TR 101 294 V5.0.0 (1998-07)

**Digital cellular telecommunications system (Phase 2+);
Subjective tests on the interoperability of the
Half Rate/Full Rate/Enhanced Full Rate (HR/FR/EFR)
speech codecs, single, tandem and tandem free operation
(GSM 06.85 version 5.0.0 Release 1996)**

GSM

**GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS**

ETSI

ETSI

Contents

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 SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available **free of charge** from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://www.etsi.fr/ipr or http://www.etsi.org/ipr).

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 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 the Special Mobile Group (SMG) of the European Telecommunications Standards Institute (ETSI).

This TR reports the subjective testing results concerning the performance of the GSM Half, Full, and Enhanced Full Rate speech codecs (including the estimated advantages obtainable by using Tandem Free Operation) for the simulated digital cellular telecommunications system. The experimental conditions adopted for this set of tests reflected, as much as possible, "realistic" mobile-to-mobile connections. The TR will be part of GSM specification series covering the half rate, full rate, and enhanced full rate speech traffic channels (and TFO).

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

Version 5.x.y

where:

5   indicates GSM Phase 2+ Release 1996;

x   the second digit is incremented for all other types of changes, i.e. technical enhancements, corrections, updates, etc.;

y   the third digit is incremented when editorial only changes have been incorporated in the specification.

1 Scope

This Technical Report (TR) contains the results obtained from an internationally co-ordinated subjective evaluation conducted by four laboratories to estimate the performance in mobile-to-mobile connections of the GSM half, full, and enhanced full rate speech codecs (including the expected advantages of the Tandem Free Operation), in case of inter-operability in the digital cellular telecommunications system. These (listening-only) tests did not include the effects of Voice Activity Detector, Comfort Noise Insertion and the Discontinuous Transmission.

2 References

References may be made to:

a) specific versions of publications (identified by date of publication, edition number, version number, etc.), in which case, subsequent revisions to the referenced document do not apply; or

b) all versions up to and including the identified version (identified by "up to and including" before the version identity); or

c) all versions subsequent to and including the identified version (identified by "onwards" following the version identity); or

d) publications without mention of a specific version, in which case the latest version applies.

A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.

[1]   GSM 01.04: "Digital cellular telecommunications system (Phase 2+); Abbreviations and acronyms".

[2]   GSM 06.01: "Digital cellular telecommunications system; Full rate speech; Processing functions".

[3]   GSM 06.10: "Digital cellular telecommunications system; Full rate speech; Transcoding".

[4]   GSM 06.11: "Digital cellular telecommunications system; Full rate speech; Substitution and muting of lost frames for full rate speech channels".

[5]   GSM 06.12: "Digital cellular telecommunications system; Full rate speech; Comfort noise aspect for full rate speech traffic channels".

[6]   GSM 06.31: "Digital cellular telecommunications system; Full rate speech; Discontinuous Transmission (DTX) for full rate speech traffic channels".

[7]   GSM 06.32: "Digital cellular telecommunications system; Voice Activity Detector (VAD)".

[8]   GSM 06.02: "Digital cellular telecommunications system; Half rate speech; Half rate speech processing functions".

[9]   GSM 06.06: "Digital cellular telecommunications system; Half rate speech; ANSI-C code for the GSM half rate speech codec".

[10]   GSM 06.07: "Digital cellular telecommunications system; Half rate speech; Test sequences for the GSM half rate speech codec".

[11]   GSM 06.08: "Digital cellular telecommunications system; Half rate speech; Performance Characterization of the GSM half rate speech codec".

[12]   GSM 06.20: "Digital cellular telecommunications system; Half rate speech; Half rate speech transcoding".

[13]   GSM 06.21: "Digital cellular telecommunications system; Half rate speech; Substitution and muting of lost frame for half rate speech traffic channels".

[14]   GSM 06.22: "Digital cellular telecommunications system; Half rate speech; Comfort noise aspects for half rate speech traffic channels".

[15]   GSM 06.41: "Digital cellular telecommunications system; Half rate speech; Discontinuous Transmission (DTX) for half rate speech traffic channels".

[16]   GSM 06.42: "Digital cellular telecommunications system; Half rate speech; Voice Activity Detector (VAD) for half rate speech traffic channels".

[17]   GSM 06.51: "Digital cellular telecommunications system; Enhanced Full Rate (EFR) speech coding functions; General description".

[18]   GSM 06.53: "Digital cellular telecommunications system; ANSI-C code for the GSM Enhanced Full Rate (EFR) speech codec".

[19]   GSM 06.54: "Digital cellular telecommunications system; Test sequences for the GSM Enhanced Full Rate (EFR) speech codec".

[20]   GSM 06.55: "Digital cellular telecommunications system; Performance Characterization of the GSM Enhanced Full Rate (EFR) speech codec".

[21]   GSM 06.60: "Digital cellular telecommunications system; Enhanced Full Rate (EFR) speech transcoding".

[22]   GSM 06.61: "Digital cellular telecommunications system; Substitution and muting of lost frames for Enhanced Full Rate (EFR) speech traffic channels".

[23]   GSM 06.62: "Digital cellular telecommunications system; Comfort noise aspects for Enhanced Full Rate (EFR) speech traffic channels".

[24]   GSM 06.81: "Digital cellular telecommunications system; Discontinuous Transmission (DTX) for Enhanced Full Rate (EFR) speech traffic channels".

[25]   GSM 06.82: "Digital cellular telecommunications system; Voice Activity Detection (VAD) for Enhanced Full Rate (EFR) speech traffic channels".

3 Definitions and abbreviations

3.1 Definitions

Definition of terms used in this TR can be found in GSM 06 Series.

3.2 Abbreviations

For the purpose of this TR, the following abbreviations apply:

BFH	Bad Frame Handling
BFI	Bad Frame Indicator
CCITT	Comité Consultatif du Téléphone et du Télégraphe
CODEC	enCOder + DECoder
DL	Down-Link
DT	Deutsch Telekom
EFR	Enhanced Full Rate
EID	Error Insertion Device
EP	Error Pattern
ETS	European Telecommunication Standard
ETSI	European Telecommunication Standardization Institute
FR	Full Rate
FTP	File Transfer Protocol

GSM	Global System for Mobile communications
HR	Half-Rate
IRS	Intermediate Reference System
ITU	International telecommunication Union
LAN	Local Area Network
M-IRS	Modified Intermediate Reference System
PCM	Pulse Coded Modulation
MNRU	Modulated Noise reference Unit
SEG	Speech Expert Group
SMG	Special Mobile Group
STL	Software Tool Library
TFO	Tandem Free Operation
TR	Technical Report
UFI	Unreliable Frame Indicator
UIT	Union Internationale des Télécommunications
UL	Up-Link

For abbreviations not given in this subclause, see GSM 01.04 (ETR 350) [1].

4 General

Clause 5 describes the testing methodologies adopted for the internationally co-ordinated subjective evaluation of the interoperability HR-FR-EFR.

Clause 6 describes the tools and procedures used by the host laboratory to produce the processed speech material (speech was added with environmental noise, when requested by the experimental design).

Clause 7 describes the results obtained from the set of tests carried out by four laboratories.

Clause 8 describes the conclusions that can be drawn from the analysis of all data that were produced over the two main phases of testing.

5 Testing methodology

A series of internationally co-ordinated subjective listening experiments were carried out in four different laboratories. These experiments were designed to investigate the performance issues surrounding interoperability across the three GSM standards; Full Rate, Half Rate, and Enhanced Full Rate GSM. The purpose of the experiments was to determine the:

1) Quality under error conditions for single encodings, providing a baseline when considering the results for tandemed connections.

2) Quality under tandem conditions, where all possible tandemings of the three GSM standards are characterised under a range of channel errors.

3) Quality with far end background noise, where both single encodings and all possible combinations of the three GSM standards are characterised under a range of channel errors.

4) Quality enhancements through the use of the TFO scheme.

5) Differences in perceived quality between the input frequency responses "flat" and "complying with ITU-T Modified IRS definition".

To investigate these aspects, the testing was split into two phases:

Phase 1 investigated all the possible tandem connections between the three GSM standards in environments free of background noise. Also, the potential quality benefits resulting from tandem-free operation were investigated, as well as the effects of input frequency response.

Phase 2 investigated all the possible tandem connections with the far end was immersed in noisy environment. The noisy environments used were street noise (with a 15dB Signal to Noise Ratio) and vehicular noise (with a 10dB Signal to Noise Ratio). Only the Modified IRS input frequency characteristic was used.

In both these phases, all of the investigations spanned the range of channel error conditions; no errors, EP1, and EP2 (corresponding to C/I = 10 dB and 7 dB, respectively, with ideal Frequency Hopping). Both phases also checked the performance of each of the three GSM standards as a single encoding (i.e. no tandeming).

MNRU reference conditions were included in all experiments. These serve two main purposes. The first is to ensure that a suitable range of qualities is presented to the subjects; not doing so can lead to unusual results. The second is to provide a calibration of the judgement scale, which allows for the possibility of comparing results across experiments and laboratories.

Previous Technical Reports on GSM standards (GSM 06.08 and 06.55) used an input level to the speech codec of -16 dBm0 (22 dB OVL). However, in light of the recent availability of updated average distributions of speech levels in fixed networks throughout the world (source ITU-T Rec. G.117), a new nominal level of -20 dBm0 (-26 dB OVL) was used in the investigations reported here. The use of this figure, sourced from information drawn from the fixed network was selected in the absence of any reliable statistics on the levels found in mobile networks. It should also be noted that no reliable information is available on the responses of microphones adopted for use in mobile terminals. It was for this reason that the Modified IRS response (again sourced from information on the fixed network) was used in the majority of these experiments.

The remainder of this clause will discuss the two phases of experimentation.

