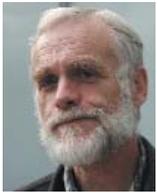# The 1987 European Speech Coding Championship

JON EMIL NATVIG

Jon Emil Natvig is Senior Research Scientist in Telenor R&D

From the outset the fundamental working assumption was that the GSM system would be based on digital transmission [1]. It was believed that digital technology would not only enable a more efficient management of scarce frequency resources, it would also provide high speech quality and advanced features such as security and data communications.

All this was unproven technology at the time and an important part of the preparatory work was to conduct a number of studies and tests to verify the feasibility of a digital system. These studies were a significant basis for the system design. The so-called "Paris competition" which compared different alternatives for the radio transmission – leading to the selection of the "narrow-band" TDMA radio access method – is well publicised as described in [2]. The radio access study however was itself done under the non-proven assumption that satisfactory speech transmission could be achieved at a bit rate around 16 kbit/s.

In 1985–86, the state-of-the-art in digital coding of speech had reached a level that made this assumption quite reasonable. At that time, ITU-CCITT had standardised speech coding at 32 kbit/s for telephony speech (ITU-G.721) and at 64 kbit/s for wideband (7 kHz) speech (ITU-G.722). Speech coding at 16 kbit/s and below was still a research subject even though computer simulations had demonstrated promising speech quality results. It was however unknown what quality could be obtained in the adverse bit error conditions expected in mobile radio. This fundamental issue was addressed in a less well-known but equally important study: the GSM speech coder selection process, which is the subject of this paper.

Creation of SCEG

The "Speech Coding Experts Group" (SCEG) was set up in October 1985 to deal with the speech processing issues of GSM. SCEG was not formally a sub-group of GSM but had to be set up under CEPT TM3, which was formally responsible for speech processing issues in the CEPT committee structure. In practice, however, SCEG acted as a normal GSM working group but had to report progress in both fora. The task given to SCEG was two-fold:

1) To demonstrate that a digital solution with speech coding at around 16 kbit/s would provide a speech quality at least as good as analogue systems;

2) Given that 1) could be fulfilled, SCEG should select and specify a speech coder for the GSM system.

At the plenary meeting of GSM in Madeira in February 1987, which is best known for the controversy over the radio access issue, SCEG could confirm that a digital solution would indeed offer superior speech QoS compared to analogue solutions. SCEG also proposed a compromise solution combining features from the two best codecs tested as the candidate for further testing and standardisation.

The SCEG work programme

The ultimate measure of speech quality must relate to the quality perception of live listeners. While using automated quality evaluation techniques have become popular lately, no such techniques were available in 1985–87, so speech quality had to be determined "the hard way", in subjective tests involving listeners. The standard for subjective speech quality assessment of telecommunication systems is the Mean Opinion Score test, or MOS for short. Listeners judge the speech quality by listening to recorded speech samples and assign scores on a five point scale: from (poor) to (excellent). The scores are assigned a numeric value from 1 to 5 and the scores are averaged over the listener group and the resulting value is the Mean Opinion Score, or MOS.

The initial stages of the GSM speech codec standardisation effort therefore consisted of the preparation of a set of basic requirements and the organisation of a large multi-national subjective test programme to compare candidate codecs with each other and with these basic requirements.

Basic requirements and rules

At the outset, more than 20 candidate codecs were proposed. The group therefore quickly agreed on a set of rules for the selection process:

1) Each country could only contribute one codec proposal. Several countries therefore had to carry out

national "qualification" rounds to arrive at one single candidate. SCEG was not involved in this process.

2) All codec proposals had to be submitted as hardware laboratory implementations operating in real time, no simulations were allowed. This requirement was meant to ensure a certain maturity of the candidate codecs – and a real commitment from the respective proponents.

The following basic requirements for candidate codecs were evaluated (details are given in [3]):

Gross bit rate

Bit rate is a less straightforward attribute than one might think. Basically, a higher bit rate should produce better speech quality for a given algorithm. However, in a mobile environment, we have to take into account bit errors. Using some of the available bit rate for protection of sensitive information will improve the average performance over the operating conditions of the codec at the expense of a certain reduction of the maximum quality in error free conditions. In the experiments, a gross bit rate including any error protection was fixed to 16 kbit/s.

Delay

Delay is involved in several difficult trade-off considerations in codec design:

a) Delay vs conversational quality: Transmission delay causes problems in two aspects: 1) long end-to-end delays disturb the flow of conversation, and 2) reflections in the network become more audible with increasing delay.

b) Delay vs complexity: Because power consumption in CMOS VLSI is almost a linear function of the clock rate of the processing machine, a relaxation of the delay requirement could make it possible to distribute the processing over a longer period with a lower clock rate, thus reducing the power consumption.

