Report, 2007 May 13

May 13, 2007

Abstract

In [2], subject responses are evaluated relative to correct outcomes (i.e., experiment design conditions). Here we include this aspect in the analysis but also carry out two "Turing test type evaluations". The latter are machine approaches, based on two algorithms, which generally (but not always) outperform the test subjects. It is not our purpose to simply outperform the subjects: our algorithms benefit strongly from prior information. Our algorithms are used to baseline the subject responses. A multivariate data analysis not only takes into account the correct outcomes, the subject performance, and our two algorithms, but also takes into account differentiation within the subject population.

1 Introduction

To baseline the test subject results we developed two simple "Turing test" type algorithms. (Recall that the Turing test proposes machine intelligence when the machine performance is indistinguishable from the human's intelligence.)

1.1 Baseline Approaches 1 and 2: Signal Differencing Maximum, and Signal Maximum

Our first algorithm, called "sys" (for system) in the figures and tables, is as follows. Following differencing, using absolute difference, of two signals we seek a relative maximum in the region where the feature is superimposed. (Cf. discussion of this algorithm on p. 20 of [2]. Even with lagged signals, actually, it works quite well. A reason for this is to be found by way of our second, "basic" algorithm: a maximum signal value, alone, is a very good indicator of feature presence. Our signal differencing is complemented by the maximum of the differenced signal.)

Experiment 1: subject vs. system				
Subject	(a)	(b)	(c)	(d)
amw	259	19	92	4
exs	210	27	51	2
gxv	279	49	52	16
mxk	256	60	51	14
pxr	262	61	52	15
sxa	0	0	1	0
axc	300	28	59	5
gwh	212	3	86	0
mxd2	215	19	38	12
njd	240	67	55	17
rdr	200	45	55	7
uyx	253	59	55	12

Table 1: Experiment 1. Left column: test subject. (a) = the subject was correct, our system was correct. (b) = the subject was correct, our system was wrong. (c) = the subject was wrong, our system was correct. (d) = both subject and our system were wrong.

Figure 1 shows all patterns in Experiment 1 differenced related to pattern 6, where the latter has a superimposed feature (in the second and third seconds of the signal) of height 30. In Figure 1, we find that our thresholding algorithm is successful for patterns 2, 3, 5, 6 (of course!), 7, 8, and 9. It is unsuccessful for patterns 1, 4, and 10. The height of the feature plays an important role. A lower height make the detection of the feature more difficult, and a height = 40, for instance, will help in some cases.

We see therefore that the height of the feature is important, and so too are the relative values in the region where the feature is to be found.

In the next subsection we will base an algorithm (referred to as: our system) around this differencing of the signals, in order to lay bare the added feature. To benchmark the results, we will also investigate a more basic algorithm (referred to as: “basic” system) which just looks for a signal maximum in the interval spanned by the feature. In both cases we are looking for the feature in the known interval. Both approaches rely on this prior knowledge.

