STQ AURORA DSR WORKING GROUP

Title: **Speech recognition performance comparison between AMR speech coding & the Advanced DSR Front-End (ETSI ES 202 050)**

Source: **Motorola**

Date: **25 March 2002**

Version: **2.0**

Introduction

Previous studies have shown the performance advantages of DSR (ES 201 108) [1] compared to the voice channel for 2G mobile networks [2,3]. For speech services in 3G networks it is of interest to quantify the recognition performance advantages of a DSR implementation compared to using the voice channel. The speech codec chosen by 3GPP is the Adaptive Multi-Rate (AMR) speech codec having 8 modes with bit rates between 4.75 and 12.2 kbps before error protection.

In this report the performance of DSR (Advanced Front End) is compared to that obtained with AMR encoded speech when using the same recognition scheme. Both the maximum (12.2kbps) and minimum (4.75kbps) bit rates available for AMR are compared when using the Advanced DSR front-end (ETSI ES 202 050) [4] (4.8kbps) and the Aurora reference back-end recogniser configurations. This like versus like comparison investigates the effect that the AMR speech codec has on performance when compared to the DSR architecture where the standardised front-end is placed at the terminal.

The ETSI STQ-Aurora databases have been used for these performance evaluations. The Aurora-2 database is connected digits recognition task with artificially added noises typical of mobile environments. The Aurora 3 databases are also for a connected digit recognition task covering 5 languages and were recorded in a vehicle. The difference in performance between DSR and AMR has been measured in terms of the increase in word error rate.

The results show that speech recognition performance is substantially better with DSR than AMR. The performance with AMR 4.75 is 50% worse than DSR on Aurora 2 and 52% worse on Aurora 3. Even with the higher data rate of AMR 12.2, AMR is 17% worse than DSR on Aurora 2 and 20% worse on Aurora 3.

Evaluation databases

Performance comparisons have been made using the small vocabulary evaluation databases developed in ETSI STQ-Aurora. These are briefly described below for anyone not already familiar with these evaluation databases.

Aurora 2: Noisy TI Digits – small vocabulary evaluation

The original high quality TIDigits database has been prepared by downsampling to 8kHz, filtering with G712 (which has frequency response representative of GSM terminal

characteristics) and the controlled addition of noise to cover a range of signal to noise ratios (clean, 20,15,10,5,0,-5dB) and 8 different noise conditions. The database consists of connected digit sequences for American English talkers and clean and multi-condition training sets are defined. A full description of the database and the test framework is given in reference [5].

There are 3 test sets; set A contains noises seen in the multi-condition training data, set B contains noises that have not been seen in the training data and set C uses M-IRS filtering and noise addition to test the combination of convolutional distortion and noise.

Aurora 3: Multilingual SpeechDat-Car Digits – small vocabulary evaluation

The purpose of these tests is to evaluate the performance of the front-end on a database that has been collected from speakers in a noisy environment. It tests the performance of the front-end with well matched training and testing as well as its performance in mismatched conditions as are likely to be encountered in deployed DSR systems. It also serves to test the front-end on a variety of languages: Finnish, Italian, Spanish, German, and Danish. It is a small vocabulary task consisting of the digits selected from a larger database collection called SpeechDat-Car. These experiments will be performed at 8kHz sampling rate. See references [6-10] for descriptions of these databases. The databases each have 3 experiments consisting of training and test sets to measure performance with:

A) Well matched training and testing - Train & test with the hands-free microphone over the range of vehicle speeds so that the training and test sets cover similar range of noise conditions.
B) Moderate mismatch training and testing - Train on only of a subset of the range of noises present in the test set. For example, hands-free microphone for lower speed driving conditions for training and hands free microphone at higher vehicle speeds for testing.
C) High mismatch training and testing - Model training with speech from close talking microphone. Hands-free microphone at range of vehicle speeds for testing.

Configuration of Simulations

For the DSR experiment the raw speech was processed using the Advanced Front-End and coder, as specified in the DSR standard for WI008 [4]. The features were then transmitted over a clean channel at 4.8kbps and the bitstream decoded. Finally the server side processing was carried out and the data endpointed. This process is shown in figure 1.

