# Voice Quality of Interconnected PCS, Japanese Cellular, and Public Switched Telephone Networks

Spiros Dimolitsas, Franklin L. Corcoran, Channasandra Ravishankar
*COMSAT Laboratories, 22300 Comsat Drive, Clarksburg, MD 20871, USA*
&
Marion Baraniecki
*INTELSAT, 3400 International Drive N.W., Washington, DC 20008, USA*

**ABSTRACT**

The non-linear nature of low-rate parametric speech coding has made it necessary to resort to formal subjective assessments for quantifying end-to-end voice quality of interconnected networks. At the same time, the rapid growth of cellular communications has highlighted the need to characterize transmission quality when cellular terminals are attached at the access or termination nodes of switched networks. In this paper the voice quality of interconnected North-American and Japanese digital cellular systems over public transmission facilities is quantified. From these assessments it was concluded that cellular networks using 8 kbit/s or 6.4 kbit/s VSELP may meet end-to-end quantization distortion criteria when interconnected with the switched network.

## 1. INTRODUCTION

Future long-distance, and especially international telephone calls, will involve an increasing number of multi-link circuits of cellular, mobile, private, and public switched telephone network (PSTN) type of connections. Calls will thus be established over multi-link connections consisting of different types of speech codecs operating at a variety of encoding rates.

This paper is intended to provide some answers with respect to the transmission quality expected in tandem connections of facilities incorporating 32 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM) [1], 16 kbit/s Low-Delay Code-Excited Linear Predictive (LD-CELP) coding [2], the 8 kbit/s Vector Sum-Excited Linear Predictive (VSELP) standard IS-54 which will be used in the North-American digital cellular network, and the 6.4 kbit/s VSELP standard which will be used in Japan [3]. Also investigated, are limited interconnections with the European 13 kbit/s Regular Pulse Excited LTP (RPE-LTP) [4] and the 8 kbit/s Conjugate Structure-CELP (CS-CELP) codecs [5], the former of which is used in the Global System for Mobile Communications (GSM), while the latter formed an early candidate for standardization by the International Telecommunication Union (ITU-T) as a wireline-quality 8 kbit/s speech coding standard. Although CS-CELP is not expected to become the ITU-T standard in the form evaluated, it can be reasonably expected that some of its characteristics, such as performance when intercon-nected with other devices might be retained in the final standard. The work reported in this paper extends prior studies which dealt with the performance of North-American networks [6].

Several foreseen network scenarios are presented which involve an international link as a part of multi-link connection (Figure 1). Network configuration A represents a general case of a call initiated from a wireline user in a foreign country that is destined to a North-American or Japanese digital cellular telephone. This type of cellular configuration is prevalent today for international connections given the wide use of Digital and Packet Circuit Multiplication Equipment (DCME and PCME) [7, 8].

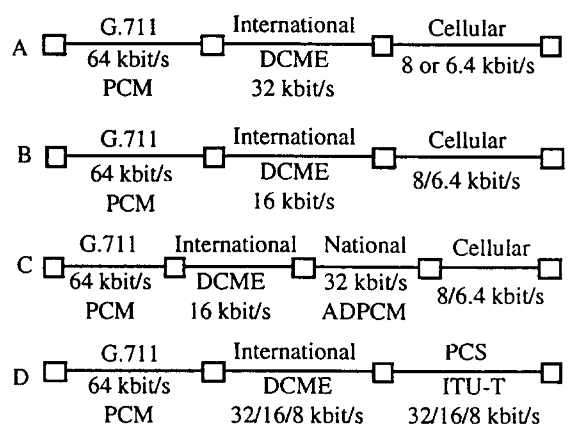


**Figure 1:** *Interconnected Network Configurations*

In the future, 32 kbit/s ADPCM-based DCME and PCME systems are likely to be replaced by 16-kbit/s LD-CELP-based DCME and PCME, and thus the remaining configurations focus on 16 kbit/s LD-CELP technology (in practice, both 32 kbit/s ADPCM and 16 kbit/s LD-CELP obey the same transmission planning rules, hence, configurations A and B are expected to give equivalent quality when the terminations are the same). Network configuration B reflects this case. Case C represents the situation where a cellular call is routed over an international link using 16 kbit/s DCME to an overseas