Results shown in Tables 1 and 2 show how the algorithm performs better

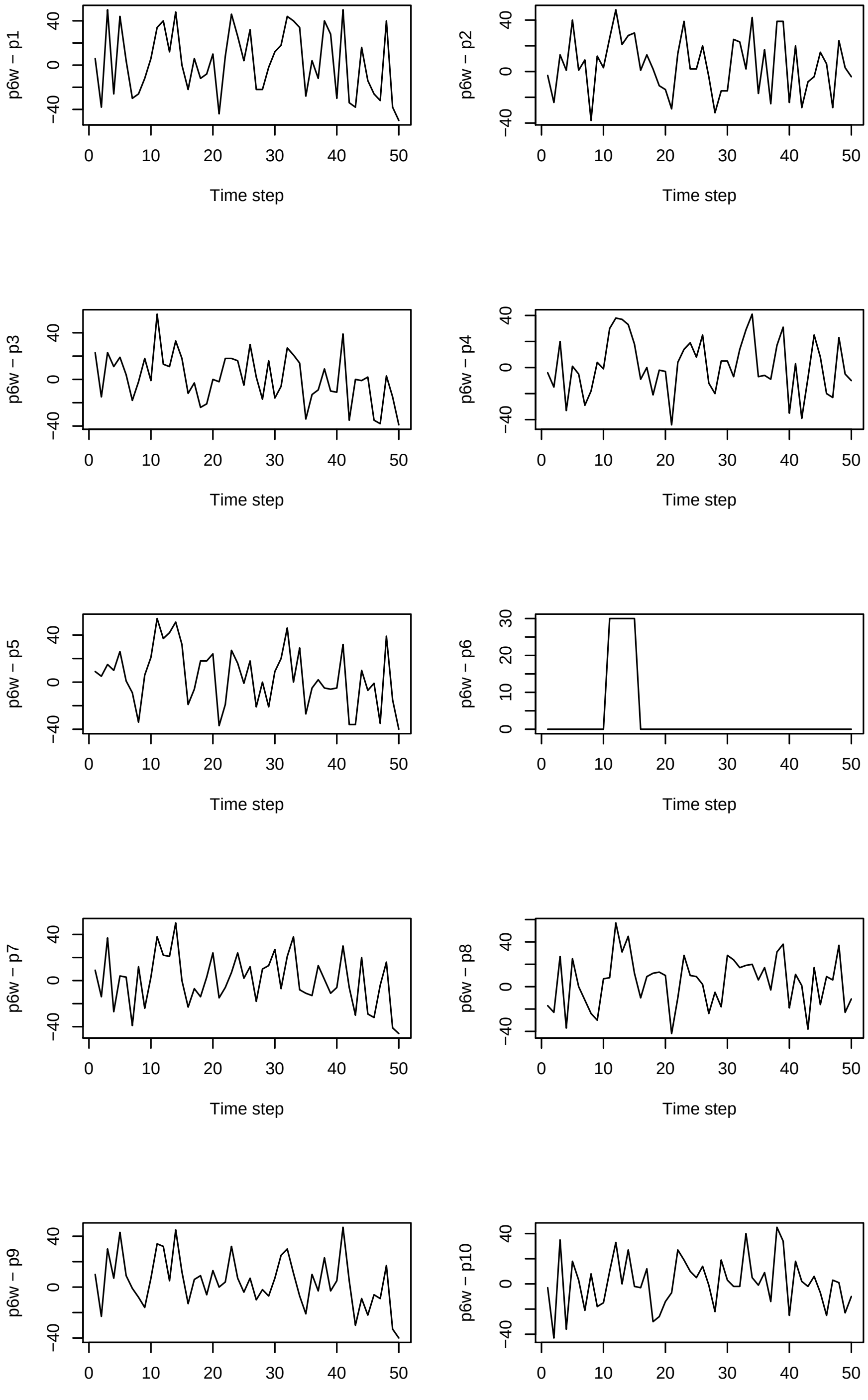

Figure 1: Pattern 6 from Experiment 1, differenced with respect to all other patterns. Pattern 6 has an added feature.

Experiment 2: subject vs. system				
Subject	(a)	(b)	(c)	(d)
s01	173	27	20	9
s02	185	9	37	5
s03	170	45	32	15
s04	196	32	40	7
s05	147	27	24	14
s06	203	9	46	6
s07	199	17	43	9
s08	171	29	35	9
s09	211	12	42	2
s10	143	46	32	15
s11	168	13	42	4

Table 2: Experiment 2. Left column: test subject. (a) = the subject was correct, our system was correct. (b) = the subject was correct, our system was wrong. (c) = the subject was wrong, our system was correct. (d) = both subject and our system were wrong.

than the human subjects. A number of remarks must be added to modulate this very evident conclusion. Firstly, our algorithm looks for a superimposed feature in a given interval of the pattern. Secondly in "blindly" looking for an overall maximum in this targeted interval, the maximum is used without regard for extent of this maximum. Evaluation of our algorithm is therefore necessarily in terms of "black box" performance.

Table 3 contrasts our system with the even more basic one, which seeks a signal maximum value in the interval where the feature is superimposed. We find considerable correspondence between our system and this more basic one.

We next ask if the more basic approach would provide an even better account of the human subjects. Table 4 shows the comparison between subject and "basic" appraoch. We find that Table 4 better accounts for the subject results, relative to our system in Table 1.

For Experiment 2, Table 5 contrasts our system with the even more basic one. Again, we find considerable correspondence between our system and this more basic one.

For Experiment 2, Table 5 shows the comparison between subject and

Experiment 1: "basic" vs. system				
Subject	(a)	(b)	(c)	(d)
amw	315	21	36	0
exs	222	27	20	0
gxv	278	50	26	0
mxk	263	51	33	6
pxr	241	51	31	6
sxa	0	0	1	0
axc	283	26	53	4
gwh	242	3	48	0
mxd2	235	29	14	2
njd	243	55	35	3
rdr	186	38	31	3
uyx	248	62	37	4

Table 3: Experiment 1. "Basic" system: taking maximum value in the signal and verifying that it appeared in the interval where a feature is expected. Left column: test subject refers. (a) = the "basic" approach was correct, our system was correct. (b) = the "basic" system was correct, our system was wrong. (c) = the "basic" system was wrong, our system was correct. (d) = both "basic" system and our system were wrong.