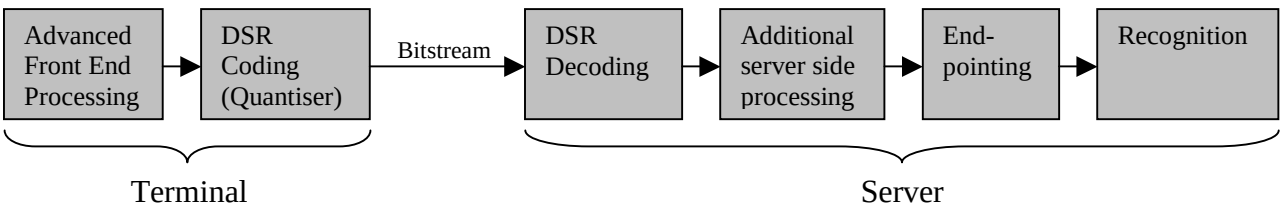

Figure 1: Processing chain for DSR

For the AMR experiments the speech data was transmitted over a clean channel using the AMR codec at 12.2 and 4.75 kbps, then processed on the same front-end (note that in this

case the feature quantisation is not applied). The server side processing was then carried out and the endpointing applied. This is shown in figure 2.

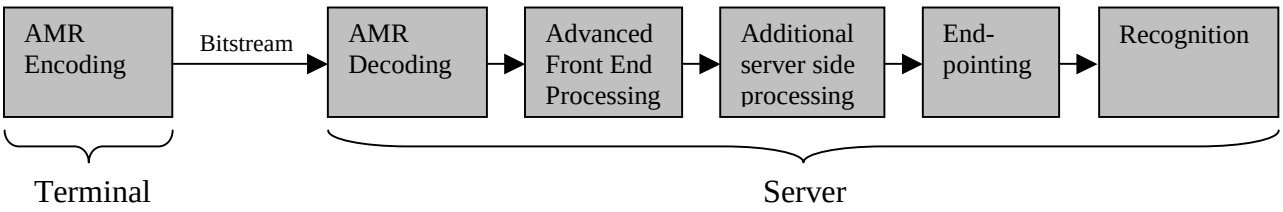

Figure 2: Processing Chain for AMR

For the purposes of obtaining the original baseline results, all of the files in the Aurora 2 and SpeechDat-Car databases were end-pointed to segment the files and a gap of 200ms was left at each end [11]. For the SpeechDat-Car databases these endpoints significantly improve performance because the data files have substantial periods of background noise before and after each utterance. As a result insertion rates were higher before applying the endpoints. For Aurora 2 the performance was not changed significantly, since the duration of silence before and after each utterance is fairly short and consistent. This endpointing was used for both the DSR and AMR simulations to prevent any inconsistency in the results due to different VAD algorithms, i.e. the same endpoints were applied to the files processed using DSR and AMR.

Results

Results are presented separately for Aurora 2 and Aurora 3. For each database the DSR performance is given, followed by the performances of AMR at each of the two coding rates. The relative performance of the AMR coding schemes as compared to the DSR performance is also given.

Aurora 2

In this section the summary results for the Aurora 2 database are given; the detailed results, which give a breakdown of the performance for different SNR's, can be found in the appendix. The results in table 1 show the absolute performance for DSR using the Advanced Front End and the rest of the processing chain as shown in figure 1. These results are then used as the baseline for later performance comparisons.

Aurora 2, Trained and tested on DSR at 4.8kbps				
Absolute performance				
Training Mode	**Set A**	**Set B**	**Set C**	**Overall**
Multicondition	92.20%	91.88%	90.68%	91.77%
Clean Only	88.15%	87.29%	86.07%	87.39%
Average	90.18%	89.58%	88.38%	**89.58%**

Table 1

The top half of table 2 shows the absolute performance that is obtained when the speech is transmitted using an AMR codec operating at 12.2kbps (maximum bit rate), before being processed by the Advanced Front End. This is the processing scheme that was shown in figure 2. The bottom half of the table shows the relative performance of this scheme when compared to the DSR baseline that was given in table 1.

