PERFORMANCE ASSESSMENT OF TANDEM CONNECTION OF CELLULAR AND SATELLITE-MOBILE CODERS †

Simão F.Campos Neto, Franklin L. Corcoran

COMSAT Laboratories, 22300 Comsat Drive, Clarksburg, MD 20871, USA

&

Ara Karahisar

Teleglobe Canada, 1000, rue de La Gauchetière ouest, Montréal, Canada H3B 4X5

ABSTRACT

In the near future, 16 and 8 kbit/s toll- or near-toll low-rate codecs are expected to be used together with 32 kbit/s digital circuit multiplication equipment, providing speech compression and digital speech interpolation. Additionally, a growing proportion of international calls originate from different digital cellular/satellite mobile (C/SM) systems. Knowledge of the end-to-end voice quality of tandem connections is fundamental in the planning of international circuits. Previous studies assessed tandem performance of cellular codecs and the fixed network, however satellite-mobile systems were not included [1,2]. This paper presents a subjective evaluation of the voice quality of tandem connections of C/SM codecs in seven basic scenarios. This study concludes that the number of codecs used in tandem should be minimized and network capacity has to be increased for a given traffic load if voice quality cannot be compromised. In extreme cases, calls originating from C/SM terminals should be transmitted using clear channels.

1. INTRODUCTION

Long-haul international circuits based on satellites and submarine cables are expected to make use of an increasing variety of low-bitrate speech codecs. These devices working at 32 kbit/s, 16 kbit/s and 8 kbit/s will be used in tandem with existing digital circuit multiplication equipment (DCME) based on ITU-T G.726 32 kbit/s ADPCM codecs and using digital speech interpolation. It is expected that 16 kbit/s and 8 kbit/s-based DCME systems will also be introduced in the next couple of years. In addition, we expect a growing proportion of international traffic, originating from cellular mobile and satellite mobile terminals, to use a multiplicity of speech codecs. The speech quality that can be obtained when these codecs are placed in tandem is of vital importance when planning the capacity of international network facilities. However, the impact on the perceived voice quality of the interconnection among these systems has not been thoroughly characterized. Previous studies assessed performance of tandem connections of cellular coders and the fixed network, but have not included satellite-mobile communication systems [1,2]. Adequately characterizing this performance, given the increasing non-linear nature of low-rate speech codecs, demands the use of direct subjective evaluation methods.

† 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 cellular and satellite mobile (C/SM) coders in 7 basic different telecommunication network scenarios. These scenarios were simulated in software and comprised transport over ITU-T G.711 pulse-code modulation 64 kbit/s channels and over DCME based on G.726 32 kbit/s ADPCM and ITU-T G.728 16 kbit/s LD-CELP coding. It should be noted that the tandem connections involving 32 kbit/s DCME systems had a 2.5% load and were simulated using a G.726 codec operating at a net bitrate of 31.2 kbit/s, instead of a fixed bit rate of 32 kbit/s. G.728-based, 16 kbit/s DCMEs were simulated by a LD-CELP at the nominal rate (16 kbit/s). These are respectively labeled DCME32 and DCME16 in this paper. The C/SM codecs considered in this study were US IS54 8 kbit/s VSELP, TIA IS96a Code Division Multiple Access QCELP, Japan's Personal Digital Cellular (PDC) Communication System 6.7 kbit/s VSELP (JVSELP), GSM 13 kbit/s RPE-LTP, Inmarsat Mini-M 4.8 kbit/s AMBE, Inmarsat M 6.4 kbit/s IMBE, Inmarsat B 16 kbit/s APC, and Inmarsat Aeronautical 9.6 kbit/s MP-LPC.

