Lecture 3 (6 November 2012)

Assignment 5: read Hayward chapter 3.

5a. Recreate Figure 3.2 on page 52: pronounce "one two three" as Americanly as possible and see if the spectrogram in Praat looks the same. Put it in your Word file (use **Paint visible spectrogram...**).

5b. Recreate the rising pitch contour of Figure 3.8 on page 62. Use **Draw visible pitch contour...**.

5c. Give the numbers from 0 to 20 in binary notation, i.e. as sequences of zeroes and ones.

The rest of this chapter lends itself quite well to an open-book test, just as in the previous lecture.

Assignment 6: factoring out the variation between speakers. A test nearly by eye: the sign test.

It is a large advantage that all eight speakers produced all three vowels. Imagine that the 24 tokens had been produced by 24 different speakers, so that you would have had to compare, for instance, the /a/ spoken by speaker F20N (a fast speaker) with the /i/ spoken by speaker F40L (a slow speaker). The standard deviation of the durations of the eight /i/ tokens is large, so that it is unlikely that the real average Dutch duration of /i/ could be computed accurately enough to decide whether it is greater (or less) than that of /ɪ/ (we will do that test in a later lecture).

These large standard deviations are due to two things: variations between speakers (some are faster, some are slower) and variations within a speaker (you'll never measure the exact duration twice). The variations between speakers can be factored out.

Factoring out the between-speaker variation is possible if we look only at what happens within speakers. For instance, we can ask **question 6a:** is each speaker's /a/ longer than his or her /i/? Now answer this question for your version of the data...

If indeed each speaker's /a/ is longer than his or her /i/, we must now compute how likely it is that such an outcome is due to chance, i.e. how reliably can we say that these speakers' /a/'s are longer than these speakers' /i/'s? As a statistician you have to turn such questions around. You first have to define a *null hypothesis*, which in this case is: "a speaker's /a/ is on average equally long as the same speaker's /i/, and any difference between tokens is due to random variation within the speaker." As a statistician, you then try to reject this null hypothesis by showing that if it were true the probability of finding so many long /a/'s would be very small.

How small would the probability be of finding eight /a/-/i/ pairs where the /a/ is longer by /i/ just by coincidence (i.e. if the null hypothesis were true)? This is like the probability of throwing eight perfect coins (null hypothesis: 50% head, 50% tail) and all of them fall on the table with the head up. The probability that a certain speaker's

/a/ token is longer than his or her /i/ token by coincidence is 50%, i.e. 1/2. The probability that all eight speakers's /a/ tokens are longer than their /i/ tokens is 1/2×1/2×1/2×1/2×1/2×1/2×1/2×1/2 = 1/256.

If this went too fast, then here is a list of the 256 possibilities for which of /a/ and /i/ is the shorter vowel token for each of the eight speakers:

F20N	F28G	F40L	F60E	M15R	M40K	M56H	M66O
i	i	i	i	i	i	i	i
i	i	i	i	i	i	i	a
i	i	i	i	i	i	a	i
i	i	i	i	i	a	i	i
i	i	i	i	a	i	i	i
i	i	i	a	i	i	i	i
i	i	a	i	i	i	i	i
i	a	i	i	i	i	i	i
a	i	i	i	i	i	i	i
i	i	i	i	i	i	a	a
i	i	i	i	i	a	i	a
...							
a	a	a	a	a	a	a	a

This table has 2×2×2×2×2×2×2×2 = 256 rows, because for each speaker there are two possibilities (either /a/ or /i/ is shorter). If the null hypothesis is true, each of these 256 possibilities is equally likely to happen, so each has a probability of 1 in 256. Only one of the 256 rows has all /i/'s (namely the top row), so the probability that all speakers have a shorter /i/ by coincidence is 1/256. Please spend some of your 11 hours trying to understand this, because this will pay off in the future. If you do not manage on your own, invoke help from your fellow students or from me.

Now how credible will the null hypothesis be if you find that each speaker's /i/ is shorter than his or her /a/? Not so credible. The chances that all /i/ tokens are shorter than all /a/ tokens in 1 in 256, and so are the chances that all /a/ tokens are shorter than all /i/ tokens. So there is a chance of 1 in 128 that one of the two vowels is consistently shorter than the other by coincidence. That's a chance of less than 1 percent. As a statistician you say that "the null hypothesis can be rejected with $p < 0.01$", or that it can be rejected with a reliability of 99 percent.

You can also fairly easily see from the table (rows 2 through 9) that the probability that 7 out of 8 /i/ tokens are shorter than the same speaker's /a/ token is 8/256 (if the null hypothesis, i.e. the hypothesis of coincidence, is true). So the probability that *at least* 7 out of 8 /i/ tokens are shorter than the same speaker's /a/ token is 8/256+1/256 = 9/256 (that's rows 1 through 9 in the table above). If the null hypothesis is that "/a/ is not longer than /i/" (one-tailed test), such a result (i.e. 7 out of 8 /i/ tokens are shorter than the same speaker's /a/) can be said to be statistically significant with $p < 0.05$, simply because $p = 9/256 = 0.035$. If the null hypothesis is that "/a/ is not different from /i/" (two-tailed test), then such a result *cannot* be said to be statistically significant at the $\alpha = 0.05$ level (i.e. with $p < 0.05$), because $p = 18/256 = 0.07$. Please understand where the 18 comes from: this number includes the cases where at least 7 out of 8 /a/ tokens are shorter than the same speaker's /i/ token! In other words 18/256 is the chance that if you throw eight coins, at least seven of them will show the same side up (heads or tails).

Here is a complete list of probabilities for the number of speakers that have a shorter /i/ than /a/, if the two vowels don't have a different duration underlyingly:

0	1/256
1	8/256
2	28/256
3	56/256
4	70/256
5	56/256
6	28/256
7	8/256
8	1/256

So it's most likely to find 4 out of 8 speakers with a shorter /i/ than /a/. You won't have expected otherwise, I think. But finding 3 or 5, or even 2 or 6, is not at all unlikely either. The distribution in this list is called a *binomial distribution* with P=1/2 (the short /i/ probability for each speaker, or the head probability for each coin) and N=8 (the number of speakers or coins).

6b. Since /a/ is expected to be longer than /i/, you can do a single-tailed sign test on your data. Compute the p value for your measurements. If all speakers agree that /a/ is longer, p must be 1/256 (but give the value as a percentage instead). If 7 out of 8 speakers agree, p must be 9/256.

6c. If only 6 out of 8 speakers agree that /a/ is longer, what will be the p value (use the list above to compute the probability that by chance *at least* 6 out of 8 agree)? Is that significant with $p < 0.05$?

6d. Then the /ɪ/-/i/ distinction. We don't know yet which of the two is supposed to be longer, so we use a two-tailed sign test (i.e. you double the p values, as above in the case of 18/256). In your measurements, can you conclude moderately reliably (i.e. with $p < 0.05$) that /i/ is the longer of the two? Or that /ɪ/ is the longer of the two? Or can you conclude neither?

In the above example you see that you can compute statistical significance (the p value) almost by eye. You can also compute them in the calculator in Praat. For instance, the probability that N coins, each with a head probability of 1/2, will yield at least K heads is `binomialQ (1/2, K, N)`.

6e. Give such formulas for the p values for the three distinctions in your measurements (/i/ versus /a/, /ɪ/ versus /i/, and /ɪ/ versus /a/), and give the outcomes (two of the outcomes have to be equal to what you computed nearly-by-eye before).