0-7803-2431-5/95 $4.00 © 1995 IEEE

destination, where the national extension is provided via facilities utilizing 32 kbit/s ADPCM. This reflects the increasing use of low-rate coding in some national networks, as well as the use of cellular technology in combination with low-rate long-haul technology to provide telephone access to developing and rural areas. For simplicity, the case of cellular-to-cellular calls is not shown in this figure, however, results are reported in Section 3.

Finally, configuration D is one that involves a PCS terminal, which currently uses 32 kbit/s ADPCM technology (as in the Digital European Cordless Telephone, DECT, and CT-2 systems), but in the future may use 16 kbit/s LD-CELP and 8 kbit/s ITU-T technology. This configuration also includes the possible use of DCME/PCME facilities using 32 kbit/s ADPCM, 16 kbit/s LD-CELP and 8 kbit/s ITU-T technologies.

## 2. EXPERIMENTAL DESIGN

In defining an experimental design for measuring end-to-end voice performance, several factors were taken into consideration including knowledge of human psychology, statistics, and the objective of the evaluation in terms of the system performance parameters sought.

**2.1 Voice Quality Assessment Approach.** It has recently been recognized, that the non-linear nature of low-rate speech coding has made necessary to resort to formal subjective assessments for quantifying *end-to-end* voice quality of interconnected networks [6]. Currently, research is underway within the ITU-T and the European Telecommunications Standards Institute (ETSI) to arrive at analytic models for estimating transmission quality [9, 10], but this work has not yet been concluded.

Since a characteristic of digital speech coding technology is the fact that link degradations are introduced in the listening transmission path only it seems suitable that tests conducted using single stimulus, listener opinion, Absolute Category Rating (ACR) Mean Opinion Score (MOS) assessments employing high quality recordings would be appropriate in this instance [11]. Listener opinion ACR-MOS assessments are usually conducted by arranging for a listener to hear a succession of groups of typically two to three sentences, with each group (or sample) being reproduced over a different circuit condition. After each sample is heard, the listener expresses an opinion with regard to his or her perception of quality of the processed speech, expressed as an excellent, good, fair, poor, or bad (5, 4, 3, 2, 1) rating. Each opinion is based on exposure to the most recently heard sample only, and the listener is typically given 5 to 10 seconds in which to cast a vote before the next sample is heard.

**2.2 Test Conditions & Experimental Design.** The experimental design developed, based on an incomplete randomized block, was selected so as to permit the measurement of the MOS for each codec-reference and codec-test condition to be made, as well as to permit a mapping to a Modulated Noise Reference Unit (MNRU) framework [12]. These conditions were arranged in eight presentation blocks with each block containing a complete set of the codec-condition combinations under evaluation. In this arrangement, any given speech codec-condition combination appeared only once in each block, and every codec-condition combination appeared in any given block spoken by a different talker. The evaluation employed 32 untrained listeners who used binaural monotic headphones to assess the processed and reference speech material. All speech material was filtered prior to processing using a compensated intermediate reference system (IRS) [13].

## 3. RESULTS & ANALYSIS

As appropriate to the type of design described above, upon collection of 256 votes for each condition, the MOS was computed for each condition together with the associated equivalent Q values. The standard errors were in general between 0.09 and 0.11 of an MOS (at a 95% level of statistical significance), and for brevity are not shown in this table.

