STT System Evaluation Summary Report

TC-1: Multilingual Support

Input:

Voice datasets for each language (English-US, English-UK, English-HK, Cantonese-HK, Mandarin), with length ranging from 4s to 10s.
Ground truth transcriptions for each clip.

Output Requirement:

WER (Word Error Rate), WRR (Word Recognition Rate) for each language.

Results:

STT Method: google

Language	Average WER (%)	Average WRR (%)
Cantonese-HK	59.42	40.58
English-US	13.32	86.68
Mandarin	66.25	55.62

TC-2: Robustness Across Accents

Input:

Voice datasets for each language (Cantonese-HK with Mandarin accent, Mandarin with Cantonese accent, English with Southeast Asian accent, English with Indian accent), with length ranging from 4s to 10s.
Ground truth transcriptions for each clip.

Output Requirement:

WER (Word Error Rate), WRR (Word Recognition Rate) and for each language.

Results:

STT Method: google

Accent/Language	Average WER (%)	Average WRR (%)
Cantonese_Mandarin_Accent	100.00	0.00
English_Indian_Accent	13.13	86.87
English_SouthEastAsian_Accent	31.05	79.53
Mandarin_Cantonese_Accent	100.00	0.00

TC-3: Domain Vocabulary Support

Input:

Voice datasets for each language (English, Cantonese-HK), with length ranging from 4s to 10s, where HSBC specific terms are mentioned.
Ground truth transcriptions for each clip.

Output Requirement:

WER (Word Error Rate), and full vocab recognition for each language.

Results:

STT Method: google

Language	Average WER (%)	Average Vocabulary Accuracy (%)
Cantonese	69.75	36.53
English	10.95	93.33
Mandarin	57.37	66.67

TC-4: Auto Punctuation Feature

Input:

Voice datasets for each language (English-US, English-UK, English-HK, Cantonese-HK, Mandarin), with length ranging from 4s to 10s, and clear punctuation syntax (periods, commas, question marks). Ground truth transcriptions for each clip.

Output Requirement:

Proportion of correct punctuation placements for each language.

Results:

STT Method: google

Language	Average Segmentation Accuracy (%)
Cantonese	2.86
Cantonese-HK	0.00
Cantonese-HK-Numbers	0.00
Cantonese-HK-Numbers\noisy_100	4.76
Cantonese-HK-Numbers\noisy_25	0.00
Cantonese-HK-Numbers\noisy_50	0.00
Cantonese-HK-Numbers\noisy_75	12.50
Cantonese-HK\noisy_100	7.29
Cantonese-HK\noisy_25	0.00
Cantonese-HK\noisy_50	6.25
Cantonese-HK\noisy_75	7.41
Cantonese_Mandarin_Accent	0.00
English-US	100.00
English-US-Numbers	100.00

Language	Average Segmentation Accuracy (%)
English-US-Numbers\noisy_100	0.00
English-US-Numbers\noisy_25	100.00
English-US-Numbers\noisy_50	33.33
English-US-Numbers\noisy_75	50.00
English-US\noisy_100	0.00
English-US\noisy_25	0.00
English-US\noisy_50	50.00
English-US\noisy_75	50.00
English_SouthEastAsian_Accent	45.31
Mandarin	30.00
Mandarin-Numbers	32.05
Mandarin-Numbers\noisy_100	0.00
Mandarin-Numbers\noisy_25	0.00
Mandarin-Numbers\noisy_50	0.00
Mandarin-Numbers\noisy_75	12.50
Mandarin\noisy_100	5.56
Mandarin\noisy_25	10.00
Mandarin\noisy_50	0.00
Mandarin\noisy_75	0.00
Mandarin_Cantonese_Accent	0.00

TC-5: Profanity Filtering

Input:

Voice datasets for each language (English-US, English-UK, English-HK, Cantonese-HK, Mandarin), with length ranging from 4s to 10s, containing profanity vocabulary.
Ground truth transcriptions for each clip.

Output Requirement:

Rate of profanity vocabulary identified for each language.

Results:

STT Method: google

No valid data to aggregate for this STT method after filtering.

TC-6: Transcription Speed and Latency

Input:

Audio clips of lengths: 5-10 seconds.

Output Requirement:

Actual latency in seconds vs system-reported latency.

Results:

STT Method: google

Metric	Value
Average Actual Latency (s)	1.930
System-Reported Latency (s)	Data not available in source CSV

TC-7: Noise Robustness

Input:

Voice datasets for each language (English-US, English-UK, English-HK, Cantonese-HK, Mandarin), with length ranging from 5s to 10s, mixed with various environment noise at different SNR levels.

Output Requirement:

WER (Word Error Rate), WRR (Word Recognition Rate).

Results:

STT Method: google

Language/Condition	Noise Level (%)	Average WER (%)	Average WRR (%)
Cantonese-HK-Numbers\noisy_100	100%	106.25	0.00
Cantonese-HK-Numbers\noisy_25	25%	89.66	10.34
Cantonese-HK-Numbers\noisy_50	50%	96.83	3.90
Cantonese-HK-Numbers\noisy_75	75%	105.75	0.50
Cantonese-HK\noisy_100	100%	98.34	1.66
Cantonese-HK\noisy_25	25%	90.79	12.38
Cantonese-HK\noisy_50	50%	104.33	4.38
Cantonese-HK\noisy_75	75%	98.42	4.75
English-US-Numbers\noisy_100	100%	70.77	29.23
English-US-Numbers\noisy_25	25%	47.12	52.88
English-US-Numbers\noisy_50	50%	71.58	28.42
English-US-Numbers\noisy_75	75%	76.34	23.66
English-US\noisy_100	100%	81.56	18.44

Language/Condition	Noise Level (%)	Average WER (%)	Average WRR (%)
English-US\noisy_25	25%	50.08	49.92
English-US\noisy_50	50%	78.04	21.96
English-US\noisy_75	75%	75.15	24.85
Mandarin-Numbers\noisy_100	100%	97.82	2.18
Mandarin-Numbers\noisy_25	25%	60.91	39.09
Mandarin-Numbers\noisy_50	50%	77.99	22.01
Mandarin-Numbers\noisy_75	75%	94.81	5.19
Mandarin\noisy_100	100%	85.67	14.33
Mandarin\noisy_25	25%	53.20	46.80
Mandarin\noisy_50	50%	78.23	21.77
Mandarin\noisy_75	75%	85.24	14.76