Aurora 2, Trained and tested on AMR at 12.2kbps				
Absolute performance				
Training Mode	**Set A**	**Set B**	**Set C**	**Overall**
Multicondition	91.61%	91.13%	90.03%	91.10%
Clean Only	85.16%	83.75%	85.58%	84.68%
Average	88.39%	87.44%	87.80%	**87.89%**
Performance relative to DSR				
Training Mode	**Set A**	**Set B**	**Set C**	**Overall**
Multicondition	-10.04%	-13.73%	-11.07%	-11.72%
Clean Only	-23.52%	-31.28%	-1.04%	-22.12%
Average	-16.78%	-22.51%	-6.05%	**-16.92%**

Table 2

These results show that there is an average performance degradation of 16.9% when using an AMR codec instead of DSR, even though the AMR bit rate is much higher than that used for DSR. This increases to an average performance degradation of 22.1% when only clean training is considered. For multicondition training the results for testing with clean speech are worse than for noisy speech, i.e. the relative performance degradation increases with the SNR. The relative degradations for clean speech and speech at 20dB SNR are 37.3% and 17.7% respectively (see the appendix for the complete detailed results).

An equivalent set of results is presented in table 3 for the AMR codec operating at 4.75kbps (minimum bit rate).

Aurora 2, Trained and tested on AMR at 4.75kbps				
Absolute performance				
Training Mode	**Set A**	**Set B**	**Set C**	**Overall**
Multicondition	90.23%	89.41%	88.19%	89.49%
Clean Only	81.97%	80.17%	82.49%	81.35%
Average	86.10%	84.79%	85.34%	**85.42%**
Performance relative to DSR				
Training Mode	**Set A**	**Set B**	**Set C**	**Overall**
Multicondition	-35.88%	-39.33%	-33.83%	-36.85%
Clean Only	-62.99%	-77.15%	-31.39%	-62.33%
Average	-49.43%	-58.24%	-32.61%	**-49.59%**

Table 3

The performance benefits of DSR can be clearly seen from these results. When using an AMR codec at a bit rate of 4.75kbps, which is equivalent to the bit rate used for DSR (4.8kbps), the average performance degradation increases to 49.6%. That is to say, the

recognition performance when using DSR is significantly better than that obtained when the speech is first transmitted through an AMR codec operating at its lowest bit rate. As before the degradation increases with SNR, so that the relative degradation for clean speech is 72.8%, and 50.6% at 20dB SNR (see appendix).

Aurora 3

In this next section equivalent results are presented for the first three experiments in the previous section, but for the five languages making up the Aurora 3 database. The row(s) in each table labelled "0.4W+0.35M+0.25H" represent the weighted average of the well matched, medium mismatch and high mismatch results as described in [12].

Following the pattern used in the previous section, table 4 shows the absolute performance for DSR using the Advanced Front End, which then serves as a baseline for the performance comparisons.

SpeechDat-Car, Trained and tested on DSR at 4.8kbps						
Absolute performance						
Training Mode	Italian	Finnish	Spanish	German	Danish	Average
Well Matched	96.45%	95.19%	96.13%	95.51%	93.00%	95.26%
Medium Mismatch	90.41%	85.98%	92.78%	89.90%	82.94%	88.40%
High Mismatch	88.64%	86.11%	88.12%	91.63%	80.26%	86.95%
0.4W+0.35M+0.25H	92.38%	89.70%	92.96%	92.58%	86.29%	**90.78%**

Table 4

The results obtained when the speech was transmitted at 12.2 kbps using the AMR codec are presented in table 5. As before, the absolute results are given in the top half of the table, while the relative performance, when compared to the DSR baseline, is shown in the bottom half.

SpeechDat-Car, Trained and tested on AMR at 12.2kbps						
Absolute performance						
Training Mode	Italian	Finnish	Spanish	German	Danish	Average
Well Matched	96.04%	95.43%	95.56%	93.89%	92.20%	94.62%
Medium Mismatch	88.17%	84.82%	90.76%	86.68%	78.12%	85.71%
High Mismatch	84.23%	84.95%	85.20%	89.59%	74.37%	83.67%
0.4W+0.35M+0.25H	90.33%	89.10%	91.29%	90.29%	82.81%	**88.77%**

Performance relative to DSR						
Training Mode	Italian	Finnish	Spanish	German	Danish	Average
Well Matched	-11.55%	4.99%	-14.73%	-36.08%	-11.43%	-13.76%
Medium Mismatch	-23.36%	-8.27%	-27.98%	-31.88%	-28.25%	-23.95%
High Mismatch	-38.82%	-8.35%	-24.58%	-24.37%	-29.84%	-25.19%
0.4W+0.35M+0.25H	-22.50%	-2.99%	-21.83%	-31.68%	-21.92%	**-20.18%**

Table 5

The average performance degradation when using 12.2 kbps AMR compared to DSR is 20.2% which is similar to the relative degradation obtained on Aurora 2.

