QUALITY AND COMPUTATION LOAD REDUCTION ACHIEVED BY APPLYING SMART TRANSCODING BETWEEN CELP SPEECH CODECS

Christophe Beaugeant, Hervé Taddei

Nokia Siemens Networks, Munich, Germany
E-mail: christophe.beaugeant,herve.taddei@nsn.com

ABSTRACT

The increasing number of standardized speech codecs implies interoperability issues between networks deploying different and sometime incompatible standards. To assure interoperability, transcoding from one codec format to another is necessary at gateways between networks. This transcoding reduces speech quality, involves computation load and additive delay. Recently, a fair amount of work has been conducted on studying alternative transcoding methods reducing complexity and delay. In this context this paper further elaborates on some alternative solutions. It presents for each of them an evaluation of quality in relation with complexity reduction. The impact of frame lost and cascading transcoding on such smart solutions is also discussed. The obtained panel of results allows clear statements on the properties and highlights the interest of smart transcoding in terms of quality, complexity and delay reduction.

1. INTRODUCTION

The last two decades have seen the development of several speech/audio codecs for different networks (e.g. Universal Mobile Telephone System (UMTS) networks with Adaptive Multi-Rate (AMR) [1], fixed networks with G.729A [2]) and/or services (e.g. conversational applications over internet). It has led to the deployment of codecs that are not interoperable with each other. The usual way to handle this issue is to decode one codec bitstream and to re-encode it into the target codec bitstream format. Such 'classical transcoding' is far from optimal as it implies computation load, decreases the speech quality and increases the algorithm delay. In order to avoid these drawbacks a new method, called smart transcoding, profits from codec similarities. This was introduced in [3]. It takes advantage that many speech codecs are based on Code Excited Linear Prediction (CELP) schemes and use very often the same parameters. In such a method, similar parameters are mapped from one codec into the target codec format. Papers [3][4][5][6][7] have shown that speech quality can be enhanced or at least be kept constant in comparison with classical transcoding. In the same time complexity and delay are reduced.

In this paper, we will especially emphasize the link existing between achieved quality and complexity reduction of such smart transcoding schemes. Compared to the works already cited, we introduce some experiments regarding robustness against packet loss and multiple transcoding. This will give us a complete figure of merit of smart transcoding in terms of quality and complexity reduction. The general principle of smart transcoding together with a short overview of CELP coding are depicted in section 2. Different smart transcoding schemes dealing with the mapping of LPC coefficients and of the pitch are discussed in section 3. Section 4 shows how these schemes can reduce complexity without compromising the quality. Sections 5 and 6 show that smart transcoding optimizations are as robust against packet loss as classical transcoding and behave properly in multiple transcoding scenarios.

2. CELP CODING AND SMART TRANSCODING PRINCIPLE

Smart transcoding described in this article deals with modification and mapping of parameters of CELP codecs. In order to introduce the context of our study, this section presents an overview of CELP codecs as well as a description of the principle of smart transcoding scheme.

2.1 Examples of CELP coding schemes: G.729A, AMR

In order to assess the properties of smart transcoding solution, our experiments study the influence of the transcoding on two widely deployed codecs: AMR [1] and G.729A [2]. The AMR codec has a frame length of 20 ms, the G.729A of 10 ms. AMR has 8 different bitrates (from 4.75 kbit/s to 12.2 kbit/s) and G.729A has one bitrate at 8 kbit/s. In both codecs, each frame is divided into subframes of 5ms (so 4 subframes for the AMR and 2 for the G.729A).

Both codecs use a 10^{th} order linear prediction filter. The Linear Prediction Coefficients (LPC), are computed for each frame by solving a linear system of equations (except for AMR mode at 12.2 kbit/s where two LPC searches are computed per frame). Linear prediction analysis uses the autocorrelation method with a windowed input signal. For example, in G.729A, a 30 ms asymmetric window is applied with a look-ahead of 5 ms. Every 10 ms, the autocorrelation coefficients of windowed speech are computed. These coefficients are used as input to the Levinson-Durbin algorithm to get the LPC coefficients.

