124-27446 AI claim verification v.1.2

25/06-2024

Tore Stegenborg-Andersen, toan@forcetechnology.com

Project overview and summary

- Jabra wishes to claim on speech transcription accuracy for three different headsets; Jabre Engage75, Jabra Evolve2 55 and Jabra Evolve2 65 Flex.

- Claim wording is:
  - "97% AI transcription accuracy"

- A total of 8 headsets were tested (3 Jabra headsets w. boom arm, 5 competitor earbuds w/o boom arm.)

- Jabra's headsets all scored ~50% better than the tested competitor earbuds.

- The claim is substantiated.

Test method – Speech transcription accuracy

o The DUT (Device Under Test) is placed on a B&K HATS 4128C Head and Torso Simulator, with artificial mouth, and connected via Bluetooth to a DUT pc.

- If a usb dongle is included with the device packaging, it is used, otherwise the built-in Bluetooth transmitter on the DUT pc is used.

o A Microsoft Teams call is established between a reference pc and the Pc.

- The measurement is performed with Default (Auto), and High microphone noise suppression settings in Microsoft Teams.

o A speech file is played through the artificial mouth, and the microphone input on the DUT is automatically transcribed by Microsoft Teams.

- A reference transcription is performed on the recorded audio by Azure Cognitive Services Speech SDK to compare and check that similar scores to Microsoft Teams are obtained.
- The transcriptions were not different enough to merit further investigation. The calculated results were within 1-2% of each other, and the ranking of the products was the same. For this reason Microsoft Teams' transcription was chosen, to reflect the user experience as closely as possible.

o The transcription is compared with the original script by calculating a similarity ratio, using the Python library difflib: https://docs.python.org/3/library/difflib.html

o Each product receives a score between 0 and 100 for each background noise x noise reduction combination.

o 4 different background noise conditions are tested:

- No background noise
- Cafeteria noise (ETSI EG 202 396-1 "Mensa")
- Callcenter (ETSI EG 202 396-1 "Work noise call center")
- Distractor (Speech from background noise reproduction system)

• The total score is the average between both noise attenuation settings, and all background noise scenarios.

Included products

Brand	Device name	Firmware version	Serial no.
Apple	Airpods Pro 2'nd gen	6F7	M45X9J6T37
Apple	Airpods 3'rd gen	6A321	W1G416FT7N
Jabra	Engage 75	5.11.0	292676362
Jabra	Evolve 2 55	1.2.13	357530262
Jabra	Evolve 2 65 Flex	2.7.7	286128417
Samsung	Galaxy Buds2 Pro	AXC2	RFAX51CMP6R / RF2X50CDJNM / RFAX51
Sony	WF-1000XM5	3.2.1	1947129-J
Beats	Powerbeats Pro	6B27	H2CM15SOPXLL

Test results (Microsoft Teams transcription)

Device Under Test	Transcription Method	Noise Scenario	No Background Noise	Mensa	Call Center	Voice Distractor	Average (Teams transcription only)		
		Teams Noise suppresion Setting							
AirPods Pro 2'nd gen	Teams Transcription	Default	96,5	37,99	84,61	49,09	67,0	66,4	
		High	96,07	42,18	87,37	37,13	65,7		
AirPods 3'rd gen	Teams Transcription	Default	99,66	55,66	80,07	48,92	71,1	70,8	
		High	99,76	34,73	92,88	54,42	70,4		
GalaxyBuds2 Pro	Teams Transcription	Default	99,59	70,39	83,54	14,84	67,1	68,3	64,1
		High	99,63	68,54	92,92	16,93	69,5		
PowerBeats Pro	Teams Transcription	Default	99,67	34,61	86,33	41,34	65,5	63,7	
		High	99,2	36,56	76,99	35,12	62,0		
Sony WF1000XM5	Teams Transcription	Default	99,38	18,83	92,37	11,53	55,5	51,3	
		High	99,53	7,29	71,21	10,38	47,1		
Jabra Engage 75	Teams Transcription	Default	99,9	99,15	99,63	99,34	99,5	99,6	
		High	99,9	99,38	99,9	99,67	99,7		
Jabra Evolve2 65 Flex	Teams Transcription	Default	99,32	96,19	98,47	71,93	91,5	91,2	96,5
		High	98,53	93,61	97,61	73,69	90,9		
Jabra Evolve2 55	Teams Transcription	Default	99,45	98,28	98,18	99,26	98,8	98,8	
		High	99,55	98,31	98,76	98,61	98,8		

Conclusions

- The three Jabra products measured on average 96,5.

- The five competitor products measured on average 64,1.

- The claim is substatiated, as the Jabra products measured an accuracy of 96,5.

Appendix

Test results (Azure Cognitive Services transcription)

Device Under Test	Transcription Method	Noise Scenario	No Background Noise	Mensa	Call Center	Voice Distractor	Average (Azure Cognitive Services transcription only)		
		Teams Noise suppresion Setting							
AirPods Pro 2'nd gen	Azure Cognitive Services	Default	97,23	36,36	87,58	33,42	63,6	67,0	
		High	97,58	45,60	88,38	49,69	70,3		
AirPods 3'rd gen	Azure Cognitive Services	Default	99,54	52,69	92,72	49,53	73,6	72,9	
		High	98,38	32,55	95,06	62,68	72,2		
GalaxyBuds2Pro	Azure Cognitive Services	Default	99,15	82,13	95,69	24,44	75,4	74,3	67,3
		High	99,69	87,70	91,08	14,71	73,3		
PowerBeats Pro	Azure Cognitive Services	Default	99,85	40,06	73,71	51,65	66,3	68,4	
		High	99,32	44,37	92,73	45,69	70,5		
Sony WF 1000XM5	Azure Cognitive Services	Default	99,21	6,15	89,07	18,16	53,1	53,8	
		High	99,79	9,69	75,04	33,22	54,4		
Jabra Engage 75	Azure Cognitive Services	Default	99,27	99,34	99,92	99,56	99,5	99,6	
		High	99,69	99,63	99,69	99,92	99,7		
Jabra Evolve2 65Flex	Azure Cognitive Services	Default	99,27	96,09	98,12	72,62	91,5	91,4	96,8
		High	99,34	94,32	98,26	73,57	91,4		
Jabra Evolve2 55	Azure Cognitive Services	Default	99,25	98,65	99,46	99,05	99,1	99,2	
		High	99,74	99,21	99,29	98,99	99,3		