

Summary Statistics	
Run ID	cued_speechUniv3
Task	Summarization
Run type	automatic
Summarization model	Supervised neural or deep learning model; Transfer learning or other pre-trained model
Summarization method	Abstractive
Number of summaries	179

Overall measures	
Summaries rated "excellent"	39 (21.8%)
Summaries rated "good"	51 (28.5%)
Summaries rated "fair"	61 (34.1%)
Summaries rated "bad"	28 (15.6%)

ROUGE score compared to episode description	
ROUGE_L P score	0.2202 (95%-conf.int. 0.21088 - 0.23003)
ROUGE_L R score	0.2053 (95%-conf.int. 0.19631 - 0.21437)
ROUGE_L F score	0.1812 (95%-conf.int. 0.17445 - 0.18817)

Qualitative judgments	
Q1: Does the summary include names of the main people (hosts, guests, characters) involved or mentioned in the podcast?	63.7%
Q2: Does the summary give any additional information about the people mentioned (such as their job titles, biographies, personal background, etc)?	44.1%
Q3: Does the summary include the main topic(s) of the podcast?	82.7%
Q4: Does the summary tell you anything about the format of the podcast; e.g. whether it's an interview, whether it's a chat between friends, a monologue, etc?	61.5%
Q5: Does the summary give you more context on the title of the podcast?	74.9%
Q6: Does the summary contain redundant information?	6.7%
Q7: Is the summary written in good English?	88.3%
Q8: Are the start and end of the summary good sentence and paragraph start and end points?	70.4%