Experiment 1: subject vs. “basic”				
Subject	(a)	(b)	(c)	(d)
amw	274	0	62	36
exs	218	0	31	20
gxv	288	0	40	26
mxk	293	0	21	39
pxr	271	0	21	37
sxa	0	0	0	1
axc	301	0	8	57
gwh	206	0	39	48
mxd2	229	0	35	16
njd	270	0	28	38
rdr	203	0	21	34
uyx	284	0	26	41

Table 4: Experiment 1. Left column: test subject. (a) = the subject was correct, “basic” system was correct. (b) = the subject was correct, “basic” system was wrong. (c) = the subject was wrong, “basic” system was correct. (d) = both subject and “basic” system were wrong.

Experiment 2: “basic” vs. system				
Subject	(a)	(b)	(c)	(d)
s01	171	24	34	8
s02	189	7	46	4
s03	156	34	35	12
s04	186	32	47	5
s05	149	24	32	12
s06	200	9	59	4
s07	201	13	69	5
s08	169	27	42	8
s09	212	13	63	2
s10	138	41	41	13
s11	162	12	47	3

Table 5: Experiment 2. Left column: test subject. (a) = the subject was correct, the “basic” system was correct. (b) = the subject was correct, the “basic” system was wrong. (c) = the subject was wrong, the “basic” system was correct. (d) = both subject and “basic” system were wrong.

“basic” appraoch. We find that Table 5 better accounts for the subject results, relative to our system in Table 2.

2 Analyses

2.1 Introduction to the Methodology

A few principles exemplified by our analysis, based on correspondence analysis, follow.

Firstly, we relate subject response not to a baseline of zero correctness of feature detection (or, for that matter, zero incorrectness of feature detection) but instead to automated, algorithmic approaches. From the work above, it is clear that these algorithmic approaches generally outperform the human subjects. What is less clear is how and where the human subject can be differentiated across the range of feature properties. A crucially important point is that the human subjects must understand the framework within which they are to indicate the presence or absence of the haptic feature, but our algorithms have much more specific prior information on which to base their decisions.

Secondly, we deal with outcomes on the aggregate level, in various conditions (related to correlation condition, or lag condition) in terms of numbers of correctness and numbers of incorrectness outcomes. This avoids sequence or chronology influence on the outcomes (or, alternatively, subject learning effects), which are all rolled into the aggregate results.

Thirdly, in our interpretation of the data, we include the effects of the subjects as well as the main conclusions to be drawn from the subject response and the outcomes furnished by our algorithms.

2.2 Subject Responses Relative to Experimental Correlation Conditions

Experiment 1 correlation conditions: 1 = positively correlated, 2 = negatively correlated, 3 = uncorrelated. Input data are shown in Table 6. A display of this data is presented in Figure 2, where the incorrect totals (denoted "F" in our labeling) are not displayed for interpretational convenience. In Figure 2, it is seen how with correlation condition 2, see lower right, the subjects and our two algorithms are not hugely different. For correlation condition 1, see upper left, our first algorithm (labeled "sys") only is close in results furnished to the subject responses. Here too, upper left, we see that for the subjects, there is not much difference between correlation conditions 1 and 3; but there is far greater difference with correlation condition 2 (which, as we have noted, is to be found in the lower right).

Figure 3 redisplays Figure 2 but this time with subjects shown. Their locations are strongly influenced by results obtained on the nearby subject responses or algorithms.

Experiment 2 correlation condition: 1 = positively correlated, 2 = uncorrelated. Table 7 show the results for Experiment 2. Figure 4 shows the overall outcomes, where the incorrect totals have been used in the analysis (they are seen in the input data table, Table 7) but are not displayed in the figures for interpretational convenience. In Figure 4 it is seen that the subjects and our "basic" algorithm agree largely. Similarity is less between the subjects and our "system" algorithm.

Figure 5 shows subject positions. Subjects s04, s07 and s08 are all close to the origin, that is to say, they are close enough to being average in their profiles. (A *profile* is the vector of values, normalized by division by column sums.) Subjects s05 and s11 are less average in their profiles, but are not greatly different from the average.

