SUBJECTIVE PERFORMANCE EVALUATION OF THE GSM HALF-RATE CODING ALGORITHM (WITH VOICE SIGNALS)

Paolino Usai
(Cselt - Centro Studi E Laboratori Telecomunicazioni)
Via G. Reiss Romoli 274 - 10148 Torino, Italy
Fax: +39 11 228 6207, Phone +39 11 228 6214
Graham Cosier
(Bt) Martlesham Heath,
Ipswich, Suffolk IP5 RE, UK
Dominique Pascal
(Cnet) Technopole Anticipa,
2 Av. Pierre Marzin, 22307 Lannion Cedex, France
Jochem Sotscheck
(Deutsche Telekom) Ringbahnstrasse 130,
12103 Berlin, Germany
Michael Kappelan
(IND - Institut Für Nachrichtengeräte Und Datenverarbeitung)
Templergraben 55, 52056 Aachen, Germany

ABSTRACT. The Pan-European cellular digital mobile radio system GSM uses a codec with a net bit rate of 13.0 kbit/s (gross bit rate including error protection 22.8 kbit/s), known as the "full rate" RPE-LTP (Regular Pulse Excitation with Long Term Prediction). GSM is now ready to dub channel capacity with the adoption of a new algorithm as an ETSI Recommendation, the candidate codec being called appropriately "half-rate" (gross bit rate 11.4 kbit/s). Internationally co-ordinated series of subjective listening experiments were planned and carried out during such exercise. Four main phases were necessary, called qualification, selection(s), optimisation and characterisation. This paper describes the tests performed and gives an outline of the performance of the codec with voice signals under realistic network conditions.

The effects on the speech performance produced by the Voice Activity Detector and related DTX system are not the main subject of this paper but information on this topic can be found in the section containing the test results from the final characterisation phase.

INTRODUCTION

In 1989, ETSI established an ad-hoc group, called TCH-HS (a list of acronyms and a glossary can be found at the end of the paper), charged with achieving, by 1995 at the latest, a Recommendation on a speech coding algorithm suitable for the implementation in the Pan-European digital cellular mobile radio system GSM of half-rate speech traffic channels.

A set of guidelines and performance requirements were already laid down by the previous Speech Coding Experts Group, called SCEG, that provided GSM with the set of recommendations for the full-rate codec RPE-LTP at 13 kbit/s. From that time TCH-HS has produced quite a number of test plans and experiments to assess the performance of the different candidate codecs. An aid in this task was a large knowledge base made available from previous CCITT (now UIT-T) and ETSI activities on codec assessment [1][2][3][4][5][6][10] [11] [12], plus the use of recommendations in the field [7][8] [9].

Evaluations of communication systems were typically conducted to measure the optimal performance of that system. However, operational performance are degraded by environmental noise, active interference, and occurring distortions in transmission media.

To model its use in a network, the half-rate algorithm had to be placed between either a G.711 PCM coder and decoder, or an Uniform PCM, which provided the necessary A/D and D/A conversions. Source files of speech could then be processed through the different experimental conditions, for presentation to subjects in a listening experiment. The host laboratory functions for the processing were provided by Aachen University of Technology (RWTH, Germany).

The various factors affecting speech transmission quality were considered, taking into account also the 'context effects', like the auditory workload, the nature of the task for the subjects and the extent of speech degradation involved in the tests.

The primary requirement was to provide an half-rate standard with speech quality approximately equivalent to the GSM full rate codec described in ETSI Recommendations of the 06.10 Series, with 1 dB of tolerance, in terms of equivalent (weighted) signal-to-quantisation distortion Q.

A practical 'indirect' method of performance comparison between different codecs is to use the Modulated Noise Reference Unit (MNRU) [8] [9] as a reference degradation in a subjective experiment including the codecs under test.[1]

The MNRU provides the additional function of normalisation across laboratories carrying out the same experiment, i.e. all

1 The MNRU is a device designed for producing speech correlated noise that sounds subjectively like the quantising 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.

0-7803-2486-2/95 $4.00 © 1995 IEEE

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 test results.