What matters for the user is the end-to-end delay, and the maximum allowed codec delay contribution in a coder-decoder back-to-back configuration set to 65 ms.

Complexity

One basic requirement for GSM was to allow for portable handsets so complexity and power consumption of the speech codec was a major concern. This criterion was not quantified but was assessed and taken into account in the selection process. In retrospect it can be seen that this attribute was more important than we knew at the time: recent analysis [4] have documented that more than 80 % of the processing power in the GSM transmission chain is used by speech encoding/-decoding.

Transcodings

Medium to low bit rate speech coding removes subjectively redundant information. Therefore the reconstructed speech is not identical to the original speech waveform. Feeding reconstructed speech into a new coding/decoding stage could result in a large increase in the distortion introduced by the first codec.

The evaluation therefore included a transcoding condition corresponding to mobile to mobile (2 codecs in tandem).

Robustness to variations in voice spectra and speech levels

To be useful in a public service, a speech codec needs to be able to work with a wide range of speakers; men, women, children and adults over a wide range of speech levels. The candidate codec tests included male and female voices and a dynamic range of speech levels of 20 dB.

Robustness to bit errors

Obviously, robustness to errors is essential in a mobile radio application.

The selection tests were carried out with random error probabilities of 0 %, 0.1 % and 1 % representing a typical range of operation.

Digital or analogue?

The actual phrasing of the speech performance requirement in GSM was:

From the subscriber's point of view, the quality of voice telephony in the GSM system shall be at least as good as that achieved by the first generation 900 MHz analogue systems over the range of practical operating conditions.

The implication of this formulation was that SCEG had to 1) consider end-to-end transmission including the public telephone network, and 2) to include an analogue 900 MHz under comparative operating conditions in the test programme.

The analogue reference system gave a substantially transparent speech quality in high C/N conditions and outperformed all codecs in error-free conditions. However, ALL digital codecs performed better on the average over the range of operating conditions than the reference FM system [3]. Thanks to error protec-
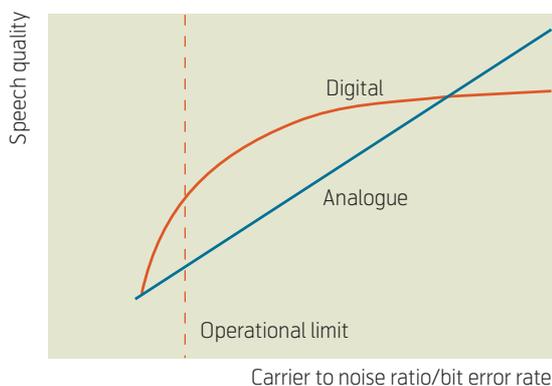
Figure 1 Analogue vs digital speech typical behavior

tion, the quality of a digital codec maintains a relatively high speech quality as transmission conditions deteriorate (until a marked breakdown point), whereas the quality of the analogue reference decreases more or less linearly. This typical behaviour is illustrated qualitatively in Figure 1.

A Solomonic decision – the RPE–LTP codec

The comparison of the codecs were organised as a large multi laboratory experiment. Running the tests in parallel in several laboratories, was essential for the credibility of subjective test results. Details of the experiments and the results are reported in [3]. Six candidate codec were presented at the first laboratory session hosted by CSELT[1] in Torino in 1987. In this session some objective measurements were made and speech material for subjective testing was recorded. Subjective tests were carried out in seven different laboratories in the respective native languages. The overall analysis (which was undertaken by TF – Televerket's Research Institute) showed that no single codec could be declared the winner in all aspects of the "competition".

However the results showed clearly that two codecs, both based on pulse excited source-filter modelling, were superior to the remaining four sub-band coders. In comparing these two solutions the development teams from IBM/France and Philips/ Germany could see the possibility of several improvements. SCEG therefore decided to allow the French and German teams to propose a compromise solution resulting in the merging of MPE-LTP and RPE-LPC into the RPE-LTP full-rate GSM codec [6] (see box).

The specification of the codec is given in GSM recommendation 06.10, which specifies the algorithm down to the bit level. This allows verification of implementations by means of a set of digital test sequences. The specification was defined in fixed point arithmetics to be implementable on fixed-point DSPs consuming less power. A public domain bit exact C-code implementation of this coder is available [7].