List of Speech Coding Acronyms Used	
ADPCM	Adaptive Differential Pulse-Code Modulation
AMBE	Advanced Multi-Band Excitation
APC	Adaptive Predictive Coding
IMBE	Improved Multi-Band Excitation
LD-CELP	Low-Delay Code-Excited Linear Prediction
MP-LPC	Multi-Pulse Linear Prediction
QCELP	Qualcomm Code-Excited Linear Prediction
RPE-LTP	Regular Pulse Excitation, Long-Term Prediction
VSELP	Vector-Sum Excited Linear Prediction

This characterization was functionally divided into two subjective listening experiments and included tandem performance for seven circuit types:

1. no tandem,
2. tandem with 32 kbit/s-based DCME,
3. tandem with 16 kbit/s-based DCME,
4. tandem with 32 kbit/s-based DCME followed by another 32 kbit/s-based DCME,
5. tandem with 16 kbit/s-based DCME followed by another 32 kbit/s-based DCME,
6. tandem with 32 kbit/s-based DCME terminated in Europe's GSM, and
7. tandem with 32 kbit/s-based DCME terminated in Japan's PDC

2. EXPERIMENTAL DESIGN

Several factors were taken into consideration for the definition of the subjective experiment design for measuring codec 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 tests were 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 experiments 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 codec-condition combinations. Two male and two female talkers were used for the two experiments, with a total of 12 sentence-pairs per talker.

The conditions, which were evaluated by 48 non-expert listeners in each experiment (96 in total), included the network configurations whose assessment was sought, as well as a number of reference systems, including Modulated Noise Reference Units (MNRU) [5], ITU-T G.728 16 kbit/s LD-CELP, ITU-T G.726 32 kbit/s ADPCM codec, and four interconnected 32 kbit/s ADPCM codecs (whose cumulative distortion is accepted as perceptually equivalent to the maximum end-to-end quantization distortion recommended by the ITU-T for international wireline connections). Listeners cast 192 votes for each of the test conditions.

3. RESULTS & DISCUSSION

Table 1 and Table 2 show the rank-ordered results respectively from Experiments 1 and 2. The codecs were grouped in circuit connections that are similar in quality.

In these tables, *MOS* represents the Mean Opinion Score. The standard errors for the test conditions, which were 0.07 in average, have not been reported due to space reasons, but their absence is compensated by presenting 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 indicates, using Dunnet's Multiple Comparison Test criterion [6], with a ">" conditions that are statistically better than the four-tandem G.726 condition, with a "<" the conditions with are worse than the four-tandem G.726 condition, and with an "=" the conditions which are statistically 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 G.726 ADPCM.

The principal observations from the MOS performance are as follows:

a) When a cellular or satellite mobile (C/SM) codec is in the front end of a tandem connection, the overall performance is not dependent on whether a DCME32 or a DCME16 is used. Therefore, for calls originated from these C/SM codecs with a single-pass on the fixed-network, a DCME32 or a DCME16 codec may in general be used interchangeably.
b) However, in the above interconnection scenario, the lower acceptability threshold was exceeded when either JVSELP or RPE-LTP codec were used in conjunction with a DCME16, the IMBE voice codec and the MP-LPC voice codec.
c) There is, in general, a perceived degradation when a C/SM codec is in the front end of a tandem connection and a DCME32 is preceded by a DCME16, if compared to the connection using two DCME32.
d) When C/SM-to-C/SM calls were studied, the deployment of DCME32 or DCME16 did not change the end-to-end quality. However, this end-to-end quality was below the four-G.726 threshold for the best part of the conditions tested.
e) In the tandem connections involving no front-end C/SM codec, the connection with three DCME16 performed statistically better than the connection involving two DCME16 and one DCME32 or the connections involving one DCME16 and two DCME32. The performance of the connection with three DCME16 was equivalent to the acceptability threshold.

f) The tandem connections involving no front-end C/SM codec and two DCME16 and one DCME32 codec or one DCME16 and two DCME32 had statistically equivalent MOS values, which were below the acceptability threshold.
g) The rank-order scores of the connections involving a DCME32 were in general higher than the rank-order scores of the connections involving DCME16, although these differences were not always statistically significant. This indicates a trend which may be considered in network planning.

These general observations can be used to classify the circuit connections studied in three categories:

Class I. Connections in Class I are those likely to provide satisfactory quality to the end user. In this study, Class I connections were in general those when the codecs were not in a tandem scenario. This was also the case for the QCELP voice codec in tandem with a single DCME32.

