ETSI TR 146 008 V5.0.0 (2002-06)

Technical Report

Digital cellular telecommunications system (Phase 2+);
Half Rate Speech;
Performance Characterization
of the GSM Half Rate speech codec
(3GPP TR 46.008 version 5.0.0 Release 5)

GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS

Reference
RTR/TSGS-0446008v500

Keywords
GSM

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

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 www.etsi.org/key .

Contents

Foreword

This Technical Specification 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

During five years of activity, the Traffic CHannel Half rate Speech (TCH-HS) Experts Group has produced a number of test plans and experiments to assess the performance of the candidate algorithms submitted for the GSM half rate standardization. An aid in this task was a large knowledge base made available from previous CCITT (now ITU-T) and ETSI activities on codec assessment (see annex A references 1) 2) 3) 4) 5)), plus the use of recommendations in the field (see annex A references 6) 7) 8)).

Here are reported 3 different phases of the standardization of the GSM half rate codec: Characterization Phase 1, Characterization Phase 2 and Verification phase. The selection of the codec candidate for the GSM half rate traffic channel was based on the results of the characterization phase 1. Test results reported hereafter are based on version 3.3 of the GSM half rate codec.

Characterization Phase 1 (Experiments 1 to 5): For characterization Phase 1, C-simulations of the candidate codecs were used as hardware implementations were not available at that time. The simulations were produced by MOTOROLA (USA) and Ericsson (Sweden) with support by MATRA (France). The following experiments were carried out:

- Experiment 1: Quality under error conditions (A-law, IRS);
- Experiment 2: Quality under error conditions (UPCM, No IRS);
- Experiment 3: Quality under tandeming conditions;
- Experiment 4: Quality under background noise conditions (ACR);
- Experiment 5: Quality under background noise conditions (DCR).

Characterization Phase 2 (Experiments 6 to 9): During Characterization Phase 2, a hardware implementation of the candidate algorithm was employed, provided by ANT (Germany). The following experiments were carried out:

- Experiment 6: Assessment of equivalent qdu;
- Experiment 7: Effect of tandeming with other standards;

- Experiment 8: Talker Dependency;
- Experiment 9: Assessment of DTX algorithm.

Verification phase: Further tests accompanied characterization Phase 1 and 2 to obtain a better knowledge of the characteristics of the GSM half rate codec and its performance under different operational conditions:

- Special background noise;
- Channel activity in DTX mode;
- Performance with DTMF tones;
- Performance with signalling tones;
- Delay;
- Frequency response;
- Complexity.

For the characterization tests, a practical "indirect" method of performance comparison between different codecs was adopted, that utilizes the Modulated Noise Reference Unit (MNRU) (see annex A reference 7)) as a reference degradation in a subjective experiment including the codecs under test.

NOTE: The MNRU is a device designed for producing speech correlated noise that sounds subjectively like the quantizing noise produced by log-companded PCM codecs. The device is subjectively calibrated for Mean Opinion Scores (MOS) against Q dB (where Q is the ratio of the speech to speech-correlated noise power). The "Equivalent Q" of the codecs under test can then be found from the corresponding MOS on the calibration curve of the MNRU.

It is well known that this procedure works as long as the reference degradation sounds similar to the degradation under test.

The MNRU provides the additional function of normalization across laboratories carrying out the same experiment, i.e. all MOS are converted to Equivalent Q (dB) and the results can be analysed statistically for differences between laboratories. An appropriate analysis of variance (ANOVA) was identified to evaluate the statistical significance of the experimental factors.

The aim was to show that the subjective performance of the GSM half rate algorithm is at least as good as that of the full rate codec over a selected set of conditions. To allow for experimental error, the half rate candidate had to perform better than 1 dB below the performance of the full rate (for the overall figure of merit) and better than 3 dB below the performance of the full rate for individual test conditions.

To model its use in a network, the half rate candidate codec had to be placed between either a ITU-T Recommendation G.711 [1] PCM coder and decoder, or a Uniform PCM, which provided the necessary A/D and D/A conversions. Source files of speech, produced either by using an "average" telephone set (called IRS - Intermediate Reference System) or a microphone showing a "flat" sending frequency characteristic (No IRS or "flat"), could then be processed through the different experimental conditions, for presentation to subjects in listening experiments. Among the different experimental conditions were error conditions at different input levels under both IRS A-Law PCM and No-IRS Linear PCM audio parts, tandeming conditions for different error patterns and background noise conditions. During all phases of testing, the host laboratory functions for the processing were provided by Aachen University of Technology (RWTH at Aachen, Germany).

The whole set of "individual" and "global" data, collected in Experiment 1 to Experiment 9 were extensively analysed and discussed within TCH-HS expert group; for each condition, the MOS (or DMOS for Experiment 5) were computed, separately for male and female speech, as well as averaged together, and the effects of different factors and their interactions were subject to analysis of variance (ANOVA). Within characterization Phase 1, conversion to Q values and weighted averages were calculated for the whole set of results, in order to assess that the global figure of merit of the GSM half rate algorithm meets the quality requirement.

1 Scope

The present document gives background information on the performance of the GSM half rate speech codec. Experimental results from the characterization and verification tests carried out during the selection process by the Traffic CHannel Half rate Speech (TCH-HS) expert group are reported to give a more detailed picture of the behaviour of the GSM half rate speech codec under different conditions of operation.

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] ITU-T Recommendation G.711: "Pulse code modulation (PCM) of voice frequencies".

[2] ITU-T Recommendation G.726: "40, 32, 24, 16 kbit/s adaptive differential pulse code modulation".