The remaining subjects are quite distinct from the average. The first factor is influenced by a polarity between s03 and s02. The second factor is

	c1	c1	c2	c2	c3	c3	c1	c1	c2	c2	c3	c3	c1	c1	c2	c2	c3	c3
	sub	sub	sub	sub	sub	sub	sys	sys	sys	sys	sys	sys	bas	bas	bas	bas	bas	bas
	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F
amw	99	27	82	25	97	44	126	0	89	18	136	5	123	3	106	1	138	3
gxv	105	24	114	21	109	23	129	0	101	34	101	31	94	35	126	9	126	6
mxk	81	21	139	24	96	20	102	0	115	49	90	26	91	11	158	6	103	13
pxr	117	29	106	21	100	17	147	0	80	50	88	29	103	44	115	15	109	8
axc	103	28	111	18	114	18	131	0	105	24	123	9	106	25	127	2	131	1
gwh	70	24	62	30	83	32	95	0	89	3	115	2	84	11	91	1	115	2
mxd2	79	18	81	16	74	16	97	0	77	20	79	11	91	6	97	0	90	0
njd	106	22	79	17	122	33	128	0	68	29	99	56	107	21	92	5	129	26
rdr	80	22	78	18	87	22	102	0	75	21	78	31	62	40	96	0	98	11
uyx	93	22	129	24	90	21	115	0	119	34	74	37	89	26	150	3	108	3

Table 6: Experiment 1 subjects, amw, gxv, ... , uyx. Correlation conditions 1, 2 and 3. Test subjects ("sub") with correct ("T") and incorrect ("F") outcomes. Our signal differencing algorithm ("sys" = system) with correct and incorect outcomes. A more basic algorithm ("bas" = basic) with correct and incorrect outcomes.

influenced by the polarity between s09 and a cluster of s06, s01 and s10.

3 Analysis of Lags in Experiment 2

Lag cases are as follows: 1 = Lag 0, implying no change to the signals; 2 = lag 200 ms (1 time step); 3 = lag 400 ms (2 time steps); and 4 = uncorrelated, implying no requirement for a lag. In our experimental work, the lag was applied to the second pattern (or signal).

Table 8 shows the numbers of responses per subject, which of course is also valid for the discussion of all other Experiment 2 data. What we observe in Table 8 is that these totals of responses are approximately similar. This means that the *masses* associated with subjects in correspondence analysis will be approximately equal. In turn then, this obviates any need for other normalization "tricks" in correspondence analysis, which are realized through data recoding. (Examples being *doubling* or the *complete disjunctive form* of coding: see [1].)

From Table 9 we see that there were 707 (= 583 + 124) responses across all subjects that were appraised for lag case 1. For lag case 2, there were 677 (= 562 + 115) responses across all subjects. For lag case 3, there were 702 (573 + 129) responses across all subjects. Finally, for lag case 4, there

	c1subT	c1subF	c2subT	c2subF	c1sysT	c1sysF	c2sysT	c2sysF	c1basT	c1basF	c2basT	c2basF
s01	166	24	34	5	152	38	36	3	177	13	39	0
s02	154	27	40	15	176	5	39	16	181	0	47	8
s03	155	35	60	12	124	66	34	38	146	44	65	7
s04	176	36	52	11	167	45	47	16	200	12	62	1
s05	129	27	45	11	113	43	41	15	140	16	54	2
s06	176	42	36	10	178	40	39	7	204	14	45	1
s07	166	40	50	12	165	43	56	6	197	11	61	1
s08	153	32	47	12	157	28	37	22	180	5	56	3
s09	153	34	70	10	170	18	74	6	180	8	79	1
s10	152	36	37	11	134	54	24	24	167	21	45	3
s11	138	35	43	11	152	22	45	9	149	25	54	0

Table 7: Experiment 2 subjects, s01 to s11. Correlation conditions 1 and 2. Test subjects ("sub") with correct ("T") and incorrect ("F") outcomes. Our signal differencing algorithm ("sys" = system) with correct and incorect outcomes. A more basic algorithm ("bas" = basic) with correct and incorrect outcomes.

s01	s02	s03	s04	s05	s06	s07	s08	s09	s10	s11
687	708	786	825	636	792	808	732	803	708	683

Table 8: Numbers of responses for 11 subjects.

lag case	1	2	3	4	1	2	3	4
correct?	T	T	T	T	F	F	F	F
subject	583	562	573	514	124	115	129	120
"system"	696	453	539	472	14	225	163	162
"basic"	646	610	665	607	64	68	37	27

Table 9: Cross-tabulation of totals of correct outcomes by subjects, and by our two algorithms, dubbed "system" and "basic". The four lag cases are labeled 1, 2, 3 and 4. Numbers of correct and incorrect outcomes are indicated by T (true) and F (false).

