Dyslexia Experiment Writeup
Jacob Rowen

Testing the Phonetic Processing Effects of Dyslexia

Abstract: In this experiment, two people diagnosed with dyslexia were asked to do a set of vowel recognition tasks in an effort to determine what components would be difficult. As expected, their performance had a huge positive correlation with the predictability of surrounding consonants. Perhaps more surprisingly, they seemed to more readily categorize by F2 than F1. This paper and all other components of the experiment can be found here.

Background and Hypothesis

One of the potential symptoms of dyslexia is difficulty detecting differences between different speech sounds in running speech. This form of dyslexia runs in my personal family; my mother's mother showed symptoms of it, and my mother, Sarah, and sister, Anne, are diagnosed with it. However, this diagnosis doesn't provide much detail as to what precise phonetic properties challenge people with this dyslexia symptom. In this experiment, I aimed to gain a better understanding to this end.

In my personal experience with Sarah and Anne, vowel sounds are the most challenging for them to distinguish in their own speech. For years of her teenage life, Anne pronounced the word "also" something like [ɛlsoʊ], and she still has trouble picking the correct sounds to distinguish "woman" from "women". Based on this anecdotal evidence, I figured I could learn the most by focusing my inquiry specifically on vowels and their properties. My general hypothesis going in was that despite having completely normally functioning ears, there was some property of vowel formants that people with this dyslexia symptom didn't process well. This was admittedly vague, but I tried to make the experiment itself robust enough and keep my analysis of the data simple enough that my results wouldn't be determined by dredging from noise.

Experiment Procedure

To keep things simple, I decided to format the experiment as a series of AX tests. This seemed extremely appropriate, since the symptom in question makes distinguishing between sounds difficult. I decided to only test on ten different vowel sounds: /a/, /æ/, /eɪ/, /ɛ/, /i/, /ɪ/, /oʊ/, /u/, /ʊ/, and /ʌ/. These are all phonemic in English, and I was comfortable producing them.

The experiment was split into four different sections, designed to have different levels and kinds of difficulty: Bare Vowels, Same Consonants, Different Consonants, and Words. For the Bare Vowel section, I recorded myself producing each vowel alone, without any consonants. For the Same Consonants section, I recorded myself producing utterances of the form /nXnX/, where X was filled in with each vowel, e.g. /nini/. For the Different Consonants section, I recorded myself producing utterances of the form /C_1XC_2X/, where in each recording C_1 and C_2 were filled in with consonants chosen by a random number generator, and X was filled in with one of the vowels. For the Words section, I recorded myself speaking monosyllabic words chosen as randomly as I could muster, each with one of the vowels as the nucleus of the lone syllable. For every section, I made two recordings per vowel; in the first two sections, each recording came out roughly the same, whereas in the last two sections, the surrounding consonants were different in each recording. I did my best to keep my intonation, loudness, and speed consistent in each recording and to control for any other external factors that would distinguish one recording from another.

The actual experiment was performed with each participant in identical conditions, one at a time. To start the experiment, they were given headphones and some sample audio to adjust the computer to their ideal volume. They were then given an explanation of the task before them, and then an explanation of the first section specifically. Once all this was explained and any clarifying questions were answered, I left the room, unable to hear what the sounds they were being given. I observed from a nearby room in case of any issues; none occurred. I had the participants stop at the end of each section so that I could explain what would happen in each section immediately before that section started.

The actual questions served to each participant was completely random, with a 50% chance to contain the same vowel in each recording and a 50% chance to contain different vowels. The participant was presented with each recording exactly once, and then asked to indicate whether the vowel in each recording was the same or different by typing the full word “same” or the full word “different” (to prevent any accidental answers). Each participant took about the same amount of time for the whole test, but I did not measure this in detail. The experiment was conducted entirely using the python script and recordings on the github page, if any additional details are desired.