[3] ITU-T Recommendation G.728: "Coding of speech at 16 kbit/s using low-delay code excited linear prediction".

3 Abbreviations

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

A/D	Analogue to Digital
ACR	Absolute Category Rating
ANOVA	ANalysis Of VAriance
C/I	Carrier-to-Interferer ratio
CEPT	Conférence Européenne des Postes et Télécommunications
CNI	Comfort Noise Insertion
D/A	Digital to Analogue
DAT	Digital Audio Tape
DCR	Degradation Category Rating
DSP	Digital Signal Processor
DTMF	Dual Tone Multi Frequency
DTX	Discontinuous Transmission for power consumption and interference reduction
EID	Error Insertion Device
ETSI	European Telecommunications Standards Institute
GBER	Average gross bit error rate
GSM	Global System for Mobile communications
IRS	Intermediate Reference System, No IRS= rather flat
HLCS	Host Laboratory Control System
ITU-T	International Telecommunication Union - Telecommunications Standardization Sector
MNRU	Modulated Noise Reference Unit
MOS	Mean Opinion Score
MS	Mobile Station
OVL	Overload point
PCM	Pulse Code Modulation

Q	Speech-to-speech correlated noise power ratio in dB
qdu	quantization distortion unit
RPE-LTP	Regular Pulse Excited codec with Long Term Prediction
SCD	Signal Conditioning Device
SFC	Sending Frequency Characteristic
SID	Silence Descriptor
SMG	Special Mobile Group
SNR	Signal to Noise Ratio
TCH-HS	Traffic CHannel Half rate Speech
TDMA	Time Division Multiple Access
UPCM	Uniform or Linear PCM
VAD	Voice Activity Detector
wMOPs	Weighted Million OPerations per second

Four different Error Patterns (EP0, EP1, EP2 and EP3) were used, where:

- EP0 without channel errors;
- EP1 C/I=10 dB; 5 % GBER (well inside a cell);
- EP2 C/I= 7 dB; 8 % GBER (at a cell boundary);
- EP3 C/I= 4 dB; 13 % GBER (outside a cell).

4 Quality under error conditions (A-law, IRS), Experiment 1

A listening-only test was chosen, adopting the Absolute Category Rating (ACR) method.

Subjective tests were carried out by BT (United Kingdom), CSELT (Italy), and Deutsche Telekom (Germany). Table 1 reports the results obtained in Experiment 1: each cell shows the difference in terms of equivalent Q values between the candidate and the full rate, negative values meaning worse performance than the full rate.

Table 1: Results from experiment 1 (A-law, IRS)

	Input Level (dB relative to OVL)		
Error Pattern	-12	-22	-32
EP0	-0,27	-0,02	0,34
EP1	-0,26	-0,86	-0,59
EP2	-0,49	-1,61	1,14
EP3	-0,39	1,79	3,80

NOTE: The figures in table 1 indicate DQ values in dB, where DQ = Q_{HR} - Q_{FR}.

In general, the candidate codec performed equally well or slightly worse than the full rate (in any case never exceeded the -3 dB limit).

5 Quality under error conditions (UPCM, No IRS), Experiment 2

A listening-only test was chosen, adopting the Absolute Category Rating (ACR) method.

Subjective tests were carried out by BT (United Kingdom), CSELT (Italy), and DEUTSCHE TELEKOM (Germany). Table 2 reports the results obtained in Experiment 2: each cell shows the difference in terms of equivalent Q values between the candidate and the full rate codec, negative values meaning worse performance than the full rate codec.

Table 2: Results from experiment 2 (UPCM, No IRS)

	Input Level (dB relative to OVL)		
Error Pattern	-12	-22	-32
EP0	-1,13	-2,90	-1,70
EP1	**-3,72**	-1,72	-1,21
EP2	-1,93	-1,79	0,69
EP3	0,79	1,49	3,59

NOTE: The figures indicate DQ values in dB, where DQ = Q_{HR} - Q_{FR}.

In general, the candidate codec performed equally well or slightly worse than the full rate (in one case, at -12 dB relative to Overload point (OVL) in EP1 condition, Q_{HR} - Q_{FR} exceeded the -3 dB limit).

6 Quality under tandeming conditions

6.1 Quality under tandeming conditions, Experiment 3

A listening-only test was chosen, adopting the Absolute Category Rating (ACR) method. Subjective tests were carried out by BT (United Kingdom), CSELT (Italy), and DEUTSCHE TELEKOM (Germany). Table 3 reports the results obtained in Experiment 3: each cell shows the difference in terms of equivalent Q values between the candidate and the full rate, negative values meaning worse performance than the full rate.

Table 3: Results from experiment 3 (Tandem Conditions)

	A-Law PCM (with IRS) DQ (dB) = (HR+HR)-(FR+FR)			**Linear PCM (No IRS) DQ (dB) = (HR+HR)-(FR+FR)**		
	Input Level (dB relative to OVL)			Input Level (dB relative to OVL)		
Error Pattern	-12	-22	-32	-12	-22	-32
EP0	-0,14	-0,56	-0,03	**-5,20**	**-5,43**	**-3,89**
EP1	-0,46	-0,75	0,49	**-4,98**	**-4,14**	-2,79

NOTE: The figures indicate DQ values in dB, where DQ = Q_{HR} - Q_{FR}.

In general, two candidate codecs in tandem performed equally well or slightly worse than two full rate codecs in tandem for the A-Law IRS audio part, while in most cases exceeded the -3 dB limit for the Uniform PCM No IRS audio part.