PHASE I: QUALIFICATION

It was decided by the SMG that the speech codec should meet the following minimum performance criteria:
Gross bit rate: 11.4 kbit/s;
Interleaving: Not greater than 8;
Type of code: No constraints;
Delay: Overall (one-way) delay less than 90 ms, i.e. not more than the full rate;
Speech quality: the aim was to maintain the quality of the full-rate traffic channel.
In general a candidate had to exhibit the following characteristics:

- Codec performance basically independent of voice characteristics as well as languages;
- For a given listening level, speech quality substantially flat over the given range of input levels;
- Speech quality substantially independent of talker sex;
- Behaviour of the half-rate TCH as a function of C/I substantially comparable to the full-rate TCH.

At the beginning of the selection process, 14 candidate codecs were studied by the participating organisations. As only 6 codecs could be logistically handled by the host laboratory, a rank order was needed, and this resulted in organisations having to perform their own tests to demonstrate that the transmission requirements, set by SMG, could be met.
A detailed subjective test procedure for assessing the average quality was described, the quality being expressed as an average value of Q (dB of signal to noise ratio of the MNRU) averaged over a number of conditions representative of practical transmission situations.
Formal national listening opinion tests using both the Absolute Category Rating (ACR) and the Degradation Category Rating (DCR) methods were conducted and elaborated in January 1991.
Initially, 5 candidates were admitted, afterwards augmented to 6 after presentation of further test information to SMG; ANT, BT, CSELT/AT&T, MATRA/ERICSSON, MOTOROLA and PKI candidates entered the selection phase.

THE HOST LAB

To assess the quality and behaviour of a codec candidate, the designated Host Lab had to perform various tests. In the case of the GSM Half-rate Codec selection and characterisation the Host Lab was organised by the "Institut für Nachrichtengeräte und Datenverarbeitung" (IND) at Aachen University of Technology (RWTH).
In the Test Configuration shown in figure 1 the individual codec candidates are connected via standardised parallel interfaces to the test equipment consisting of three DSP systems.
The speech samples are transferred from the Host Lab Control System (HLCS) hard disk to the Signal Conditioning Device (SCD), which is realised with DSP system 1 (three DSP 56001 and one DSP 96002) to perform real-time operations such as

- PCM filtering
- interpolation / decimation
- quantization (linear and A-law)
- conversion to DAT format

Then the samples are processed via the codec candidate (configuration 1) with the possibility to introduce channel errors with the second DSP system (Error Insertion Device, EID). Finally the samples are fed back to the SCD and stored on the HLCS hard disk and then transferred off-line to the SUN workstation with Exabyte tape drive.
The second DSP system consists of two DSP 96002, two DSP 56001, one DSP 32C, one DSP 16 and can also be used as reference system (configuration 2) to realise codecs such as G.711, G.726, G.728, GSM full rate (with EID) and MNRU (Modulated Noise Reference Unit).
During the Characterisation Phase speech conversations were directly (digital signal) read from DAT tapes and sent to the codec chain (configuration 3) in order to produce the processed material that was used by expert listeners in order to assess the quality of the implemented DTX algorithm. This task has been performed by the third DSP system which consists of four DSP 56001 to obtain

- digital DAT interfaces
- interpolation / decimation
- delay compensation and level measurement

Since several input filter characteristics and background noises are essential for codec testing, these databases are generated on a workstation whereas the Host Lab Control System stores all information necessary to control the complete test equipment.
During the third selection phase about 35000 different speech samples of 8.5 s each had to be processed. Since the data was stored with a sampling frequency of 16 kHz in order to allow asynchronous multiple transcodings, the resulting data in this test was about 10 GBytes.

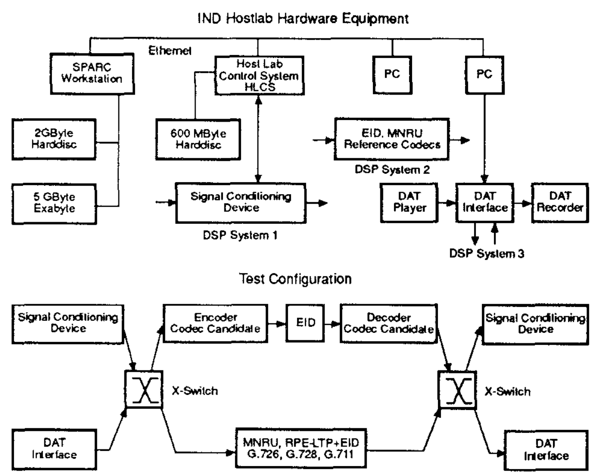

Fig. 1 - Hostlab System

PHASE 2 TESTING: SELECTION

The primary function of this phase of testing was to produce realistic test plans in order to select a codec suitable for GSM applications from the 6 candidates.