Class II. Class II connections are those which are equivalent to the four-G.726 tandem threshold. This was the case for:
- QCELP codec in tandem with a two 32 kbit/s DCME;
- AMBE, VSELP, RPE-LTP, and JVSELP in tandem with a one or a two DCME32
- QCELP, AMBE, and VSELP in tandem with a DCME16;
- AMBE codec in tandem with a DCME32;
- AMBE in tandem with a DCME16;
- VSELP in tandem with a DCME32 and JVSELP;
- APC in tandem with a DCME32 or a DCME16;
- APC in tandem with either DCME32 or DCME16 followed by a DCME32;
- Single transcoding of the IMBE codec;
- Three DCME16 in tandem.

In general, C/SM connections in tandem with a DCME32 fell in this category as did some of the C/SM connections.

Class III. These are connections whose end-to-end quality were considered to be below the four-G.726 tandem threshold and which may cause the user to complain. This was the case for:
- JVSELP or RPE-LTP in tandem with one DCME16;
- QCELP, AMBE, VSELP, RPE-LTP, and JVSELP in tandem with a DCME16 followed by a DCME32;
- VSELP in tandem with DCME16 followed by either the RPE-LTP codec or the JVSELP codec;
- QCELP in tandem with DCME32 or DCME16, followed either by RPE-LTP or JVSELP;
- AMBE in tandem with DCME16 and followed by a DCME32;
- IMBE or APC in tandem with either a DCME32 or a DCME16;
- IMBE or MP-LPC in tandem with either a DCME32 or a DCME16 followed by a DCME32;
- Single transcoding of the MP-LPC codec;
- Two DCME16 in tandem with one DCME32;
- Two DCME32 in tandem with one DCME16.

A part of C/SM connections in single tandem with DCME16 fell in this category. In general, the C/SM tandem connections with DCME16 associated with other speech codecs also fell in this category.

It should be noted that some effects of deployment in a real network were not considered in Experiments 1 and 2: transmission delay, echo, environmental noise, speech level mismatch, etc. When these impairments are present, the overall quality figures observed in this test are likely to be optimistic. As a result, the configurations in these experiments that obtained scores below the four-G.726 threshold should be considered as the least desirable. On the other hand, the configurations that met the four-G.726 threshold must be carefully implemented. In addition, other psychological factors may influence the decision on which connections should be avoided. In particular, when C/SM calls are deployed, there is a user expectation factor which reduces quality requirements in exchange for mobility (e.g. mobile car phones) and accessibility (e.g. remote region, airplane, or maritime calls). In this case, the additional impairments mentioned above, and not included in this study, are likely to be tolerated. Therefore, the overall conclusions regarding the acceptability of several connections may be used in view of the additional allowance imposed by its application context.

4. CONCLUSIONS

It can be seen that most of the connections involving the satellite mobile codecs exceed 4 G.726 tandem threshold, as did all the connections between US national cellular systems and the European and Japanese full-rate cellular systems (GSM and PDC). However, we feel that users will accept a lower quality in exchange for mobility. The results also show significant degradation with the use of a 16 kbit/s-based DCME, when compared to the performance of a 32-kbit based DCME, which in some cases also cause the threshold to be exceeded. Network planners should minimize the number of transcodings, be ready to swap increased network capacity in exchange for better speech quality, and even consider the use of clear channels for cellular and satellite mobile calls.