In operating networks, A-law coding and decoding is performed between both speech processing steps in both mobile to mobile calls. Therefore, the results of real network configurations are expected to be somewhere in between the figures obtained using the A-law input speech material and those obtained using the linear PCM speech material for each condition.

6.2 Effect of tandeming with other standards, Experiment 7

The experiment was conducted in two different laboratories: BT (UK) and CNET (France).

The following standards were tandemed with the half rate codec in this experiment: half rate, full rate, ITU-T Recommendation G.726 [2] (at 32 kbit/s) and G.728 [3]. Both possible orders of tandeming were tested for each of these cases, in both error free and EP1 conditions. The error pattern EP1 was only applied to the full and half rate codecs.

The main conclusion that can be drawn is that the performance is always better when the half rate codec follows the other codec in the tandeming chain. This effect is most pronounced at the higher speech input level (12 dB below overload point).

7 Quality under background noise conditions

7.1 Experiments 4 and 5

International subjective test programs have been conducted in the past, by both the ITU and ETSI, to investigate the effects of environmental noise. This has proved to be a difficult area to evaluate, and more satisfactory methodologies are continually being sought to improve the accuracy of these tests. Several methodologies have been used recently to investigate this factor:

a) the ACR (Absolute Category Rating) method using the classical Quality scale (second selection phase of the GSM half rate speech coding algorithm candidate, 1992);

b) the ACR method using the Listening Effort scale (second pre-selection test of the GSM Half Rate candidate, 1992);

c) the DCR (Degradation Category Rating) method such as in the ITU-T test methodology for the 16 kbit/s and 8 kbit/s speech coders which is an adapted version of the standard DCR procedure (described in ITU-T Recommendation P.80) and where several types of noise at different Signal-to-Noise ratios were evaluated in a unique experiment;

d) the DCR procedure adapted such as in the first pre-selection phase of testing for the GSM half rate candidates in 1991, where only one distinct noise has been tested in the same experiment in order to prevent the noise from being the predominant factor within the test; two experiments were, then, designed to take into account two types of noise: babble noise at a SNR of 30 dB and vehicle noise at a SNR of 10 dB.

Analysis of results gathered from these four experimental designs led to the conclusion that the last procedure - DCR test per noise (d) - is the most appropriate one to study the effects of environmental noise on a codec's behaviour.

For the final characterization phase of testing, it was decided to follow up two methodologies: the ACR and the DCR methods, i.e. to formally compare two distinct modes of collecting the subjects' responses with exactly the same experimental test plan (four 24 x 24 interleaved graeco-latin squares). The following environmental noises were considered of interest: office babble, vehicular, and traffic.

A listening-only test was chosen, adopting, for Exp. 4, the Absolute Category Rating (ACR) method, and subjective tests were carried out by BT (United Kingdom) and DEUTSCHE TELEKOM (Germany), while a modified version of the Degradation Category rating (DCR) was agreed for Exp. 5, and subjective tests were carried out by CNET (France) and CSELT (Italy).

Table 4 and 5 report the results obtained in experiment 4 and 5, respectively: each cell shows the difference in terms of equivalent Q values between the candidate and the full rate, negative values meaning worse performance than the full rate.

Table 4: Results from experiment 4 (ACR)

Noise	Office Babble	Vehicular	Traffic
Low noise	-0,78	-2,19	-1,06
High Noise	-1,75	-0,87	-1,25
Low Noise Tandem	-1,75	-2,38	-2,66
High Noise Tandem	-2,99	**-4,10**	**-3,09**

NOTE: The figures indicate DQ values in dB, where DQ = Q_{HR} - Q_{FR}.

Table 5: Results from experiment 5 (DCR)

Noise	Office Babble	Vehicular	Traffic
Low noise	-2,10	-2,96	**-4,53**
High Noise	-2,79	-2,83	-2,04
Low Noise Tandem	**-4,03**	**-4,39**	**-5,31**
High Noise Tandem	**-4,96**	**-5,85**	**-5,68**

NOTE: The figures indicate DQ values in dB, where DQ = Q_{HR} - Q_{FR}.

The main conclusion that can be drawn is that the performance of the half rate codec is (always) worse than that of the full rate, the amount of perceived degradation, in terms of DQ in dB, depending on the method chosen for the test (DCR being clearly more discriminant than ACR). Such background noise effect is most pronounced in tandem conditions.

7.2 Special background noise

7.2.1 Introduction

Some informal listening sessions were carried out to further investigate background noise effects. Speech samples from four different talkers were electronically mixed (at 3 different Signal-to-Noise Ratios; 5 dB, 10 dB, and 20 dB) with a wide range of different background noises, reflecting the following types of environment:

- Industrial Setting;
- Babble (offices and public places such as airports);
- Trains;
- Cars and Lorries;
- Roadside.

These were processed through a simulation of the Half Rate codec (with no DTX) and were listened to (on an informal basis) under controlled listening conditions using headphones.

No formal method of voting or opinion collation was employed; observations were simply noted.

7.2.2 Observations

At the lower Signal-to-Noise Ratios, the speech was often unintelligible without considerable concentration and effort on the part of the listener. In some cases, even where the listener was familiar with the speech material, it was impossible to understand some parts of the speech.

The codec had the effect of making the background noises sound "babbley", which, for example, made most background noises sound more "busy". This effect was particularly bad at 5 dB SNR. At 10 dB, the listening was more comfortable although parts of it were still difficult to understand. At 20 dB the speech was clearly understandable, although the noise was still "babbley".

For the -12 dB and -22 dB input levels, peak clipping also distorted the speech. Understandably, this effect was worse for the higher input level and for the higher Signal-to-Noise Ratios.