More and more sophisticated test plans were finalised and subjective experiments were designed, including all procedures and analysis, then allocated to various testing laboratories.
RWTH acted always as host laboratory, processing source (speech) material through all the conditions defined in the test plans.
Each of the experiments was conducted in several languages drawn from the following set of six: Dutch, English, French, German, Italian, and Swedish.
Three years (and three rounds) of selections were necessary, lasting from January 1991 to December 1993.
Every year a reduction of the number of candidates was achieved. After the first selection (and a new pre-selection) the number passed from 6 to 4 (ANT, AT&T/PKI/CSELT, MOTOROLA and TFL); after the second selection, from 4 to 2 (ANT and MOTOROLA), and, after the third selection, the Motorola candidate was chosen as the basic algorithm for the following optimisation phase.
Full data from all participating laboratories were available and an extensive analysis was carried out.
For the third selection phase, three criteria were examined by 3 experiments:

Exp. 1 i) the effect of bit errors at different input levels under IRS A-Law PCM audio part.
Exp. 2 ii) the effect of bit errors at different input levels under No IRS (flat) and Linear PCM audio part.
Exp. 3 iii) the effect of asynchronous tandeming with two bit error patterns (error free and C/I=10 dB) at different input levels under both IRS A-Law and No IRS Linear PCM audio parts.

A listening-only test was chosen, adopting the Absolute Category Rating (ACR) method for all the criteria. The experiments employed four 24 x 24 interleaved Graeco-Latin Squares, first designed by BT, to achieve a balance over all the conditions and talkers. Each listener cast 96 votes, i.e. 24 conditions x 4 talkers x 1 listening level.
The ACR method used employs a five-points quality scale (Excellent, Good, Fair, Poor and Bad or Unsatisfactory) assigned to corresponding values 5 to 1, for computing MOS. After normalisation to Equivalent Q, via the MNRU data, a global analysis showed no significant differences between different laboratories carrying out the same experiment, enabling the averaging over laboratories to improve precision.
Tables 1, 2, 3 & 4 report final results from the third selection.
Table 1 summarises the overall 'speech quality' performance. Tables 2 to 4 contain the codecs' partial performance for each listed item.
Given the measured performance, it was necessary to improve the Motorola algorithm to enable up to 2 codecs to operate in tandem, with the No IRS Linear PCM audio part.

Phase 2: Experimental results from selection

codec	dQ (dB)
ANT	-1.9
Motorola	-0.7

Table 1: Overall differential Q ratings

Note: a negative number in this text context indicates lower performance than the Full-Rate.
The values obtained in this analysis are not directly comparable to previous results, since the codecs, the listeners, and the condition sets used in the experiments are different.

Dependence on specific conditions

The weighted average dQ(dB) value was computed over a number of subsets of the test conditions, these results are shown in Tables 2 to 4 below. These values are averaged over all input levels.

Codec	Single encoding conditions				Tandem		All	
	EP0	EP0/1	EP0/1/2	EP0/1/2/3	EP0	EP0/1	exc. EP3	All
ANT	-1.42	-1.57	-1.67	-1.74	-1.40	-1.51	-1.63	-1.69
Motorola	-0.43	-0.38	-0.16	+0.47	-0.40	-0.34	-0.20	+0.31

Table 2: A-Law IRS

Codec	Single encoding conditions				Tandem		All	
	EP0	EP0/1	EP0/1/2	EP0/1/2/3	EP0	EP0/1	exc. EP3	All
ANT	-1.61	-1.62	-1.70	n/a	-3.87	-3.30	n/a	-2.10
Motorola	-2.13	-1.66	-1.32	n/a	-4.25	-3.70	n/a	-1.91

Table 3: No IRS Linear PCM

Codec	Single encoding conditions				Tandem		All	
	EP0	EP0/1	EP0/1/2	EP0/1/2/3	EP0	EP0/1	exc. EP3	All
ANT	-1.52	-1.60	-1.68	-1.72	-2.63	-2.41	-1.86	-1.87
Motorola	-1.28	-1.02	-0.74	-0.30	-2.32	-2.02	-1.06	-0.68