The final set of results, in table 6, is for speech transmitted using the AMR codec at 4.75kbps.

SpeechDat-Car, Trained and tested on AMR at 4.75kbps						
Absolute performance						
Training Mode	Italian	Finnish	Spanish	German	Danish	Average
Well Matched	94.40%	94.65%	94.65%	93.09%	89.54%	93.27%
Medium Mismatch	83.90%	81.74%	88.93%	85.51%	70.70%	82.16%
High Mismatch	76.85%	80.92%	81.02%	85.62%	68.12%	78.51%
0.4W+0.35M+0.25H	86.34%	86.70%	89.24%	88.57%	77.59%	85.69%

Performance relative to DSR						
Training Mode	Italian	Finnish	Spanish	German	Danish	Average
Well Matched	-57.75%	-11.23%	-38.24%	-53.90%	-49.43%	-42.11%
Medium Mismatch	-67.88%	-30.24%	-53.32%	-43.47%	-71.75%	-53.33%
High Mismatch	-103.79%	-37.37%	-59.76%	-71.80%	-61.50%	-66.84%
0.4W+0.35M+0.25H	-72.80%	-24.42%	-48.90%	-54.72%	-60.26%	-52.22%

Table 6

Again the Aurora 3 results are similar to those for Aurora 2. The average relative degradation for AMR as compared to DSR is 52.2%. So the recognition performance when using DSR is significantly better than that obtained when the speech is first transmitted through an AMR codec operating at a low bit rate.

Conclusion

This report investigated a like versus like comparison between using an AMR codec to transmit speech prior to sever-side only recognition, and the DSR architecture, where the standardised front-end is placed in the terminal. The comparison was carried out using the Advanced Front End and the Aurora reference back-end recogniser for both the maximum and minimum bit rate of AMR.

The results show that speech recognition performance is substantially better with DSR than AMR. The performance with the minimum bit rate of AMR (4.75 kbps) is 50% worse than DSR (4.8kbps) on Aurora 2 and 52% worse on Aurora 3 [average degradation over the two databases 51%]. Even with the much higher maximum bit rate of AMR (12.2kbps), AMR is 17% worse than DSR on Aurora 2 and 20% worse on Aurora 3 [average degradation over the two databases 19%].

Coder	A2	A3	Average
AMR 12.2	17%	20%	19%
AMR 4.75	50%	52%	51%

Table 7 Summary of Results (Average relative degradation compared to DSR)

The results on the Aurora 2 database show a trend of increasing degradation from AMR as the SNR increases. For clean testing (with multicondition training) the degradation is 37.3 % for AMR 12.2 and 72.8 % for AMR 4.75.

In conclusion, using the DSR architecture gives substantial gains in recognition performance compared to using speech that has been transcoded with an AMR codec.

References

[1] ETSI standard document, "Speech Processing, Transmission and Quality aspects (STQ); Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithm", ETSI ES 201 108 v1.1.2 (2000-04), April 2000.

[2] Petri Haavisto, "Speech recognition for mobile communications" in Proc. IEEE workshop on Robust Methods for Speech Recognition in Adverse Conditions pp 15-18, Tampere, Finland, May 1999.

[3] Imre Kiss, "A comparison of Distributed and Network Speech Recognition for Mobile Communication systems", ICSLP 2000, Beijing, China, Sept 2000.

