ETSI TR 126 943 V12.0.0 (2014-09)

Digital cellular telecommunications system (Phase 2+);
Universal Mobile Telecommunications System (UMTS);
LTE;
Recognition performance evaluations of codecs
for Speech Enabled Services (SES)
(3GPP TR 26.943 version 12.0.0 Release 12)

Reference
RTR/TSGS-0426943vc00

Keywords
GSM,LTE,UMTS

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://ipr.etsi.org).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**may not**", "**need**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
  - 1 presented to TSG for information;
  - 2 presented to TSG for approval;
  - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

SA4 has been working on the selection of a codec to recommend for Speech Enabled Services since October 2002 under the WID for SES [9]. The usual process of agreeing "design constrains" [10], "test and processing plan" [7] and "recommendation criteria" [8] was followed and completed before evaluating the candidates.

Two candidate codecs were proposed and evaluated:

1. ETSI Standard for the DSR Extended Advanced Front-end (ES 202 212)
2. AMR and AMR-WB audio codec

The performance evaluations were conducted by two leading companies in the area of speech recognition, IBM and Scansoft. Results from these evaluations were presented at SA4#30 in February 2004 and are summarised here. The "recommendation criteria" have been applied and SA4 recommends the DSR codec for Speech Enabled Services. SES codecs are introduced in packet switched conversational services in Technical Specifications 26.235 & 26.236 [5,6].

1 Scope

This technical report provides information on the recognition performance of the DSR Extended Advanced Front End conducted by speech recognition vendors IBM and Scansoft for the selection of a codec for Speech Enabled Services. The performance results are provided both as absolute word error rates for DSR and AMR-NB/AMR-WB on a range of extensive evaluation databases and as relative word error rate reductions when compared to both the AMR-NB and AMR-WB codecs.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TS 22.234: "Speech recognition framework for automated voice services; Stage 1".

[2] 3GPP TR 22.977: "Feasibility study for speech enabled services".

[3] ETSI ES 202 050: "Distributed Speech Recognition; Advanced Front-end Feature Extraction Algorithm; Compression Algorithm".

[4] ETSI ES 202 212: "Distributed Speech Recognition; Extended Advanced Front-end Feature Extraction Algorithm; Compression Algorithm, Back-end Speech Reconstruction Algorithm".

[5] 3GPP TS 26.235: "Packet switched conversational multimedia applications; Default codecs".

[6] 3GPP TS 26.236: "Packet switched conversational multimedia applications; Transport Protocols".

[7] TD S4-030543 "Test and Processing plan for default codec evaluation for speech enabled services (SES)", SA4

[8] TD SP-030440 "Recommendation Criteria for Default Codec for Speech Enabled Services (SES)", TSG SA.

[9] TD SP-020687 WID Codec Work to Support Speech Recognition Framework for Automated Voice Services (Rel-6), TSG SA.

[10] TD S4-030248 "Design Constraints for default codec for speech enabled services (SES)", SA4.

Note: Annex A lists all the key SA4 SES selection phase documents. Temporary Documents are attached to this specification in a separate .zip file.

3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

AFE	Advanced Front-end
AMR	Adaptive Multi-Rate
AMR-NB	AMR Narrowband
AMR-WB	AMR Wideband
BLER	Block Error Rate

DSR	Distributed Speech Recognition
EDGE	Enhanced Data for GSM Evolution
ETSI	European Telecommunications Standards Institute
GSM	Global System for Mobile communications
SES	Speech Enabled Services
SNR	Signal To Noise Ratio
VAD	Voice Activity Detector
X-AFE	eXtended Advanced Front-end

4 General

4.1 Project History

Table 1 below shows the progress and timeline of the project. In particular the creation of permanent documents; identification of candidate codecs and test organisations; running of the performance evaluations by test organisations; selection at SA4; verification; and the approval of CRs and TS at SA. Key milestones are highlighted in bold.