Table 1
Rank-ordered MOS Presentation of Scores for Experiment 1

<table>
<tr><th>Test Condition</th><th>MOS</th><th>HSD</th><th>D</th></tr>
<tr><td>QCELP+1 DCME 32k</td><td>3.48</td><td>|</td><td>></td></tr>
<tr><td>AMBE+1 DCME 32k</td><td>3.24</td><td>| |</td><td>=</td></tr>
<tr><td>VSELP+1 DCME 32k</td><td>3.21</td><td>| |</td><td>=</td></tr>
<tr><td>RPE-LTP+1 DCME 32k</td><td>3.20</td><td>| |</td><td>=</td></tr>
<tr><td>JVSELP+1 DCME 32k</td><td>3.20</td><td>| |</td><td>=</td></tr>
<tr><td>VSELP+1 DCME 16k</td><td>3.13</td><td>| |</td><td>=</td></tr>
<tr><td>QCELP+1 DCME 16k</td><td>3.08</td><td>| |</td><td>=</td></tr>
<tr><td>AMBE+1 DCME 16k</td><td>3.02</td><td>| |</td><td>=</td></tr>
<tr><td>JVSELP+1 DCME 16k</td><td>2.90</td><td>|</td><td><</td></tr>
<tr><td>RPE-LTP+1 DCME 16k</td><td>2.89</td><td>|</td><td><</td></tr>
<tr><td>QCELP+2 DCME (32k+32k)</td><td>3.41</td><td>|</td><td>=</td></tr>
<tr><td>VSELP+2 DCME (32k+32k)</td><td>3.15</td><td>| |</td><td>=</td></tr>
<tr><td>JVSELP+2 DCME (32k+32k)</td><td>3.13</td><td>| |</td><td>=</td></tr>
<tr><td>AMBE+2 DCME (32k+32k)</td><td>3.05</td><td>| |</td><td>=</td></tr>
<tr><td>RPE-LTP+2 DCME (32k+32k)</td><td>3.02</td><td>| |</td><td>=</td></tr>
<tr><td>VSELP+2 DCME (16k+32k)</td><td>2.92</td><td>| | |</td><td><</td></tr>
<tr><td>QCELP+2 DCME (16k+32k)</td><td>2.90</td><td>| | |</td><td><</td></tr>
<tr><td>AMBE+2 DCME (16k+32k)</td><td>2.87</td><td>| | |</td><td><</td></tr>
<tr><td>RPE-LTP+2 DCME (16k+32k)</td><td>2.81</td><td>| |</td><td><</td></tr>
<tr><td>JVSELP+2 DCME (16k+32k)</td><td>2.69</td><td>|</td><td><</td></tr>
<tr><td>VSELP+1 DCME 32k+JVSELP</td><td>2.97</td><td>|</td><td>=</td></tr>
<tr><td>QCELP+1 DCME 32k+JVSELP</td><td>2.90</td><td>| |</td><td><</td></tr>
<tr><td>QCELP+1 DCME 16k+JVSELP</td><td>2.81</td><td>| | |</td><td><</td></tr>
<tr><td>VSELP+1 DCME 16k+JVSELP</td><td>2.78</td><td>| | |</td><td><</td></tr>
<tr><td>QCELP+1 DCME 32k+RPE-LTP</td><td>2.68</td><td>| | |</td><td><</td></tr>
<tr><td>VSELP+1 DCME 32k+RPE-LTP</td><td>2.64</td><td>| |</td><td><</td></tr>
<tr><td>QCELP+1 DCME 16k+RPE-LTP</td><td>2.56</td><td>|</td><td><</td></tr>
<tr><td>VSELP+1 DCME 16k+RPE-LTP</td><td>2.53</td><td>|</td><td><</td></tr>
<tr><td>APC</td><td>4.18</td><td>|</td><td>></td></tr>
<tr><td>G.728</td><td>4.07</td><td>|</td><td>></td></tr>
<tr><td>G.726</td><td>4.01</td><td>| |</td><td>></td></tr>
<tr><td>QCELP</td><td>3.99</td><td>| |</td><td>></td></tr>
<tr><td>VSELP</td><td>3.91</td><td>| | |</td><td>></td></tr>
<tr><td>JVSELP</td><td>3.74</td><td>| | |</td><td>></td></tr>
<tr><td>AMBE</td><td>3.61</td><td>| |</td><td>></td></tr>
<tr><td>RPE-LTP</td><td>3.57</td><td>|</td><td>></td></tr>
<tr><td>4xG.726</td><td>3.18</td><td>|</td><td>=</td></tr>
<tr><td>IMBE</td><td>3.12</td><td>|</td><td>=</td></tr>
<tr><td>MP-LPC</td><td>2.83</td><td>|</td><td><</td></tr>
</table>