Table 4: A-Law IRS, No IRS Linear PCM

Note 1: EP0: without channel errors
EP1: C/I=10dB; 5% GBER (well inside a cell)
EP2: C/I= 7dB; 8% GBER (at a cell boundary)
EP3: C/I= 4dB; 13% GBER (outside a cell)
where: (GBER: = average gross bit error rate)
and (C/I: = carrier to interferer ratio)

Figure 2 shows the main results, in terms of average Q ratings, obtained from the Motorola candidate in the Phase 2 of tests.
An important feature to note from these graphs is the influence of the sending frequency characteristic (IRS vs. No IRS) and of the I/O interface (A-law PCM vs. Linear PCM); it was confirmed that for the tandemed connections of the codecs, the quality was worse than the acceptable individual limits set by SMG (full rate - 3 dB), with the No IRS Linear PCM audio part.
On grounds of results from informal verification tests conducted during this phase by TCH-HS, SMG also requested some improvements on the behaviour of the final candidate with different background noise conditions and as concerns talker dependency.

PHASE 3 TESTING: OPTIMISATION

Rather than change the measurement criteria, the designers of the TCH-HS carried out an upgrade programme on the codec trying to achieve what they believed to be a significant improvement with two different modifications of the coder, the latter foreseen for improvement in background noise conditions; this resulted in two versions of the candidate to be tested, HC1 or version 3.3 and HC2 or version 3.5.
The codec candidates were tested with C-simulations due to the lack of existing hardware implementations. The simulations

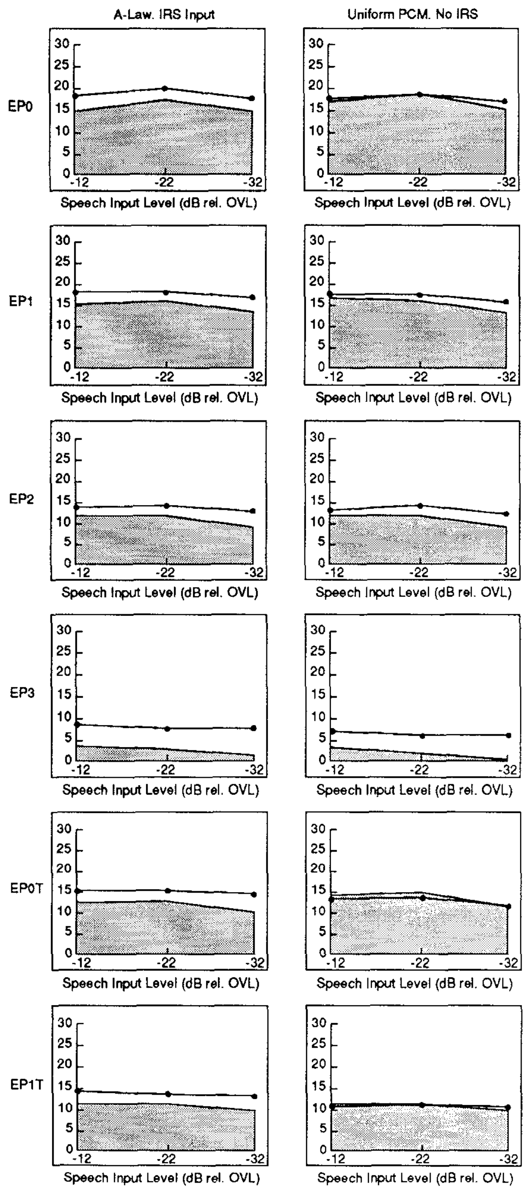

Fig. 2 - Phase 2 (selection) results for MOTOROLA candidate

were performed by MOTOROLA (USA) and ERICSSON (Sweden) with support by MATRA (France).

The objective of the Phase 3 tests was to ensure that desired improvements had been achieved, at no cost to performance in other areas. Five experiments were designed.

The first three experiments were a replication of test plan for the third selection and examined the same three criteria:

i) and ii) the effects of bit errors at different input levels under IRS A-Law and NO IRS Linear PCM audio parts;

iii) the effect of asynchronous tandeming under both IRS A-Law and No IRS Linear PCM audio parts.

Experiments 4 & 5 were added to investigate a fourth criteria,

iv) the effect of environmental noise.