Table 1: SES project timeline

Meeting	Status of progress in activities
SA4 #23 (30 Sept - 4 Oct 2002)	▪ **Draft WID and work plan**
SA4 #24 (11-15 Nov 2002)	• Permanent documents o Design Constraints V1.0 o Test & Processing Plan V0.8 o Recommendation Criteria V0.1
Intermediate deadline on SA4 reflector 31.12.2002	• Submission of specification of additional databases as candidate for testing as part of test and processing plan.
Intermediate deadline on SA4 reflector 31.12.2002	▪ Any company which would possibly like to submit a candidate will indicate before 31.12.2002. Later indications will not be considered.
SA4 #25 (20-24 Jan 2003)	▪ List of testing organisations ▪ Permanent documents o Design Constraints V1.1 o Test Plan & Processing Plan V1.0 o Recommendation Criteria V0.3
SA4 #25 bis (24-28 Feb 2003)	▪ **List of testing organisations (IBM & SpeechWorks)** ▪ **List of candidate codecs (DSR X-AFE & AMR-NB/AMR-WB)** ▪ Permanent documents o **Design Constraints V2.0** o Test Plan & Processing Plan V1.3

	o Recommendation Criteria V0.3
SA4 SQ SES ad-hoc 1-2 April 2003 Basingstoke, UK	▪ Permanent documents o Test & Processing Plan V1.4 o Recommendation Criteria V0.3
SA4 #26 (5-9 May 2003)	▪ Permanent documents o **Test & Processing Plan V2.0** o Recommendation Criteria V0.6
SA4 #27 (7-11 July 2003)	Approval of permanent docs o **Test & Processing Plan V2.2** o **Recommendation Criteria V2.0**
ASR vendor evaluations start. Aug 2003	▪ **ASR vendors start tests.**
Deliverables from candidates: (31 October 2003)	▪ Fixed point complexity assessment ▪ Drafts of new 3GPP TSs (for new codecs), or existing specifications for information (codecs already in standards) ▪ Justification document of having met the Design Constraints
SA4 #29 (24-28 Nov 2003)	▪
Preparation for verification	▪ Agree verification plan by correspondence (19 Dec) ▪ Complete any legal agreements (NDAs) that are needed (15 Feb) ▪ Verification labs to obtain any databases needed (15 Feb)
Informative speech quality listening tests	▪ Nokia and Ericsson to supply listening test speech files to Motorola (5th Dec) ▪ Motorola to process listening test speech files supplied by Nokia and Ericsson (15 Jan) ▪ Nokia and Ericsson conduct listening tests
Completion of ASR vendor evaluations (31 Jan 2004)	▪ **Results from ASR vendor evaluations to ETSI representative**
SA4 #30 (23-27 Feb 2004)	**SES Selection meeting** ▪ Results from evaluator tests available ▪ Make recommendation ▪ Prepare TSs for approval @ SA#23

	▪ Prepare CRs for approval @ SA#23
SES Verification (1 March)	▪ Verification of selected codec (ST-Micro). ▪ Discussion of results of verification conference call March.
SA #23 (15-17 March 2004)	▪ TSs for information ▪ CRs for information
SA4 #31 (17-21 May 2004)	• **Verification report**
SA #24 (7-10 June 2004)	▪ **TSs approval (**TS 26.243) • **CRs approval** (TS 26.235 & TS 26.236)

4.2 Overview of the speech recognition framework for automated voice services work item

The work item covered the evaluation of candidate codecs for use in a speech recognition framework for automated voice services. The 3GPP speech recognition framework enables the use of conventional codecs (e.g. AMR) or DSR optimised codecs to distribute in the network the speech engines that process speech input or generate speech output.

The aim of the work item is, through objective evaluation, to recommend a single codec for speech enabled services based on a speech recognition framework.

4.3 Presentation of the following sections

The following sections provide a summary of the Selection Phase test results, including the results of the objective performance measurements, and a record of other relevant information for the selected candidate algorithm.

- Section 5 describes the Recommendation Criteria defined for the Selection Phase
- Section 6 defines the means used to measure the performance of each of the candidates
- Section 7 summarises the recognition evaluation results

5 Recommendation criteria

5.1 Overview

The set of databases used for the evaluations are defined in the Test and Processing Plan [7]. Each of these databases contains different types of speech material covering a variety of tasks, environments and languages. Recommendation was based on a score obtained from the recognition performance measured on each of these different databases. Section 5.3 describes how the scores from all the individual databases are combined using a weighting table.