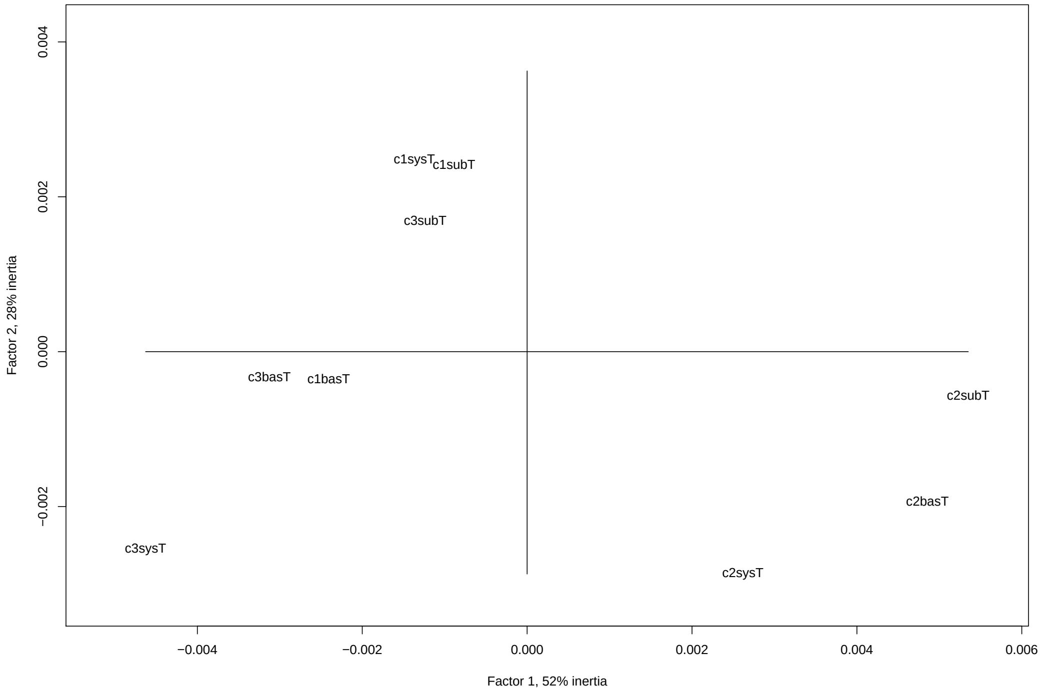

Figure 2: Principal factor plane of correspondence analysis, Experiment 1 results, using correlation conditions 1, 2 and 3. The false outcomes are not displayed, for clarity of presentation. The positive outcomes related to: subject ("sub"), our system ("sys"), and another basic system ("bas").