It must be particularly remembered when considering these results that the listening was informal and used headphones, not a handset. Also, the use of electrically summed speech and noise will not give the same results as would have been obtained if the speech used had actually been recorded in the noisy environment.

8 Assessment of equivalent qdu, Experiment 6

The experiment on the assessment of qdu was designed to assess the half rate codec performance, in error free conditions, in terms of Equivalent Quantization Distortion Units (qdu) as defined by the ITU-T. Two laboratories performed the experiment (CSELT and DEUTSCHE TELEKOM) and the following conclusions were drawn from their results:

a) For single encoding, the half rate codec was judged to be statistically equivalent to the full rate. Similar planning rules could therefore be applied to both algorithms if the configuration is not mobile-to-mobile. The figure of equivalent qdu for the half rate codec was found to lie somewhere between 8 and 16 qdu. A more precise figure could not be determined due to differences in the results from the two laboratories.

   (It is reminded that for the full rate an "average" figure of 7-8 qdu was indicated by SCEG to GSM, after considering test results showing values between a minimum of 4-5 qdu and a maximum of 21-22 qdu).

b) For tandemed conditions, a statistically significant difference in performance between the half and full rate codecs was detected in one of the two laboratories. The results confirmed that a noticeable degradation in speech quality in mobile-to-mobile connections is likely.

Generally, both the Half- and Full-Rate showed a worse performance than the other standards (ITU-T Recommendations G.711 [1], G.726 [2], at 32 kbit/s, and G.728 [3]) included in the experiment.

9 Talker dependency, Experiment 8

From the results obtained in the two laboratories which conducted this experiment, the performance of any given condition undoubtedly varies from talker to talker.

The existence of this talker dependency has been confirmed by a further analysis applied to the results from the first phase of characterization testing.

Under error free conditions, it was shown in the tests carried out, that the talker dependency for the half rate codec is similar to that for the full rate.

10 DTX System

10.1 Assessment of DTX algorithm, Experiment 9

The four laboratories who performed the subjective evaluation of DTX functions concentrated their expert listening on the following effects, using conversational speech;

- Voice Activity Detection (VAD); and
- Comfort Noise Insertion (CNI).

For this, the speech material available was monitored for the following effects:

- speech clipping;
- noise quality;
- noise contrast.

The tests showed that malfunctions of the VAD and the CNI were only predominant with low SNRs. The VAD functions appeared to work well in most situations (i.e. rather little clipping). In many situations, the Comfort Noise Insertion did not operate properly, being poorly matched in terms of quality and/or level. The DTX performed better with hand-held terminals relative to its performance with hands-free.

10.2 Channel activity in DTX mode

10.2.1 Test procedure

Speech material recorded during testing of the full rate DTX system was processed through the codec/DTX hardware. This material comprised real conversations in the English, French, German and Italian languages. The activity of the VAD algorithm was measured for all 480 conversations. The mean channel activity was then calculated by means of a software simulation of the TX DTX handler.

10.2.2 Speech channel activity

The percentage of speech frames scheduled for transmission by the radio sub-system (subsequently referred to as the speech channel activity) varied significantly between conversations. Speech channel activities ranged from 35 % to 85 % for individual sides of a conversation. For this reason, it was not possible to identify any significant trends in the results with regard to terminal type and environmental conditions. The mean speech channel activity, measured over all 480 conversations, was approximately 55 %.

10.2.3 Level compensation

During the expert listening, it was found that the speech material had been processed at a level 6,5 dB below the original recorded level. However, the activity of the basic VAD algorithm rises approximately 0,5 % per dB increase in input level. To compensate for this, a factor of 3 % must be added to the speech channel activity estimate.

10.2.4 SID update rate

The DTX handler simulation used a SID update period of 480 ms. The SID update rate has subsequently been reduced to 240 ms. This modification will raise the speech channel activity by approximately 2 %.

10.2.5 Interleaving compensation

The channel measurements were calculated on a signal frame basis. However, the use of interleaving (depth 4) implies that the TDMA activity will be approximately 2 % higher than the signal frame activity.

10.2.6 Estimated mean TDMA channel activity

The estimated mean TDMA channel activity is shown in table 6.

Table 6: Calculation of mean TDMA channel activity

speech channel activity	55 %
level compensation	3 %
240 ms SID update period	2 %
interleaving compensation	2 %
total TDMA channel activity	62 %

11 Performance with DTMF tones

11.1 Introduction

In the fixed telephone system, DTMF (Dual Tone Multi Frequency) signals are transmitted in the speech channel for signalling. This has led to the use of DTMF tones for applications such as the control of answering machines and mail/messaging boxes. In the GSM system, the handling of these signalling tones is dependant on the direction the signal is travelling. If it is in the uplink (from the mobile station to the network), the signalling channel is used, rather than the speech channel. In the downlink (from the network to the mobile station), these tones are carried in the speech channel. Even though it was not a requirement for the half rate speech channel to be able to carry these tones in the downlink, their transmission was tested.

11.2 Test set-up

16 DTMF signals are defined representing the 10 numeric keys, the characters "A", "B", "C", "D", "*" and "#". Each digit consists of two sine signals of distinctive frequencies, one chosen out of 4 values from the low frequency group (or row frequency), and one out of the 4 values from the high frequency group (column frequency). Both frequencies are sent simultaneously ideally with same amplitude and at exact frequency values. For practical use, certain tolerances of the frequencies and of the signal amplitudes are specified.