5.2 Scoring on individual databases

For each database the reference performance is measured as the word error rate obtained from the ASR vendor"s system. This is the performance obtained from a state-of-the-art system from the ASR vendor assuming a transparent channel.

The performance (word error rate) on a given database is also measured with the ASR vendors system for a codec under test as described in the test and processing plan [7].

Scoring for tests performed with channel BLER were also computed in a similar way. Note that only BLER of 1% and 3% were considered as part of the recommendation criteria[8].

5.3 Performance metric over all databases

The overall performance was determined by averaging the absolute word error rate using the weightings presented in the detailed tables of Appendix 2 of the recommendation criteria document [8]. The result of this weighted average is an overall measure of the average word error rate for each codec. This metric is called the "average word error rate".

5.4 Comparisons between codecs

5.4.1 Low data-rate codec comparison

The two codecs under consideration at low data-rate are AMR 4.75 and DSR AFE with extension (5.6kbit/s). Only 8 kHz sampling rate is considered since there is no AMR-WB codec at low data rate.

Table A2.1 in Appendix 2 of the recommendation criteria document [8] shows the list of databases that will be tested and the weightings to be given to the scores obtained for each of these databases.

5.4.2 High data-rate codec comparison

At high data-rates the comparisons are made separately at 8 kHz and 16 kHz sampling rates.

5.4.2.1 8 kHz sampling rate

The two codecs under consideration at high data-rate at 8 kHz sampling are AMR 12.2 & DSR X-AFE (5.6kbit/s).

5.4.2.2 16 kHz sampling rate

The two codecs under consideration at high data-rate at 16 kHz sampling are AMR-WB 12.65 & DSR X-AFE (5.6 kbit/s).

5.5 Detailed recommendation comparisons

To justify the introduction of a new codec for SES it was considered that DSR would need to demonstrate substantial improvements compared to the existing AMR voice codec already mandated for voice communications. The following performance requirements were agreed by 3GPP SA4:

For the low data-rate comparison:

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR 4.75 kbps codec is more than 35% then the DSR codec and its extension will be recommended.

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR 4.75 kbps codec is less than 20% then the DSR codec will not be recommended.

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR 4.75 kbps codec is less than 20% then AMR will be recommended.

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR 4.75 kbps codec is between 20% and 35% then the performance results will be further considered by SA4 and if there is no consensus the results will be passed to SA for decision on what recommendation to make.

For the high data-rate comparison at 8 kHz:

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR 12.2 kbps codec is more than 30% then the DSR codec and its extension will be recommended.

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR 12.2 kbps codec is less than 20% then the DSR codec will not be recommended.

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR 12.2 kbps codec is less than 20% then AMR will be recommended.

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR 12.2 kbps codec is between 20% and 30% then the performance results will be further considered by SA4 and if there is no consensus the results will be passed to SA for decision on what recommendation to make.

For the high data-rate comparison at 16 kHz:

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR-WB codec is more than 25% then the DSR codec and its extension will be recommended.

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR-WB codec is less than 15% then the DSR codec will not be recommended.

· If the relative reduction in average word error rate for the DSR AFE codec and its extension compared to the AMR-WB codec is less than 15% then AMR-WB will be recommended.

6 Performance evaluation method

6.1 Introduction

Codec evaluation was based on a framework which includes databases, codecs and speech recognition engine. The evaluators were requested to use the same recognition engine for all codecs.

The evaluation framework for codec test is shown in Figures 1 and 2 below. Fig 1 applies for codecs with speech interface like a conventional speech codec and figure 2 applies for codecs with feature data interface like DSR optimised codecs.

The evaluation framework contains 2 processing stages:

· The candidate codec

· The speech recogniser from the evaluator

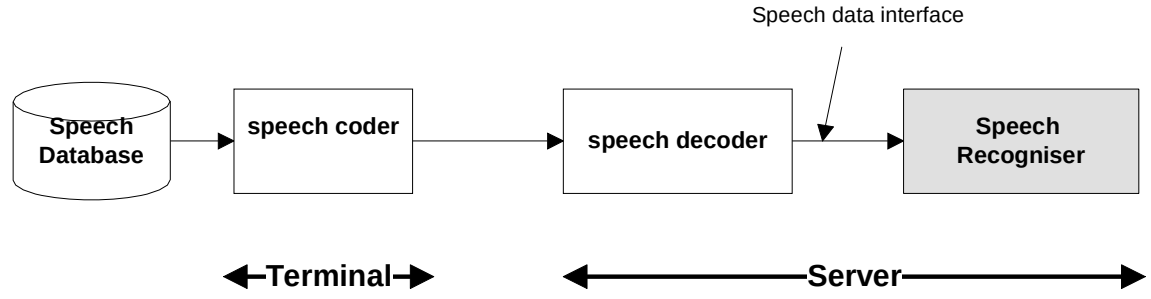

Figure 1: Evaluation framework for speech codec (note that in this case the speech recogniser includes front-end and back-end decoder)

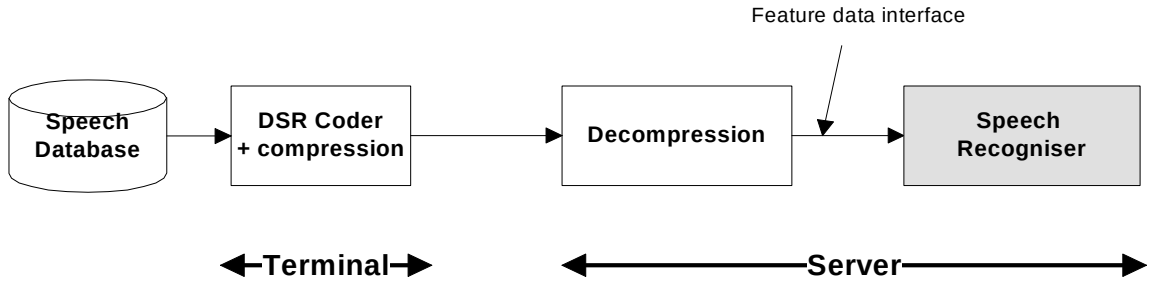

Figure 2: Evaluation framework for DSR optimised codec (note that in this case the speech recogniser is back-end decoder only)

6.2 Recognition engines

ASR vendors performed the evaluations. Each ASR vendor was provided with the databases for the evaluation consisting of defined training and test sets (3GPP supplied databases). In addition ASR vendors proprietary databases were used as well (ASR Vendor Supplied databases). Each ASR Vendor ran performance tests on these databases considering both the AMR codec chain shown in figure 1 and the DSR optimised codec chain as shown in figure 2. ASR vendors had a free choice over the recogniser back-end configuration.

6.2.1 Recognizer for speech codecs based proposals

As the AMR and AMR WB Codec can operate at several bitrates, a selection of bitrates had to be done for each test. Simulation of all AMR and AMR WB modes with all databases leads to practically unfeasible tests, therefore the number of Modes which were evaluated was limited. For each selected bitrates the complete evaluation was run on all databases. That means the training and test was performed with that bitrate on the whole database. Table 2 below shows the test conditions for AMR and AMR WB.

Table 2: Test conditions for AMR and AMR WB Codec

Bitrate	Codec	Sampling rate
4.75 kbps	AMR	8 kHz
12.2 kbps	AMR	8 kHz
12.65 kbps	AMR WB	16 kHz

6.2.2 Training and testing

The training was done using the coded & decoded speech data processed at the tested AMR bit rates as shown in table 2 above.

After the speech decoder, any speech signal processing, e.g. compensating the coding artefacts or calculating the tonal language parameters, can be applied to the speech signal before calculating the actual recognition features.

6.2.3 Recognizer for DSR

Figure 2 shows the processing chain for a DSR front-end. The Advanced DSR Front-end (AFE) can operate with 8 or 16 kHz sampling rates. The feature extraction produces 12 mel-cepstral features (C1-C12), the zeroth order cepstral feature (C0) and log energy parameter (logE) at a 10ms frame rate. Recognisers may make use of either C0 or logE or both. The feature extraction is described in the ETSI standard document for ES 202 050 [1]. The static feature vector may be subject to further processing of the evaluators choice to produce dynamic features. The software for the DSR standard contains an example implementing the recommended way of derivative calculation although evaluators were free to use their own alternatives.