After filtering the input signal by the LPC filter, a residual is obtained for each subframe. This residual is quantized and transmitted for reconstruction of speech to the decoder. To do so, first an adaptive codebook search is performed on each subframe leading to a pitch delay and an adaptive gain value. Pitch is computed in two steps. First, an open-loop pitch value T_0 is obtained by minimizing the mean square value of a residual signal, i.e. the output of the LPC analysis filter. Secondly, a closed-loop search is performed minimizing the error between the original signal and a reconstructed signal obtained by filtering the past excitation at delay T_k through the LPC filter. In order to reduce the computation load, the delay values T_k are restricted to an interval around the pitch value T_0 determined by the open loop search.

By subtracting the excitation of the adaptive codebook multiplied with its respective gain a new target signal is obtained. This signal is used to process the fixed codebook search (fixed codebook index and fixed gain value). Finally the LPC coefficients, the pitch delays, the fixed codebook indexes and both fixed and adaptive gains are transmitted to the decoder.

The AMR and G.729A decoders perform the synthesis of the speech using the transmitted parameters. The adaptive excitation is found by interpolating the past excitation multiplied by the adaptive gain. The fixed excitation is obtained by multiplying the fixed codebook vector with the fixed codebook gain. Both excitations are then summed up and enter the LPC synthesis filter. Finally, a post-processing algorithm is applied to enhance the quality of the reconstructed speech.

2.2 Smart transcoding principle

Usually, when transcoding from a bitstream format of codec A into a bitstream format of codec B, bitstream A is first decoded. The obtained decoded signal is then encoded in target format B by encoder B. In case both codecs are CELP codecs, bitstreams A and B transmit similar set of parameters, i.e. the ones depicted in section 2.1. The key idea of smart transcoding consists in avoiding the computation of parameters already available. An intelligent mapping and quantization of the parameters available in bitstream A into bitstream parameters B allow skipping many functions and hence reduce the computation load of the transcoding. As depicted in Fig. 1, only a partial decoding is necessary, to extract the parameters from bitstream A. Their mapping as well as a partial encoding builds the accurate bitstream B.

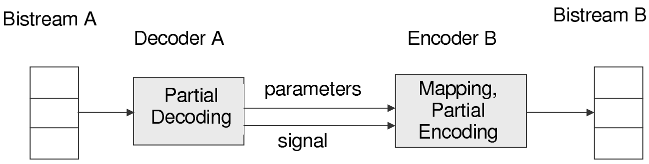

Figure 1: Principle of smart transcoding

Smart transcoding schemes were already tested with success on different CELP parameters as for example the LPC coefficients, pitch values and the gains (see [3]-[7]). Fixed codebook index was also smartly transposed in [4, 7] by restricting the search of the fixed codebook at encoder B around the value given by decoder A. In addition to this mapping of parameters, suppression of redundant functions, for example a high pass filtering done in encoder A and B, can further reduce the computation load.

3. SMART TRANSCODING SOLUTIONS

In this paper we only considered the mapping of LPC coefficients and of the pitch lag. We also checked the influence of skipping high-pass filtering, low-pass filtering and up/downscaling. In our proposed schemes, the look-ahead at decoder B is skipped, as a result no extra delay is added as when compared to the classical transcoding. We studied the following transcoding schemes:

- Scheme T1: we skip in decoder A and encoder B the possible downscaling. We also skip the low-pass, high-pass filtering at encoder B depending on the one already applied at encoder A.
- Scheme T2: the Levinson Durbin algorithm in encoder B is skipped and we mapped the LPC coefficients from decoder A to encoder B. Then quantization of the LPC at encoder B is performed.
- Scheme T3: the open loop pitch search is skipped at encoder B. The pitch is extracted from decoder A and used as input for the closed loop pitch search of encoder B.
- Scheme T4: open loop search and closed loop search are both skipped at encoder B. The pitch values from decoder A are used as pitch values for encoder B. Only pitch quantization is processed.
- Scheme T5: All processes of T1, T2 and T3 are applied.
- Scheme T6: All processes of T1, T2 and T4 are applied.

