PERFORMANCE ASSESSMENT OF TANDEM CONNECTION OF ENHANCED CELLULAR CODERS †

Simão F.Campos Neto, Franklin L. Corcoran
COMSAT Laboratories, 22300 Comsat Drive, Clarksburg, MD 20871, USA
&
Ara Karahisar
Teleglobe International, 1000, rue de La Gauchetière ouest, Montréal, Canada H3B 4X5

ABSTRACT

The growth and increased competition in the second-generation (digital) cellular communication market has led service providers to improve the speech quality in their systems by introducing enhanced speech coders. Advancements in speech coding allowed designers to aim at toll-quality for these enhanced coders, and investigation of the impact of speech coders on the end-to-end quality of the public switched telephone network (PSTN) is necessary. This paper presents the continuation of a series of studies on the impact of tandem connection of cellular systems, where the quality of the enhanced cellular coders for major systems in use today is studied in the context of PSTN interconnection. A major conclusion of this study is that deployment of enhanced coders in second-generation cellular systems makes possible a substantial increase in quality of the cellular connections when in tandem with other speech coders in long haul international networks.

1. INTRODUCTION

Long-haul international circuits are making use of an increasing variety of low-bitrate speech coders. These devices, working at 32 kbit/s, 16 kbit/s and 8 kbit/s, are being used in tandem with a variety of cellular coders. A growing proportion of international traffic is originating from cellular mobile and satellite mobile terminals, in particular from second-generation cellular systems. While the original second-generation cellular coders were benchmarked against their (first-generation) analog counterparts, the enhanced cellular coders were specified to deliver better quality than the original digital systems. The speech quality that can be obtained when these coders are placed in tandem is of vital importance when planning international network facilities. The performance of the enhanced second generation cellular coders when in tandem with other speech coders in long-haul international networks was not assessed in earlier studies [1,2]. Since these coders are non-linear in nature, direct subjective evaluation methods are necessary to adequately characterize their performance.

† This work was sponsored by Teleglobe Canada.

This paper presents the results of an evaluation of the voice quality of the tandem connection of a number of original and enhanced time and code division multiple access (TDMA, CDMA) cellular coders in six basic telecommunication network scenarios. These scenarios were simulated in software and comprised transport over ITU-T G.711 64 kbit/s pulse-code modulation channels and over DCME based on G.726 32 kbit/s ADPCM, ITU-T G.728 16 kbit/s LD-CELP, and ITU-T G.729 8 kbit/s CS-ACELP coding. It should be noted that the tandem connections involving 32 kbit/s DCME systems had a 2.5% overload and were simulated using a G.726 coder operating at a net bitrate of 31.2 kbit/s, instead of a fixed bit rate of 32 kbit/s. G.728- and G.729-based DCMEs were simulated by the coder at the nominal rate (16 and 8 kbit/s, respectively). These are labeled DCME32, DCME16, and DCME8 in this paper. The cellular coders considered in this study were US TIA 7.4 kbit/s IS-641 ACELP enhanced TDMA speech coder, 9.6 kbit/s IS-127 RCELP enhanced variable rate CDMA (EVRC) speech coder, 12.2 kbit/s GSM-06.60 enhanced full rate (EFR) ACELP speech coder, and the original 13 kbit/s GSM 06.10 full-rate (FR) RPE-LTP speech coder. IS-641 and IS-127 are the enhanced versions of the IS-54 8 kbit/s VSELP and IS96a 8-1 kbit/s QCELP standards, respectively, which were not included in this study due to test size constraints.

List of Speech Coding Acronyms Used	
ADPCM	Adaptive Differential Pulse-Code Modulation
ACELP	Algebraic Code-Excited Linear Prediction
CS-ACELP	Conjugate Structure ACELP
RCELP	Residual CELP
LD-CELP	Low-Delay CELP
QCELP	Qualcomm Code-Excited Linear Prediction
RPE-LTP	Regular Pulse Excitation, Long-Term Prediction
VSELP	Vector-Sum Excited Linear Prediction

One subjective listening experiment was designed, which included tandem performance for six circuit types:

1. No tandem connection.
2. Tandem connections with only DCME coders.
3. Tandem connections involving the IS-641 Enhanced TDMA Cellular coder and DCME coders.
4. Tandem connections involving the IS-127 Enhanced CDMA Cellular coder and DCME coders.
5. Tandem connections involving the GSM-FR coder and DCME coders.
6. Tandem connections involving the GSM-EFR coder and DCME coders.