A listening-only test was again chosen, adopting the Absolute Category Rating (ACR) method for the first three experiments. The experiments were the same already used in the third selection, designed by BT, and employed four pairs of 24 x 24 interleaved Graeco-Latin Squares to achieve a balance over all the conditions and talkers. Each listener cast 96 votes, i.e. 24 conditions x 4 talkers x 1 listening level.

For the environmental noise experiment(s) it was proposed to adopt, besides the ACR, also a modified version of the DCR method, that was expected to be more sensitive to differences in performance between the codecs. Each of the noise experiments followed the same designed structure as in Exp. #1 to Exp. #3 (4 interleaved 24 x 24 Graeco-Latin squares). Subjective tests were carried out by BT (UK), CNET (France), CSELT (Italy), and Deutsche Telekom (Germany). The whole set of Phase 3 individual and global data were extensively analysed and discussed within TCH-HS; for each condition, the MOS were computed, separately for male and female speech, as well as averaged together.

The effects of different factors and their interactions were subject to analysis of variance (ANOVA). Conversion to Q values and weighted averages were again calculated for the whole set of results.

Tables 5 to 8 report the results obtained in Phase 3 experiments.

Phase 3: Experimental results from optimization

		A-Law PCM (with IRS)			Uniform PCM (No IRS)		
		Input Level (dB rel. OVL)			Input Level (dB rel. OVL)		
Error pattern	Codec Version	-12	-22	-32	-12	-22	-32
EP0	3.3	-0.27	-0.02	0.34	-1.13	-2.90	-1.70
	3.5	-0.95	0.10	0.24	-1.97	-2.21	-1.40
EP1	3.3	-0.26	-0.86	-0.59	-3.72	-1.72	-1.21
	3.5	-0.55	-1.58	0.29	-3.82	-1.82	-0.94
EP2	3.3	-0.49	-1.61	1.14	-1.93	-1.79	0.69
	3.5	0.11	-1.03	1.20	-1.69	-2.03	1.10
EP3	3.3	-0.39	1.79	3.80	0.79	1.49	3.59
	3.5	-0.20	1.94	3.82	0.59	0.79	3.48
EP0 Tandem	3.3	-0.14	-0.56	-0.03	-5.20	-5.43	-3.89
	3.5	-0.22	-0.41	0.04	-5.71	-6.28	-3.55
EP1 Tandem	3.3	-0.46	-0.75	0.49	-4.98	-4.14	-2.79
	3.5	-0.42	-1.04	0.32	-4.36	-4.35	-2.86

Table 5 - Results for Experiments 1, 2, and 3

	Codec Version	Office	Vehicular	Traffic
Low Noise	3.3	-0.78	-2.19	-1.06
	3.5	-1.34	-1.86	-1.10
High Noise	3.3	-1.75	-0.87	-1.25
	3.5	-1.82	-0.80	-1.05
Low Tandem	3.3	-1.75	-2.38	-2.66
	3.5	-1.66	-3.12	-2.52
High Tandem	3.3	-2.99	-4.10	-3.09
	3.5	-1.89	-3.11	-2.48

Table 6 - Results for Experiment 4

	Codec Version	Office	Vehicular	Traffic
Low Noise	3.3	-2.10	-2.96	-4.53
	3.5	-1.73	-3.67	-4.07
High Noise	3.3	-2.79	-2.83	-2.04
	3.5	-2.92	-3.28	-3.28
Low Tandem	3.3	-4.03	-4.39	-5.31
	3.5	-3.89	-4.82	-4.52
High Tandem	3.3	-4.96	-5.85	-5.68
	3.5	-4.63	-4.90	-5.68

Table 7 - Results for Experiment 5

Note 1: The results in table 5 to 7 are differential Q ratings relative to the full rate codec.

Note 2: Negative numbers denote lower performance of the half rate codec component to the full codec.