These schemes can be applied to many standardized CELP codecs but special attention has to be paid for codecs with different structures (e.g. different number of subframes). To check the relevance and the properties of the retained schemes T1-T6, we tested them on the pair AMR/G729A. It leads to several concrete implementations.

T1 was implemented by skipping the high-pass and low pass filtering, as both encoders apply similar filters. Redundant downscaling/upscaling were also bypassed. The mapping of the LPC coefficients (scheme T2) between the AMR at 12.2 kbit/s and G.729A can easily be done as the LPC coefficients are computed every 10ms in both codecs. For the other AMR modes where the computation is made each 20 ms, interpolation or extrapolation of the LPC coefficients is applied. For instance, if codec A is the AMR 7.4 kbit/s, its decoder provides 4 sets of interpolated LPC coefficients (1 set per subframe of 5 ms). In our implementation, the 1^{st} and 3^{rd} set of AMR LPC coefficients are given as input to the quantization function of G.729A encoder. For the inverse scenario, encoder B = AMR 7.4 kbit/s, for each AMR frame of 20 ms, the G.729A decoder provides two sets of LPC coefficients. Only the first set is used as input to the LPC quantization function of the AMR encoder.

Regarding T3 and T4, the pitch resolution is different between AMR 12.2 kbit/s and the G.729A (resolution of 1/6 and 1/3 respectively). A correct mapping is done by looking for the lower closest value (from AMR to G.729A) and by keeping the same value (from G.729A to AMR). For other mode of the AMR, the pitch resolution is the same as for G.729A, so a direct mapping is possible.

These concrete processes were implemented to study the properties of schemes T1-T6 in the different scenarios we introduce in the following sections.

4. QUALITY VS. COMPUTATION LOAD

Smart transcoding has a triple influence over classical transcoding: impact on delay, on computation load and on speech quality.

As already mentioned, smart transcoding avoids a complete decoding and re-encoding and also avoids the look-ahead at the encoder due to the LPC analysis window. Accordingly, any of the solution T1-T6 provides a decrease of 5 ms delay compared to normal transcoding.

Scheme	T1	T2	T3	T4	T5	T6
CPU reduction	7%	20%	10%	24%	37%	49%

Table 1: Complexity reduction of smart transcoding schemes

Table 1 gives an estimation of typical complexity reduction that can be obtained with the schemes presented in section 3. These average complexity values are obtained from the analysis of the C fixed point reference code of the AMR codec. In this table, normal transcoding is the reference (100% of complexity). These values are in accordance with the complexity reduction mentioned in [4, 7].

To evaluate the quality of our smart transcoding scheme, we performed experiments using PESQ (Perceptual Evaluation of Speech Quality) [8]. PESQ produces values which can be assimilated to a Mean Opinion Score (MOS) of the signal. Using this tool, we proceed the following experiments:

- Measurement of PESQ values for speech files obtained when applying normal transcoding and smart transcoding. Our speech database contained 48 files of 8 s each at -26 dBov mixing equally French and German voices as well as male and female voices.
- Computation of the difference between the PESQ mean score obtained on signal processed by the smart transcoding and the PESQ mean score obtained on signal processed by a classical transcoding (i.e. complete decoding and encoding).

Fig 2 and 3 present the results for transcoding between G.729A and three AMR modes (7.4, 7.95, 12.2 kbit/s). They plot the difference of PESQ for the schemes (T1-T6) with regard to their complexity reduction (according to Table 1).