Legend: HSD: Tukey's Honestly Significant Difference; D: Dunnet's Multiple Comparison Criterion

Table 2
Rank-ordered MOS Presentation of Scores for Experiment 2

<table>
<tr><th>Test Condition</th><th>MOS</th><th>HSD</th><th>D</th></tr>
<tr><td>APC+1 DCME 32k</td><td>3.32</td><td>|</td><td>=</td></tr>
<tr><td>APC+1 DCME 16k</td><td>3.20</td><td>|</td><td>=</td></tr>
<tr><td>AMBE+1 DCME 16k</td><td>3.08</td><td>|</td><td>=</td></tr>
<tr><td>AMBE+1 DCME 32k</td><td>3.08</td><td>|</td><td>=</td></tr>
<tr><td>IMBE+1 DCME 32k</td><td>2.48</td><td>|</td><td><</td></tr>
<tr><td>IMBE+1 DCME 16k</td><td>2.43</td><td>|</td><td><</td></tr>
<tr><td>MP-LPC+1 DCME 32k</td><td>2.35</td><td>|</td><td><</td></tr>
<tr><td>MP-LPC+1 DCME 16k</td><td>2.33</td><td>|</td><td><</td></tr>
<tr><td>APC+2 DCME (32k+32k)</td><td>3.29</td><td>|</td><td>=</td></tr>
<tr><td>APC+2 DCME (16k+32k)</td><td>3.02</td><td>| |</td><td>=</td></tr>
<tr><td>AMBE+2 DCME (32k+32k)</td><td>3.00</td><td>|</td><td>=</td></tr>
<tr><td>AMBE+2 DCME (16k+32k)</td><td>2.82</td><td>|</td><td><</td></tr>
<tr><td>IMBE+2 DCME (32k+32k)</td><td>2.48</td><td>|</td><td><</td></tr>
<tr><td>MP-LPC+2 DCME (32k+32k)</td><td>2.34</td><td>| |</td><td><</td></tr>
<tr><td>IMBE+2 DCME (16k+32k)</td><td>2.25</td><td>| |</td><td><</td></tr>
<tr><td>MP-LPC+2 DCME (16k+32k)</td><td>2.17</td><td>|</td><td><</td></tr>
<tr><td>3 DCME (16k+16k+16k)</td><td>3.08</td><td>|</td><td>=</td></tr>
<tr><td>3 DCME (16k+32k+16k)</td><td>2.87</td><td>|</td><td><</td></tr>
<tr><td>3 DCME (32k+16k+32k)</td><td>2.85</td><td>|</td><td><</td></tr>
<tr><td>APC</td><td>4.17</td><td>|</td><td>></td></tr>
<tr><td>G.728</td><td>4.10</td><td>| |</td><td>></td></tr>
<tr><td>QCELP</td><td>3.85</td><td>| |</td><td>></td></tr>
<tr><td>VSELP</td><td>3.83</td><td>| |</td><td>></td></tr>
<tr><td>AMBE</td><td>3.70</td><td>|</td><td>></td></tr>
<tr><td>JVSELP</td><td>3.57</td><td>|</td><td>></td></tr>
<tr><td>RPE-LTP</td><td>3.56</td><td>|</td><td>></td></tr>
<tr><td>4xG.726</td><td>3.13</td><td>|</td><td>=</td></tr>
<tr><td>IMBE</td><td>3.01</td><td>|</td><td>=</td></tr>
<tr><td>MP-LPC</td><td>2.67</td><td>|</td><td><</td></tr>
</table>

Legend: HSD: Tukey's Honestly Significant Difference; D: Dunnet's Multiple Comparison Criterion

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] Dimolitsas S.et al., *"Impact of Network Routing on the End-to-End Transmission Quality of Cellular and PCS Connections";* Submitted to IEEE PCS Vehicular Transactions

[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.