Telenor/Televerket contributions

TF's strong involvement in the development of GSM in general was founded on experience and competencies built up over more than ten years of activities in several research "threads" in the mobile communication area, both conventional land mobile and satellite mobile communication.

TF had been studying satellite communication and digital communication almost since it was established in 1967, and TF scientists were strongly involved when Televerket developed the NORSAT satellite system to provide telecommunications to the Ekofisk offshore oil platform in the North Sea. This was one of the first domestic satellite systems in the world and probably the first system worldwide to be based on digital transmission – a bold choice at the time. The speech encoding used 32 kbit/s adaptive delta modulation – an advanced quantisation scheme rather than a speech compression method – combined with advanced error correction. It turned out that this system provided clearly better frequency utilisation than alternative analogue solutions based on FM.

The experience from the NORSAT system convinced TF scientists that the future of wireless mobile communication lay in digital transmission. In the 1970s TF was very active in ship-to-shore satellite mobile communication in the framework of CCIR SG 8 and in the development of the MARISAT system. In this work, TF proposed a digital solution following the experiences from NORSAT, but did not succeed in convincing the committee at this time. Digital speech transmission for ship-to-shore systems was introduced later by INMARSAT (Standard B) and TF also played an active part in the studies that led to the introduction of that system.

In all this work, TF established broad know-how in system and network design, QoS aspects of speech transmission, as well as subjective speech quality, which also included laboratory facilities for subjective evaluation of speech codecs.

1) *CSELT – Centro Studi e Loboratori Telecommunicazioni.*

GSM speech coding in a nutshell
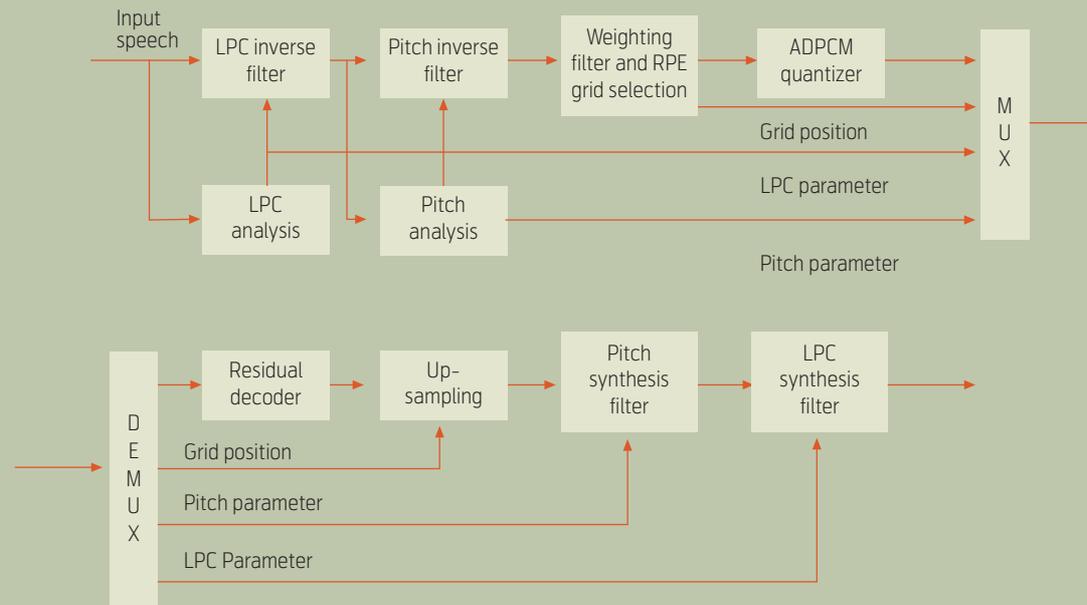
The 13 kbit/s coding scheme selected for the full-rate GSM standard is an example of a so-called analysis by synthesis system.

The sampling rate is 8 kHz and each sample is quantized at 13 bit/sample.

The GSM full-rate codec operates on time frames of 20 ms (160 samples). This corresponds to one pitch period for someone with a very low voice or ten periods for a very high-pitched voice. This is a very short time during which the speech can be assumed to be stationary. The speech frame is analysed using the source-filter model, which assumes that the speech signal is the result of a time varying linear filter excited by a source signal. For each frame, parameters are determined for a synthesis filter that models the effect of the vocal and nasal tracts. The short-term inverse filter flattens the spectrum envelope of the residual speech signal.