2. Experimental Design

The test structure used in this study was the same as in [2], in order to facilitate cross-experiment comparisons, and compensate for the lack of IS-54 and IS-96a coders in the test. Several factors were taken into consideration for the definition of the subjective experiment design for measuring coder voice performance, including knowledge of human psychology, statistics, experiment size, and the objective of the evaluation in terms of the system performance parameters sought. The listener-opinion test was conducted using an Absolute Category Rating (ACR) (single-stimulus) 5-point Mean Opinion Score (MOS) transmission quality scale [3] to quantify the performance in the different scenarios, for ITU-T P.48 IRS-weighted speech [4]. The design of the listening experiment was based on a balanced block structure, and provided for arranging the conditions in presentation blocks, where each block contained a complete set of the coder-condition combinations. Two male and two female talkers were used, with a total of 12 sentence-pairs per talker.

The test conditions were evaluated by 48 non-expert listeners. In total, listeners cast 192 votes for each of the test conditions. The test included the network configurations defined above, and a number of reference systems. These reference systems comprised of Modulated Noise Reference Units (MNRU) [5], which are included to provide a continuum of quality in the test, and of several reference coders: ITU-T G.726 32 kbit/s ADPCM coder, ITU-T G.728 16 kbit/s LD-CELP, ITU-T G.729 8 kbit/s CS-ACELP, and four interconnected 32 kbit/s ADPCM coders. The cumulative distortion produced by the latter is usually accepted as perceptually equivalent to the maximum end-to-end quantization distortion recommended by the ITU-T for international wireline connections.

3. Results & Discussion

Table 1 shows the rank-ordered results grouped by front-end coder. Table 2 contains the same data as Table 1 but regrouped by tandem length (some of the test conditions are repeated for easier comparison). In these tables, *MOS* represents the Mean Opinion Score. The standard errors for the test conditions, which averaged 0.07, were not reported due to space reasons, although their absence is compensated by the presentation of the relevant statistical tests. The *HSD* column indicates which test conditions can be considered equivalent by the Tukey-Kramer honestly significant difference (HSD) criterion (indicated by contiguous vertical lines within each test factor) for a given impairment. Column *D* presents statistical significance tests according to Dunnet's Multiple Comparison Test criterion [6], with ">", "<", and "=" indicating conditions that are statistically better, worse, and equivalent to the four-tandem G.726 condition. The HSD criterion is used to compare multiple pairs against each other while the Dunnet criterion compares a number of test conditions against a control condition, here the acceptability threshold represented by four tandem 32 kbit/s G.726 ADPCM.

The test results show that most of the tandem connections with cellular coders studied had a performance that was either equivalent to or better than the acceptability threshold for the fixed network. This indicates that the enhanced cellular coders are likely to provide a much higher end-to-end performance than the previous generation of the European GSM (GSM 06-10) and North-American TDMA (IS-54) and CDMA (IS-96a) systems. The following general results and observations can also be made:

a) For a given number of tandem coders, the DCME coders in this study were preferred according to the formula

$$\underset{(31.2\text{ kbit/s})}{\text{G.726}} \geq \underset{(8\text{ kbit/s})}{\text{G.729}} \geq \underset{(16\text{ kbit/s}).}{\text{G.728}}$$

This trend also indicates that, in this study, superior performance was obtained with a G.729-based system than with a G.728-based system.

b) For a given tandem connection, the mobile coders were in general preferred according to the formula

$$\text{GSM-EFR} = \text{IS-127} = \text{IS-641} > \text{GSM-FR}.$$

Additionally, it should be noted that the overall quality of the GSM-EFR and IS-641 coders were in general equivalent to that of G.729. This is not surprising, since both coders are based on the same technology as G.729.