5.1 Phase 1: Investigation under conditions free of environmental noise.

Phase one was designed to characterise GSM performance in noise free environments in order to provide information on the optimum performance figures likely to be obtained across the three different standards. Single encodings as well as tandemed connections between the available GSM standards, both with and without the use of the TFO scheme were investigated. The methodology used the ACR (Absolute Category Rating) method based on a 5-point rating scale, as given in ITU-T Recommendation P.800. To ensure that the maximum amount of information could be extracted from the experiment, the following precautions were taken in the experiment design:

To ensure that all the subjects started the experiments with similar levels of expectation, a standard set of instructions was given to each subject, and they were all given the same set of practice trials at the start of the experiment. These practice trials covered a range of conditions representative of the qualities in the main body of the experiment. The scores from the practice trials are recorded, but are not used in any further analysis, and hence are not reported on in this TR.

It is known, particularly in ACR tests, that the order in which the test material is presented to the subjects can have an effect on the results due to carry-over effects. Three different presentation orders were therefore used to minimise this.

Subjects can experience fatigue if listening sessions are too long, leading to an increase in the inherent variability of the results. The phase 1 tests, required subjects to spend approximately one hour in the experiment. For the reasons given, this was split into three sessions, each of approximately 20 minutes separated by short breaks.

The results from this phase are discussed in clauses 7.1 (single encoding), 7.2 (tandem operation), 7.4 (TFO enhancements), and 7.5 (input frequency response).

5.2 Phase 2: Investigation under conditions with far end environmental noise.

Phase two was designed to evaluate the GSM algorithms with regard to transmission quality in the presence of far end background noise. Two types of noise were investigated, street noise at a signal to noise ratio of 15dB and vehicle noise at a signal to noise ratio of 10dB. The phase was split into two experiments, each of which investigated one of the noise types. Other than this difference, these experiments were identical in design. The methodology used the DCR (Degradation Category Rating) method, where the opinions given by subjects are a measure of the perceived difference

between a quality reference and a degraded sample. To ensure that the maximum amount of information could be extracted from the experiment, the following precautions were taken in the experiment design:

> When investigating far end background noise, the effects of the noise itself can dominate the results to such an extent that the influence of the factors of interest are hidden. The Degradation Category Rating method was selected to minimise this, as was the decision to place each of the two different types of noise in separate experiments.

> To ensure that all the subjects started the experiments with similar levels of expectation, a standard set of instructions was given to each subject, and they were all given the same set of practice trials at the start of the experiment. These practice trials covered a range of conditions representative of the qualities in the main body of the experiment. The scores from the practice trials are recorded, but are not used in any further analysis, and hence are not reported on in this TR.

> Three different presentation orders were used to minimise the possibility of carry-over effects.

> Subjects can experience fatigue if listening sessions are too long, leading to an increase in the inherent variability of the results. The phase 2 tests, required subjects to spend approximately one hour 15 minutes in the experiment. For the reasons given, this was split into four sessions, each of just over 20 minutes separated by short breaks.

The results from this phase are discussed in clause 7.3.

6   Test set-up

Overview of the Host Lab Simulation Tool

In this section the different software components that are used by the Host Lab simulation. We present also the way they are used by the simulator.

The software components

The UIT-T STL96

In 1990 a group was set up within the CCITT to develop common software tools to help the development of speech coding standards. The first formal release of the library was the STL92. Matra Communication had used "IUT-T Software Tool Library Manual," ITU-T Users' Group on Software Tools, Geneva, May 1996, to develop the Software Tool used for the optimization of the Half Rate GSM codec.

It includes many different tools but more precisely all the tools needed to properly condition speech signals. It was mainly used for pre-processing and post-processing. However it was also used for A-Law PCM and MNRU.

The Speech codecs

The three speech codecs were simulated using 16-bit fixed point libraries that are part of the codecs' source code.

The test vectors of the three GSM speech codecs were successfully processed by their respective codecs.

<u>The Full Rate GSM</u>

The Full-Rate GSM simulation used (ETSI GSM Recommendation 06.10, "GSM full-rate speech transcoding") has been developed in Matra's Speech Processing Department. It was found easier to use this version rather than the one which is provided in STL96. Bad Frame Handling was included in the channel decoder function delivered by Nokia.

<u>The Half-Rate GSM</u>

The Half-Rate simulation used is based on that described in ETSI GSM Recommendation 06.20, "GSM half-rate speech transcoding". It was checked that it corresponds to the latest version available on the ETSI ftp server. The Bad Frame Handling simulation was that provided in ETSI GSM Recommendation 06.06, "ANSI-C code for the GSM half rate speech codec".

<u>The Enhanced Full Rate GSM</u>

The Enhanced Full Rate (ETSI GSM Recommendation 06.60, "GSM enhanced full-rate speech transcoding") simulation was based on versions 5.0 and 5.1 of ETSI GSM Recommendation 06.53, , "ANSI-C code for the GSM enhanced full rate codec"). Version 5.0 was upgraded taking into account the changes brought by version 5.1, i.e. initialization in the pulse search and modification of the basic operators. The pre-channel coding and the post-channel decoding of version 5.0 were kept.

The G.711 PCM codec

A-Law PCM coding was performed using the functions provided in the STL96 ["IUT-T Software Tool Library Manual," ITU-T Users' Group on Software Tools, Geneva, May 1996].

The Channel codecs

The Full Rate GSM

It was agreed that the channel codec and above all the error Concealment would have to be "state-of-the-art" as it had already been done for the pre-qualification test of the EFR. In consequence it was asked to Nokia if they could, as they had already done for the pre-qualification tests, deliver such a simulation to the Host Lab.

An NDA was signed between Nokia and Matra Communication to enable the use of these functions. Since Sun workstation are used in both companies, it was decided to exchange object code compiled for such systems.

The Half Rate GSM

An NDA was signed between Motorola and Matra Communication in order to use the Channel codec simulation that has already been used during the complexity estimation and the Optimization of the Half Rate GSM.

The interfaces were updated with the help of Eric Winter of Motorola.

The Enhanced full Rate GSM

Matra Communication received together with the source code of the EFR v1.4 speech codec the source code of the channel codec. This was the one that has been used so far to measure the performances of the EFR codec in error conditions.

Therefore it was decided to use it in the Host Lab simulation.

The Modulated Noise Reference Unit (MNRU)

The so-called "Duo-MNRU" [11][12] tool of STL96 was included in the Host LAB simulation.

6.1 Host Lab processing for the evaluation of HR-FR-EFR interoperability (phase 1)

The processing carried out by Matra Communication under contract of the ETSI were made using a Host Lab simulator. This section contains the description of the Host Lab simulator.

Modified IRS filtering procedure

The first problem in phase 1A was due to a wrong interpretation of the test plan by the Host Lab. The « HQN » key word of the command line was wrongly interpreted as "the spectral characteristic must be flat".

We replace the « HQN » key word by « MIRS » in the test plan file and we created a pre-processing procedure to filter and adjust the level of the speech material. This procedure is based on the function available in the STL96. Figure 1 shows the scheme of this pre-processing.
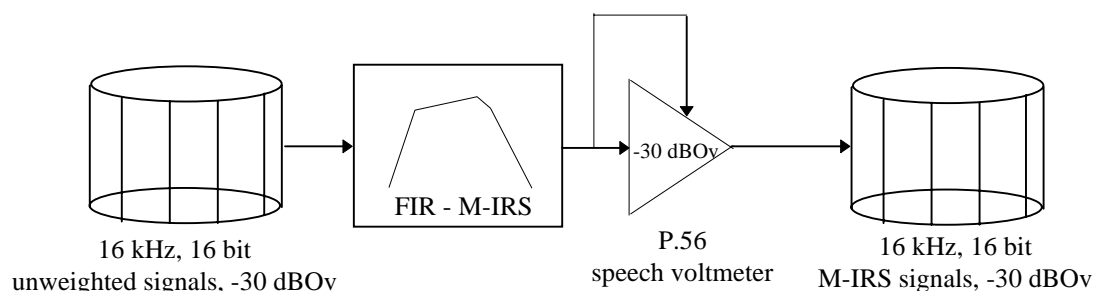
Figure 1. Scheme of the pre-processing procedure: M-IRS filtering and level adjustment.

6.2 Error concealment for TCH/FS

The second problem in phase 1A concerns the Error Concealment function included in the TCH - FS Channel Codec Libraries. During phase 1A this function has been implicitly switched off by the provider. As decided during the SEG #8 meeting the provider sent to Matra a correct version of the libraries. This version works correctly for all the possible channel simulations (with or without noise insertion). In the demo tape delivered by Matra (cf. SMG11 Tdoc 18/96) we provided some examples of sentences in EP1 and EP2 conditions obtained by using the old and the new versions of these libraries.

