

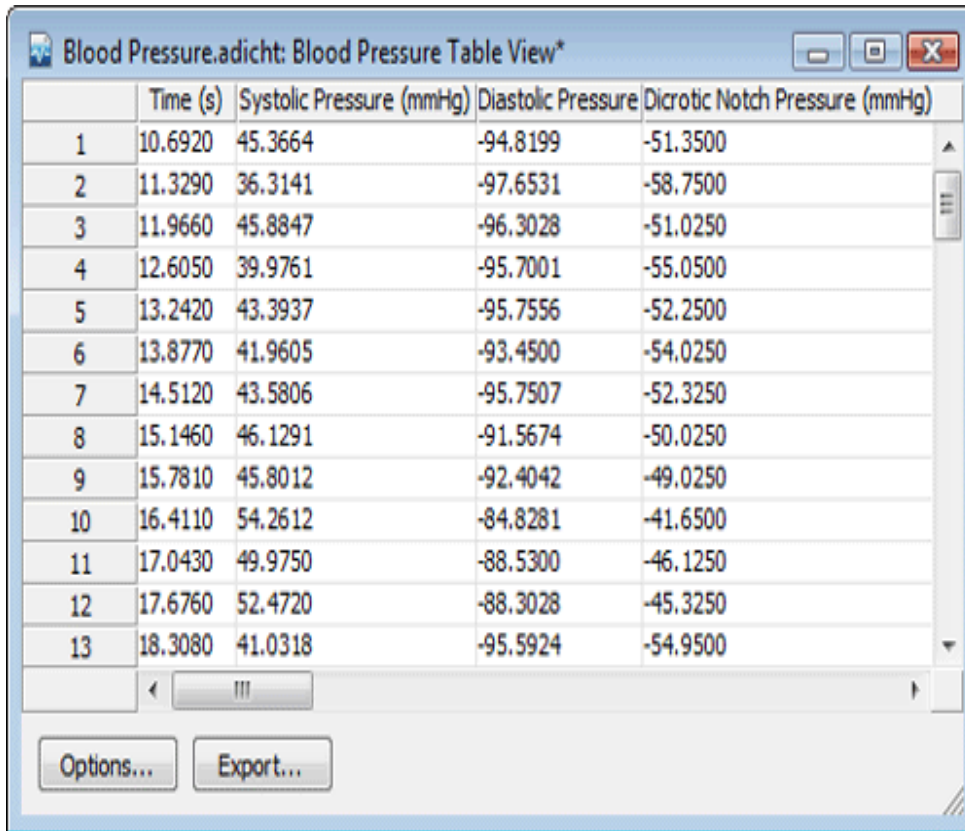
Vagueness in referring expressions: audience effects

Matt Green and Kees van Deemter

NLG group

University of Aberdeen

Input



	Time (s)	Systolic Pressure (mmHg)	Diastolic Pressure	Dicrotic Notch Pressure (mmHg)
1	10.6920	45.3664	-94.8199	-51.3500
2	11.3290	36.3141	-97.6531	-58.7500
3	11.9660	45.8847	-96.3028	-51.0250
4	12.6050	39.9761	-95.7001	-55.0500
5	13.2420	43.3937	-95.7556	-52.2500
6	13.8770	41.9605	-93.4500	-54.0250
7	14.5120	43.5806	-95.7507	-52.3250
8	15.1460	46.1291	-91.5674	-50.0250
9	15.7810	45.8012	-92.4042	-49.0250
10	16.4110	54.2612	-84.8281	-41.6500
11	17.0430	49.9750	-88.5300	-46.1250
12	17.6760	52.4720	-88.3028	-45.3250
13	18.3080	41.0318	-95.5924	-54.9500

Options... Export...

Output

“The patient’s blood pressure remained within acceptable limits.”

or

“The patient’s blood pressure had a range of 112/64, 120/70.”

	Time (s)	Systolic Pressure (mmHg)	Diastolic Pressure (mmHg)	Diastolic Notch Pressure (mmHg)
1	10.6920	45.3664	-94.8199	-51.3500
2	11.3290	36.3141	-97.6531	-58.7500
3	11.9660	45.8847	-96.3028	-51.0250
4	12.6050	39.9761	-95.7001	-55.0500
5	13.2420	43.3937	-95.7556	-52.2500
6	13.8770	41.9605	-93.4900	-54.0250
7	14.5120	43.5806	-95.7907	-52.3250
8	15.1460	46.1291	-91.5674	-50.0250
9	15.7810	45.8012	-92.4042	-49.0250
10	16.4110	54.2612	-84.8281	-41.6500
11	17.0430	49.9750	-88.5300	-46.1250
12	17.6760	52.4720	-88.3028	-45.3250
13	18.3080	41.0318	-95.5924	-54.9500

“...acceptable”



“...112/64”

How does the NLG system decide which output to produce?

- Software Engineer’s intuition? ✗
- Audience data? ✓

How do we get useful data from the audience?

- Present people with system outputs
- Measure response

Scenario (HCI style)?

- More naturalistic
- Range of possible response data

Controlled experiment?

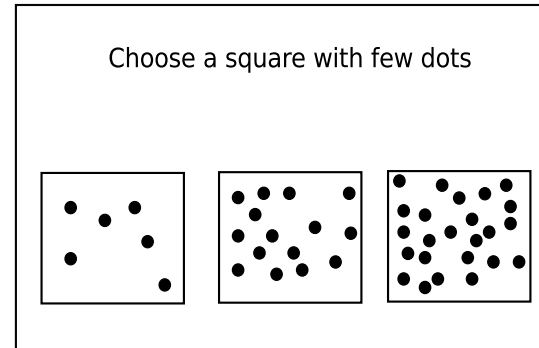
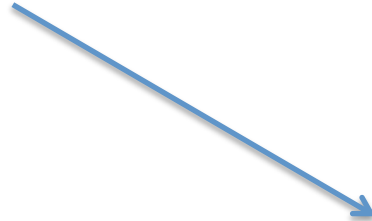
- More confidence that output is causally linked to input
- Fine-grained numerical response

Operationalising a controlled experiment

Present ps with stimulus and task



Measure response



Is reaction time a good indicator of cognitive demands?

Central contention:

- Faster accurate responses means "better", "less effortful"

Disadvantages

- Glosses over depth of processing: no info on level of comprehension, recall, retention
- Speed/accuracy tradeoff can be interpreted differently by different participants
- Not always clear how to treat erroneous responses

Advantages

- Can attribute response directly to manipulation
- Especially useful when speed is a good indicator of quality of response

Why use reaction time for systems evaluation?

Main advantage is twofold

1. RT gets at cognitive processes that are below the threshold for self-report...
compares favourably with questionnaire, interview
2. ... without too much reliance on linking hypotheses.
compares favourably with functional neuroimaging (fMRI);
electroencephalography (EEG)