A rank-ordered presentation of the mean opinion scores for each condition assessed is presented in Table 1. (In the second column of this table, unless explicitly noted otherwise, 16, 32, 13, and 6.4 are used to denote the 16 kbit/s LD-CELP, 32 kbit/s ADPCM, 13 kbit/s RPE-LTP, and 6.4 kbit/s VSELP codings, respectively; and a "+" sign is used to indicate interconnected devices of the appropriate bit-rate). The 8 kbit/s VSELP and 8 kbit/s CS-CELP are denoted by V(8) and C(8), respectively.

In assessing, whether the measured "end-to-end" transmission distortion met or exceeded recommended maximum quantization noise distortion allocations, it was found necessary to establish a lower limit for the network end-to-end MOS. Since it is known that the accumulated distortion of four interconnected 32 kbit/s ADPCM systems equals the maximum end-to-end distortion allocation of digital networks [14], it was assumed that the minimum MOS value corresponding to 17.5 dBQ could be used to establish a lower threshold of acceptable performance. Using the MNRU reference system, this value can be found to correspond to an MOS of 2.98 (highlighted in Table 1 by a line between conditions 21 and 22, and shown in Figure 2).

**Table 1:** *MOS for Conditions Evaluated*

| *Cond.* | *Circuit Condition* | *MOS* | *Q Value* |
|---|---|---|---|
| 1 | Direct | 4.34 | |
| 2 | MNRU=30 dBQ | 4.21 | 30.00 |
| 3 | MNRU=36 dBQ | 4.19 | 36.00 |
| 4 | 64 kbit/s PCM | 4.16 | 29.75 |
| 5 | MNRU=24 dBQ | 3.89 | 24.00 |
| 6 | 16 kbit/s LD-CELP | 3.77 | 23.25 |
| 7 | 8 kbit/s CS-CELP | 3.77 | 23.25 |
| 8 | 32 kbit/s ADPCM | 3.66 | 22.25 |
| 9 | 64+16+C(8) | 3.65 | 22.20 |
| 10 | 8 kbit/s VSELP | 3.57 | 21.50 |
| 11 | 64+C(8)+16 | 3.45 | 20.75 |
| 12 | 6.4 kbit/s VSELP | 3.40 | 20.30 |
| 13 | 64+C(8)+C(8) | 3.38 | 20.25 |
| 14 | 16+16+16 | 3.37 | 20.25 |
| 15 | 64+32+V(8) | 3.36 | 20.10 |
| 16 | 32+16+C(8) | 3.27 | 19.50 |
| 17 | 64+16+V(8) | 3.15 | 18.65 |
| 18 | 32+C(8)+C(8) | 3.13 | 18.60 |
| 19 | MNRU=18 dBQ | 3.06 | 18.00 |
| 20 | V(8)+16+64 | 3.00 | 17.70 |
| 21 | 64+16+6.4 | 2.98 | 17.50 |
| 22 | 64+C(8)+V(8) | 2.88 | 16.80 |
| 23 | 64+C(8)+6.4 | 2.86 | 16.70 |
| 24 | 6.4+16+13 | 2.50 | 15.00 |
| 25 | V(8)+16+13 | 2.48 | 14.10 |
| 26 | MNRU=12 dBQ | 2.17 | 12.00 |
| 27 | MNRU=6 dBQ | 1.41 | 6.00 |
| 28 | MNRU=0 dBQ | 1.13 | 0.00 |

From conditions 10 and 12 it can be seen that the 8 kbit/s VSELP performs somewhat better than the 6.4 kbit/s VSELP, while both perform worse than the 8 kbit/s CS-CELP (condition 7). Furthermore, 6.4 kbit/s VSELP performs equivalently to two interconnected 8 kbit/s CS-CELP or three interconnected 16 kbit/s LD-CELP devices (conditions 13 and 14).