The next step is the long-term filter, which operates on 5 ms subframes (40 bits). In the case of voiced speech, the input signal will be a periodic impulse train where the pulses correspond to the glottal pulses. This periodicity is removed using a long-term predictive (LTP) filter, resulting in a weak signal. If the speech is part of an unvoiced phoneme, the residual is noisy and does not need to be transmitted precisely.

In the last step, the resulting residual signal is down-sampled by a factor three. This is done by selecting every third sample of each 5 ms sub-frame, starting at sample 0, 1, 2 and 3, resulting in 4 candidate sequences. The algorithm picks the sequence with maximum energy for quantisation and transmission

The *decoder* starts with the transmitted sub-sampled sequence padding the missing samples with zeroes. The resulting residual pulses are fed into the long term synthesis filter which has had its parameters updated, reconstructing the short-term residual. These impulses are then passed through the short-term synthesis filter simulating the vocal tract and thereby producing the speech signal.

More information can be found in [6] and details are given in [10].

One important basis for TF's contributions to GSM was the close collaboration between TF and ELAB. [2] gives an account of the work in the radio area. Already in 1979 – one year *before* the NMT network was launched – TF awarded its first research contract in the speech coding area to ELAB. The task given to ELAB was to study low bit rate speech coding for mobile radio applications. In the following years, a series of contracts were awarded and a fruitful collab-

oration evolved between TF and ELAB which established a competence environment in digital speech coding. The division of responsibilities was such that ELAB covered the work on speech algorithms and hardware implementations while TF dealt with speech system aspects, QoS and subjective testing.

In 1983 the FMK (Future Mobile Communication) committee was established in the framework of the

Nordic Council to plan for a future digital land mobile radio system to follow after NMT in the Nordic countries. FMK also established a special sub-group (NR-FMK-T) to look into the speech coding issue. When SCEG was set up in 1985, the Nordic countries had already completed a coordinated subjective experiment to compare possible codecs for mobil radio.

A major result of the TF/ELAB collaboration [8,9] was a sub-band coder with sophisticated adaptive bit allocation as a candidate for GSM. Unfortunately, due to the very tight time schedule for the delivery of a hardware version at the CSELT laboratory session, some error sneaked into the algorithm. Considering that none of the sub- and coder could match we believe that this would not have changed the selection anyway.

References

1 Haug, T. How it all began. *Telektronikk*, 100 (3), 155–158, 2004. (This issue)

2 Maseng, T. *Telektronikk*, 100 (3), 161–164, 2004. (This issue)

3 Natvig, J E. Evaluation of six Medium Bit-Rate Coders for the Pan-European Digital Mobile Radio System. *IEEE Journal on Selected Areas in Communications*, 6 (2), 324–331, 1988.

4 Thierry, T, Tennehouse, D. *Estimating the computational requirements of a software GSM base station*. July 1, 2004. [online] – URL: http://citeseer.ist.psu.edu/336927.html

5 Natvig, J E, Hansen, S, de Brito, J. Speech Processing in the pan-European Digital Mobile Radio System (GSM) – System Overview. In: *IEEE GLOBECOM 1989*, November 1989.

6 Vary, P et al. Speech Codec for the European Mobile Radio System. *Proc. ICASSP 1988*, New York, 227–230.

7 Degener, J. *Digital Speech Compression – Putting the GSM 06.10 RPE-LTP algorithm to work*. July 8, 2004. [online] – URL: http://www.ddj.com/documents/s=1012/ddj9412b

8 Ramstad, T A. Sub-band coder with a simple adaptive bit allocation algorithm. A possible candidate for digital mobile telephony? *ICASSP Paris*, May 3–5, 1982.

9 Hergum, R, Perkis, A. *16 kbit/s speech coder for the GSM system*. Trondheim, ELAB, 1986. (ELAB report STF44 F86133)

10 ETSI. *GSM full rate speech transcoding*. Sophia Antipolis, 1992. (ETSI recommendation 06.10.)

Jon Emil Natvig (58) graduated from the Norwegian Institute of Technology (NTH) as siv.ing. in 1970 and Dr. Ing. in 1975. He has worked with Televerket/Telenor since 1972. In the period 1975–1990 his main work area was in speech QoS aspects, mostly related to digital voice transmission in satellite and land mobile systems. From 1985 to 1989, he chaired the Speech Coding Experts Group responsible for selecting and specifying the speech processing functions in the GSM system. Since 1990 his main work has been in the field of voice user interfaces. He is currently working with the usability and accessibility issues in the Products and Markets research program at Telenor R&D.

email: jon-emil.natvig@telenor.com