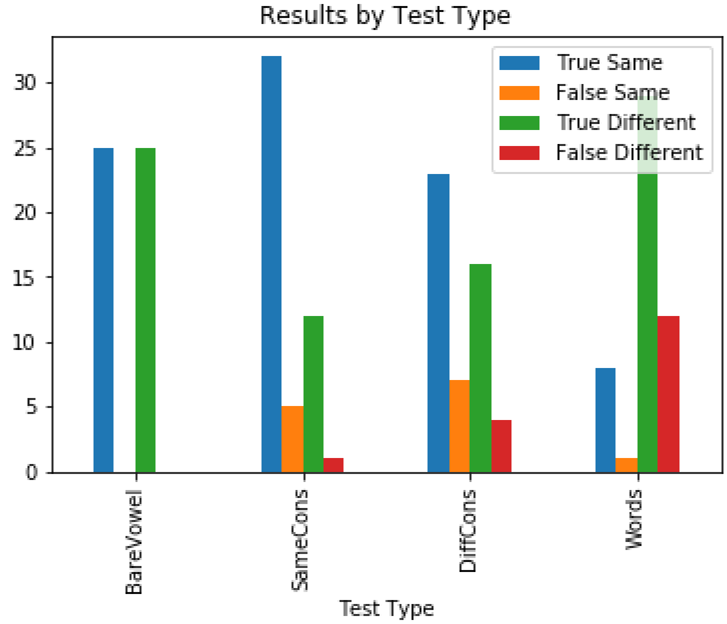
Results

Three different people participated in the experiment. First, my father (Larry) participated as a control group. He got 99/100 of the questions right and confirmed that my testing apparatus was sufficiently clear and straightforward to use. The full results of his run can be seen in Larry.csv. Then Sarah (my mother) and Anne (my sister) took the experiment, with far more interesting results. Their full data sets can be seen Sarah.csv and Anne.csv, respectively. My father also confirmed that, as a fairly typical candidate, getting almost every single question right was extremely easy for him and took almost no effort, confirming that wrong answers from the other participants can be attributed to their dyslexia. From this point forward, I will be excluding his results from the data and analysis, as his results would only serve to dilute the results from Sarah and Anne.

Sarah got a total of 13/100 answers wrong, and Anne got a total of 17/100 answers wrong. Each of them reported that they found the test highly challenging. Sarah reported that she was surprised at how challenging it was, and that she was exhausted after trying so hard on the whole thing. She was even more surprised to hear how easy it was for my father and for me. Anne reported that she had taken a lot of tests that she found similar when she was being diagnosed with dyslexia. She said she had long since accepted that doing well on these tests was essentially impossible for her. I think the four point discrepancy between them might be accounted for by a difference in effort and expectations based on their attitudes exiting the test.