In addition to the cepstral features the DSR AFE extension provides a pitch feature that may optionally be used as a feature to assist recognition when processing tonal languages. The raw pitch feature may be subject to further processing of the evaluators choice to produce tonal features to supplement the cepstral feature vector (e.g. smoothing or derivative calculation).

6.2.4 Training and testing

Training should be performed with the features after compression and decompression with an error free channel. The same feature post-processing should be used for training as for recognition.

6.3 Usage of VAD for frame dropping

For the purpose of these performance evaluations no voice activity detector was used for frame dropping either for discontinuous transmission at the terminal or at the recognition engine at the server.

6.4 Codec evaluations

Codec evaluations were conducted over a range of tasks as described in the following sections.

6.4.1 Recognition experiments under error-free channel

1. Connected digit recognition task
    - Aurora-2
    - Aurora-3
    - Vendor 2 In-car Japanese, German, US English
    - Vendor 1 US English in-car
    - Vendor 1 Mandarin Embedded corpus (digits)

2. Sub-word trained model recognition task
    - Nokia Mandarin Chinese name dialling (tone recognition ignored in performance scoring)
    - Vendor 2 In-car
        - Japanese
        - German,
        - US English
    - Vendor 1 Mandarin Embedded Corpus (names /street names /organization names/commands)
    - Vendor 1 US English in car (commands, addresses, radio-controls, navigation, lifestyle information services and points-of-interest)
3. Tone confusability task
    - Nokia Mandarin Chinese name dialling (tone recognition taken into account in performance scoring)
4. Channel error task.
    - Aurora-3 Italian

Table 3: Table of databases for 8 kHz evaluations

Database Source	Database	Evaluator
3GPP supplied	Aurora-2	Vendor 2
	Aurora-3 German	Vendor 2
	Aurora-3 Spanish	Vendor 2
	Mandarin Name Dial	Vendor 1
	Aurora-2	Vendor 1
	Aurora-3 Spanish	Vendor 1
	Aurora-3 Italian	Vendor 1
ASR Vendor supplied	Mandarin Embedded PDA	Vendor 1
	US English In-Car	Vendor 1
	US English In-Car	Vendor 2
	German In-Car	Vendor 2
	Japanese In-Car	Vendor 2

Table 4: Table of databases for 16 kHz evaluations

Database Source	Database	Evaluator
3GPP Supplied		
	Aurora-3 Spanish	Vendor 2
	Mandarin Name Dial	Vendor 1
	Aurora-3 Spanish	Vendor 1
	Aurora-3 Italian	Vendor 1
ASR Vendor Supplied	Mandarin Embedded PDA	Vendor 1
	US English In-Car	Vendor 1
	US English In-Car	Vendor 2
	German In-Car	Vendor 2
	Japanese In-Car	Vendor 2

6.5 Recognition experiments under channel errors

For the purposes of testing under channel errors the Aurora-3 Italian database with the well-matched training and testing condition was used.

Each codec was tested under error free channel and with average channel BLERs of 1%, 3%.

Recognition tests were conducted by SpeechWorks and IBM using the supplied test sets. Models for these tests were trained on the error free training data.

SES are planned for use with PSS over UTRAN, EGPRS and GPRS channels. The BLER error masks were generated by Alcatel using a network simulation model to have representative distributions with the following considerations and specific conditions:-

EGPRS (/GPRS) channel:

Simulations for GPRS and EGPRS were combined as the coding schemes for CS1 ..CS4 and MCS1 .. MCS4 are equivalent. Thereby the use of a EGPRS channel was sufficient.

The following parameters were used in the model:

- Typical Urban condition
- Scenarios: pedestrian with 3 km/h speed
- no FH
- unacknowledged mode

- One 20msec Frame per RTP/UDP Packet
- One RTP/UDP Packet per RLC/MAC Block

3 BLER patterns for EGPRS were provided, namely EG_EP1, EG_EP2 and EG_EP3

EG_EP1 = error condition in very good channel (mean BLER ~ 1 %)

EG_EP2 = error condition in good channel.(mean BLER ~ 3 %)

EG_EP3 = error condition in bad channel.(mean BLER ~ 10 %)

UTRAN Channel:

Error situation for UTRAN channel will be better (fast power control) than in EGPRS channel. The UTRAN channel is here approximated using the EG_EP1 error mask of the EGPRS channel.

