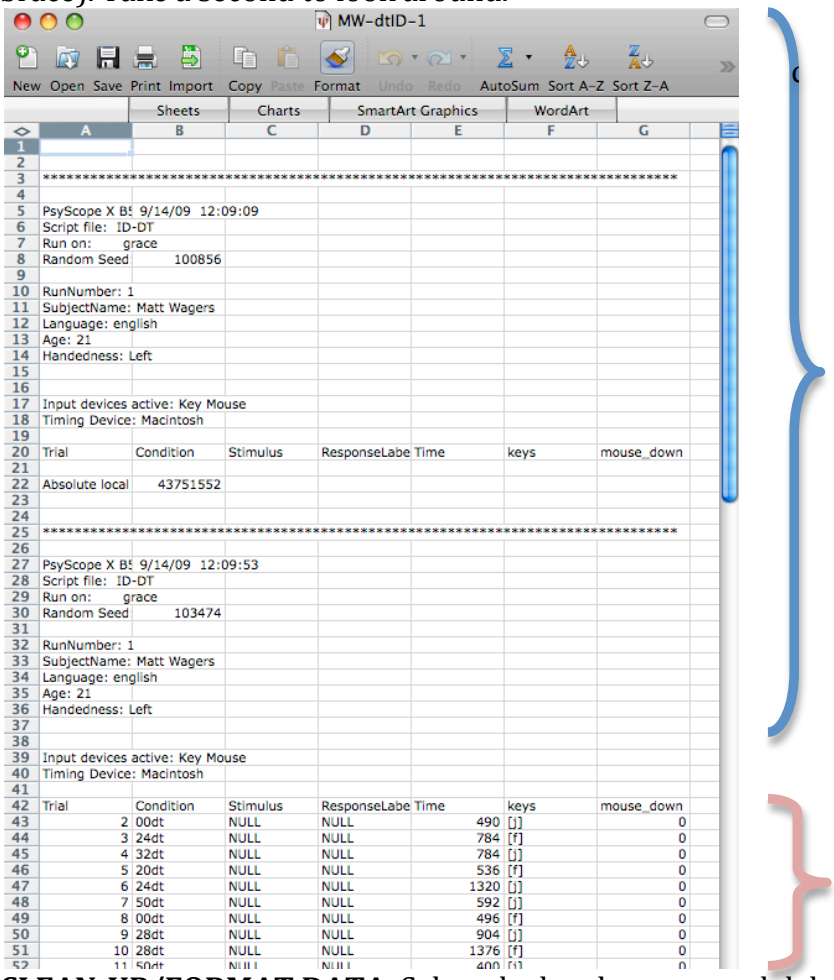
1. Once the experiments are completed, you should find 4 data files with your initials inside the *Data Files* folder, one for each of the 4 experiments. It should be easy to match up data files with experimental conditions, using either the name of the data file, or the header information inside the text file.
2. If you are using a lab computer, email the files to yourself, or copy it to a personal USB key. [If you want to use a Windows PC for analysis, it is recommended to add the '.txt' filename extension].
3. **IMPORTANT.** Select your four data files and compress them (within the Finder window, File >> Compress). Send the compressed data file to the class scribe (email address will be updated) by the deadline (7 October, 5pm).

Basic Analysis Instructions

The following text contains tips on how to analyze the data from your speech lab using Microsoft Excel. I work through a basic analysis, where some of the steps are done by hand, though many can be automated. If you have not manipulated some simple data before in the past, I recommend taking your time to work through this procedure as I outline it.

If you have expertise here, feel free to use whatever appropriate techniques you know, (including using other analysis/graphing programs). For example, you can analyze your data in just a few minutes using Excel's PivotTable feature. Explaining how to do so goes beyond the scope of these instructions, but you may already know, or you may want to seek out a tutorial/help file.

1. **GET STARTED.** Open the data file in Excel. (Select the 'Delimited' text option). As you'll see, the file contains header information (blue curly brace below), followed by a table of the trials you completed in the experiment (pink curly brace). Take a second to look around.