6.3 Control of codec simulation

All the codecs used in this simulation have been tested using the test vectors defined in the GSM recommendations. When the Host Lab simulator is used in the « test mode » the pre-processing, post-processing and channel coding procedure are bypassed as illustrated by figure 2. No errors have been observed in the output file.
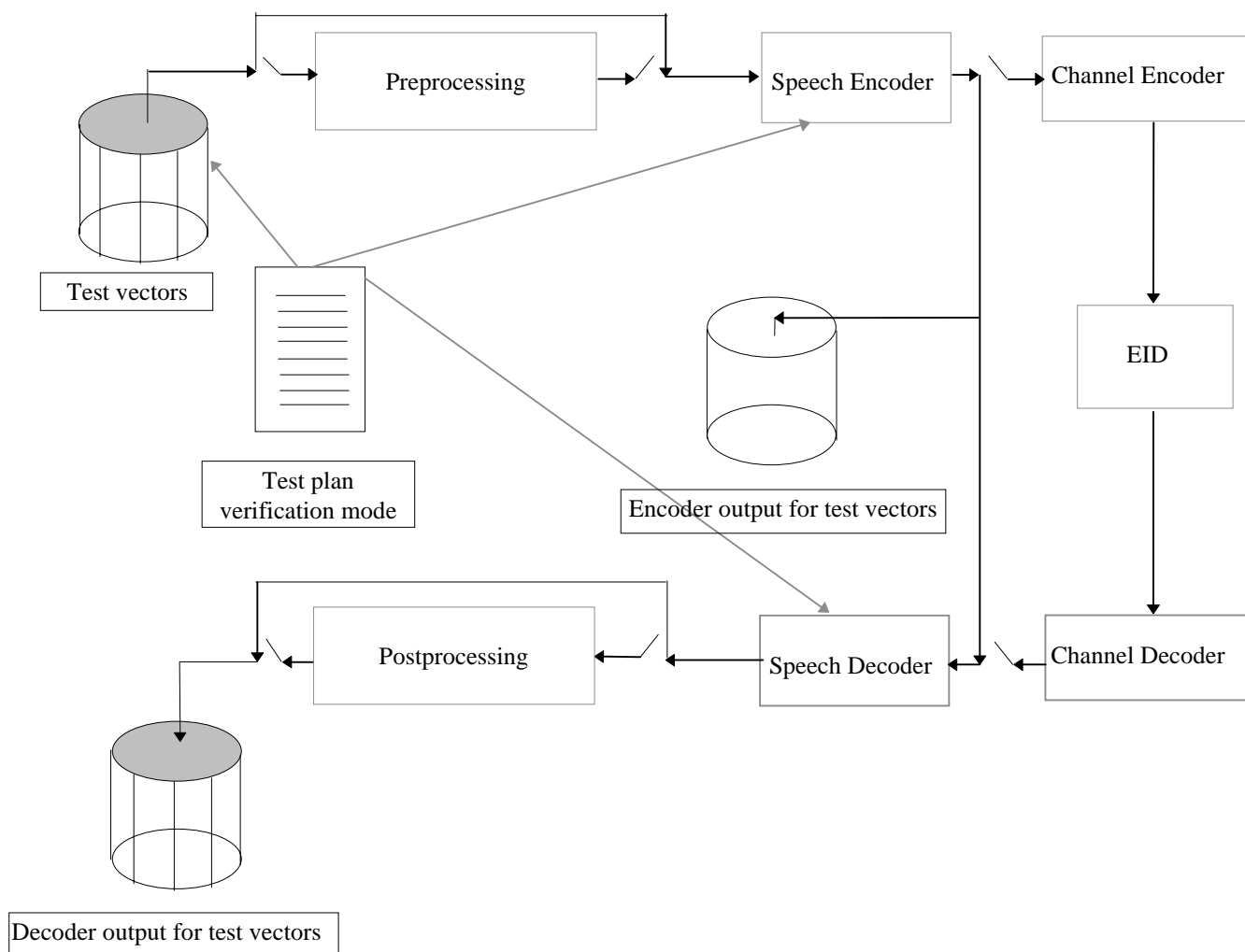
Figure 2. Scheme of the testing procedure using the test vectors.

6.4   Pre-processing and post-processing procedure

Figure 3 and figure 4 show the frequency response of the M-IRS filtering module and the frequency response for the pre and post processing. The behaviours of these frequency responses are closed to the frequency responses presented in the STL96 documents.
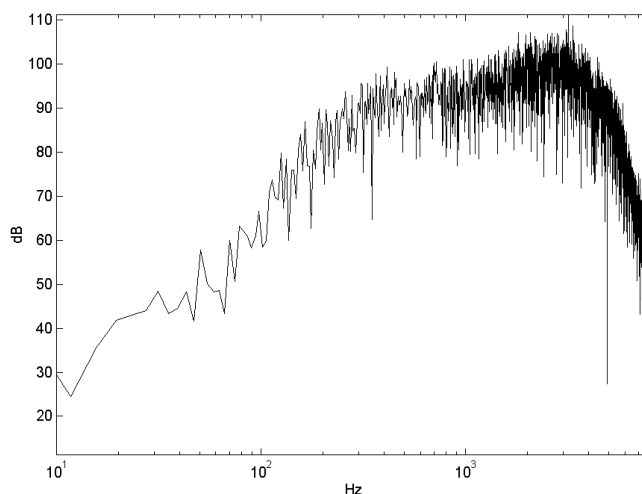
Figure 3. Frequency response of the M-IRS filtering module.
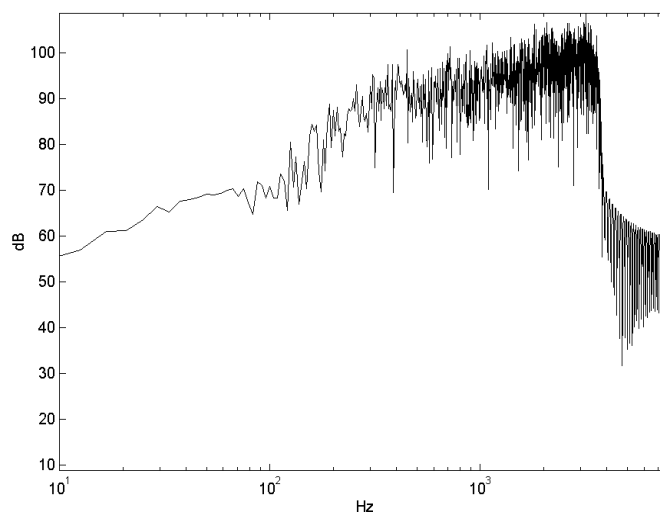
Figure 4. Frequency response of the global simulator filtering modules.

6.5 Conclusion of Phase 1

All the accessible components of the Host Lab simulator have been tested by Matra. No particular problem was observed during this phase of test.

6.6 Host Lab processing for the evaluation of HR-FR-EFR interoperability (phase 2)

The phase 2 of interoperability test have been carried out to evaluate the performances of GSM speech codecs in environmental noise conditions. This document describes the noise mixing procedure used to produce the speech material.

Noise mixing process

The original speech material used during phase 1 of test has been the starting point to produce the input databases for phase 2. The procedure used to produce these samples is the following (cf. figure 6). First of all the sources (speech and

noise) are filtered using the FIR version of Modified IRS filter (M-IRS). Then they are gain-scaled to obtain a -30 dBOv level for speech samples and the (-30-SNR) dBOv level for noise source. The value of SNR is dependent on the noise characteristic: this value is fixed to 15 dB for street noise and 10 dB for car noise.
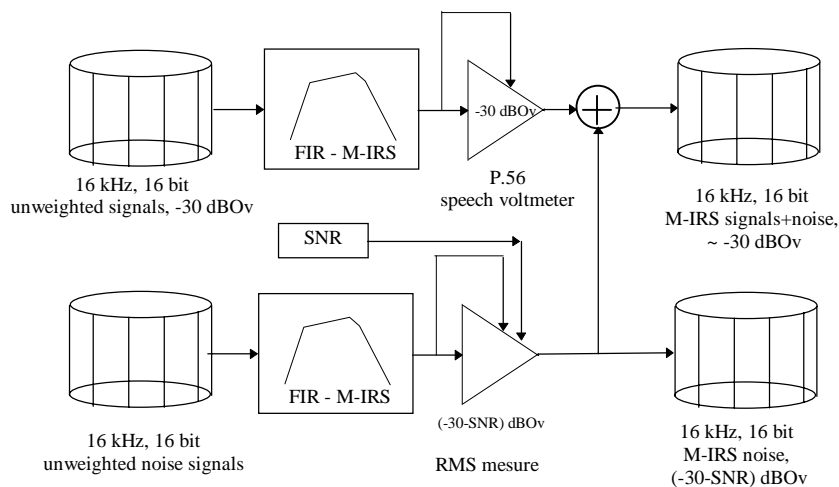

Figure 6. Noise and speech mixing procedure

The gain scaling values are obtained using the STL'96 speech voltmeter for the speech samples (according with ITU-T Recommendation. P.56) and using RMS measure for noise files.

The files filtered and scaled are then mixed by summing the samples. An automatic control of overload is performed during the mixing procedure. If a saturation occurred an expert listener checked if the quality of the speech + noise sample is acceptable. If the clipping effect was annoying the level of the speech after the M-IRS filter was down-scaled in a maximum range of 1 dB and the mixing procedure is restarted. During the mixing procedure any saturation has been observed.

6.7 Processing

The data-bases obtained by the noise mixing procedure are processed according to the test using the Host-Lab simulator.

6.8 Conclusion of Phase 2

During phase 2, test have been performed without any problem. Quality of processed speech has been checked by MATRA speech expert. The processed files were sent to the listening test laboratories in January 1997, according to the time schedule defined during the first SMG11 meeting.

7 Summary of results

The test results obtained from the subjective evaluation of the interoperability HR-FR-EFR, considering the different combinations of the three GSM standards, and also the so called TFO (Tandem-Free or Transcoding-Free Operation), are summarised in the following sub-clauses.

7.1 Quality under error conditions

Statistical analysis and significance tests (at 95% confidence level) on the full set of raw data were performed in terms of Mean Opinion Scores (MOS) and confidence intervals (C. I.), that were calculated to interpret the statistical significance of differences between the given scores.

An analysis of variance (using the General Linear Model) of all opinion scores was carried out to show which factors under test (laboratories, talkers, order of presentation, experimental conditions, and listeners) did account for the

subjective judgements. The analysis showed that all factors were significant; anyway, the 'common behaviour' showed by all the individual results, justified the decision of amalgamating the data coming from the different laboratories.

A statistical method based on a simple Tukey's Honestly Significant Difference (HSD) procedure, by utilising the measure MSE obtained from the analysis of variance, at the 95% confidence level, was used to accept or reject the hypothesis that the average score obtained by a certain condition was "significantly" (in statistical sense) worse than another one.