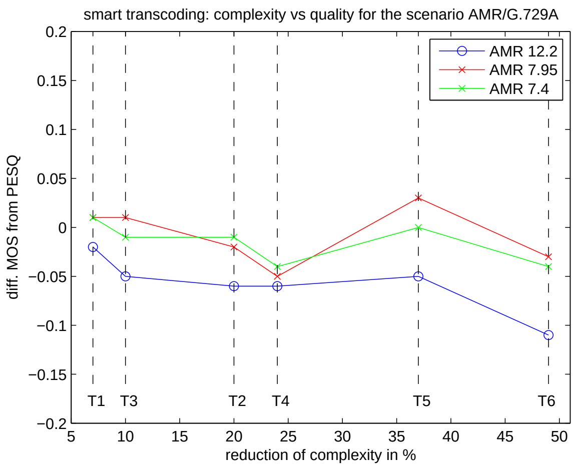

Figure 2: PESQ differences for the scenario AMR / G.729A.

The curves show that all solutions provide a quality comparable to the normal transcoding. It can be noted that the schemes skipping the closed-loop search seem to be the most critical ones. Nevertheless, their quality is relatively the same as for the normal transcoding case (differences less than 0.15 MOS in average). Smart transcoding applied on lower AMR modes seems to slightly enhance the signal quality.

Another remark deals with the trade-off between reduction of complexity and quality. In the studied solutions, there is no clear correlation between complexity reduction and quality reduction. For instance, solution T2 leads generally to a relatively lower quality than solution T5. It seems that mapping a single parameter is sub-optimal when compared to mapping several. In fact, T5 maps the LPC coefficients and pitch, so that the processing at encoder B is reduced to the computation of the gains and of the fixed codebook. The computation of encoder B is more directed than with T2, where only the LPC coefficients are mapped. It seems that this decision directed approach brings better results or in other words that encoder B needs to be driven closely and properly by the mappings to assure quality.

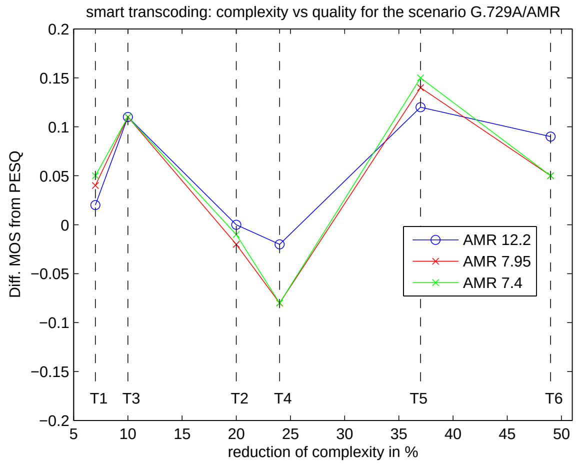

Figure 3: PESQ differences for the scenario G.729A / AMR.

Informal listening tests are in line with the results obtained with PESQ for schemes T1, T2, T3 and T5 (i.e. equivalence between normal transcoding and smart transcoding). They also show that for lower AMR bitrates, the LPC mapping is the key point for good quality. Listeners had the feeling that a better sharpening of the spectral envelope was obtained. The processed signal sounds more intelligible, more present. This effect is clearly noticeable in the case encoder A = AMR 7.4/7.95 kbit/s and encoder B = G.729A.

PESQ results give unclear trends with regard to the influence of skipping the closed loop search (solutions T4, T6). In Fig. 2, T4 is shown as good as T3 and T6 as good as T5, whereas in Fig. 3, T4 is worse than T3 and T6 is worse than T5. Informal listening tests clarify these rankings. T4 was assessed to bring worse results than T3, and T6 worse results than T5. Some artifacts are added in solutions skipping the closed loop analysis (slight distortion for the AMR 12.2 kbit/s; audible additive noise during voiced periods for mode 7.4 kbit/s in particular for female samples).

For the lowest AMR bitrates, quality impacts of smart transcoding are more noticeable. The quality enhancement due to the LPC mapping is more significant. The artifacts due to skipping the closed loop search are more annoying. In concrete terms, it means that for low bitrates codec, we may expect that smart transcoding not only allow decreasing the complexity, but also has a significant impact on quality. Solution T5 is a good example of such property as it leads to a reduction of 37% of complexity with slight speech quality enhancement.