	A	B	C	D	E	F	G
1							
2							
3	**********						
4							
5	PsyScope X B5	9/14/09 12:09:09					
6	Script file:	ID-DT					
7	Run on:	grace					
8	Random Seed	100856					
9							
10	RunNumber: 1						
11	SubjectName: Matt Wagers						
12	Language: english						
13	Age: 21						
14	Handedness: Left						
15							
16							
17	Input devices active: Key Mouse						
18	Timing Device: Macintosh						
19							
20	Trial	Condition	Stimulus	ResponseLabe	Time	keys	mouse_down
21							
22	Absolute local	43751552					
23							
24							
25	**********						
26							
27	PsyScope X B5	9/14/09 12:09:53					
28	Script file:	ID-DT					
29	Run on:	grace					
30	Random Seed	103474					
31							
32	RunNumber: 1						
33	SubjectName: Matt Wagers						
34	Language: english						
35	Age: 21						
36	Handedness: Left						
37							
38							
39	Input devices active: Key Mouse						
40	Timing Device: Macintosh						
41							
42	Trial	Condition	Stimulus	ResponseLabe	Time	keys	mouse_down
43	2	00dt	NULL	NULL	490	[j]	0
44	3	24dt	NULL	NULL	784	[f]	0
45	4	32dt	NULL	NULL	784	[j]	0
46	5	20dt	NULL	NULL	536	[f]	0
47	6	24dt	NULL	NULL	1320	[j]	0
48	7	50dt	NULL	NULL	592	[j]	0
49	8	00dt	NULL	NULL	496	[f]	0
50	9	28dt	NULL	NULL	904	[j]	0
51	10	28dt	NULL	NULL	1376	[f]	0
52	11	50dt	NULL	NULL	400	[j]	0

2. **CLEAN-UP/FORMAT DATA.** Select the header rows and delete them (Edit >> Delete). Leave the row that has the column names for the data table ('Trial',

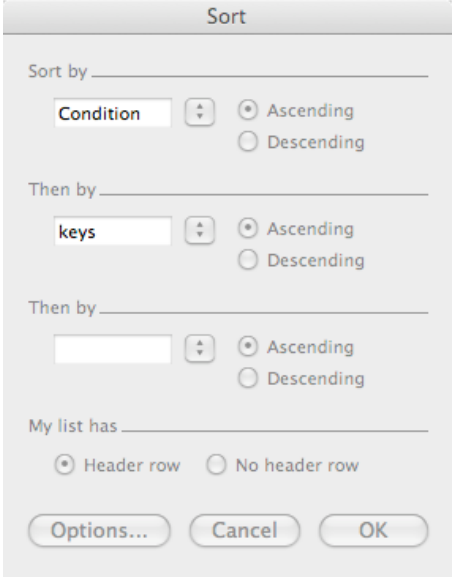
'Condition', etc.). Several of the columns are extraneous ('Stimulus', 'ResponseLabel','mouse_down'), and you can delete those.

3. **SORT DATA.** To complete the analysis, you will want to sort the data by the Condition (i.e., the independent variable). In the English Identification experiment, the Condition label is just the VOT value + 'dt'. So during a trial whose condition label is '24dt', you heard a stimulus with a 24 ms VOT. In the Russian Identification experiment, the Condition label is the VOT preceded by the string 'ADA'. By default, the VOT is negative (i.e. pre-release), e.g., 'ADA20' = VOT of -20 ms. For conditions in which it is positive, the condition name is annotated, e.g. 'ADA+2' = VOT of 2ms.

   In the Discrimination experiments, the relation between labels and stimulus type is given in the following table:

English Discrimination (Label: VOT1/VOT2)		**Russian Discrimination**	
1S**:** 0 / 0	**SAME**	1RS**:** +8 /+8	**SAME**
2S: 7 / 7		2RS: +2 /+2	
3S: 14 / 14		3RS: -4 /-4	
4S: 21 / 21		4RS: -10 /-10	
5S: 28 / 28		5RS: -16 /-16	
6S: 35 / 35		6RS: -24 /-24	
7S: 42 / 42		7RS: -28 /-28	
8S: 49 / 49		8RS: -34 / -34	
9S: 56 / 56		9RS: -40 / -40	
1: 0 / 14	**DIFFERENT**	1R: +8/-4	**DIFFERENT**
2: 7 / 21		2R: +2/-10	
3: 14 / 28		3R: -4/-16	
4: 21 / 35		4R: -10/ -24	
5: 28 / 42		5R: -16/-28	
6: 35 / 49		6R: -24/-34	
7: 42 / 56		7R: -28/-40	
1L: 7 / 35		1RL: +2/-16	
2L: 14 / 42		2RL: -4/-24	
3L: 21 / 49		3RL: -10/-28	
1L5: 0 / 35		1RL5: +8/-24	
2L5: 7 / 42		2RL5: +2/-28	
3L5: 14 / 49		3RL5: -4/-34	
4L5: 21 / 56		4RL5: -10/-40	

   Sort the data by selecting any cell in the dataset and then the sort command (Data >> Sort). You will see a dialog box, which allows you to sort the table along several (ordered) dimensions. The responses in the example below indicate the data should first be sorted by 'Condition', and within condition, by 'keys' (i.e., the key pressed).