The number of votes cast by each listener was: 4 (talkers) x 48 (conditions) = **192**. There were 24 listeners (e.g. 12 male and 12 female) per experiment, organized in a number of sub-groups, e.g. 3 sub-groups made up of 8 persons each, each sub-group of listeners hearing the sequence in a different order. The number of votes per condition was: 4 (talkers) x 4 (languages/countries) x 24 (subjects) = **384**.

A primary elaboration of data was performed for the individual raw data, averaging Male and Female talkers (see Annex C for Tables of experimental results).

The following figure was obtained for the "reference conditions" inserted in the tests, i.e. the MNRU (Modulated Noise Reference Unit) and the "Direct" signal. The figure correlates Mean Opinion Scores (MOS, Yq) with Equivalent Q (dB) values, and is usually considered the "transfer curve" to convert data coming from different countries, i.e. taking into account cultural, or language, or other differences among laboratories conducting the "same" subjective test.
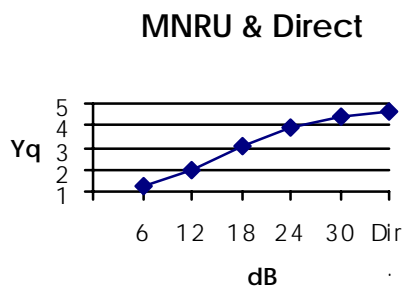
The effect of errors on the three algorithms, HR FR, EFR, in single encoding and for quiet conditions is depicted in the following figure.

Single encoding
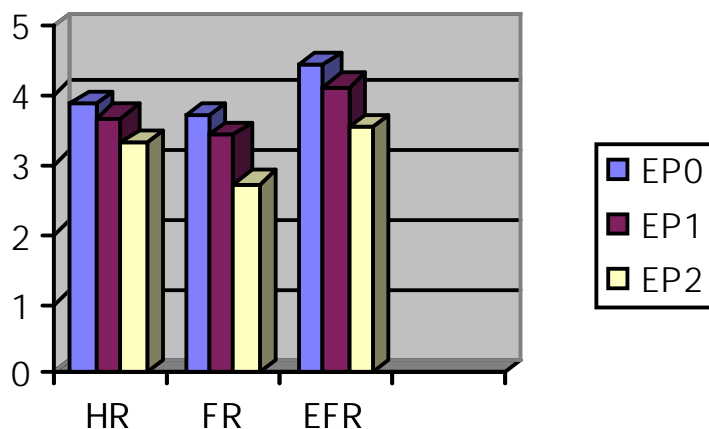
7.2   Quality under tandeming conditions

Tandeming conditions were tested in EP0, EP1 and EP2 error conditions (see Annex C for tables of results). The following figures clearly indicate the performance of HR, FR, and EFR when interworking with each of the other standards (i.e. the combinations of codecs in the figures should be read: Coder in the Legenda is tandemed with codec in the abscissa). The following figures were obtained from the experiments that adopted the Modified IRS input frequency characteristics.
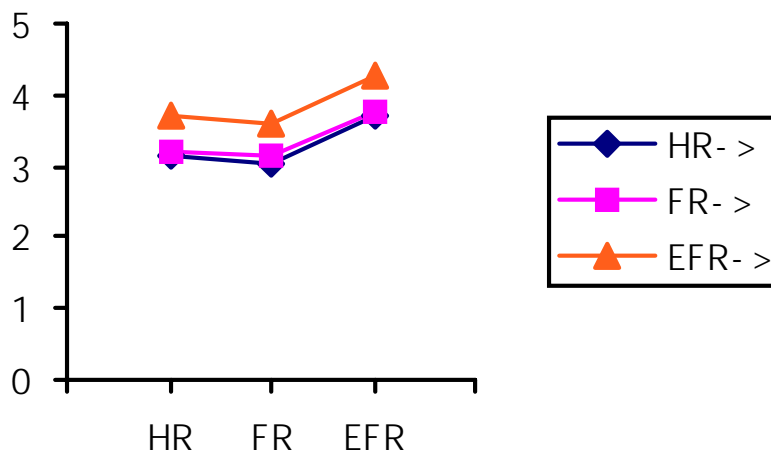
EP0 Tandem
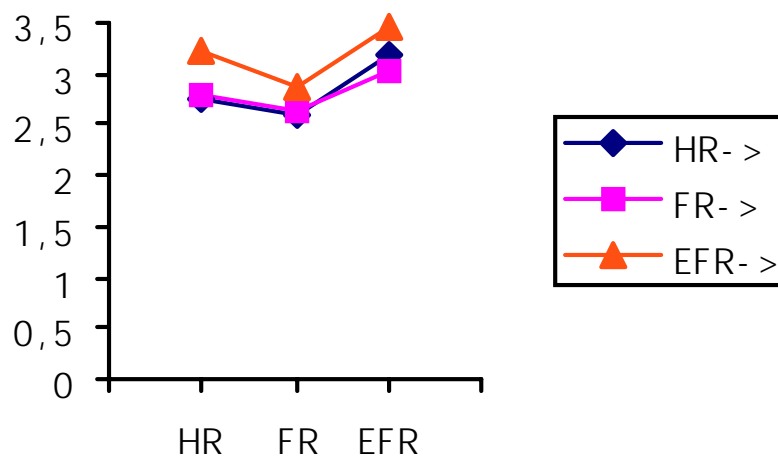
EP1 Tandem
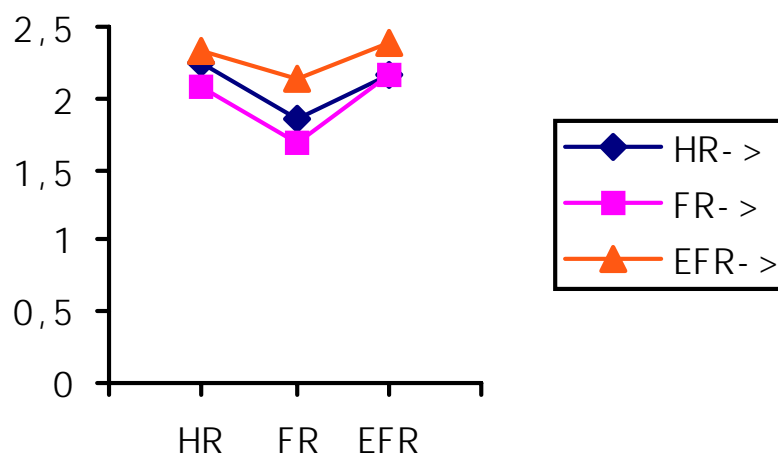
EP2 Tandem

The following figures were obtained from the experiments adopting the flat input frequency characteristics.

EP0 Tandem
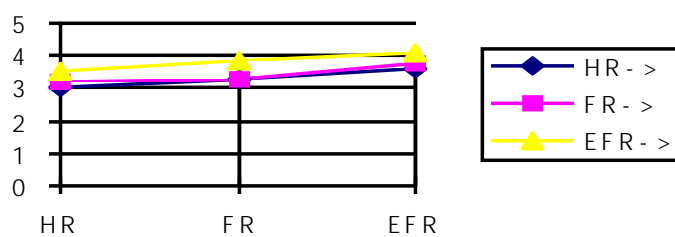
EP1 Tandem
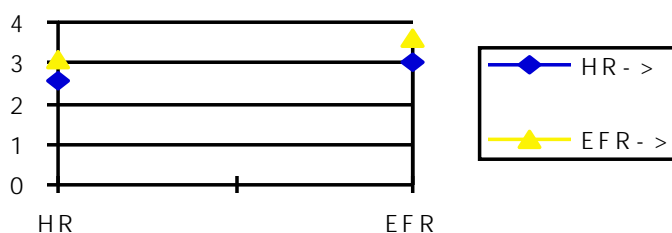
EP2 Tandem
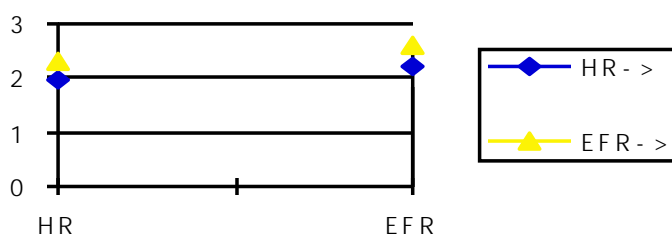
7.3 Quality with far end background noise

Any of the listening laboratories had to report their results from Exp#1 or Exp#2 in a way that the effects of environmental noise, error pattern and transcoding of the different standards are documented as DMOS values with their confidence interval (two-tailed t-distribution with a=0.05).

An individual analysis of data was conducted by each participant laboratory (see Bibliography). Tables of "averaged" results are given in DMOS, and confidence intervals. An analysis of variance revealed that conditions, talkers and groups of listeners were highly significant factors (p<0.0001) in both experiments.

A global analysis over the laboratories was performed and results were produced in a similar form to the one produced for global analysis for Phase 1 of testing (see Annex D for Tables of results).

In Exp. 1 the two laboratories were not significantly different each other, while in Exp. 2 the LAB factor was a highly significant one. The Tukey's minimum significant difference between DMOS values was around 0.35 in both experiments, such value being used to group conditions that were not "statistically" different in the same "class". The following figures illustrate the results obtained during Phase 2 of testing.