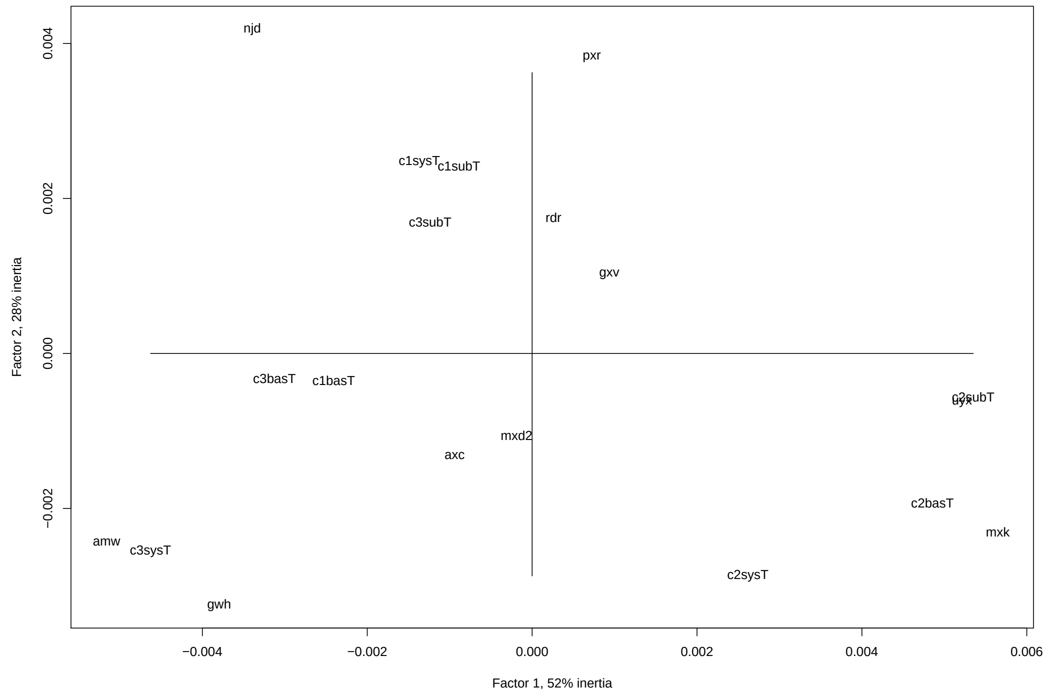

Figure 3: As Figure 2 but with subjects also displayed.

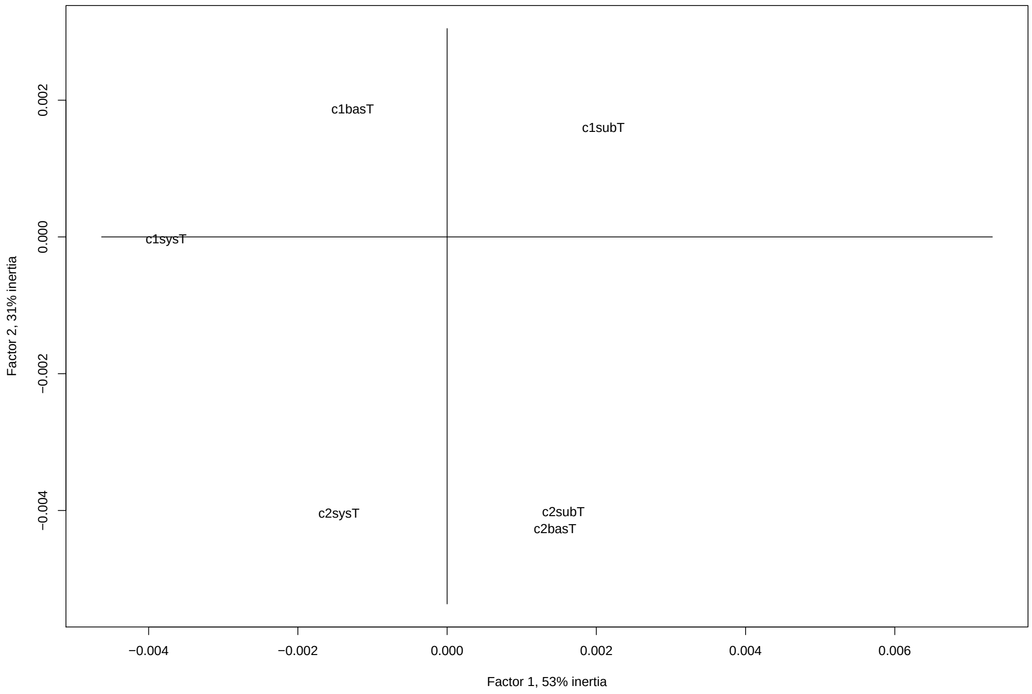

Figure 4: Principal factor plane of correspondence analysis, Experiment 2 results, using correlation conditions 1 and 2. The false outcomes are not displayed, for clarity of presentation. The positive outcomes related to: subject, our system, and a basic system.