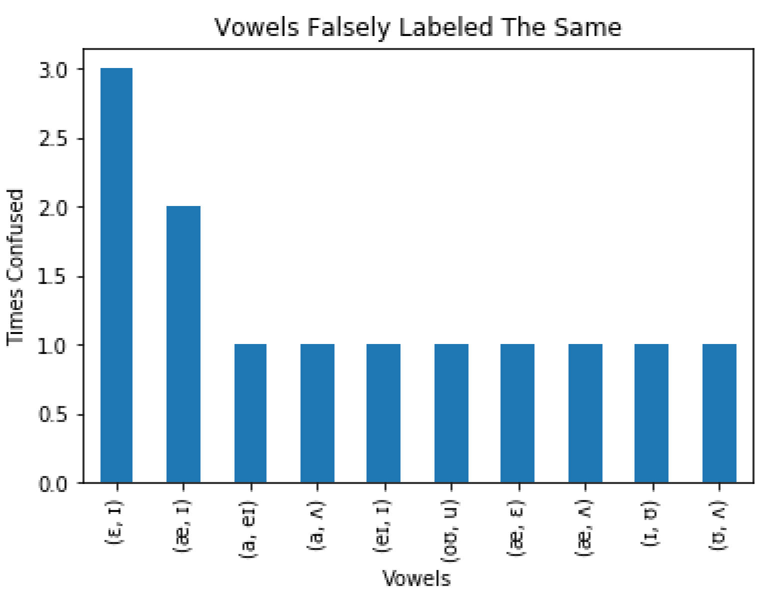
Dyslexia Experiment Writeup
Jacob Rowen

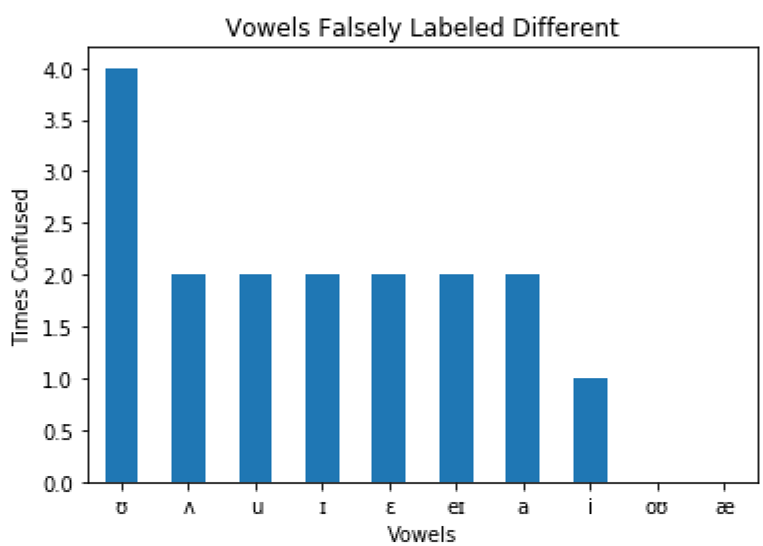
Here is a full breakdown of their combined results, split by section:



“True Same” refers to when the vowel sounds in each recording were actually the same, and the participant answered as such. “True Different” refers to when the vowel sounds were different, and the participant answered that they were different. “False Same” refers to when actually different vowels were answered as being the same, and “False Different” refers to when the same vowel was answered as being different from itself.

Here’s an aggregated breakdown of the vowels in their incorrect answers:





Dyslexia Experiment Writeup
Jacob Rowen

Analysis

Immediately, it jumps out from the data that most of the incorrect answers were in the latter sections. As I expected when designing the various sections, both participants reported that the last two sections were harder than the first two. On top of that, though, they each reported that they found the third section impossibly challenging, and that a lot of their answers were nearly random guesses. This is borne out in the data to some extent, but due to the 50/50 nature of the AX testing, it's a bit hard to tell how lucky they got.

Regardless, it is also immediately noticeable that the wrong answers in the Same Consonants and Different Consonants sections were predominantly False Sames, whereas the wrong answers in the Words section were predominantly False Differents. This can probably be attributed to another piece of feedback independently given by both Sarah and Anne. In the Words section, faced with such uncertainty about the vowel sounds, they each independently decided to try to spell the words they were hearing and just answer whether the words had the same vowel letters. Even this was confounded by their chronic difficulties with spelling that also arise from their dyslexia, though. Regardless, this dramatic change in strategy they each made for this section might explain the differences in the results.

Briefly, it's worth going over the results from the other sections individually. Both participants reported, and the data bears out, that the first section was not very challenging at all; all of the answers from the first section were correct. This is good, as it shows there was nothing inherently challenging about the experiment setup, and each participant was able to hear everything properly and answer properly. It also shows that the dyslexia doesn't get in the way of these isolated judgements, without any other sounds around to increase the processing load. The Same Consonants section caused almost exclusively False Sames; the participants reported not having too much trouble with this section relatively, though they didn't necessarily do all that well on it. This disconnect is interesting, and it might be ripe for further study, but delving into it out of the scope of this experiment. Regardless, they still did better on the Same Consonants section than either of the last two sections, which confirms the expectation that a greater processing load makes vowel idenitification more challenging for people with this dyslexia symptom.

Looking at the particular vowel sounds, there didn't seem to be any particular vowel grouping that really stood out among the ones confused for each other in False Sames or the ones in False Differents, with the possible slight exception of /ʊ/ in the False Differents, though even this seems slight enough to not point to much. Perhaps a better way of confirming this would be to look at the average formant values[1] in hertz for each type of correct and incorrect answer:

	F1	F2	F3
True Same	503.427791	1677.298617	2578.756461
False Same	525.230721	1568.985872	2561.303527
True Different	510.817117	1599.091735	2513.850159
False Different	508.858198	1571.073796	2512.380914

[1] Calculated in formant_analysis.py, which finds the formants at the peak of intensity in each recording

Relatively speaking, the average formant values for all of these answer types were extremely similar. Thus, no particular vowel or type of vowel was particularly easy or difficult for the participants to categorize, e.g. high vowels or back vowels or unrounded vowels.