[4] ETSI draft standard document, "Speech Processing, Transmission and Quality aspects (STQ); Distributed speech recognition; Advanced Front-end feature extraction algorithm; Compression algorithm", ETSI ES 202 050 v0.1.0 (2002-04), April 2002. (in preparation)

[5] H G Hirsch & D Pearce, "The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions", ISCA ITRW ASR2000 "Automatic Speech Recognition: Challenges for the Next Millennium"; Paris, France, September 18-20, 2000

[6] Aurora document no. AU/225/00 "Baseline Results for subset of SpeechDat-Car Finnish Database for ETSI STQ WI008 Advanced Front-end Evaluation", Nokia, Jan 2000

[7] Aurora document no. AU/237/00 "Description and baseline results for the SpeechDat-Car Italian Database", Alcatel, April 2000

[8] Aurora document no. AU/271/00 "Spanish SDC-Aurora Database for ETSI STQ Aurora WI008 Advanced DSR Front-End Evaluation: Description and Baseline Results", UPC, Nov 2000

[9] Aurora document no. AU/273/00 "Description and Baseline Results for the Subset of the Speechdat-Car German Database used for ETSI STQ Aurora WI008 Advanced DSR Front-end Evaluation", Texas Instruments, Dec 2001

[10] Aurora document no. AU/378/01 "Danish SpeechDat-Car Digits Database for ETSI STQ-Aurora Advanced DSR", Aalborg University, Jan 2001.

[11] Aurora document no. AU/374/01 "Small vocabulary evaluations: Baseline Mel-Cepstrum Performances with Speech Endpoints: version 2", Oct 2001

[12] Aurora document no. AU/371/01 "Advanced DSR Front-end: Definition of required performance characteristics, Version 4", Dec 2001

Appendix – Detailed Aurora 2 Results

DSR Absolute Results - Multicondition Training														
	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	99.14	99.09	99.22	99.41	99.22	99.14	99.09	99.22	99.41	99.22	99.17	99.12	99.15	99.20
20 dB	98.53	98.34	98.69	98.55	98.53	98.68	98.07	99.02	99.07	98.71	98.53	98.37	98.45	98.59
15 dB	97.61	98	98.45	98.06	98.03	97.73	97.4	98.12	97.99	97.81	97.54	97.64	97.59	97.85
10 dB	95.49	95.53	97.02	96.17	96.05	95.49	96.28	96.57	96.61	96.24	95.95	95.92	95.94	96.10
5 dB	91.19	90.78	93.56	90.9	91.61	88.76	90.48	92.57	91.27	90.77	90.42	89.33	89.88	90.93
0 dB	77	70.47	82.82	76.92	76.80	70.68	77.06	78.26	77.45	75.86	72.03	71.07	71.55	75.38
-5dB	45.26	32.77	48.14	48.66	43.71	34.14	43.74	46.85	49.21	43.49	37.33	36.43	36.88	42.25
Average	91.96	90.62	94.11	92.12	**92.20**	90.27	91.86	92.91	92.48	**91.88**	90.89	90.47	**90.68**	91.77

DSR Absolute Results - Clean Training														
	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	99.29	99.21	99.31	99.54	99.34	99.29	99.21	99.31	99.54	99.34	99.32	99.24	99.28	99.33
20 dB	98.04	98.46	98.78	98.33	98.40	98.43	97.88	98.72	98.61	98.41	97.64	97.88	97.76	98.28
15 dB	96.9	97.07	98.18	96.98	97.28	95.82	96.67	97.76	97.1	96.84	96.19	96.7	96.45	96.94
10 dB	92.63	91.87	96.27	94.01	93.70	91.5	92.74	94.96	94.82	93.51	92.23	92.44	92.34	93.35
5 dB	85.6	81.8	91.05	86.52	86.24	79.37	85.01	86.43	86.83	84.41	83.11	83.25	83.18	84.90
0 dB	66.41	53.63	73.64	66.8	65.12	54.19	65.48	64.06	69.36	63.27	60.79	60.49	60.64	63.49
-5dB	34.97	20.89	36.62	37.55	32.51	21.43	33.56	32.03	36.69	30.93	30.61	30.8	30.71	31.52
Average	87.92	84.57	91.58	88.53	**88.15**	83.86	87.56	88.39	89.34	**87.29**	85.99	86.15	**86.07**	87.39