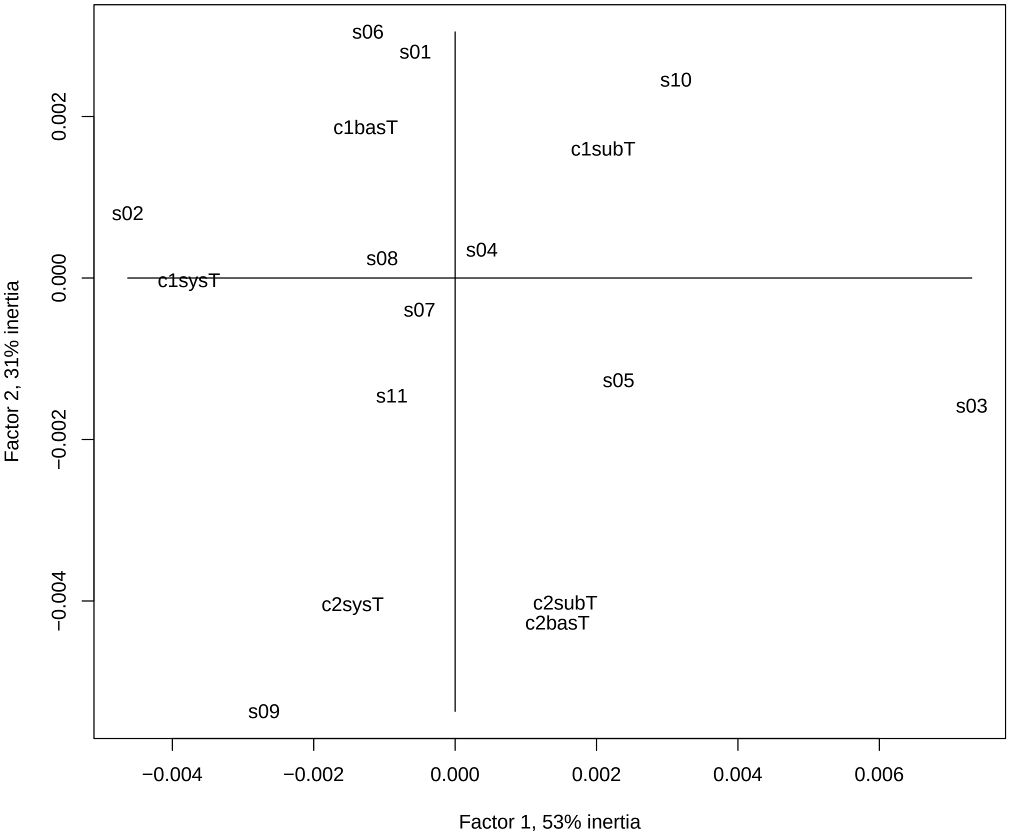

Figure 5: As Figure 4 but with subjects also displayed.

were 634 (514 + 120) responses across all subjects. Determining these totals using our two algorithms (rows 2 and 3 of Table 9) leads to the same set of totals.

[BUT THEY DON'T: LAG CASES 3 AND 4 ARE FINE. BUT LAG CASES 1 AND 2 GIVE: 710, 678. CHECK THE DATA – SOMEWHERE THERE ARE 3 MULTIPLE ENTRIES FOR CASE 1 AND 1 MULTIPLE ENTRY FOR CASE 2, THAT I HAVE DOUBLE COUNTED.]

Across all lag cases the totals of responses are close enough not to warrant further data coding for the correspondence analysis. That is to say, lag cases will have close masses in the correspondence analysis.

Figure 6 displays the results.

T (= True) indicates a correct outcome, and F (= False) indicates an incorrect outcome. We consider as before, the subject responses ("sub"), baselined against our first and second algorithms (resp., "syst" = system, and "basic" = basic). The four possible lags (lag = 1 and = 4 implying no lag in time step; lag = 2 and 3 implying, resp., a lag of 1 and of 2 time steps). Note that [2] takes the lag in a random way (uniformly distributed on the left and right patterns), whereas we take it always on the right pattern.

In Figure 6 we see that for lag = 1, the results of the subjects, and of our two algorithms, are close: see positive side of factor 1. With less proximity, but at least not scattered, we also find these results for the subjects, and the two algorithms, for lag = 2, in the lower left quadrant. We can conclude that the better result, i.e. more consistent across subjects and baselining algorithms, of the lag = 1 case has given way to the somewhat less consistent, but nonetheless coherent, result for the lag = 2 case.

The case of lag = 3 is, if anything more coherent looking than the lag = 2 case. The case of lag = 4 is less clearcut. We find that the subjects, and the "basic" algorithm, concur very well: see subT4 and basT4 in the upper right quadrant. But the "system" algorithm, sysT4, at positive end of factor 2, is somewhat separated.

Factor 1 expresses a polarity between lag case 1 vis-à-vis lag case 3 and lag case 2. Factor 2 expresses a certain polarity, with a number of reservations though, between lag case 4 and lag case 2.

Figure 7 indicates the subject influence on these results.

References

[1] F. Murtagh, Correspondence Analysis and Data Coding with Java and R, Chapman and Hall/CRC Press, 2006.

