Supplementary Material: "I'm Done": Describing Human Reactions to Successive Robot Failure

SHANNON LIU, MARIA TERESA PARREIRA, and WENDY JU, Cornell University

1 CODEBOOK

Table 1 displays the codes used for annotating the successive error dataset. Annotations were created using labelstud.io [1].

2 STATISTICAL ANALYSIS

We additionally extracted features for the participants' external behavior. Taking as input the HelperBot POV videos of participants (30 fps), we used OpenFace [1] for facial expressions (intensity of activation, 17 features), Opensmile [3] for acoustic features (25 features) and Openpose [2] for body position changes, only taking upper body features since participants were sitting behind a table (24 features). The total number of features is 66, in a total of 14148 data samples from participants reactions to Errors I, II, and III. We conducted a statistical analysis to quantify the relationship between the extracted features and participant reactions to Errors I, II, and III. We list the 39 features deemed statistically significant on Table 2. The code will be provided in the study repository.

REFERENCES

[1] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. 59–66. https://doi.org/10.1109/FG.2018.00019

[2] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

[3] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM International Conference on Multimedia* (Firenze, Italy) *(MM '10)*. Association for Computing Machinery, New York, NY, USA, 1459–1462. https://doi.org/10.1145/1873951.1874246

[1] https://labelstud.io/

Authors' address: Shannon Liu; Maria Teresa Parreira; Wendy JuCornell University.

Table 1. Codebook used for annotation of successive error dataset. Colors represent coding categories: verbal responses, verbal tone changes and emotional displays.

Code	Definition	Examples
prompt	the first dialogue spoken to Nodbot	"I'm done" "I'm finished" "The survey is complete"
more specific / longer prompt	participant's prompt included more words and description than the previous prompt (exclude filler words)	previous prompt: "I'm done"; current prompt: "I've finished the survey" previous prompt: "The survey is finished"; current prompt "I've completed the survey"
swapping terms in a prompt	participant's prompt has same meaning as the previous prompt but some words are substituted with similar meaning words	*previous prompt: "I've finished the survey";* *current prompt: "I've completed the questionnaire"* *previous prompt: "The survey is complete";* *current prompt: "The survey is finished"* *previous prompt: "I'm done";* *current prompt "I'm finished"*
simpler prompt	participant's prompt included less words and/or syllables than the previous prompt	previous prompt: "I've finished the survey"; current prompt: "I've finished" previous prompt: "The survey is submitted"; current prompt: "Done" previous prompt: "I've completed the survey"; current prompt: "survey completed"
slower	participant's prompt was spoken slower or choppier than the previous prompt	previous prompt: "I've finished the survey"; current prompt: "I've… finished…the…survey" previous prompt: "The survey is complete"; current prompt "sur… vey…is…com…plete"
demanding tone	participant's prompt was directed at Nodbot (participant was facing Nodbot) and was spoken louder	
interrogative tone	participant's prompt has rising intonation (ends with a higher pitch)	
filler words	participant uses filler words at the beginning of prompt	"Oh…[prompt]" "Uh…[prompt]" "Um…[prompt]"
looking at tablet	participant is likely checking whether the survey is actually submitted	*overlaps with confusion or frustration*
amusement / humor	smile, chuckle, speaking to Nodbot with humor	"I want to go home" "Don't be like that, be nice"
frustration	frown, pursed lips, scrunched face, clenched jaw, utterance, sigh, eye-rolling, annoyed, glaring, looking away quickly	
confusion	awkward smile (corners of lips pulled to side), head tilt, furrowed or raised eyebrow, darting or widened eyes, looking up and rightward, staring at camera, looking around room	
quitting	participant stops interacting with Nodbot	participant says "okay" and stops interacting or giving prompts to Nodbot

Table 2. Features deemed statistically significant across the 3 error instances, for Kruskal-Wallis non-parametric test, with adjusted p-values using Bonferroni correction. A Dunn's post hoc test was also carried out to investigate differences across error numbers. Adjusted p-values: p <0.05 *, p <0.01 **, p <0.001 ***, N.S. not significant

Feature	Adjusted p-value	Dunn test Errors I-II (adjusted p-value)	Dunn test Errors I-III (adjusted p-value)	Dunn test Errors II-III (adjusted p-value)
AU01_r	3.337E-08	***	***	N.S.
AU02_r	3.534E-08	**	*	***
AU04_r	2.548E-17	N.S.	***	***
AU05_r	0.0003	**	N.S.	***
AU06_r	2.127E-10	N.S.	***	***
AU07_r	1.562E-10	N.S.	***	***
AU09_r	2.783E-61	*	***	***
AU10_r	2.877E-51	***	***	**
AU12_r	5.224E-39	***	***	**
AU14_r	1.442E-36	N.S.	***	***
AU15_r	7.737E-43	N.S.	***	***
AU17_r	2.194E-13	***	***	N.S.
AU20_r	1.958E-20	***	N.S.	***
AU23_r	6.097E-78	N.S.	***	***
AU25_r	5.588E-21	***	N.S.	***
AU26_r	6.268E-07	***	***	N.S.
AU45_r	9.777E-23	***	***	***
Loudness_sma3	1.425E-47	***	***	***
alphaRatio_sma3	6.338E-36	***	***	***
hammarbergIndex_sma3	6.465E-44	***	***	***
slope500-1500_sma3	6.305E-25	***	***	***
spectralFlux_sma3	4.585E-48	***	***	***
mfcc1_sma3	8.151E-30	***	***	***
mfcc3_sma3	4.692E-30	***	***	***
F0semitoneFrom27.5Hz_sma3nz	1.275E-33	**	***	***
jitterLocal_sma3nz	2.096E-29	*	***	***
shimmerLocaldB_sma3nz	8.471E-32	**	***	***
HNRdBACF_sma3nz	1.638E-07	N.S.	***	***
logRelF0-H1-H2_sma3nz	0.034	N.S.	***	N.S.
logRelF0-H1-A3_sma3nz	3.384E-28	**	***	***
F1frequency_sma3nz	7.428E-48	***	***	***
F1bandwidth_sma3nz	3.554E-48	***	***	***
F1amplitudeLogRelF0_sma3nz	1.384E-24	***	***	***
F2frequency_sma3nz	6.924E-49	***	***	***
F2bandwidth_sma3nz	1.440E-48	***	***	***
F2amplitudeLogRelF0_sma3nz	7.263E-41	***	***	***
F3frequency_sma3nz	4.749E-48	***	***	***
F3bandwidth_sma3nz	1.348E-48	***	***	***
F3amplitudeLogRelF0_sma3nz	2.792E-50	***	***	***