Street noise added to speech (S/N = 15 dB)

MNRU & Direct

Vehicular noise added to speech (S/N = 10 dB)

MNRU & Direct

Street noise added to speech (S/N = 15 dB)

Single encoding (DMOS)
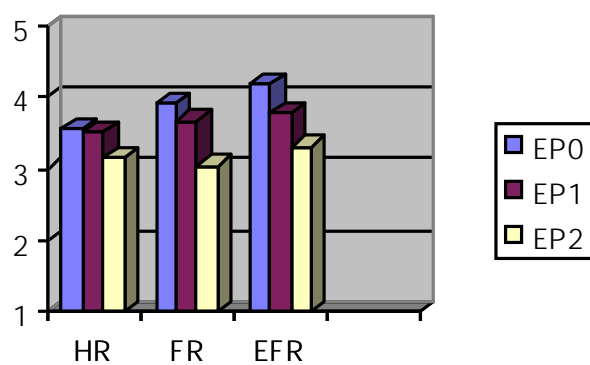
Vehicular noise added to speech (S/N = 10 dB)

Single encoding (DMOS)
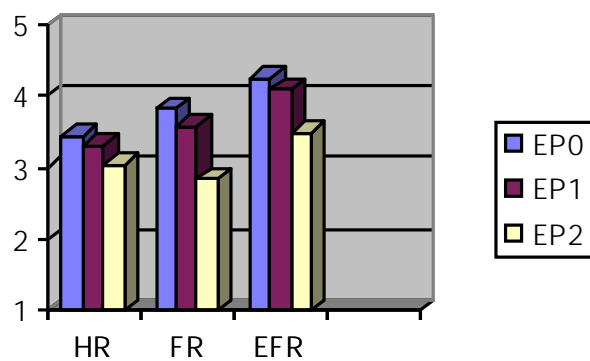
Street noise added to speech (S/N = 15 dB)
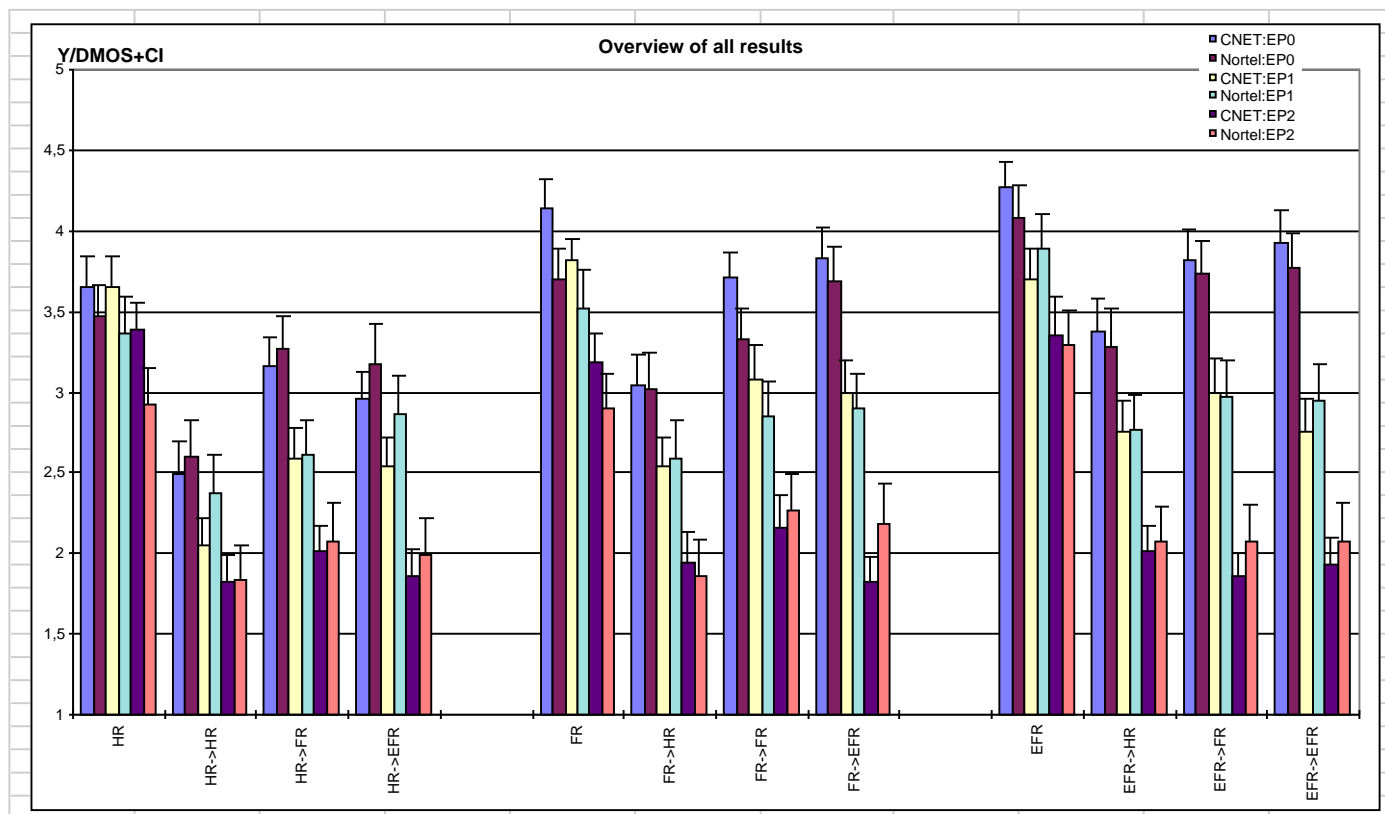
Vehicular noise added to speech (S/N = 10 dB)

Street noise added to speech (S/N = 15 dB)

