	Metrics	Approaches			p-values ANOVA	p-values 1^{st} vs. 2^{nd} (t-test)
		Finite-state automata	Factored statistical model	Rule-structured model		
	Objective metrics:					
1.	Average number of repetition requests per dialogue	18.68	12.24	**0***	9×10^{-19}	1×10^{-16}
2.	Average number of confirmation requests per dialogue	9.16	10.32	**5.78***	1.7×10^{-4}	0.001
3.	Average number of repeated instructions per dialogue	3.73	7.97	**2.78**	3.8×10^{-9}	0.18
4.	Average number of “Disconfirm” acts per dialogue	**2.16**	2.59	2.59	0.65	0.33
5.	Average number of physical movements per dialogue	**26.68**	29.89	27.08	0.13	0.80
6.	Average number of (user and system) turns between movements	3.63	3.1	**2.54***	$4.\times10^{-4}$	2.2×10^{-4}
7.	Average number of user turns per dialogue	78.95	77.3	**69.14**	0.17	0.11
8.	Average number of system turns per dialogue	57.27	54.59	**35.11***	4.4×10^{-9}	5.6×10^{-8}
9.	Average duration of each dialogue (in minutes)	6’18	7’13	**5’24***	1.4×10^{-4}	0.02
	Subjective metrics: *“Did you feel that ...*					
10.	*... the robot correctly understood what you said?”*	3.32	2.92	**3.68***	1.3×10^{-4}	0.03
11.	*.. the robot reacted appropriately to your instructions?”*	3.70	3.32	**3.86**	7.6×10^{-3}	0.23
12.	*... the robot asked you to repeat or confirm your instructions ...”*	2.16	2.19	**3.3***	1.7×10^{-9}	4.7×10^{-7}
13.	*... the robot sometimes ignored when you were speaking?”*	3.24	2.76	**3.43**	6.7×10^{-3}	0.21
14.	*... the robot sometimes thought you were talking when you were not?”*	3.43	3.14	**4.41***	3.4×10^{-6}	4.7×10^{-5}
15.	*... the interaction flowed in a pleasant and natural manner?”*	2.97	2.46	**3.32***	8.6×10^{-4}	0.03

Table 8.7: Empirical results obtained for the user evaluation with a total of 37 participants, based on a set of 15 metrics (9 objective and 6 subjective). The $*$ symbol indicates results that outperform the two other approaches with a level of statistical significance $\alpha = 0.05$.