Laboratory	Exp1	Exp2	Exp3	Exp4	Exp5
BT	HC1 = HC2 HC1 = FR HC2 < FR	HC1 = HC2 HC1 < FR HC2 < FR	HC1 = HC2 HC1 < FR HC2 < FR	HC1 = HC2 HC1 < FR HC2 < FR	X
CNET	X	X	X	X	HC1 = HC2 HC1 < FR HC2 < FR
CSELT	HC1 = HC2 HC1 = FR HC2 = FR	HC1 = HC2 HC1 = FR HC2 = FR	HC1 = HC2 HC1 = FR HC2 < FR	X	HC1 = HC2 HC1 < FR HC2 < FR
Deutsche Telekom	HC1 = HC2 HC1 = FR HC2 = FR	HC1 = HC2 HC1 = FR HC2 = FR	HC1 = HC2 HC1< FR HC2 < FR	HC1 = HC2 HC1 < FR HC2 < FR	X
Global	HC1 = HC2 HC1 = FR HC2 = FR	HC1 = HC2 HC1 < FR HC2 < FR	HC1 = HC2 HC1 < FR HC2 < FR	HC1 = HC2 HC1 < FR HC2 < FR	HC1 = HC2 HC1 < FR HC2 < FR

KEY

Symbol	Definition
=	no significant difference at the 95% confidence level
HC1	codec version 3.3
HC2	codec version 3.5
FR	full rate codec
X	experiment not performed by laboratory
HC1 < FR	HC1 significant worse than FR at the 95% confidence level

Table 8 - Codec comparison summary (from ANOVA)

The two versions of the candidate codecs performed equally well or slightly worse than the full rate for most cases. An advantage in quality in favour of the full rate was confirmed with UPCM No IRS audio part, significant in tandeming conditions, as already shown for Phase 2. In environmental noise conditions, the still unsatisfactory performance was quantified for the first time, by means of formal tests that used two different methods, ACR and DCR.

Differential Q (dB) values of each candidate from the Full-Rate are more pronounced with the DCR procedure than with the ACR procedure, the corresponding differences in terms of DMOS scores resulting more often statistically validated. Results confirm that the DCR methodological procedure is more discriminating than the ACR classical one.

In Phase 3 TCH-HS significantly improved the methodology for measuring subjectively the performance of candidate codecs. Since the most important requirements set by SMG and tested by TCH-HS were met by the optimised algorithm, SMG approved the optimised codec.

PHASE 4 TESTING: CHARACTERISATION

The results from Phase 3 of testing were also utilised for the characterisation of the performance of the algorithm candidate to the ETSI standardisation: Phase 3 became the Characterisation Phase I, and Phase 4 of testing (the hardware implementation) was called the Characterisation Phase II, with the following characteristics listed below:

1) the characterisation tests were based on the characterisation of only <u>one</u> codec (the Half-rate Candidate codec),
2) some of the test conditions adopted during Phase 2 were included in the characterisation Phase,
3) the Characterisation Phase II included other four experiments,
   (i) Exp. 6: Assessment of equivalent qdu,
   (ii) Exp. 7: Tandeming with other standards,
   (iii) Exp. 8: Talker dependency,
   (iv) Exp. 9: Assessment of DTX algorithm,
4) the analysis of results was made on the Mean Opinion Scores (MOS) data, in order to get the minimum manipulation of the results and the maximum reliability.

Additional informal expert listening tests were made using conversation speech processed through the DTX system.

Characterisation Phase Results

Four laboratories participated in the assessment program[10] [11]: BT, CNET, CSELT, and Deutsche Telekom. The results, made available by the end 1994, are summarised in Table 9.

Subject under investigation	qdu	Tandeming with other Standards	Talker Dependency	DTX Functions
Laboratory				
BT		HR+any<any+HR	see full text	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
Deutsche Telekom	HR = FR HR+HR=FR+FR		see full text	DTX fairly satisfactory, concerns over comfort noise quality

KEY

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 significant worse than FR at the 95% confidence level
any	all tested codecs, except HR (G.726, G.728, and FR)

Table 9 - Summary of Characterisation Phase II Results

Assessment of qdu

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 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 that performed the test.
(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 indicate 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 (G.711, G.726 @ 32 kbit/s, and G.728) included in the experiment.

Effect of tandeming with other standards

The following standards were tandemed with the half-rate codec in this experiment; half-rate, full-rate, G.726 (at 32 kbit/s), and G.728. 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 experiment was conducted in two different laboratories. The main conclusion that could be drawn was that the performance showed to be 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).

Talker Dependency

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 characterisation testing. From the tests carried out it was shown that the talker dependency for the half-rate codec is similar to that for the full-rate under error free conditions.

Subjective evaluation of DTX functions

The four laboratories that performed this work 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 (mutilation)
- noise quality (faithful transmission)
- noise contrast (noticeable differences)

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.