EP2

Legend: CNET, Nortel

Categories: HR, HR->HR, HR->FR, HR->EFR, FR, FR->HR, FR->FR, FR->EFR, EFR, EFR->HR, EFR->FR, EFR->EFR
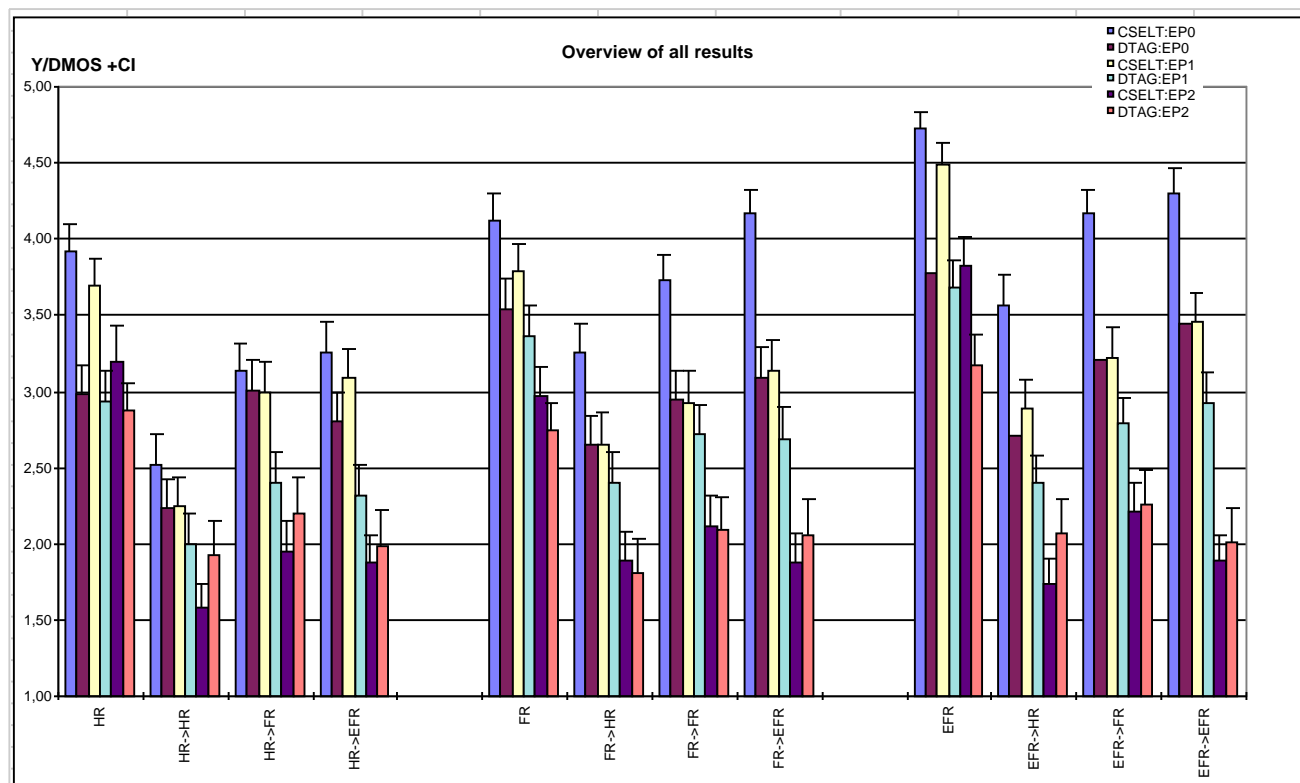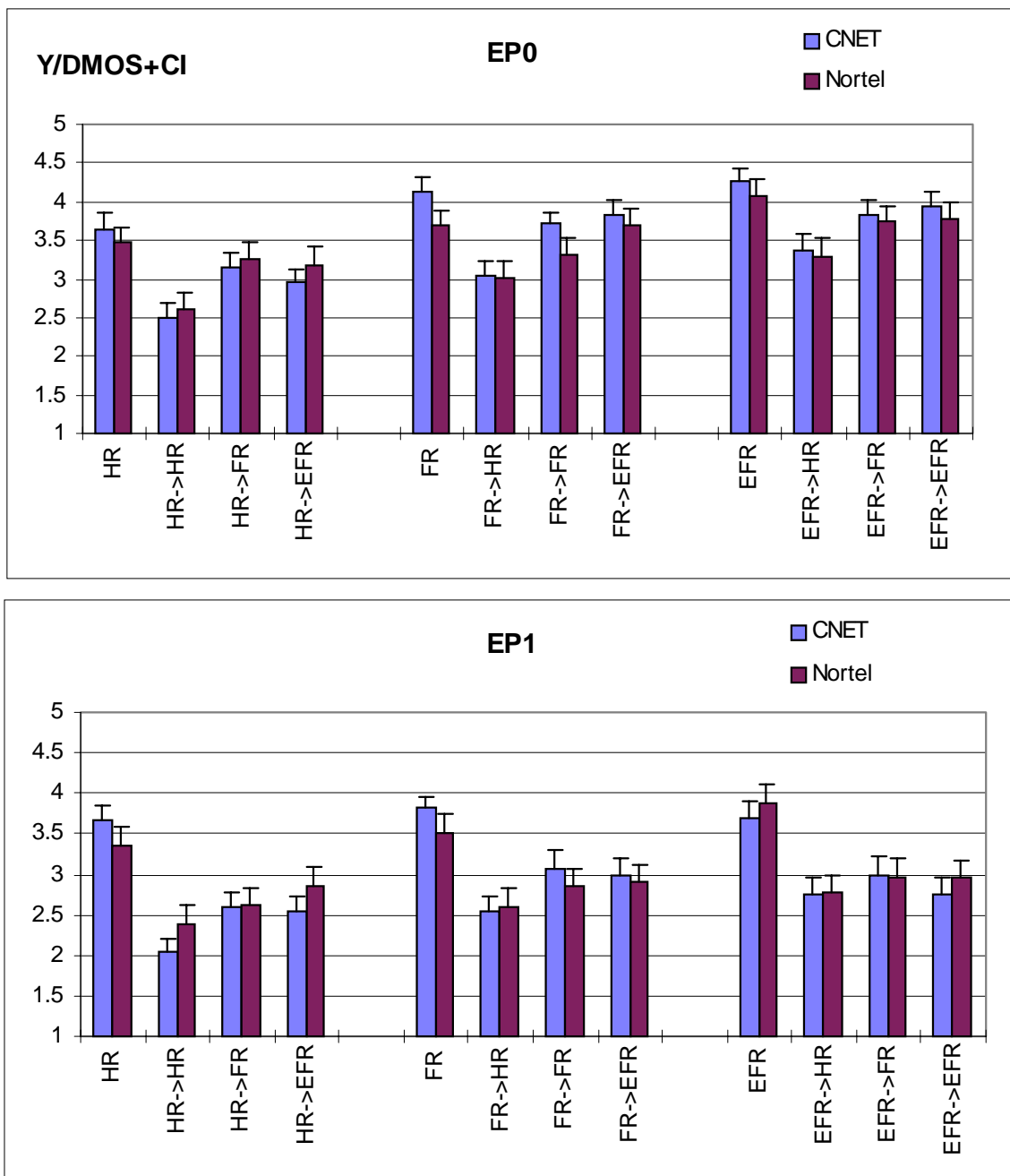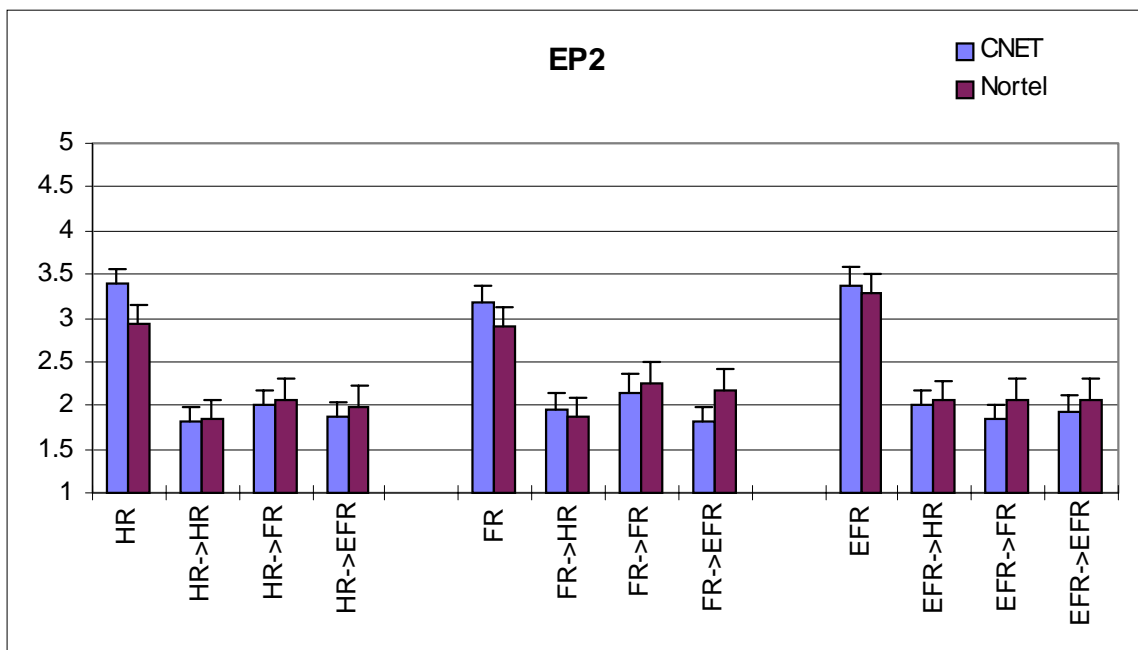
Vehicle noise added to speech (S/N = 10 dB)

EP0

Y/DMOS +CI

Legend: CSELT, DTAG

Categories: HR, HR->FR, FR->HR, FR->EFR, EFR, EFR->FR

Y/DMOS +CI — EP1 (CSELT / DTAG) chart with conditions HR, HR->FR, FR->HR, FR->EFR, EFR, EFR->FR

Y/DMOS +CI — EP2 (CSELT / DTAG) chart with conditions HR, HR->FR, FR->HR, FR->EFR, EFR, EFR->FR
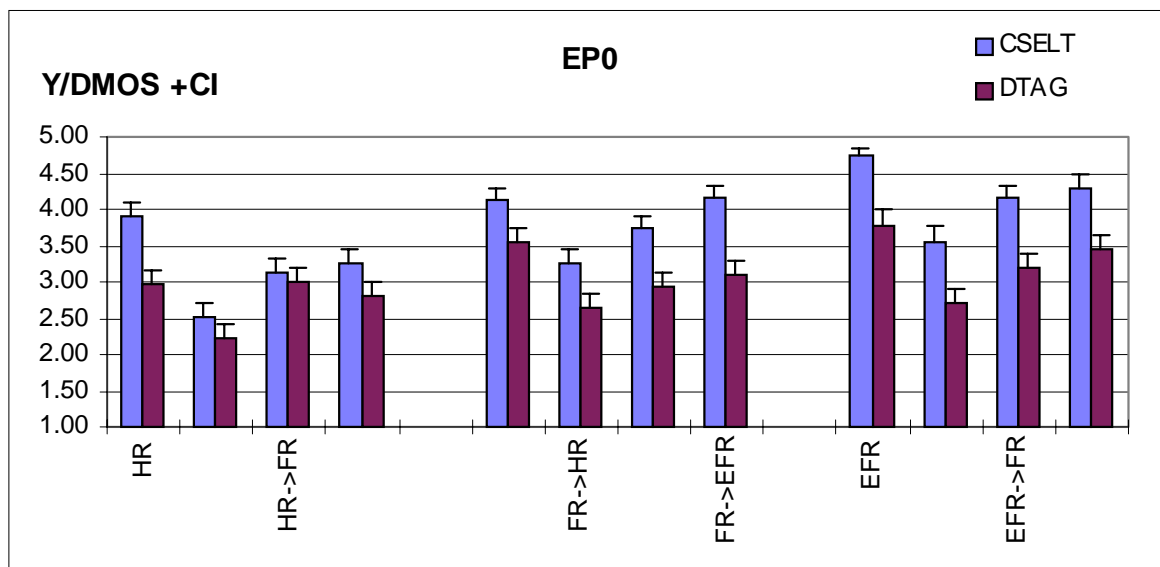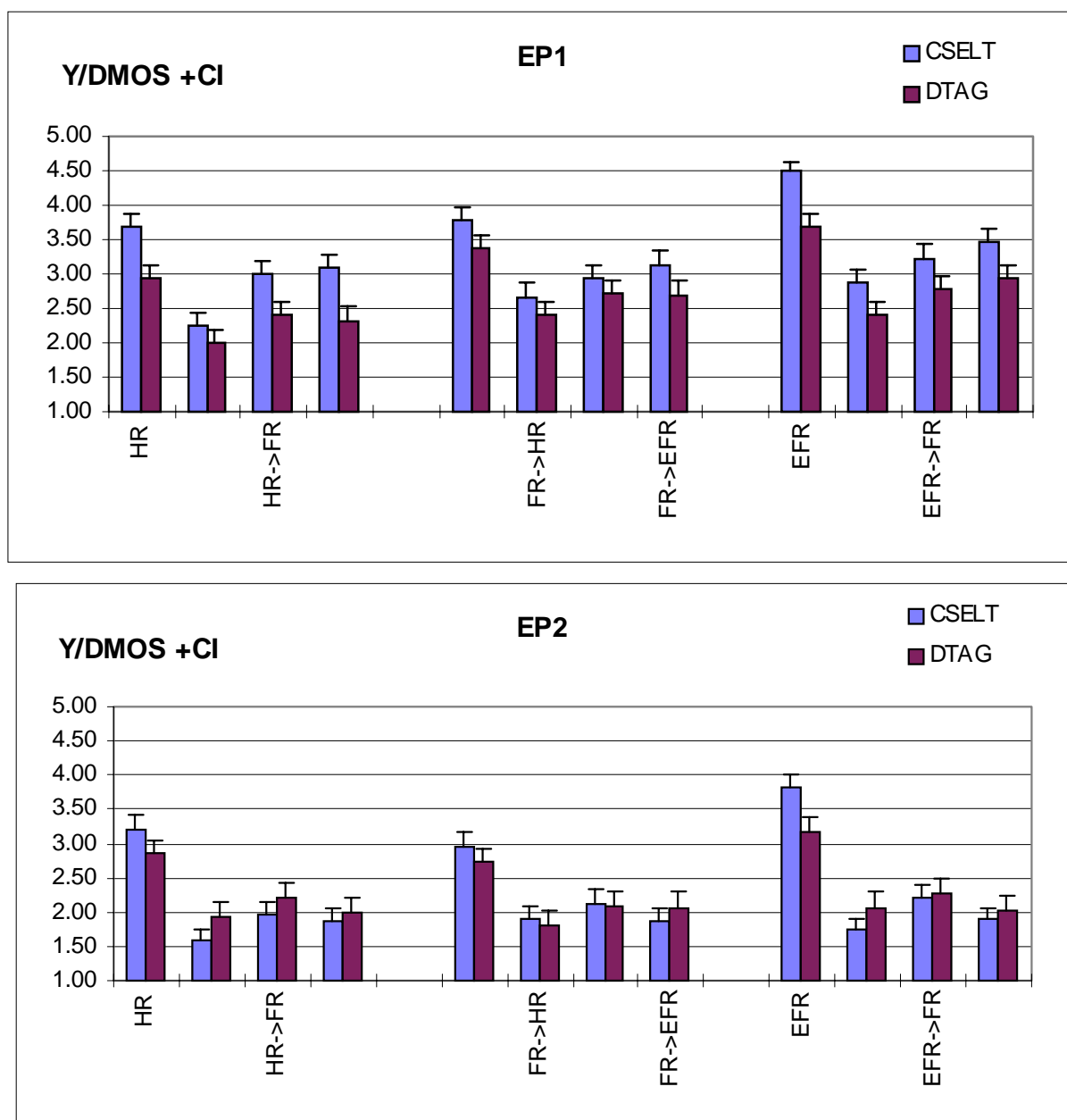
7.4   Quality enhancement using TFO techniques