5. ROBUSTNESS AGAINST FRAME LOST

Frame Erasure Concealment (FEC) algorithms are usually provided in most standardized codecs. They process an estimation of a lost frame by extrapolating it from previously received parameters. Introducing the parameter mappings as described in section 3 leads to questions with regard to robustness of transcoding. In worst cases, if a frame is lost, we could set all the parameters to zero and transcode them into bitstream B format. But it would produce noticeable artifacts. Introducing FEC at decoder A permits to estimate a decoded signal and accordingly an accurate computation of these parameters at encoder B.

This section presents results obtained with the couple G.729A and AMR when random frame loss occurs on codec A side. In all scenarios, FEC algorithm of decoder A was activated. We also take care that the recovery of LPC and pitch done by the FECs at decoder A are done before the mapping of these parameters to decoder B. Fig 4-5 present the results we obtained for two typical scenarios. Other AMR modes and other scenarios (AMR/G729, G729/AMR) were also studied. They produced similar figures. We used the same processing database as the one from section 4.

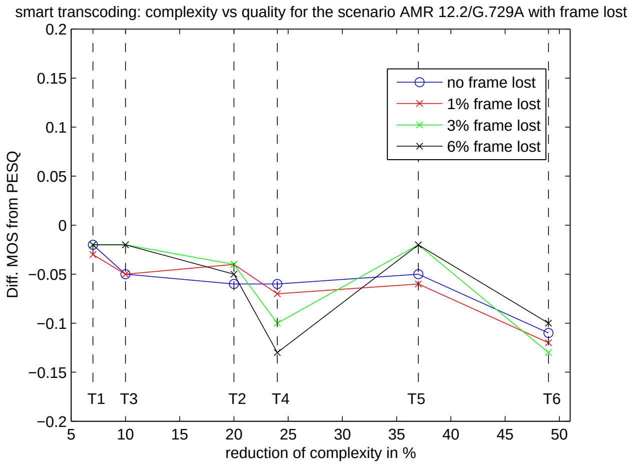

Figure 4: PESQ differences with random frame lost, scenario AMR 12.2 / G.729A .

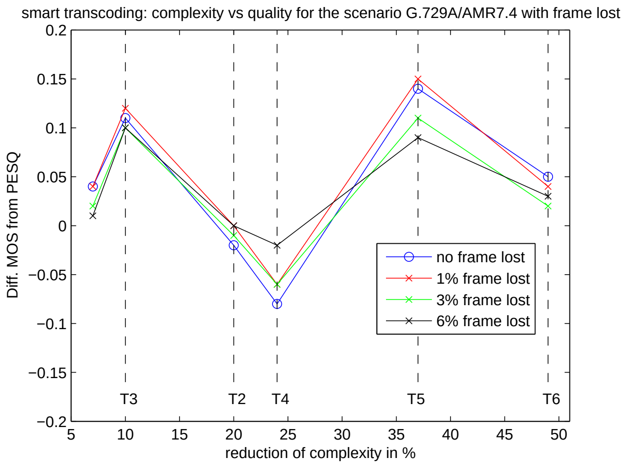

Figure 5: PESQ differences with random frame lost, scenario G.729A / AMR 7.4 .

The objective measurements indicate that smart transcoding is not more disturbed than normal transcoding. The quality impact of frame lost is about the same for smart transcoding as for normal transcoding. One can see that the difference PESQ values stay almost constant with regards to error rate except for solutions T4 and T5. Informal listening tests show that the overall quality of smart transcoding is as good as normal transcoding, whatever the error rate is, contradicting the results of PESQ for T4 and T5. Quality is even found better for solutions T3 and T5. There is no increase of audible artifacts when using smart transcoding under error condition as when compared with the disturbance introduced by normal transcoding. In fact, the FEC algorithm included in the AMR and G.729A estimate the LPC and pitch values. As we take care to run the FEC algorithm before doing any mapping, it is normal to get similar results as for the normal transcoding.