c) In the context of this study, it can be seen that the overall performance of the G.729 coder was that of a "toll-quality" coder. This result agrees with the subjective assessments performed by the ITU [7]. Although the G.729 coder does not reproduce the background noise as naturally as G.726, these tests showed that both had equivalent quality. For digital speech interpolation (DSI) applications, it should be noted that the voice activity detection and comfort noise generation (VAD/CNI) algorithm standardized for use with the G.729 algorithm (G.729 Annex B) has a more natural reproduction of background noise than the current G.726-based DCME systems. The G.729 VAD/CNI algorithm works very closely with the speech coder (and maximally exploits its properties), while the G.726 VAD/CNI algorithm is based on a generic approach that models any background noise as Gaussian-like noise. Although not yet finalized, the VAD/CNI approach for the G.728 algorithm will be very similar to the G.726-based DCME approach. Therefore, speech quality performance of a DSI system based on G.729 has the potential of outperforming G.726- and G.728-based DSI systems.

d) Tandem connections involving two coders had, in this test, a performance significantly above the acceptability threshold for the fixed network, while most of the connections involving three coders (either three DCME coders or an enhanced cellular coder and two DCME coders) were assessed as equivalent to this threshold. In particular, it should also be noted that the quality of three tandem G.729 and of three tandem G.728 were both equivalent to the acceptability threshold for the fixed network.

In our previous study [2], three categories were established based on the MOS results for the network connections studied. *Class I* are connections that are likely to provide satisfactory quality to the end user. *Class II* are connections that are equivalent to the acceptability threshold. *Class III* connections are those whose end-to-end quality was considered to be below the acceptability threshold, and which may cause the user to complain. Most of the connections in this work fell in Class I. The exceptions were tandem connections with the full-rate GSM coder, and most connections with three low-bit rate coders in tandem, which were classified as Class II connections. None of the connections here were classified in Class III. In contrast, tandem connections of the original 2nd generation coders in [2] were split between Classes II and III, which again indicates the improvement in overall perceived quality for the enhanced GSM, TDMA, and CDMA coders.

It should be noted, however, that since several of the coders used in this study may in fact be better than their field implementations, some Class III connections may, in practice, occur. On the other hand, the acceptability threshold for the fixed network should not be considered absolute when cellular services are involved, because users in general may be willing to compromise speech quality with mobility (e.g. mobile car phones) and accessibility (roaming calls, etc.). In this case, additional implementation- and system-originated impairments (not included in this study) are likely to be tolerated to some extent. Therefore, the overall conclusions regarding the acceptability of several connections may be moderated in view of the additional tolerance allowed by the application context.

4. CONCLUSIONS

It can be seen that most of the connections involving the enhanced coders perform better than the four tandem G.726 threshold, indicating a substantial improvement from the original second generation coders studied in [2]. Also, the tandem performance of the enhanced cellular coders was equal to or better than the four tandem G.726 threshold, indicating that significantly better quality performance can be obtained with the enhanced full-rate cellular systems. It should be noted that since this study considered only the speech coders involved, other system aspects may cause additional degradation.

This study also shows significant degradation with 16 kbit/s-based DCMEs, when compared to the performance of a 32 kbit/s and an 8 kbit/s-based DCME. In some cases, the 16 kbit/s-based DCME also caused performance worse than the threshold. This, in addition to the VAD/CNI approach standardized for G.729, indicates that future 8 kbit/s-based DCME systems will be a suitable evolution from the current 32 kbit/s-based DCME systems.