The following table summarises the relevant results for the TFO conditions, i.e. the average improvement achievable with TFO technique (DMOS, **bold** means significant, **NA** stands for NOT APPLICABLE). Values were calculated using the Honestly Significant Difference (HSD) Tukey's test at the 95% significance level. The Table was derived from the experiments that used the flat input characteristics for recording speech.

	EP0	EP1	EP2
HR	**.85**	**.68**	**.39**
FR	**.53**	**NA**	**NA**
EFR	**.32**	**.46**	.19

The following figures summarise the relevant results for the TFO conditions, i.e. the average improvement achievable with TFO technique. The figures was derived from the experiments that used the Modified IRS input characteristics for recording speech.

EP1 Tandem TFO/No TFO
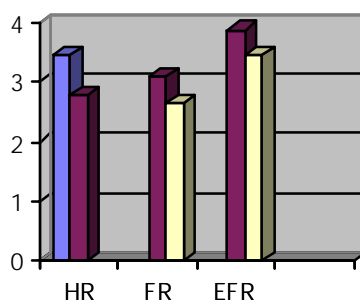
EP2 Tandem TFO/No TFO
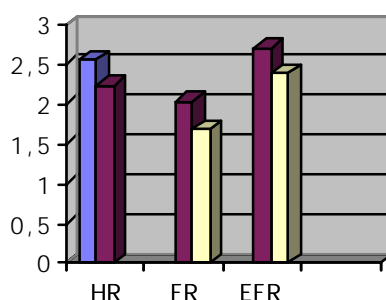
7.5 Effects of input frequency response

The results show that the "Modified IRS" input frequency characteristic was slightly preferred to the "flat" one.

8 Conclusion

A number of experimental conditions for the performance characterisation of GSM interworking standards, including codecs in single, tandem, and tandem-free operation with and without errors, plus some MNRU conditions and Direct, have been tested subjectively.

This document summarised the subjective test results obtained by CNET, CSELT, DT AG and NORTEL.

The similar or slightly lower quality behaviour showed by the FR codec in all conditions, in comparison with the HR codec, suggested to consider the analysis of results from previous exercises pertaining to those experimental conditions, and to undertake proper investigations on their reliability. Even if the results seem to be in line with the content of GSM 06.08, it is not advised to draw any conclusion on the relative performances of the HR and FR codecs from the limited test results contained in this technical report. The HR characterization test results (GSM 06.08) provides a more complete picture of the HR performances in multiple environment. Specifically, GSM 06.08 shows that the HR is more sensitive than the GSM FR to the type of filtering (Modified IRS or Linear 'flat') used for the pre-processing of the speech samples (see Section 6.1, Table 3 of GSM 06.08).

On grounds of the estimated Mean Opinion Scores, and considering the statistically significant differences among the simulated network configurations (computed using the Honestly Significant Difference (HSD) Tukey's test at the 95% significance level), the main positive results from this test are in general the rather good performance showed by the EFR and the quantification of improvement achievable by implementing the proposed TFO schemes, and by replacing the present FR by the EFR codec.

For the HR-HR tandem and EP1, the TFO scheme showed a significant improvement of 0.7 MOS; the same significant improvement (0.7 MOS) was obtained by EFR codec for both error-free (EP0) and EP1 (C/I=10 dB) conditions.

Based on the analysis of experimental results, the following conclusions can be drawn, in general, for speech mixed to street noise at SNR = 15 dB, or to vehicular noise at SNR = 10 dB:

1) In single encoding, the Enhanced Full-Rate codec performs significantly better than the Full-Rate and Half-Rate codecs, respectively.

2) In tandem encoding, for EP0 and EP1 error conditions (i.e. error-free and C/I=10 dB), the ranking order EFR, FR and HR is still kept, with HR performance significantly worse, while differences between Full-Rate and Enhanced Full-Rate were in a few cases "statistically" NOT significant.

3) In tandem encoding, for EP2 error conditions (i.e. C/I= 7 dB), the performance differences between the codecs decline.

Annex A:
Bibliography

[1] UIT-T Recommendation P.800**:** " Methods for subjective determination of transmission quality", Rev. September 1995.

[2] UIT-T Recommendation P.830**:** " Subjective performance assessment of telephone-band and wide band digital codecs", 1992.

[3] CCITT Recommendation P.810: "Modulated Noise Reference Unit", Blue Book, Volume V, 1988 and proposed revisions (SG12 meeting, 1994).

[4] D.J. Finney "Probit Analysis- A statistical analysis of the Sigmoid Response Curve", Cambridge University Press.

[5] R.E. Kirk "Experimental Design: Procedures for the Behavioral Sciences", 2nd ed., Brooks/cole Publishing Co., California, 1982.

Annex B:
Deliverables to ETSI

Other Deliverables to ETSI from subjective tests on the interoperability HR/FR/EFR & TFO:

Tdoc	SMG2-SEG	Source
34/96R	Test plan for the evaluation of the interoperability HR-FR-EFR (Phase 1).	CSELT
92/96	Evaluation of the interoperability HR-FR-EFR (Phase 1)	STSG of SEG
102/96	Subjective tests on the interoperability HR/FR/EFR & TFO: global analysis of results (Phase 1a of testing).	CSELT
103/96	Subjective tests on the interoperability HR/FR/EFR & TFO: (Phase 1a of testing). Results from CNET.	CNET
104/96	Subjective tests on the interoperability HR/FR/EFR & TFO: (Phase 1a of testing). Results from CSELT.	CSELT
105/96	Subjective tests on the interoperability HR/FR/EFR & TFO: (Phase 1a of testing). Results from DT AG.	DT AG
106/96	Subjective tests on the interoperability HR/FR/EFR & TFO: (Phase 1a of testing). Results from NORTEL.	NORTEL
107/96	Subjective tests on the interoperability HR/FR/EFR & TFO: (Phase 1a of testing). Report of the host laboratory session.	MATRA
127/96	Demostration to SEG (Phase 1a)	MATRA
Tdoc	SMG11	
7/96	Test results Phase 1b of testing	CSELT
8/96	Test results Phase 1b of testing	Nortel
9/96	Global analysis Phase 1b of testing	CSELT
13/96 R2	Test plan Phase 2 of testing	DT and FT/CNET
14/96	Test results Phase 1b of testing	CNET
17/96	Host laboratory processing for the evaluation of HR/FR/EFR interoperability (Phase 1b)	Matra
18/96	Demonstration tape to SMG11 (TFO Phase 1b of testing)	Matra
14/97	Phase 2 of testing, report from CNET	CNET
15/97	Phase 2 of testing, report from CSELT	CSELT
16/97	Phase 2 of testing, report from DeTeBerkom	DeTeBerkom
17/97	Phase 2 of testing, report from NORTEL	Nortel
18/97	Phase 2 of testing, report from host laboratory	MATRA
19/97	Phase 2 of testing, global analysis	CSELT

Annex C:
Phase 1: Tables of experimental results

Phase 1 of testing: experimental conditions and results.