However, that is not to say that there were no relevant generalizations to be made about vowel formants and the participants' ability to correctly identify different vowels. Let's look at the average *differences* between the formants of each recording in each question, split by type of correct or incorrect answer:

	F1	F2	F3
True Same	30.112704	104.988122	168.788264
False Same	121.948080	251.660524	241.921538
True Different	179.544871	463.075380	346.896550
False Different	49.438788	198.189236	299.122028

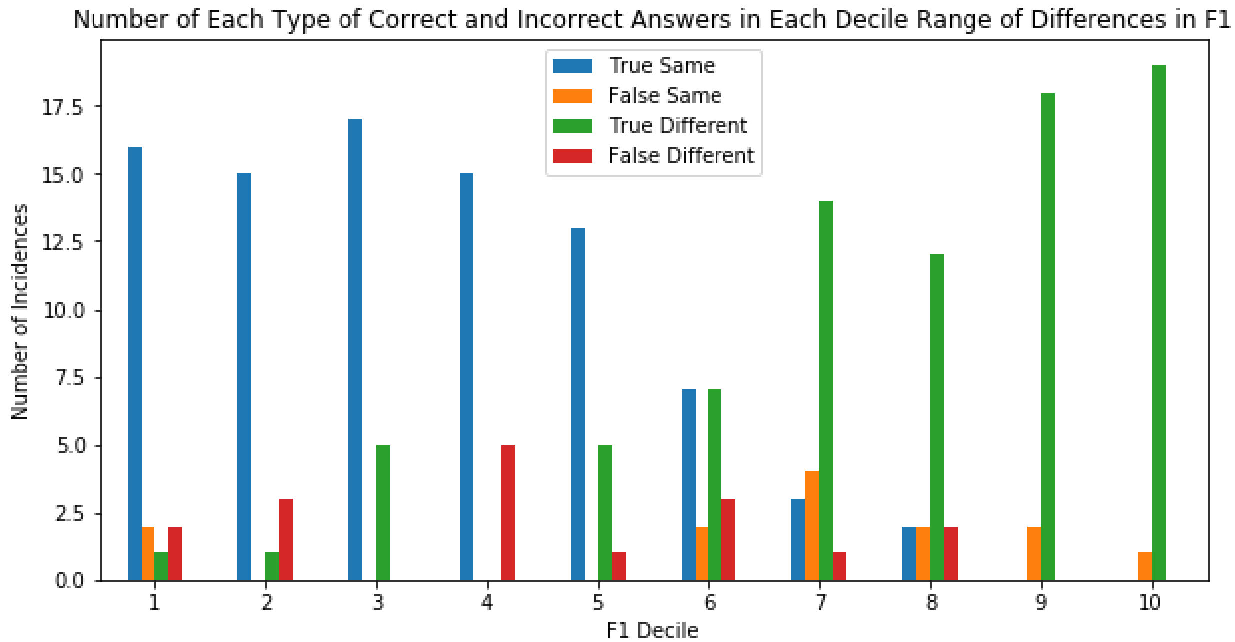
As one might expect, generally speaking of the averages, True Sames had smaller differences in formants than False Differents, and True Differents had larger differences in formants than False Sames. In other words, the more accoustically similar the recordings, the more likely the participants labeled those recordings as having the same vowel, whether it actually did or didn't. This shouldn't be shocking; on its own, all it confirms is that, like anyone, people with this dyslexia symptom are more likely to label similar things as being similar and different things as being different.