7 Recognition Performance Evaluation Results

This section presents the results of the evaluations performed by ASR vendors IBM and Scansoft. The ASR vendors requested to keep their individual results anonymous and are therefore documented here as Result A and Result B.

Results are provided in three sets of tables.

- The first set presents the results of the low data rate comparison between the DSR AFE codec and the AMR-NB codec operating at 4.75 kbps with 8 kHz sampling rate.
- The second set presents the results of the high data rate comparison between the DSR AFE codec and the AMR-NB codec operating at 12.2 kbps with 8 kHz sampling rate.
- The third set presents the results of the high data rate comparison between the DSR AFE codec and the AMR-WB codec operating at 12.65 kbps with 16 kHz sampling rate.

The tasks are split into the task categories shown in the tables: ie digits, subword, tone confusability and channel errors.

Finally an overall relative reduction in word error rate figure is given for each comparison.

For each test database the absolute performance in terms of word error rate is given and the relative improvement of DSR compared to AMR is shown.

The relative improvement is computed for each database as (word error rate for AMR – word error rate for DSR)/word error rate for AMR.

The overall word error rate improvement is the weighted average of the improvement for each of the categories. Ie Sum of task category weight x improvement for category.

Low Data Rate comparison

Sampling rate = 8kHz
AMR mode = AMR-NB 4.75

		word error rate		
		AMR-NB 4.75	DSR	Relative Improvement
Digits	Aurora-2 (result B)	11.73	9.62	17.99%
	Aurora-2 (result A)	16.1	12.4	22.98%
	Aurora-3 German	18.27	13.83	24.30%
	Aurora-3 Spanish (Result A)	9.23	4.86	47.35%
	Aurora-3 Spanish (Result B)	13.93	4.86	65.11%
	Aurora-3 Italian	21.68	6.15	71.63%
	US English In-Car (digits test)	19	12	36.84%
	German In-Car (digit test)	11.4	8.3	27.19%
	Japanese In-Car (digit test)	16.2	9	44.44%
	US English In-Car (digits test)	4.49	2.44	45.66%
	Mandarin Embedded PDA (digit test)	2.57	1.66	35.41%
0.3	**Average improvement on digits tasks**			**39.90%**
Subword	Mandarin Embedded PDA	4.09	2.52	38.39%
	US English In-Car	4.25	2.78	34.59%
	US English In-Car	14.2	9.5	33.10%
	German In-Car	12	10.1	15.83%
	Japanese In-Car	18	13	27.78%
	Mandarin Name dialling (baseform test)	0.83	0.58	30.12%
0.4	**Average improvement on subword tasks**			**29.97%**
Tone Confusability	Mandarin Name dialling (tone confusion test)	3.59	3.06	14.76%
0.1	**Average improvement on tone confusability**			**14.76%**
Channel errors	1% BLER (result A)	5.67	2.39	57.85%
	1% BLER (result B)	9.4	6.7	28.72%
	3% BLER (result A)	6.51	2.38	63.44%
	3% BLER (result B)	17.6	6.8	61.36%
0.2	**Average improvement with channel errors**			**52.84%**
	OVERALL RELATIVE REDUCTION IN WORD ERROR RATE			**36%**