4. **RECODE THE DATA.** The resulting sorted data should make now make it to easy to see -- for any given condition -- how many 'f's and how many 'j's were pressed. You could count these manually to create a summary table, but here's a tip that will help automate the analysis. Use a formula to re-code 'f's and 'j's. Here's how:

   In the Identification experiment, 'f' and 'j' correspond to /dae/ and /tae/ labels, respectively. The goal of analysis in that experiment is to determine, for a given VOT, what percentage of responses are /dae/ (or, /tae/). If we re-code 'f' as the number 100, and 'j' as the number 0, then the average of those numbers will give the percentage of /dae/ responses were given. Take a second to make sure you under the logic here. In Excel, you can recode stimuli by using an *if-then* formula. *If-then* formulae are structured as follows:

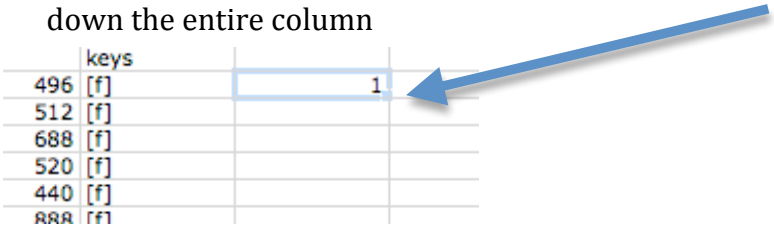
   ```
   =IF(P,T,F)
   ```

   Where P sets a condition (the *if* part), T returns a value for the cell if P is true (the *then* part), and F returns the value for the cell is P is false (*else*). Suppose you have information about what key was pressed in cell D2, then put the following formula in E2:

   ```
   =IF(D2="[f]",100,0)
   ```

   In plain English: "if the contents of cell D2 is "[f]", then put a 100 in this cell; otherwise put a 0." *Note*: it is important to enclose [f] in double quotation marks.

Once you type in the formula once (and press Enter), copy it to the remaining cells in that column. In Excel, the easiest way to do this is by grabbing hold of the cell by the small box in the bottom-right corner of the cell and dragging that box down the entire column

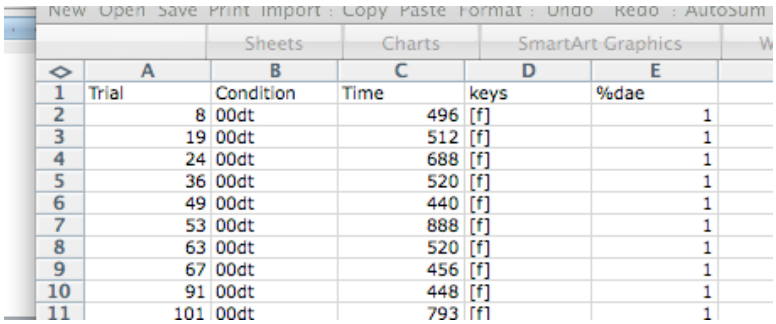


	keys	
496	[f]	1
512	[f]	
688	[f]	
520	[f]	
440	[f]	
888	[f]	

In the Discrimination experiment, the measure of interest isn't what percentage of trials was assigned a certain label. Instead, we want to know how many sound pairs were correctly discriminated. The stimuli (objectively) are either the same or different, so a response is either (objectively) right or wrong. Therefore, we are interested in knowing percentage correct. Since the '[f]' key is used for 'same' responses in the experiment, the formula above will recode 'same' responses as 100, and different responses as '0'. For conditions in which the two sounds are identical ('AA' trials), this will code correct responses as 100; but for conditions in which the sounds are actually different ('AB' trials), this will code incorrect responses as 100. You will therefore want to modify the formula above, depending on condition type. Alternately, you can do the analysis in terms of '% same', but remember in the averaging and summary steps below to convert '%same' to '%correct' for the AB trials. Take a moment, and you should see how to do this straightforwardly.

5. **AVERAGE THE DATA.**

If your formulae succeeded, then your table should have a new column, filled with numbers instead of characters. Give it a column label, like %dae.