A DTMF receiver must be capable of detecting these tones. It should detect all the DTMF tones even under noisy conditions or when speech is present. Also, it should not interpret other signals from the voice band as valid tones. The tones can only be distinguished by their specific frequency and amplitude composition so it is important, if they are to be recognized by the half rate system, that they conform to the CEPT recommendation T/CS 46-02 (1985). Among others, the difference in the amplitudes of the 2 components (twist) shall not exceed 6 dB. The minimum signal length from sending unit is 75 ms while a 40 ms signal should be detected at the receiver side. Pauses from the generator shall last 65 ms while the receiver shall detect 20 ms.

The DTMF tests were done at nominal frequencies with different pulse and pause duration and different amplitude levels on a PC based set-up. DTMF signal files were generated by means of a DTMF software package for the 16 signals with 10 samples for each tone. After processing with the HR-codec software, the result files were input to a DSP based hardware with a standard DTMF recognition S/W meeting CEPT requirements. All experiments were done also with modified DTMF receiver software. The tables in subclause 11.5 list the number of recognized tones.

All dB values mentioned are for each individual component of the DTMF signal, with reference to the overload point.

11.3 Results

The results of the test with a standard DTMF detector are shown in table 7. Even at ideal conditions with nominal DTMF signal frequencies, no additional signals in the speech band, and error free transmission, the recognition is poor after processing. Only with a relatively high level of -12 dB and a tone length of 80 ms is a 100 % recognition achieved. Under all other conditions at least one tone shows severe problems. There is no linearity in this experiment, e.g. "4" is recognized well at -18 dB level but very poor at -22 dB while "7" shows the opposite behaviour. Also, when the twist is reversed, the results differ in ways which depend on the code being transmitted. The recognition of very short tones (40 ms) is not acceptable, and the longer tones (120 ms) are problematical too.

A reason for the poor behaviour might be a time dependent twist generated by the GSM Half Rate codec when one of the two components develops differently from the other due to the non-linear behaviour of the codec. For more than 40 ms the twist at certain DTMF tones was observed to be greater than 6 dB and thus out of the allowed range of the specification of standard DTMF receivers. In experiment (h) with -12/-18 dB signals and 120 ms tones, 10 inputs of "A" resulted in 12 recognitions. A slow oscillation of the signal amplitudes may have generated a twist of more than 6 dB for longer than 20 ms. This made the detector observe a valid pause and a new tone, increasing the number of detected tones above the number of input tones. This might have happened also for other tones under the condition (h) where e.g. 10 detected tones may result of 8 correct detections, a double detection from one input and one failure. The test equipment could not decide such effects - as also in practise just the result counts. At 80 ms twist signals, such slow oscillations do not have the same effect because under no condition a valid 2nd tone can be detected (40+20+40>80).

Table 8 shows the results of the same experiments as described above with a DTMF detector tuned for recognition in GSM half rate speech codec transmission. Using knowledge of the possible reasons for detection errors in the tuned detector, the detection rate was improved. However, even at the still ideal signal conditions as described above, the results were not satisfactory where there was severe twist or short (40 ms) tones. Also, the modifications may well increase the acceptance of non-DTMF signals as valid DTMF tones. This, however, was not tested.

11.4 Conclusions

With the standard detector the recognition rate averaged over all experiments was 74 %. The tuning of the detector for the half rate channel characteristics could improve the detection rate to 92 %. As all experiments still had rather ideal conditions, in real application an even lower rate for detection has to be assumed, also due to the misinterpretation of other signals in the modified detector.

In conclusion, a serious commercial application using DTMF in the speech channel should not be supported with the GSM half rate codec.

11.5 Result tables of experiments with standard and modified DTMF detectors

The tables below list the numbers of detected tones from 10 input signals at each tested condition. For twist conditions, the pair of level figures indicate the level of row frequencies and column frequencies respectively.

Table 7: Summary of DTMF tests with standard DTMF detector

Condition\Tone	**1**	**2**	**3**	**A**	**4**	**5**	**6**	**B**	**7**	**8**	**9**	**C**	*****	**0**	**#**	**D**	**Total**
(a) -12 dB, 40 ms	5	2	9	10	0	8	1	5	10	5	3	0	0	10	10	9	87
(b) -12 dB, 80 ms	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	160
(c) -18 dB, 80 ms	10	10	10	10	10	10	8	10	1	10	10	10	10	10	10	10	149
(d) -22 dB, 80 ms	10	10	10	10	4	10	10	10	10	10	10	10	10	10	10	10	154
(e) 3 dB twist, -12/-15 dB, 80 ms	10	10	10	10	10	10	10	10	7	10	10	10	0	10	10	10	147
(f) 6 dB twist, -12/-18 dB, 80 ms	0	1	0	10	0	10	0	10	0	10	0	4	0	0	10	0	55
(g) 6 dB tw. reverse, -18/-12 dB, 80 ms	10	10	4	10	6	3	0	0	0	0	10	10	0	0	6	10	79
(h) 6 dB twist, -12/-18 dB, 120 ms	8	3	9	12 note	7	10	5	10	0	10	7	9	3	10	10	9	122
Total	63	56	62	82	47	71	44	65	38	65	60	63	33	60	76	68	953

NOTE: 10 input signals in this test case resulted in 12 recognized tones. An explanation is given in subclause 11.3.