AMR 122 Absolute Results - Multicondition Training														
	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	98.86	98.7	98.9	99.11	98.89	98.86	98.7	98.9	99.11	98.89	99.08	98.88	98.98	98.91
20 dB	98.34	98.34	98.72	98.21	98.40	98.56	97.76	98.57	98.77	98.42	98.16	98.13	98.15	98.36
15 dB	97.05	97.55	98.27	97.62	97.62	97.24	97.13	97.97	97.78	97.53	97.42	97.46	97.44	97.55
10 dB	94.9	95.59	96.9	95.65	95.76	95.06	95.37	96.03	96.39	95.71	95.21	95.31	95.26	95.64
5 dB	90.3	90.45	93.2	90.16	91.03	88.52	90.02	91.2	90.77	90.13	89.71	88.45	89.08	90.28
0 dB	76.33	70.07	80.73	73.84	75.24	68.53	74.43	77.04	75.38	73.85	69.82	70.62	70.22	73.68
-5dB	42.4	30.83	42.71	41.72	39.42	31.93	40.39	44.11	42.55	39.75	31.96	35.28	33.62	38.39
Average	91.38	90.40	93.56	91.10	**91.61**	89.58	90.94	92.16	91.82	**91.13**	90.06	89.99	**90.03**	91.10

Multicondition Training - Relative Performance														
	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	-32.56%	-42.86%	-41.03%	-50.85%	-41.82%	-32.56%	-42.86%	-41.03%	-50.85%	-41.82%	-10.84%	-27.27%	-19.06%	-37.27%
20 dB	-12.93%	0.00%	2.29%	-23.45%	-8.52%	-9.09%	-16.06%	-45.92%	-32.26%	-25.83%	-25.17%	-14.72%	-19.95%	-17.73%
15 dB	-23.43%	-22.50%	-11.61%	-22.68%	-20.06%	-21.59%	-10.38%	-7.98%	-10.45%	-12.60%	-4.88%	-7.63%	-6.25%	-14.31%
10 dB	-13.08%	1.34%	-4.03%	-13.58%	-7.34%	-9.53%	-24.46%	-15.74%	-6.49%	-14.06%	-18.27%	-14.95%	-16.61%	-11.88%
5 dB	-10.10%	-3.58%	-5.59%	-8.13%	-6.85%	-2.14%	-4.83%	-18.44%	-5.73%	-7.78%	-7.41%	-8.25%	-7.83%	-7.42%
0 dB	-2.91%	-1.35%	-12.17%	-13.34%	-7.44%	-7.33%	-11.46%	-5.61%	-9.18%	-8.40%	-7.90%	-1.56%	-4.73%	-7.28%
-5dB	-5.22%	-2.89%	-10.47%	-13.52%	-8.02%	-3.36%	-5.95%	-5.16%	-13.11%	-6.89%	-8.57%	-1.81%	-5.19%	-7.01%
Average	-12.49%	-5.22%	-6.22%	-16.24%	**-10.04%**	-9.94%	-13.44%	-18.74%	-12.82%	**-13.73%**	-12.73%	-9.42%	**-11.07%**	-11.72%

AMR 122 Absolute Results - Clean Training

	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	99.26	98.82	99.02	99.38	99.12	99.26	98.82	99.02	99.38	99.12	99.29	99.06	99.18	**99.13**
20 dB	97.45	97.94	98.9	98.12	98.10	97.61	97.97	98.24	97.9	97.93	97.57	98.07	97.82	**97.98**
15 dB	95.7	96.04	97.91	96.05	96.43	94.32	96.7	96.66	96.51	96.05	96.28	96.98	96.63	**96.32**
10 dB	91.5	90.66	95.68	92.66	92.63	87.84	92.26	91.59	93.03	91.18	92.05	92.02	92.04	**91.93**
5 dB	79.92	75.88	89.11	83.65	82.14	71.2	82.74	79.15	83.09	79.05	82.04	82.26	82.15	**80.90**
0 dB	57.02	43.35	65.79	59.89	56.51	45.07	59.25	54.1	59.86	54.57	58.37	60.13	59.25	**56.28**
-5dB	24.93	11.67	25.29	25.05	21.74	13.88	26.33	21	23.42	21.16	27.82	30.47	29.15	**22.99**
Average	84.32	80.77	89.48	86.07	**85.16**	79.21	85.78	83.95	86.08	**83.75**	85.26	85.89	**85.58**	**84.68**