Condition	Error Pattern	Codec / Reference	M.O.S Male + Female Talkers	Confidence Interval Male + Female Talkers
1	EP0	HR -> HR	3.15	0.10
2	EP0	HR -> FR	3.03	0.08
3	EP0	HR - > EFR	3.72	0.08
4	EP0	FR -> HR	3.18	0.08
5	EP0	FR -> FR	3.13	0.08
6	EP0	FR - > EFR	3.79	0.08
7	EP0	EFR -> HR	3.72	0.08
8	EP0	EFR -> FR	3.60	0.09
9	EP0	EFR - > EFR	4.29	0.08
10	EP1	HR -> HR	2.77	0.09
11	EP1	HR -> FR	2.59	0.08
12	EP1	HR - > EFR	3.17	0.09
13	EP1	FR -> HR	2.82	0.09
14	EP1	FR -> FR	2.64	0.09
15	EP1	FR - > EFR	3.03	0.09
16	EP1	EFR -> HR	3.22	0.09
17	EP1	EFR -> FR	2.89	0.08
18	EP1	EFR - > EFR	3.45	0.09
19	EP2	HR -> HR	2.24	0.09
20	EP2	HR -> FR	1.85	0.08
21	EP2	HR - > EFR	2.15	0.09
22	EP2	FR -> HR	2.07	0.08
23	EP2	FR -> FR	1.68	0.08
24	EP2	FR - > EFR	2.15	0.09
25	EP2	EFR -> HR	2.34	0.09

(continued)

Phase 1 of testing (concluded): experimental conditions and results.

Condition	Error Pattern	Codec / Reference	M.O.S	Confidence Interval
26	EP2	EFR -> FR	2.13	0.09
27	EP2	EFR - > EFR	2.38	0.09
28	EP1	HR -> HR TFO	3.46	0.09
29	EP1	FR -> FR TFO	3.10	0.08
30	EP1	EFR -> EFR TFO	3.87	0.08
31	EP2	HR -> HR TFO	2.55	0.10
32	EP2	FR -> FR TFO	2.03	0.09
33	EP2	EFR -> EFR TFO	2.71	0.10
34	EP0	HR	3.85	0.08
35	EP0	FR	3.71	0.08
36	EP0	EFR	4.43	0.07
37	EP1	HR	3.68	0.08
38	EP1	FR	3.41	0.08
39	EP1	EFR	4.12	0.07
40	EP2	HR	3.30	0.09
41	EP2	FR	2.73	0.09
42	EP2	EFR	3.53	0.09
43		MNRU = 6 dB	1.26	0.06
44		MNRU = 12 dB	1.93	0.08
45		MNRU = 18 dB	3.03	0.08
46		MNRU = 24 dB	3.94	0.08
47		MNRU = 30 dB	4.35	0.07
48		Direct	4.61	0.06

Analysis of Variance

General Linear Models Procedure

Number of observations in data set = 18432

Dependent Variable: SCORE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	76	12655.81	166.52	261.80	0.0001
Error	18355	11674.99	0.63		
Corrected Total	18431	24330.81			

R-Square	C.V.	Root MSE	SCORE Mean
0.52	26.08	0.80	3.06

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LAB	3	310.00	103.33	162.46	0.0001
COND	47	11153.85	237.31	373.10	0.0001
TALKER	3	352.45	117.48	184.70	0.0001
GROUP	2	82.13	41.06	64.56	0.0001
LISTENER (GROUP)	21	757.38	36.07	56.70	0.0001

General Linear Models Procedure

Tukey's Studentized Range (HSD) Test for variable: SCORE

> NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 18355 MSE= 0.636

Critical Value of Studentized Range= 5.62

Minimum Significant Difference= 0.2287

Means with the same letter are not significantly different.

Tukey Grouping				Mean	Condition No.	Condition Description
		A		4.61	48	Direct
B		A		4.43	36	EP0 EFR
B				4.35	47	MNRU = 30 dB
B		C		4.29	9	EP0 EFR-EFR
D		C		4.12	39	EP1 EFR
D		E		3.94	46	MNRU = 24 dB
F		E		3.87	30	EP1 EFR-EFR TFO
F		E		3.85	34	EP0 HR

(continued)

Tukey Grouping				Mean	Condition No.	Condition Description
F		E	G	3.79	6	EP0 FR-EFR
F	H	E	G	3.72	7	EP0 EFR-HR
F	H	E	G	3.72	3	EP0 HR-EFR
F	H		G	3.71	35	EP0 FR
F	H		G	3.68	37	EP1 HR
	H	I	G	3.60	8	EP0 EFR-FR
	H	I		3.53	42	EP2 EFR
J		I		3.46	28	EP1 HR-HR TFO
J		I	K	3.45	18	EP1 EFR-EFR
J		I	K	3.41	38	EP1 FR
J		L	K	3.30	40	EP2 HR
M		L	K	3.22	16	EP1 EFR-HR
M		L		3.18	4	EP0 FR-HR
M		L		3.17	12	EP1 HR-FR
M		L		3.15	1	EP0 HR-HR
M		L		3.13	5	EP0 FR-FR
M		L	N	3.10	29	EP1 FR-FR TFO
M		O	N	3.03	45	MNRU = 18 dB
M		O	N	3.03	15	EP1 FR-EFR
M		O	N	3.03	2	EP0 HR-FR
P		O	N	2.89	17	EP1 EFR-FR
P		O	Q	2.82	13	EP1 FR-HR
P		R	Q	2.77	10	EP1 HR-HR
P		R	Q	2.73	41	EP2 FR
P		R	Q	2.71	33	EP2 EFR-EFR TFO
		R	Q	2.64	14	EP1 FR-FR
S		R		2.59	11	EP1 HR-FR
S		R	T	2.55	31	EP2 HR-HR TFO

(continued)

ETSI

(concluded)

Tukey Grouping				Mean	Condition No.	Condition Description
S		U	T	2.38	27	EP2 EFR-EFR
V		U	T	2.34	25	EP2 EFR-HR
V		U	W	2.24	19	EP2 HR-HR
V	X	U	W	2.15	24	EP2 FR-EFR
V	X		W	2.15	21	EP2 HR-EFR
V	X		W	2.13	26	EP2 EFR-FR
	X	Y	W	2.07	22	EP2 FR-HR
	X	Y	W	2.03	32	EP2 FR-FR TFO
	X	Y		1.93	44	MNRU = 12 dB
Z		Y		1.85	20	EP2 HR-FR
Z				1.69	23	EP2 FR-FR
		A		1.26	43	MNRU = 6 dB

Annex D:
Phase 2: Tables of experimental results

HR-FR-EFR Interoperability Phase 2 of testing: Exp. 1 & 2 average results.

N	Experimental condition	Street Noise SNR = 15 dB	(CNET & Nortel)	Vehicle Noise SNR = 10 dB	CSELT & DeTeBerkom
		DMOS	C.I.	DMOS	C.I.
1	HR>HR_EP0_S+N	2.55	0.15	2.38	0.14
2	HR>FR_EP0_S+N	3.21	0.14	3.07	0.13
3	HR>EFR_EP0_S+N	3.07	0.15	3.03	0.14
4	FR>HR_EP0_S+N	3.03	0.14	2.95	0.14
5	FR>FR_EP0_S+N	3.52	0.13	3.34	0.14
6	FR>EFR_EP0_S+N	3.76	0.14	3.63	0.15
7	EFR>HR_EP0_S+N	3.33	0.16	3.14	0.15
8	EFR>FR_EP0_S+N	3.78	0.14	3.69	0.14
9	EFR>EFR_EP0_S+N	3.85	0.14	3.87	0.14
10	HR>HR_EP1_S+N	2.21	0.14	2.12	0.14
11	HR>FR_EP1_S+N	2.60	0.14	2.69	0.15
12	HR>EFR_EP1_S+N	2.70	0.15	2.71	0.15
13	FR>HR_EP1_S+N	2.57	0.15	2.53	0.15
14	FR>FR_EP1_S+N	2.96	0.15	2.82	0.14
15	FR>EFR_EP1_S+N	2.95	0.14	2.91	0.15
16	EFR>HR_EP1_S+N	2.76	0.15	2.64	0.14
17	EFR>FR_EP1_S+N	2.98	0.15	3.00	0.14
18	EFR>EFR_EP1_S+N	2.85	0.15	3.19	0.14
19	HR>HR_EP2_S+N	1.83	0.14	1.76	0.14
20	HR>FR_EP2_S+N	2.04	0.15	2.07	0.16
21	HR>EFR_EP2_S+N	1.93	0.14	1.93	0.15
22	FR>HR_EP2_S+N	1.91	0.14	1.85	0.15
23	FR>FR_EP2_S+N	2.21	0.15	2.10	0.15
24	FR>EFR_EP2_S+N	2.00	0.15	1.96	0.15
25	EFR>HR_EP2_S+N	2.04	0.13	1.90	0.14

(continued)

HR-FR-EFR Interoperability Phase 2 of testing (concluded): Exp. 1 & 2 average results.

N	Experimental condition	Street Noise SNR = 15 dB	(CNET & Nortel)	Vehicle Noise SNR = 10 dB	CSELT & DeTeBerkom
26	EFR>FR_EP2_S+N	1.96	0.14	2.23	0.15
27	EFR>EFR_EP2_S+N	2.00	0.15	1.95	0.14
28	HR_EP0_S+N	3.56	0.14	3.45	0.15
29	HR_EP1_S+N	3.51	0.15	3.32	0.14
30	HR_EP2_S+N	3.16	0.14	3.04	0.15
31	FR_EP0_S+N	3.92	0.14	3.83	0.14
32	FR_EP1_S+N	3.67	0.14	3.58	0.14
33	FR_EP2_S+N	3.04	0.14	2.85	0.13
34	EFR_EP0_S+N	4.18	0.13	4.25	0.14
35	EFR_EP1_S+N	3.79	0.15	4.08	0.13
36	EFR_EP2_S+N	3.32	0.16	3.49	0.15
37	MNRU=6dB	1.59	0.12	1.81	0.12
38	MNRU=12dB	2.92	0.14	3.12	0.12
39	MNRU=18dB	4.08	0.11	4.07	0.11
40	MNRU=24dB	4.66	0.10	4.70	0.07
41	MNRU=30dB	4.76	0.10	4.77	0.06
42	Direct,clean	4.81	0.09	4.81	0.06
43	Direct+Noise	4.35	0.12	4.42	0.12

History

Document history		
V5.0.0	July 1998	Publication