An alternative solution would be to indicate to decoder B when a frame is lost, e.g. by mapping the bad frame indication from decoder A to encoder B. Then decoder B would run its FEC algorithm. In this scenario decoder A should nevertheless activate its FEC algorithm so that its working memories are kept up to date. This solution could enhance the quality as it, in some way, avoids to “transcode” the output of an FEC algorithm. It presents the advantage of reducing the computation load as encoder B does not need to be run.

6. MULTIPLE TRANSCODING

In communication system, it can occur that transcoding is applied more than once. A plausible scenario is a communication between two WLAN (Wireless Local Area Network) phones using of the AMR codec on different local networks. Communication between the local networks may be done through a connection using G.729 as codec. It results that the speech signal is first encoded with the AMR, transcoded from AMR to G.729 in the gateway between the local network and the fixed network, and transcoded once more from G.729 to AMR at the gateway of the far-end WLAN network. This leads to a cascading of two transcodings.

In this section we study the behavior of the smart transcoding solutions in such cascaded scenario. Several experiments have been conducted depending on the different codec combinations and on the number of smart transcodings. Regarding this number, there are actually three cases: when smart transcoding is applied once at the first transcoding step, when it is applied once but at the second transcoding step and when it is applied twice at both transcoding steps.

Our quality reference is the one determined when no smart transcoding is applied. Similar experiments as depicted in section 3 were conducted. We considered the AMR at 12.2, 7.95 and 7.4 kbit/s, the G729A, the same speech database and we processed PESQ for the different scenarios. Fig 6-7 show the results for two typical scenarios. Other scenarios (with different smart transcoding locations and different AMR modes) provide the same kind of results.

All our objective measurements show that in a cascading scenario, introducing transcoding at any place lead almost to the same quality or even slightly better quality than normal transcoding. Applying smart transcoding twice leads to the best results. Informal listening tests confirm this general trend, i.e. smart transcoding provides a better speech clarity compared to normal transcoding. We also felt that normal transcoding has a comparable effect as smoothing the speech. On the contrary, speech obtained with smart transcoding seems to a certain extend more present, less smoothed.

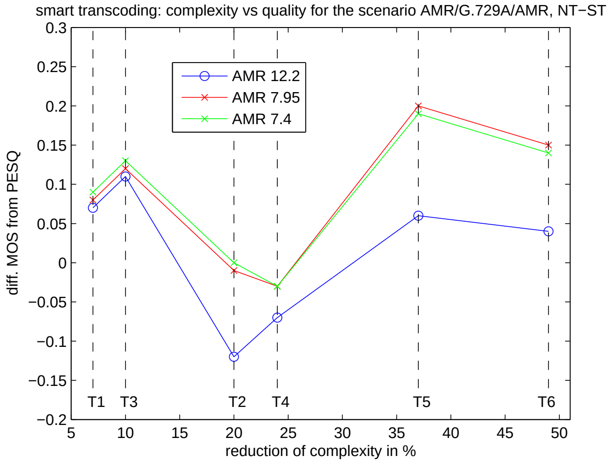

Figure 6: PESQ differences in scenario AMR / G.729A / AMR, smart transcoding applied between G.729A /AMR.

It seems that cascading underlines advantages and drawbacks of the smart transcoding. The properties mentioned in section 4 are more significant. Listening tests show for instance that skipping the pitch closed loop search decreases significantly the quality: a perceptual noise is added to the speech and some artifacts during voiced periods can be heard.

It is also important to notice that in scenarios where smart transcoding is applied twice, solution T5 provides significantly better quality than normal transcoding (0.25 PESQ increase for the best cases). Informal listening tests show that the intelligibility and the clarity of speech processed with T5 are indeed enhanced compared to the normal transcoding. Nevertheless, an additive noise during speech periods can also be heard for low AMR bitrates. This noise is not present in normal transcoding mode. If the impact on quality of the additive noise is subjectively less important than the enhancement obtained on the speech intelligibility, it puts into perspective the high enhancement assessed by the objective measurements.