Table 8: Summary of DTMF tests with modified DTMF detector

Condition\Tone	**1**	**2**	**3**	**A**	**4**	**5**	**6**	**B**	**7**	**8**	**9**	**C**	*****	**0**	**#**	**D**	**Total**
(a) -12 dB, 40 ms	9	10	8	1	0	8	4	10	10	10	8	9	8	10	6	8	119
(b) -12 dB, 80 ms	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	160
(c) -18 dB, 80 ms	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	160
(d) -22 dB, 80 ms	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	160
(e) 3 dB twist, -12/-15 dB, 80 ms	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	160
(f) 6 dB twist, -12/-18 dB, 80 ms	10	10	10	10	10	10	6	10	0	10	6	10	0	10	10	0	122
(g) 6 dB tw. reverse, , -18/-12 dB, 80 ms	10	10	10	10	10	10	10	10	7	10	10	10	7	0	10	10	144
(h) 6 dB twist, -12/-18 dB, 120 ms	10	8	10	10	10	10	8	10	8	10	10	10	8	10	10	10	152
Total	79	78	78	71	70	78	68	80	65	80	74	79	63	70	76	68	1177

12 Performance with signalling tones

The capability of the codec to transmit network information tones was assessed with 5 French signalling tones following the IUT-T recommendation: "Warn" tone, "Busy" tone, "Ring" tone, "Inf" tone and "Pay" tone (tones of length 200 ms to 1s and silences between tones of length 30 ms to 4 s).

All signalling tones are recognized. However, the half rate codec introduces a very audible distortion and performs significantly worse than the full rate codec.

None of the tones is perturbed by the VAD/DTI system.

13 Delay

[tbd]

14 Frequency response

The frequency response of the GSM half rate codec, has been evaluated by computing the logarithmic gain.

The codec has been tested in error free condition only, without DTX associated, by independently processing 198 sine waves files spaced by 20 Hz and spanning the range between 50 to 3 990 Hz. Each file had a duration of 8 seconds and the input signal level was fixed at -22 dB (V_{max} = 2 603).

The gain of the codec has been calculated by means of the formula:

$$gain = 10 * \log_{10}\left(\frac{\sum_i (out_i)^2}{\sum_i (inp_i)^2}\right)$$

Figures 1 and 2 report the logarithmic gain for the whole range of tones considered and for telephone bandwidth respectively.

Both figures show that the codec provide a flat frequency response in the telephone bandwidth, with the algorithmic gain confined in the range ±0,2 dB with a very few outliers.

The highest attenuation observed is 0,65 dB and occurs at 1 150 Hz.

It shall be noted that small deviations from these figures can be observed by using different levels and/or different initial phases for the sinewave signals.