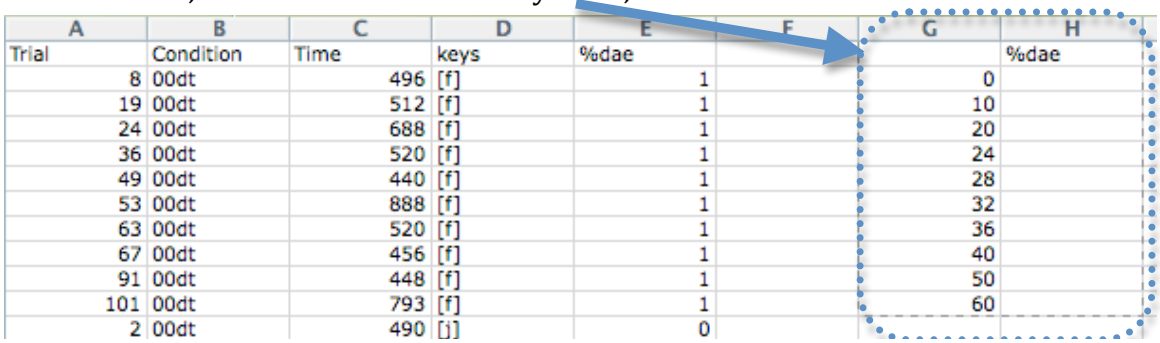
New Open Save Print Import : Copy Paste Format : Undo Redo : AutoSum

Sheets Charts SmartArt Graphics

<>	A	B	C	D	E
1	Trial	Condition	Time	keys	%dae
2	8	00dt	496	[f]	1
3	19	00dt	512	[f]	1
4	24	00dt	688	[f]	1
5	36	00dt	520	[f]	1
6	49	00dt	440	[f]	1
7	53	00dt	888	[f]	1
8	63	00dt	520	[f]	1
9	67	00dt	456	[f]	1
10	91	00dt	448	[f]	1
11	101	00dt	793	[f]	1

Average the numbers in the %dae column for each condition. Here's how I did it:

Off to the side, I created a data summary table, with the VOT as row names.



A	B	C	D	E	F	G	H
Trial	Condition	Time	keys	%dae			%dae
8	00dt	496	[f]	1		0	
19	00dt	512	[f]	1		10	
24	00dt	688	[f]	1		20	
36	00dt	520	[f]	1		24	
49	00dt	440	[f]	1		28	
53	00dt	888	[f]	1		32	
63	00dt	520	[f]	1		36	
67	00dt	456	[f]	1		40	
91	00dt	448	[f]	1		50	
101	00dt	793	[f]	1		60	
2	00dt	490	[j]	0			

I identified the cells associated with the 0 VOT. In my spreadsheet, it was cells E2 through E12. For the 0 VOT cell in my data summary table, I then entered the formula

```
=average(e2:e12)
```

The name of this formula is 'average', and its argument is the range of cells containing the data you wish to average. Notice how the colon is used to separate the beginning and ending cells in the range.
Repeat (b) and (c) for each VOT.

6. **Graph the data.** Your data summary table should now be complete. Create a chart to visualize the data. I recommend a line or scatter plot. The independent variable should be on the x-axis (e.g. VOT in milliseconds, pair in the discrimination) and the dependent variable on the y-axis (what are the right units for a given experiment?). For the discrimination experiment, there are several intervals of discrimination pair (+14ms, +28ms, +35ms). Use your judgment in determining how to best order them on the x-axis; alternately, make three separate plots for each interval type.

Write-up

Write a report for this experiment with the following information:

- A brief introduction, with enough background information to orient the reader. Use information from the in-class lecture, discussion, and readings. What sort of hypotheses would have led people to do this kind of experiment in the first place?
- **Set-up.** Describe the two types of experiments you will be reporting. Typically a research report would then include details about the design of the experiment, how the materials are made, etc. Since I have provided the lab before, you can simply say you obtained the materials and the scripts from me [see the introductory footnote about where *I* got them]. But you should describe how a experimental session proceeds. Also indicate what order you ran the sub-experiments in.
- **Results.** Describe the pattern you saw in each experiment. Why do you think certain conditions showed the accuracy they did – particular for English?
- What were the differences between your performance on English and Russian contrasts? Why do you think you observed these?
- Be sure you include graphs for each of the four experiments.
- Is there anything else about the data worth mentioning? Anything surprising, or is the pattern totally predictable?
- How 'clean' does the data look to you? Is there anything that looks like an artifact of the experiment? What could be done to improve the procedure or get better data?
- Recommend an informative follow-up.

Format:
Use common sense to organize the paper in a logical fashion. If you cite papers, use in-text citations in parentheses – like "… (Eimas et al., 1971)." – and include a "Works Cited" list at the end of the paper. If you want to know the real nitty-gritty about properly formatting a paper, see this helpful resource on the APA style: http://owl.english.purdue.edu/owl/resource/560/01/. I am not grading on adherence to a style-sheet – but organization and readability count.

If you use a word processor other than Microsoft Word, try to convert it to a PDF [usually via the "Print" or "Save" dialog boxes].