By examining conditions 15–17, it can also be seen that when interconnected with 32 kbit/s ADPCM and 16 kbit/s ADPCM, one 8 kbit/s VSELP device is probably equivalent to one 8 kbit/s CS-CELP codec tandemed with one 32 kbit/s ADPCM or 16 kbit/s LD-CELP device. Because of the transmission planning equivalency of 32 kbit/s ADPCM and 16 kbit/s LD-CELP, it can also be concluded (from condition 18) that two interconnected 8 kbit/s CS-CELP devices perform equivalently to one 8 kbit/s VSELP codec.

Examining now conditions 20 and 21, it can be seen that when interconnected with 16 kbit/s LD-CELP, 8 kbit/s VSELP and 6.4 kbit/s VSELP perform equivalently, and just meet the end-to-end distortion criterion cited above. The equivalency conclusion appears to hold also when VSELP technology is interconnected with 8 kbit/s CS-CELP (conditions 22 and 23); or when interconnected with the 13 kbit/s RPE-LTP codec (conditions 24 and 25). From this it can be concluded that, *when in isolation*,

$$\text{MOS}(1 \times \textit{8 kbit/s VSELP}) > \text{MOS}(1 \times \textit{6.4 kbit/s VSELP}),$$

*while, when interconnected with other standard codecs*:

$$\text{MOS}(1 \times \textit{8 kbit/s VSELP + Codecs}) \approx \text{MOS}(1 \times \textit{6.4 kbit/s VSELP +Codecs})$$

where ≈ denotes equivalency. However, conditions 24 and 25 also highlight that the end-to-end transmission performance guidelines set by the ITU-T for public switched telephony will be violated when placing cellular-to-cellular calls over international transmission facilities. This conclusion holds even when placing such calls from an 8 kbit/s VSELP terminal to another 8 kbit/s VSELP terminal, as the results of earlier studies have revealed [6].

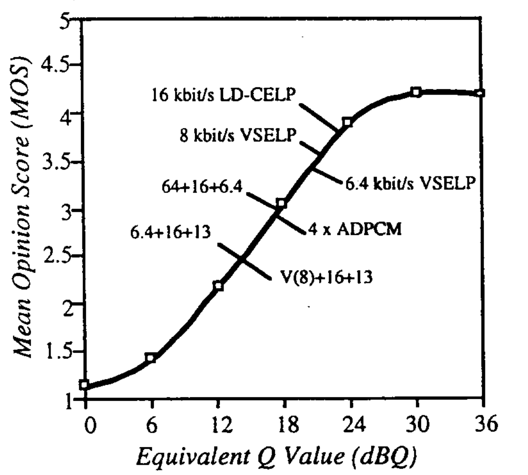


**Figure 2:** *Performance of Various Interconnected Codec Configurations*

From the above results, it is also possible to make some preliminary assessments on the transmission performance symmetry of multiple low-rate encoded connections. In particular, by examining conditions 9 and 11 as a pair, and conditions 17 and 20 as a pair, it can be seen that the placement order of low-rate codecs appears to influence the performance of an end-to-end connection, albeit not greatly. However, in both instances it appears that lower performance is realized when low-rate encoding is encountered first, rather than last, in a connection. This is a somewhat surprising result, given that lower-rate speech codecs tend to be less robust to ambient noise than higher-rate codecs. Although extensive data were not obtained in this experiment, it appears that a more thorough investigation is warranted to identify whether the following preliminary conclusion can be generalized:

MOS (*higher-rate codec + lower-rate codec*)≥
MOS(*low-rate codec + higher-rate codec*)

These results do indicate, however, that transmission planning becomes increasingly difficult with the introduction of low-rate voice codecs. Not only does the type and number of codecs need to be considered, but also the symmetry of the connection may have to be explicitly taken into account.

## 4. CONCLUSIONS

From the results obtained it can be seen that the performance of both 8 kbit/s and 6.4 kbit/s VSELP appears to be sufficiently high so that, under clear-channel conditions, interconnectivity with DCME and PCME-based international networks can be validated without exceeding the end-to-end quantization distortion allocation limits (these limits are only defined for clear channel conditions). This conclusion is tenable as long as the far-end termination does not comprise another cellular interconnection. Further, it was found that 8 kbit/s and 6.4 kbit/s VSELP yield equivalent performance when interconnected with other devices, provided that the number of low-rate encoded links are constrained to those included in this study.