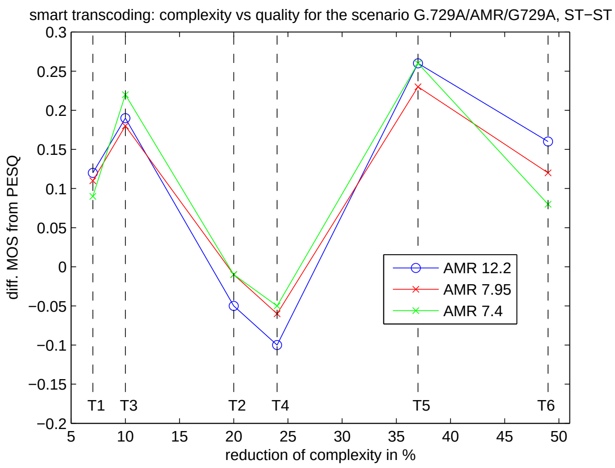

Figure 7: PESQ differences in scenario G.729A / AMR / G.729A, smart transcoding applied twice.

7. CONCLUSION

In this article, we studied the interest to map CELP parameters in order to avoid decoding and re-encoding bitstreams when transcoding. Several options were presented and an overview of the ratio quality vs. complexity reduction was conducted by studying a typical set of codecs. We focused on LPC mapping, pitch mapping and skipping pre/post processing. The retained solutions permit to reduce the delay and to keep the same quality as with normal transcoding (i.e. decoding and re-encoding) while reducing up to 37% the computation load. 49% of complexity reduction is even possible to the extend of the acceptance of slight degradation on speech. Such conclusions are in accordance with the ones proposed in [3]-[7]. They precise them by giving a detailed presentation of the properties of different mappings. We point out that from a quality point of view, mapping of several parameters leads to better results compared to the mapping of a single one. Experiments on the robustness against frame erasures and in cascading scenarios complete our study. It permits to conclude that smart transcoding is a really interesting option. It drastically reduces the computation load without compromising the quality. Moreover, for low-bit rate codecs, quality enhancement through correct mapping of parameters is also expected. It lets us foresee that advanced mapping of parameters could cope with even better quality, leading to an optimum situation where high complexity decrease and quality increase are jointly possible.

REFERENCES

[1] 3GPP TS 26.090, “Adaptive Multi-Rate (AMR) Speech Codec: Transcoding functions (release 6), v6.0.0,” Dec. 2004.

[2] ITU-T Rec. G.729A, “Coding of speech at 8 kbit/s using conjugate structure algebraic-code-excited linear-prediction (CS-ACELP). Reduced complexity 8 kbit/s CS-ACELP speech codec,” Nov. 1996.

[3] H.G. Kang, H.K. Kim, and R.V. Cox, “Improving transcoding capability of speech coders in clean and frame erasured channel environments,” in *IEEE Worshop on Speech Coding*, 2000, pp. 78–80.

[4] S.M. Tsai and J.F. Yang, “GSM to G.729 speech transcoder,” in *8th IEEE Int. Conf. on Electronics, Circuits and Systems*, 2001, pp. 485–488.

[5] H.G. Kang, H.K. Kim, and R.V. Cox, “Improving the transcoding capability of speech coders,” *IEEE Trans. on Multimedia*, vol. 5, no. 1, pp. 24–33, 2003.

[6] J.K. Choi, C.H. Lee, H.G. Kang, Y.C. Park, and D.H. Youn, “Improvement issues on transcoding algorithms: for the flexible usage to the various pairs of speech codec,” in *ICASSP*, 2004.

[7] M. Ghenania and C. Lamblin, “Low-cost smart transcoding algorithm between ITU-T G.729 (8 kbit/s) and 3GPP NB-AMR (12.2 kbit/s),” in *EUSIPCO*, 2004.

[8] ITU-T Rec. P.862, “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” Feb. 2001.