High Data Rate comparison at 8kHz

Sampling rate = 8kHz
AMR mode = AMR-NB 12.2

		word error rate AMR-NB 12.2	DSR	Relative Improvement
Digits	Aurora-2 (result B)	10.28	9.62	6.42%
	Aurora-2 (result A)	14.2	12.4	12.68%
	Aurora-3 German	15.9	13.83	13.02%
	Aurora-3 Spanish (Result A)	7.7	4.86	36.88%
	Aurora-3 Spanish (Result B)	11.95	4.86	59.33%
	Aurora-3 Italian	19.04	6.15	67.70%
	US English In-Car (digits test)	15.6	12	23.08%
	German In-Car (digit test)	8.6	8.3	3.49%
	Japanese In-Car (digit test)	11	9	18.18%
	US English In-Car (digits test)	3.37	2.44	27.60%
	Mandarin Embedded PDA (digit test)	2.57	1.66	35.41%
0.3	**Average improvement on digits tasks**			**27.62%**
Subword	Mandarin Embedded PDA	3.14	2.52	19.75%
	US English In-Car	3.29	2.78	15.50%
	US English In-Car	12.9	9.5	26.36%
	German In-Car	9.7	10.1	-4.12%
	Japanese In-Car	12.8	13	-1.56%
	Mandarin Name dialling (baseform test)	0.84	0.58	30.95%
0.4	**Average improvement on subword tasks**			**14.48%**
Tone Confusability	Mandarin Name dialling (tone confusion test)	3.81	3.06	19.69%
0.1	**Average improvement on tone confusability**			**19.69%**
Channel errors	1% BLER (result A)	4.73	2.39	49.47%
	1% BLER (result B)	7.1	6.7	5.63%
	3% BLER (result A)	6.33	2.38	62.40%
	3% BLER (result B)	12.6	6.8	46.03%
0.2	**Average improvement with channel errors**			**40.88%**

OVERALL RELATIVE REDUCTION IN WORD ERROR RATE **24%**

High Data Rate comparison at 16kHz

Sampling rate = 16kHz
AMR mode = AMR-WB 12.65

		word error rate AMR-WB	DSR	Relative Improvement
Digits	Aurora-3 Spanish (Result A)	7.5	4.6	38.67%
	Aurora-3 Spanish (Result B)	7.39	3.47	53.04%
	Aurora-3 Italian	14.77	5.62	61.95%
	US English In-Car (digits test)	17.8	12.3	30.90%
	German In-Car (digit test)	9.2	7.3	20.65%
	Japanese In-Car (digit test)	11.3	8.4	25.66%
	US English In-Car (digits test)	2.04	1.78	12.75%
	Mandarin Embedded PDA (digit test)	1.8	1.14	36.67%
0.35	**Average improvement on digits tasks**			**35.04%**
Subword	Mandarin Embedded PDA	2.29	1.63	28.82%
	US English In-Car	2.35	2.31	1.70%
	US English In-Car	13.2	7.8	40.91%
	German In-Car	10.7	7.1	33.64%
	Japanese In-Car	12.3	10.8	12.20%
0.45	**Average improvement on subword tasks**			**23.45%**
Channel errors	1% BLER (result A)	2.74	1.84	32.85%
	1% BLER (result B)	7.4	4.8	35.14%
	3% BLER (result A)	3.44	1.84	46.51%
	3% BLER (result B)	10.9	5	54.13%
0.2	**Average improvement with channel errors**			**42.16%**
	OVERALL RELATIVE REDUCTION IN WORD ERROR RATE			**31%**

It is noted that for the evaluations at 16 kHz one of the ASR vendors performed downsampling to 8 kHz.

Annex A (informative): Key selection phase documents

All the following documents can be found on the 3GPP FTP site (and are attached in a .zip file to this specification).

- TD SP-020687 WID Codec Work to Support Speech Recognition Framework for Automated Voice Services (Rel-6), TSG SA
- TD S4-030248 "Design Constraints for default codec for speech enabled services (SES)", SA4
- TD S4-030543 "Test and Processing plan for default codec evaluation for speech enabled services (SES)", SA4
- TD SP-030440 "Recommendation Criteria for Default Codec for Speech Enabled Services (SES)", TSG SA
- TD S4-040145 "SES Evaluation from ASR vendors (spreadsheet and informative data)", SA4
- TD S4-040153 "Verification plan for SES DSR v1.0", SA4
- TD S4-040152 "Results of error resilience for SES codec candidates**",** France Telecom, SA4
- TD S4-040336 "SES Verification report", SA4

Annex B (informative): Change history

Change history							
Date	**TSG SA#**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2004-12	26	SP-040834			Version 1.0.0 approved at TSG SA#26	1.0.0	6.0.0
2007-06	36				Version for Release 7	6.0.0	7.0.0
2008-12	42				Version for Release 8	7.0.0	8.0.0
2009-12	46				Version for Release 9	8.0.0	9.0.0
2011-03	51				Version for Release 10	9.0.0	10.0.0
2012-09	57				Version for Release 11	10.0.0	11.0.0
2014-09	65				Version for Release 12	11.0.0	12.0.0

History

Document history		
V12.0.0	September 2014	Publication