Table 1
Rank-ordered MOS grouped by front-end coder

Front Coder	DCME	MOS	HSD	D
—	32+8	3.56		>
	8+8	3.41		>
	16+8	3.36		>
	32+8+32	3.34		>
	16+16	3.30		>
	32+8+8	3.29		>
	16+8+8	3.19		=
	16+16+16	3.18		=
	16+8+16	3.05		=
	8+8+8	3.01		=
IS641	32	3.64		>
	8	3.48		>
	16	3.25		>
	8+32	3.24		>
	32+32	3.23		>
	8+8	3.10		=
	8+16	3.02		=
IS127	8	3.56		>
	8+32	3.33		>
	8+8	3.02		=
	8+16	3.02		=
GSM-FR	8	3.05		=
	8+8	2.89		=
	8+32	2.87		=
	8+16	2.79		=
GSM-EFR	32	3.57		>
	8	3.53		>
	2 32	3.35		>
	8+32	3.34		>
	8+8	3.21		>
	16	3.21		>
	8+16	3.16		=
IS127	-	4.29		>
IS641	-	4.24		>
GSM-EFR	-	4.22		>
G.729	-	4.15		>
G.726	-	3.95		>
G.728	-	3.89		>
GSM-FR	-	3.37		>
4-Tandem G.726	-	2.96		—

Legends: HSD: Tukey's Honestly Significant Difference; D: Dunnet's Multiple Comparison Criterion; 32, 16, and 8 are DCME 32, DCME 16 and DCME 8.

Table 2
Rank-ordered MOS grouped by tandem length

Grouping	Circuit	MOS	HSD	D
1 coder **+** 1 tandem *or* 0 coder **+** 2 tandem	IS641+32	3.64		>
	GSM-EFR+32	3.57		>
	32+8	3.56		>
	IS127+8	3.56		>
	GSM-EFR+8	3.53		>
	IS641+8	3.48		>
	8+8	3.41		>
	16+8	3.36		>
	16+16	3.30		>
	IS641+16	3.25		>
	GSM-EFR+16	3.21		=
	GSM-FR+8	3.05		=
1 coder**+** 2 tandems *or* 0 coder**+** 2 tandems	32+8	3.56		>
	8+8	3.41		>
	16+8	3.36		>
	GSM-EFR+32+32	3.35		>
	GSM-EFR+8+32	3.34		>
	IS127+8+32	3.33		>
	16+16	3.30		>
	IS641+8+32	3.24		=
	IS641+32+32	3.23		=
	GSM-EFR+8+8	3.21		=
	GSM-EFR+8+16	3.16		=
	IS641+8+8	3.10		=
	IS641+8+16	3.02		=
	IS127+8+8	3.02		=
	IS127+8+16	3.02		=
	GSM-FR+8+8	2.89		=
	GSM-FR+8+32	2.87		=
	GSM-FR+8+16	2.79		=
0 coder**+** 3 tandems	32+8+32	3.34		>
	32+8+8	3.29		>
	16+8+8	3.19		=
	16+16+16	3.18		=
	16+8+16	3.05		=
	8+8+8	3.01		=
1 coder **+** 0 tandem	IS127	4.29		>
	IS641	4.24		>
	GSM-EFR	4.22		>
	G.729	4.15		>
	G.726	3.95		>
	G.728	3.89		>
	GSM-FR	3.37		>
	4 x G.726	2.96		—

Legends: same as for Table 1.

5. REFERENCES

[1] Dimolitsas S. et al.; "Voice Quality of Interconnected North-American Cellular, European Cellular, and Public Switched Telephone Networks"; *45th IEEE Vehicular Technology Conference (VTC'95)*; Chicago, Illinois, USA; 1995

[2] Campos Neto S. et al.; "Performance assessment of tandem connection of cellular and satellite-mobile coders"; *ICASSP'97*, pp.1379-1382.

[3] ITU-T,"Methods for Subjective Determination of Transmission Quality," Rec.P.800, May, 1996.

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

[5] ITU-T, "Modulated Noise Reference Unit," Rec. P.810, Geneva, 1996.

[6] Glass, G.V. and Hopkins, K.D., "Statistical Methods in Education and Psychology", 2nd Edition, Allyn and Bacon, Boston, USA.

[7] Perkins M.E. et al. "Characterizing the subjective performance of the ITU 8 kb/s speech coding algorithm – ITU-T G.729"; *IEEE Communications Magazine*, Vol.35 No.9, September 1997; pp.74-81.