Clean Training - Relative Performance

	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	-4.23%	-49.37%	-42.03%	-34.78%	-32.60%	-4.23%	-49.37%	-42.03%	-34.78%	-32.60%	-4.41%	-23.68%	-14.05%	**-28.89%**
20 dB	-30.10%	-33.77%	9.84%	-12.57%	-16.65%	-52.23%	4.25%	-37.50%	-51.08%	-34.14%	-2.97%	8.96%	3.00%	**-19.72%**
15 dB	-38.71%	-35.15%	-14.84%	-30.79%	-29.87%	-35.89%	0.90%	-49.11%	-20.34%	-26.11%	2.36%	8.48%	5.42%	**-21.31%**
10 dB	-15.33%	-14.88%	-15.82%	-22.54%	-17.14%	-43.06%	-6.61%	-66.87%	-34.56%	-37.77%	-2.32%	-5.56%	-3.94%	**-22.75%**
5 dB	-39.44%	-32.53%	-21.68%	-21.29%	-28.73%	-39.60%	-15.14%	-53.65%	-28.40%	-34.20%	-6.34%	-5.91%	-6.12%	**-26.40%**
0 dB	-27.95%	-22.17%	-29.78%	-20.81%	-25.18%	-19.91%	-18.05%	-27.71%	-31.01%	-24.17%	-6.17%	-0.91%	-3.54%	**-20.45%**
-5dB	-15.44%	-11.65%	-17.88%	-20.02%	-16.25%	-9.61%	-10.88%	-16.23%	-20.96%	-14.42%	-4.02%	-0.48%	-2.25%	**-12.72%**
Average	-30.31%	-27.70%	-14.45%	-21.60%	**-23.52%**	-38.14%	-6.93%	-46.97%	-33.08%	**-31.28%**	-3.09%	1.01%	-1.04%	**-22.12%**

AMR 475 Absolute Results - Multicondition Training

	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	98.77	98.4	98.48	98.77	98.61	98.77	98.4	98.48	98.77	98.61	98.93	98.16	98.55	**98.59**
20 dB	97.51	97.67	98.03	97.93	97.79	98.37	97.85	98.06	98.36	98.16	97.7	97.76	97.73	**97.92**
15 dB	97.18	97.1	97.49	96.82	97.15	97.3	96.55	96.87	97.44	97.04	97.33	97.04	97.19	**97.11**
10 dB	94.53	94.71	96.03	94.08	94.84	94.14	94.59	94.48	94.66	94.47	93.89	93.77	93.83	**94.49**
5 dB	89.68	88.3	91.35	87.81	89.29	85.72	88.42	88.52	88.71	87.84	86.52	86.79	86.66	**88.18**
0 dB	72.89	65.96	77.01	72.51	72.09	63.52	72.61	71.7	70.32	69.54	65.06	66.02	65.54	**69.76**
-5dB	40.8	27.78	37.67	40.94	36.80	26.5	39.39	36.15	36.29	34.58	28	30.89	29.45	**34.44**
Average	90.36	88.75	91.98	89.83	**90.23**	87.81	90.00	89.93	89.90	**89.41**	88.10	88.28	**88.19**	**89.49**

Multicondition Training - Relative Performance

	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	-43.02%	-75.82%	-94.87%	-108.47%	-80.55%	-43.02%	-75.82%	-94.87%	-108.47%	-80.55%	-28.92%	-109.09%	-69.00%	**-78.24%**
20 dB	-69.39%	-40.36%	-50.38%	-42.76%	-50.72%	-23.48%	-11.40%	-97.96%	-76.34%	-52.30%	-56.46%	-37.42%	-46.94%	**-50.60%**
15 dB	-17.99%	-45.00%	-61.94%	-63.92%	-47.21%	-18.94%	-32.69%	-66.49%	-27.36%	-36.37%	-8.54%	-25.42%	-16.98%	**-36.83%**
10 dB	-21.29%	-18.34%	-33.22%	-54.57%	-31.86%	-29.93%	-45.43%	-60.93%	-57.52%	-48.45%	-50.86%	-52.70%	-51.78%	**-42.48%**
5 dB	-17.14%	-26.90%	-34.32%	-33.96%	-28.08%	-27.05%	-21.64%	-54.51%	-29.32%	-33.13%	-40.71%	-23.81%	-32.26%	**-30.93%**
0 dB	-17.87%	-15.27%	-33.82%	-19.11%	-21.52%	-24.42%	-19.40%	-30.17%	-31.62%	-26.40%	-24.92%	-17.46%	-21.19%	**-23.41%**
-5dB	-8.15%	-7.42%	-20.19%	-15.04%	-12.70%	-11.60%	-7.73%	-20.13%	-25.44%	-16.23%	-14.89%	-8.71%	-11.80%	**-13.93%**
Average	-28.73%	-29.18%	-42.73%	-42.86%	**-35.88%**	-24.77%	-26.11%	-62.01%	-44.43%	**-39.33%**	-36.30%	-31.36%	**-33.83%**	**-36.85%**