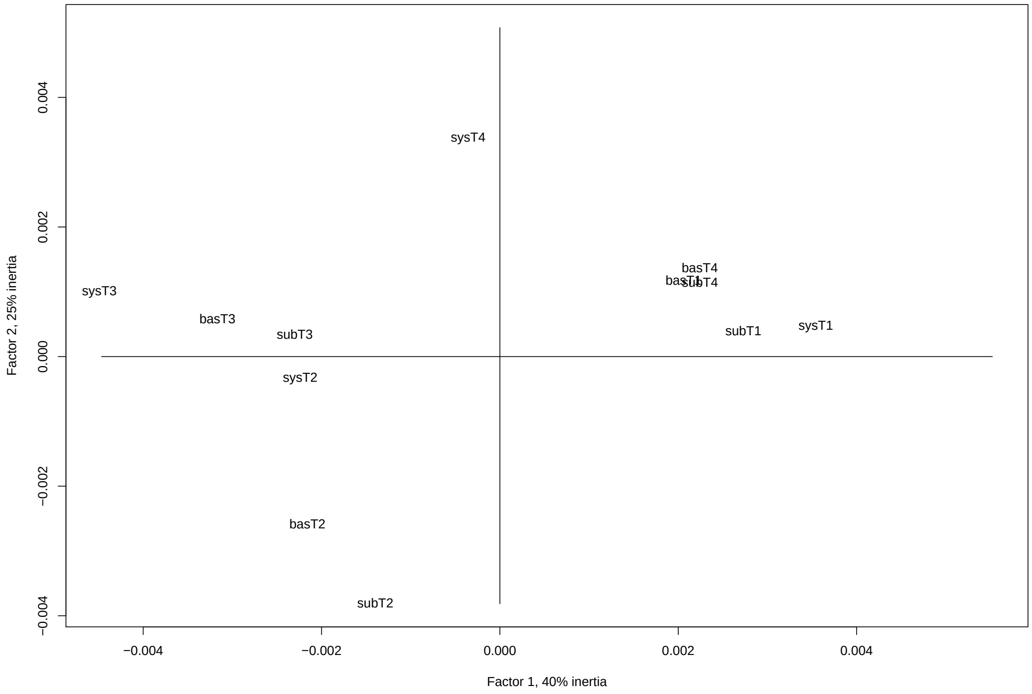

Figure 6: Principal factor plane of correspondence analysis, Experiment 2 results, using lag cases 1, 2, 3 and 4. The false or incorrect outcomes are not displayed, for clarity of presentation. The positive outcomes related to: subject, our system, and a basic system.

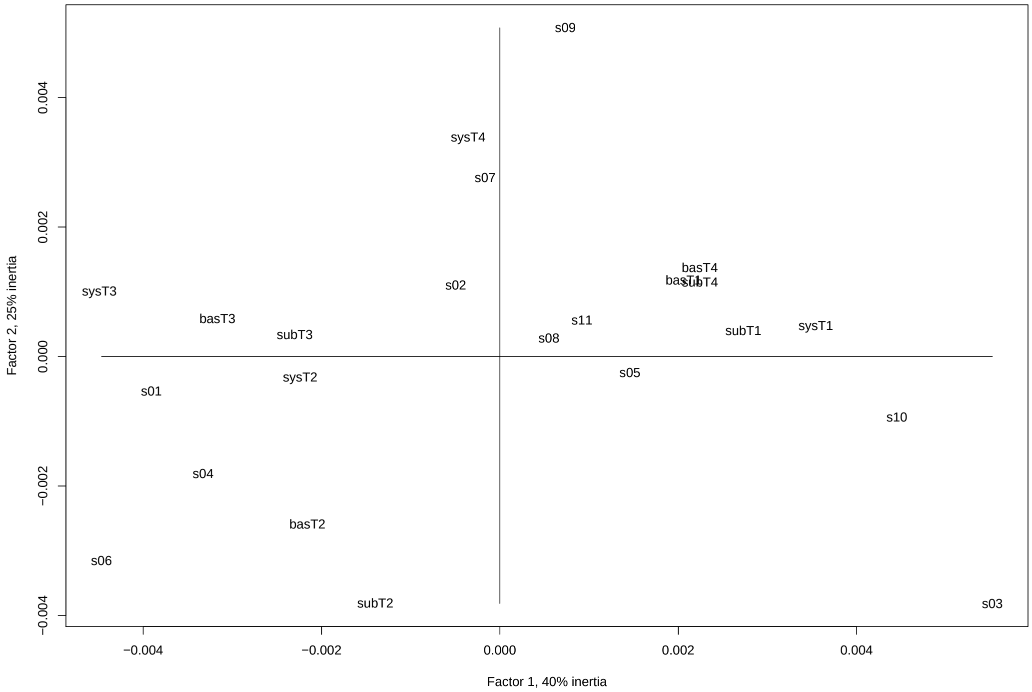

Figure 7: As Figure 4 but with subjects also displayed.

[2] R.D. Roberts, G.W. Humphreys and A.M. Wing, “Temporal constraints on ineractions across kinaesthetic channels”, 2005.