In interpreting the results presented in this paper, it should be noted that speech codecs are expected to encounter digital impairments which were not considered here. This does not restrict the validity of this study, however, as network planning is normally confined to consideration of stationary transmission impairments.

Finally, it should not be forgotten that the network configurations given in Section 1 represent a mere sub-set of possible network configurations exploiting the same type of voice coding technology. For example, it should be noted that although ADPCM and LD-CELP are found in international connections, these types of encoding can also be encountered in many private networks. Therefore, the conclusions drawn from this evaluation can be expanded to apply beyond international public switched telephone networks, provided that mapping of the cellular/private network topology can be made within the conditions evaluated in this study.

## 5. REFERENCES


[1] CCITT Rec. G.726. "40-, 32-, 24-, and 16-kbit/s Adaptive Differential Pulse Code Modulation."

[2] J.-H. Chen, R.V. Cox, Y.-C. Lin, N. Jayant, and M.J. Melchner, "A Low-Delay CELP Coder for the CCITT 16 kb/s Speech Coding Standard", *IEEE J. Select. Areas in Comm.*, **J-SAC-10**, No. 5, pp. 830–849, June 1992.

[3] I.A. Gershon, M.A. Jasiuk, "Vector-Sum Excited Linear Prediction - VSELP Speech Coding at 8 kbit/s," Proceedings, *IEEE Intl. Conference on Acoust., Speech and Signal Proc. (ICASSP'90)*, April 1990, pp. 461-464.

[4] K. Hellwig, P. Vary, D. Massaloux, J. P. Petit, C. Galand and M. Rosso, "Speech Codec for the European Mobile Radio System", Proceedings, *IEEE Global Communications Conference, (GLOBECOM'89)*, pp. 1065–1069, 1989.

[5] A. Kataoka, T. Moriya, S. Hayashi. "Conjugate Structure CELP Coder for The CCITT 8-kbit/s Standardization Candidate," Proceedings, *IEEE Workshop on Speech Coding for Telecommunications*, pp. 25–26, Sainte-Adèle, Québec, Canada, Oct. 13–15, 1993.

[6] S. Dimolitsas, F. L. Corcoran, and M. Baraniecki, "Transmission Quality of North-American Cellular, Personal Communications, and Public Switched Telephone Networks," *IEEE Trans. on Vehicular Techn.*, Vol. **32**, No. 2, pp. 245–251, May, 1994.

[7] CCITT Rec. G.763. "Digital Circuit Multiplication Equipment Using 32 kbit/s ADPCM and Digital Speech Interpolation."

[8] CCITT Rec. G.765. "Packet Circuit Multiplication Equipment."

[9] Ellemtel. "Considerations of Low-Bit Rate Codecs in Transmission Planning." ITU-T SG 12 Contr. D.18. Geneva, February 15-25, 1994.

[10] Ellemtel. "Some Comparison Between Results from the ETSI Model and From Other Models and Subjective Tests." ITU-T SG 12 Contr. D.19. Geneva, February 15-25, 1994.

[11] S. Dimolitsas. "Subjective Quality Quantification of Digital Voice Communications Systems," *IEE Proc.*, Part **I**, *Comm., Speech and Vision*, **138**, (6), pp. 585–595, Dec. 1991.

[12] CCITT Rec. P.81. "Modulated Noise Reference Unit (MNRU)."

[13] CCITT, "Specification of an Intermediate Reference System," Rec. P.48, *Blue Book*, Vol. V, pp. 81–86, Melbourne, 1988.

[14] CCITT Rec. G.113. "Transmission Impairments,"