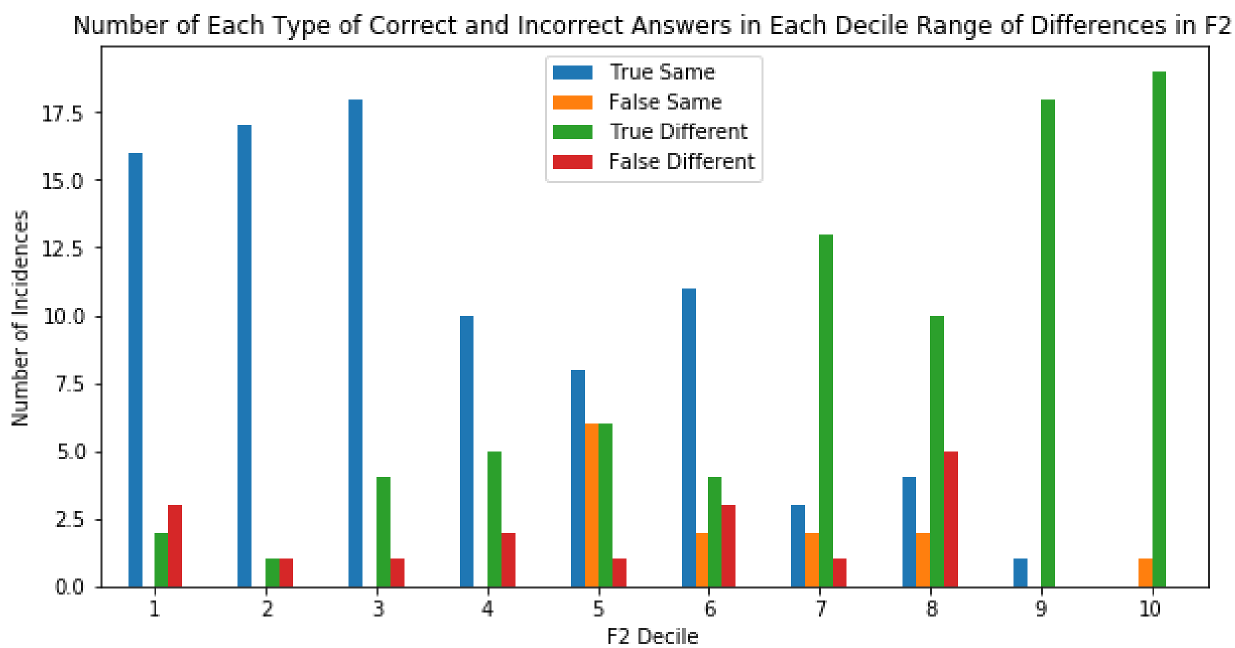
But digging a bit deeper into the details of the formant differences reveals something a bit more interesting; see the graphs on the following page. These graphs are based on cutting the formant differences into even deciles. Thus, the recording pairs that landed in the first few deciles for F1 had very similar F1s, while the recording pairs that landed in the eigth or ninth or tenth deciles for F1 had very different F1s. The precise calculations used to generate these graphs, and all other calculations and graphs herein, can be found in [this jupyter notebook] (or, if you just want to see a pdf with the output already generated, that can be found [here]).

The first thing that must be remarked upon is the clear trend in the correct answers that the recordings of the same vowels are very common in the lower deciles and dwindle through the higher deciles, and the recordings of different vowels are very common in the higher deciles and dwindle through the lower ones. This is precisely what one would expect. However, the False Differents are distributed somewhat randomly among each graph; this is probably because most of them came from the Words section, where the participants weren't making their answers based on phonetic judgements.

The False Sames, though, are a much different story. In F1, they're fairly evenly distributed through the higher deciles, even up through the ninth and tenth. Crucially, this shows that when the pariticipants incorrectly labeled two recordings as having the same vowel, they did so just as much no matter how big the accoustic difference was between the recordings. In other words, the participants simply weren't all that sensitive to the difference in F1 between the two recordings, be it high or low.

Dyslexia Experiment Writeup
Jacob Rowen





This is in sharp contrast with the F2 graph. In that one, there is a much stronger presence of False Sames where the actual F2 values were more similar to each other, with a huge spike in the fifth decile and almost none in the ninth and tenth deciles. This shows a sensitivity to the accoustic differences in F2; the participants were far more likely to have False Sames when the sounds were actually accoustically similar. And, for what it's worth, if the False Differents have any trend at all, they follow the corresponding one; the False Differents are greater in the lower deciles of F1 and higher deciles of F2. Each of these trends are at least a little bit shaken by the fact that F2 is just a bit more volatile than F1, as shown by the markedly less clean trends in the correct answer types.

Conclusions

This experiment, with its population of exactly one control and exactly two highly biased participants, hardly even qualifies as evidence of anything. Even with that set aside, the biggest conclusion, that the participants were more sensitive to F2 than to F1, was arrived at by scanning the results for interesting trends, not by designing an experiment meant to test that specific generalization. Not to mention that a better experiment design would absolutely have gathered specific data about the time spent on each question and the answerer's confidence in their answers, to ascertain more directly how challenging each judgement was to make.

Yet, the trends in the data clearly do exist. At the very least, it is evidence that there should be proper experiments done on whether people with this kind of dyslexia are putting undue emphasis on their F2 judgements over F1. If this really is the case, then it's not hard to make a plausible story of what happens when they try to identify vowels. With the vowels laid bare, they can easily spot them out, but in contexts with more going on, they aren't able to grasp all the accoustic phenomena at once, so they lean on F2 to tell them what they're hearing. But F2 is not enough to make good categorical judgements, and it's often somewhat inherently volatile and hard to pin down. If people with this kind of dyslexia are using F2 to categorize what they're hearing at the expense of F1, then the fact that they find it so difficult is no wonder at all.