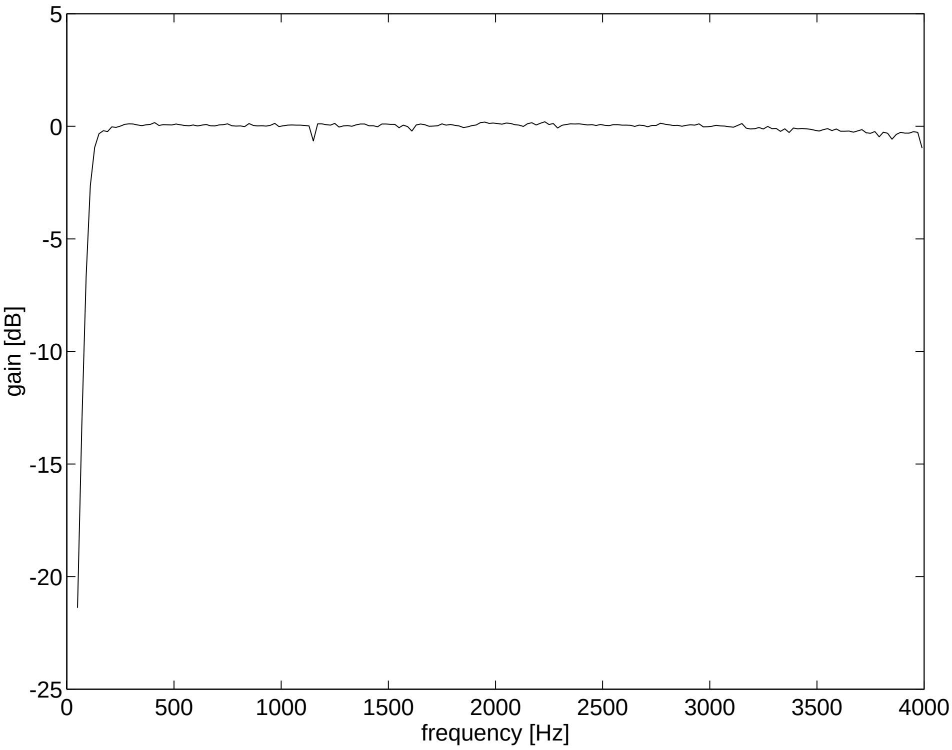

Figure 1: Frequency response for the whole bandwidth considered

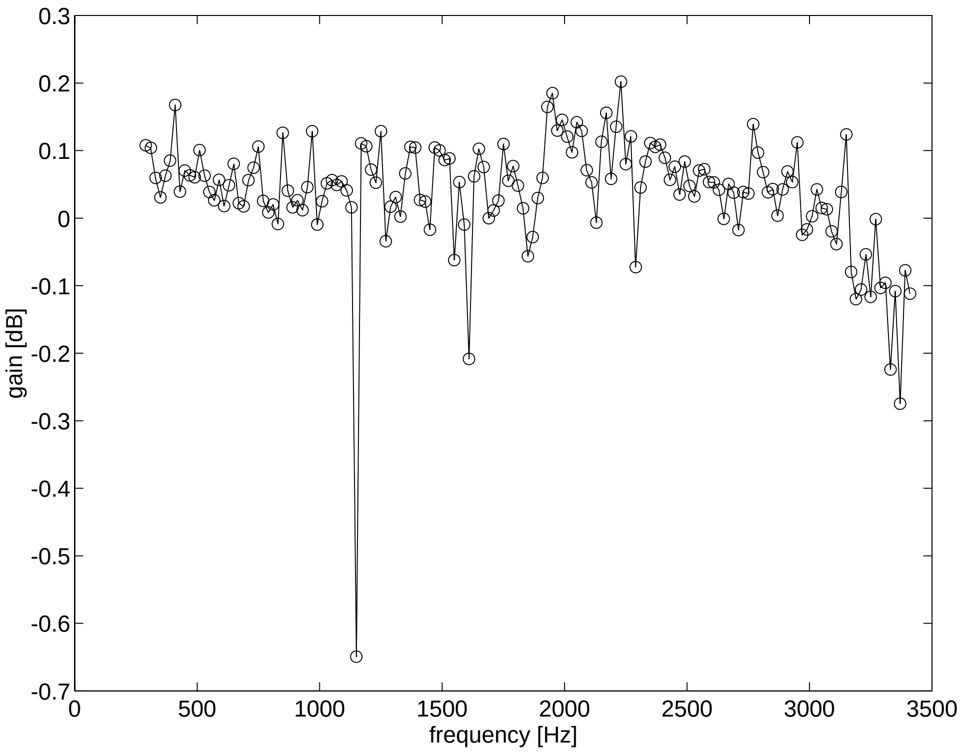

Figure 2: Frequency response in the telephone bandwidth

15 Half Rate codec complexity

The complexity of the half rate codec is characterized by the 3 following items:

- the number of cycles;
- the data memory size;
- the program memory size.

The values of these different figures depend on a specific DSP implementation. Nevertheless, the results obtained by the C description analysis can be used as references.

The speech transcoding functions are specified using a set of basic arithmetic operations. The wMOPs figure quoted is a weighted sum of the operations required to perform transcoding. The weight assigned to each operation is representative of the number of instruction cycles required to perform that operation on a typical DSP device.

The complexity range of the half rate codec is approximately 4,5 times that of the full rate codec.

The number of cycles required by the half rate algorithm is highly dependent on the values of input samples. The execution time of an average and an extreme input case may differ by up to 20 %.

That is why, to evaluate the complexity, it is necessary to compute the theoretical worst case, i.e. the maximum possible number of cycles, and not just observe the results of a simulation.

The principal figures of this evaluation are the following:

Table 9: Principal figures of evaluation

	Theoretical worst case wMOPs	Data RAM (note) (16 bits words)	Data ROM (constants) (16 bits words)	Program ROM (assembly instructions)
Speech and channel half rate codec (excluding DTX functions)	21,2	5 002	8 781	8 000-12 000
Ratio half rate vers. full rate	4,5	2,4	9,7	4

NOTE: The Data RAM figure can be split in 2 parts: the static variables: 2 100 words; and the dynamic variables (i.e. local to a procedure): 2 900 words.

16 Summary of results from characterization Phase 1 and 2

The whole set of individual and global data were extensively analysed and discussed within the TCH-HS expert group. The effects of different factors and their interactions were subject to analysis of variance (ANOVA). Tables 10 to 14 report the results obtained in 9 experiments.

16.1 Summary of Results From Characterization Phase 1

The whole set of individual and global data were extensively analysed and discussed within the TCH-HS expert group. The effects of different factors and their interactions were subject to analysis of variance (ANOVA). Tables 10 to 14 report the results obtained in 9 experiments.

Table 10: Summary of Characterization Phase 1 Results - Differential Q values

Diff Q (dB)			
exc. (UPCM, No-IRS) Exp. 1 - 3	**EP3** (A law-IRS and UPCM, No-IRS) Exp. 1 - 3	**Noise only** Exp. 4 and 5	**All** Exp. 1 - 5
-1,09	-1,45	-3,01	-2,29

NOTE: The figures indicate DQ values in dB averaged over input level,

where $DQ = Q_{HR} - Q_{FR}$.

Table 11: Summary of Characterization Phase 1 Results (Exp. 1, 2 and 3)

	Single	Encoding	Conditions		Tandeming	Conditions	All	All
Audio part	**EP0**	**EP0/1**	**EP0/1/2**	**EP0/1/2/3**	**EP0**	**EP0/1**	**exc. EP3**	
1.A-Law IRS	+0,01	-0,32	-0,43	+0,12	-0,32	-0,34	-0,41	+0,02
2.NoIRS, LinearPCM	-2,16	-2,13	-1,82	-0,90	-4,98	-4,50	-2,49	-1,62
1 and 2.	-1,08	-1,22	-1,12	-0,39	-2,65	-2,42	-1,45	**-0,80**

NOTE: Dependence on Specific Conditions without Background Noise. The figures indicate DQ values in dB averaged over input level,

where $DQ = Q_{HR} - Q_{FR}$.

Table 12: Summary of Characterization Phase 1 Results (Exp. 4 and 5)-

Audio part	Office Babble	Vehicle	Traffic	No Tandeming	With Tandeming	All
A-Law IRS	-2,64	-3,19	-3,20	-2,10	-3,93	-3,01

NOTE: Differential Q values in Noise Conditions. The figures indicate DQ values in dB averaged over input level,

where $DQ = Q_{HR} - Q_{FR}$.

Table 13: Summary of Characterization Phase 1 Results - Significant differences (Experiment 1 to Experiment 5)

Laboratory	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
BT	HR = FR	HR < FR	HR < FR	HR < FR	x
CNET	x	x	x	x	HR < FR
CSELT	HR = FR	HR = FR	HR = FR	x	HR < FR
Deutsche Telekom	HR = FR	HR = FR	HR < FR	HR < FR	x
Global	HR = FR	HR < FR	HR < FR	HR < FR	HR < FR

NOTE: See legend in subclause 16.2 for symbol explanation.