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 UIT-T Series P Recommendations, and that have shown to be suitable for characterising both the GSM full rate and half-rate algorithm performance.
Results of tests conducted by several organisations showed consistency when normalised 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.
It was noticed that the performance of the half-rate candidate depends on the audio part, sending frequency characteristics (IRS vs. No IRS) and I/O interface (A-Law PCM vs. Linear PCM), and that for tandeming connections, with the No IRS Linear PCM audio part, the quality is worse than the 3 dB of tolerance compared with the Full-rate RPE-LTP.
For network planning purposes, the same rules will be adopted as for RPE-LTP, which will be adequate for most applications.

REFERENCES

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

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

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

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

[5] SQEG CCITT SG XII 'Subjective Test Methodology for a 16 kbit/s Speech Coder', V.3.2 March 1990.

[6] SQEG UIT-T SG 12 'Subjective Test Methodology for a 8 kbit/s Speech Coder', V.3.1 Nov. 1993.

[7] SQEG SG XII CCITT: "Methods for subjective determination of transmission quality", Revised Recommendation P.80, 1992.

[8] CCITT: "Modulated Noise Reference Unit", Recommendation P.81, Blue Book, Volume V, 1988.

[9] SQEG SG XII CCITT: "Subjective performance assessment of telephone-band and wide band digital codecs", Draft Recommendation P.83, 1992.

[10] ETSI/TM5/TCH-HS Doc. 94/32 "Test Plans for Characterisation Phase I of the GSM Half-Rate Speech Coding Algorithm", 1994.

[11] ETSI/TM5/TCH-HS Doc. 94/57 Rev. 2 "Test Plans for Characterisation Phase II of the GSM Half-Rate Speech Coding Algorithm", 1994.

[12] C.R. South, J. Rugelbak, P. Usai, N. Kitawaki, H. Irii, J. Rosenberger, J.R. Cavanaugh, C.A. Adesanya, D. Pascal, N. Gleiss, G.J. Barnes: "Subjective performance assessment of CCITT's 16 kbit/s speech coding algorithm", Speech Communication, Volume 12, pp. 113-133, 1993.

LIST OF ACRONYMS AND GLOSSARY

A/D	Analogue to Digital
ACR	Absolute Category Rating
ANOVA	ANalysis Of VAriance
C/I	Carrier-to-Interferer ratio
D/A	Digital to Analogue
DAT	Digital Audio Tape
DCR	Degradation Category Rating
DTX	Discontinuous Transmission (for power consumption and interference reduction)
EID	Error Insertion Device
ETSI	European Telecommunications Standards Institute
GSM	Global System for Mobile communications
IRS	Intermediate Reference System
No IRS	= rather flat sending frequency characteristics
HLCS	Host Laboratory Control System
MNRU	Modulated Noise Reference Unit
MOS	Mean Opinion Score
PCM	Pulse Code Modulation
Q	Speech-to-speech correlated noise power ratio in dB
QDU (or qdu)	Quantization Distortion Unit
RPE-LTP	Regular Pulse Excited codec with Long Term Prediction
SCD	Signal Conditioning Device
SFC	Sending Frequency Characteristic
SMG	Special Mobile Group
TCH-HS	Traffic CHannel Half-rate Speech
UIT-T	International Telecommunication Union - Telecommunications Standardisation Sector
UPCM	Uniform or Linear PCM

ACKNOWLEDGEMENT

The authors would like to express their gratitude to TCH-HS Chairman, R. Montagna, to the organisations that conducted the host lab processing and the subjective tests, and to all members of the ETSI/TM/TM5 TCH-HS. In particular the authors thank John Beerends, Frank Muller, Bob van Ravesteijn and Michael Emons (Royal Dutch PTT, NL), Christiane Antweiler and Rainer Martin (RWTH, D), Nigel Hoult (Racal, UK), Norman Gleiss and Ove Till (Telia Research), Rupert Voelcker, Brian Surtees and Keith Evans (BT, UK), Jon Natvig and Jon Rugelbak (NTA, Norway), Gerhard Schroeder, Stefan Feldes, Wolfgang Koch and Edmund Steinke (Deutsche Telekom, Germany), Franco Perosino, Angelo Franceschini and Claudia Grilli (CSELT, Italy), that, among many others, deserve special credit for their contributions and suggestions throughout the whole exercise.