AMR 475 Absolute Results - Clean Training

	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	98.96	98.85	98.72	99.17	98.93	98.96	98.85	98.72	99.17	98.93	99.08	98.7	98.89	**98.92**
20 dB	96.87	97.28	98.12	97.41	97.42	96.62	97.4	97.17	96.88	97.02	97.18	96.98	97.08	**97.19**
15 dB	95.06	94.14	96.84	94.23	95.07	92.14	95.62	94.42	95.09	94.32	95.03	94.98	95.01	**94.76**
10 dB	88.67	87.24	93.89	89.63	89.86	83.39	90.3	87.41	89.85	87.74	90.08	89.6	89.84	**89.01**
5 dB	76.11	69.62	84.67	77.85	77.06	64.75	78.48	74.02	77.41	73.67	77.43	78.23	77.83	**75.86**
0 dB	49.19	38.63	59.59	54.4	50.45	38.26	55.74	47.42	50.97	48.10	50.6	54.75	52.68	**49.96**
-5dB	17.26	9.16	22.37	20.98	17.44	11.39	24.12	16.19	19.04	17.69	19.99	25.03	22.51	**18.55**
Average	81.18	77.38	86.62	82.70	**81.97**	75.03	83.51	80.09	82.04	**80.17**	82.06	82.91	**82.49**	**81.35**

Clean Training - Relative Performance

	A					B					C			
	Subway	Babble	Car	Exhibition	Average	Restaurant	Street	Airport	Station	Average	Subway M	Street M	Average	Average
Clean	-46.48%	-45.57%	-85.51%	-80.43%	-64.50%	-46.48%	-45.57%	-85.51%	-80.43%	-64.50%	-35.29%	-71.05%	-53.17%	**-62.23%**
20 dB	-59.69%	-76.62%	-54.10%	-55.09%	-61.38%	-115.29%	-22.64%	-121.09%	-124.46%	-95.87%	-19.49%	-42.45%	-30.97%	**-69.09%**
15 dB	-59.35%	-100.00%	-73.63%	-91.06%	-81.01%	-88.04%	-31.53%	-149.11%	-69.31%	-84.50%	-30.45%	-52.12%	-41.28%	**-74.46%**
10 dB	-53.73%	-56.95%	-63.81%	-73.12%	-61.90%	-95.41%	-33.61%	-149.80%	-95.95%	-93.69%	-27.67%	-37.57%	-32.62%	**-68.76%**
5 dB	-65.90%	-66.92%	-71.28%	-64.32%	-67.11%	-70.87%	-43.56%	-91.45%	-71.53%	-69.35%	-33.63%	-29.97%	-31.80%	**-60.94%**
0 dB	-51.27%	-32.35%	-53.30%	-37.35%	-43.57%	-34.77%	-28.22%	-46.30%	-60.02%	-42.33%	-25.99%	-14.53%	-20.26%	**-38.41%**
-5dB	-27.23%	-14.83%	-22.48%	-26.53%	-22.77%	-12.78%	-14.21%	-23.30%	-27.88%	-19.54%	-15.30%	-8.34%	-11.82%	**-19.29%**
Average	-57.99%	-66.57%	-63.22%	-64.19%	**-62.99%**	-80.88%	-31.91%	-111.55%	-84.25%	**-77.15%**	-27.45%	-35.33%	**-31.39%**	**-62.33%**