16.2 Summary of Results From Characterization Phase 2

Table 14: Summary of Characterization Phase 2 Results (Experiment 6 to Experiment 9)

Subject:	qdu	Tandeming with other Standards	Talker Dependency	DTX Functions
Laboratory	Experiment 6	Experiment 7	Experiment 8	Experiment 9
BT		HR+any < any+HR	see clause 9	DTX operation appears to be satisfactory.
CNET		HR+any < any+HR		DTX fairly satisfactory, concerns over CNI.
CSELT	HR = FR HR+HR < FR+FR			DTX satisfactory, concerns over CNI and comfort noise quality.
DBP	HR = FR HR+HR = FR+FR		see clause 9	DTX fairly satisfactory, concerns over comfort noise quality.

Legend

Symbol **Definition**

=	no significant difference at the 95 % confidence level
HR	Half rate codec
FR	Full rate codec
x	Experiment not performed by laboratory
HR<FR	HR significantly worse than FR at the 95 % confidence level.
any	All tested codecs, except HR (G.726 [2], G.728 [3], and FR)

The candidate codec performed equally well or slightly worse than the full rate for most cases, the overall figure of merit being **-0,8 dB** (weighted) signal-to-quantization distortion (without taking into account the noise conditions). The requirement was to provide a half rate standard with speech quality approximately equivalent to the GSM full rate codec with 1 dB of tolerance in terms of equivalent (weighted) signal-to-quantization distortion.

Under UPCM No IRS audio part conditions, particularly when tandemed, the full rate performed consistently better than the half-rate.

In environmental noise conditions, formal tests using two different methods, ACR and DCR were used to determine the difference in performance between the full and half-rate systems. It was found that differential Q (dB) values (comparing full and half-rate codecs) are more pronounced when using the DCR procedure than when using the ACR procedure, leading to a larger measured difference between the systems in the DCR experiments. The half-rate always performed worse than the full rate under the noise conditions, often with the difference in performance falling outside the -3 dB limit.

TCH-HS significantly improved the methodology for measuring subjectively the performance of candidate codec. Since the most important requirements set by SMG and tested by TCH-HS were met by the optimized algorithm, SMG approved the optimized codec.

Further information can be found in annex A reference 9).

16.3 Conclusion

A subjective test methodology for the quality assessment of ETSI's half rate algorithm has been implemented, based on listening opinion tests.

The test methodology reflected international telephony assessment methods that are described extensively in the ITU-T Series P Recommendations, and that have shown to be suitable for characterizing both the GSM full rate and half rate algorithm performance. Results of tests conducted by several organizations showed consistency when normalized to Equivalent Q, in terms of the relative performance of the half rate algorithm and full rate RPE-LTP, removing the effect of differences in absolute performance, due to different languages, interpretation of quality scales, etc.

By considering the average performance across all countries, it was concluded that the half rate algorithm performance was comparable to RPE-LTP in all the experimental conditions tested, except for tandeming and background noise conditions, and met the initial requirements set out by SMG.

The results confirmed that the performance of the half rate codec falls short of that normally experienced on the PSTN.

For network planning purposes, it is proposed that the same rules will be adopted as for RPE-LTP, which will be adequate for most applications.

Annex A: Bibliography

1) IEEE Proceedings. GLOBECOM'84, Vol. 2, pp.778-785, G. Williams and H. Suyderoud (1984): "Subjective performance evaluation of the 32-kbit/s ADPCM algorithm".

2) IEEE Proceedings GLOBECOM '86, Houston, Texas, pp. 599-604, G. Modena, A. Coleman, P. Coverdale, P. Usai (1986): "Subjective performance evaluation of the 64 kbit/s (7 kHz) audio coder".

3) A. Coleman, N. Gleiss, P. Usai, Speech Communication, Vol. 7, , pp.151-166, (June 1988): "A subjective testing methodology for evaluating Medium Rate Codecs for Digital Mobile Radio applications".

4) IEEE Proceedings GLOBECOM '89, Dallas, Texas, pp. 1075-1079, A. Coleman, N. Gleiss, J. Sotscheck, H.Scheuermann, P. Usai (1989): "Subjective performance evaluation of the RPE-LTP codec for the Pan-European cellular digital mobile radio system".

5) C. South et alii (1993), Speech Communication Vol. 13, pp.113-132: "Subjective Test Methodology for a 16 kbit/s Speech Coder".

6) ITU-T Recommendation P.80 (1992): "Methods for subjective determination of transmission quality".

7) ITU-T Recommendation P.81 (1988): "Modulated Noise Reference Unit",

8) ITU-T Recommendation P.83 (1992): "Subjective performance assessment oftelephone-band and wideband digital codecs"

9) ICC '95, Session 2.2.2, P. Usai et alii (June 1995): "Subjective performance evaluation of the GSM Half Rate Speech Coding Algorithm (with voice signals)".

Annex B:
Change history

Change history					
SMG No.	**TDoc. No.**	**CR. No.**	**Section affected**	**New version**	**Subject/Comments**
SMG#16				4.0.0	ETSI Publication
SMG#20				5.0.0	Release 1996 version
SMG#27				6.0.0	Release 1997 version
SMG#29				7.0.0	Release 1998 version
SMG#31				8.0.0	Release 1999 version

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
03-2001	11				Version for Release 4		4.0.0
06-2002	16				Version for Release 5	4.0.0	5.0.0

History

Document history		
V5.0.0	June 2002	Publication