VoIP quality improvement with a rate-determination algorithm

Demóstenes Zegarra Rodríguez
Electronic Systems Engineering Department
Escola Politécnica, University of São Paulo
demostenes@lps.usp.br

Miguel Arjona Ramírez
Electronic Systems Engineering Department
Escola Politécnica, University of São Paulo
miguel@lps.usp.br

***Abstract*— This article proposes a rate-determination algorithm (RDA) for the G.726 speech coder recommended for use in packetized speech systems. It determines the bit rate based on the MOS value at the destination end point obtained by Recommendation P.563 for improvement of speech quality and bandwidth efficiency in a communication scenario with time-varying network load. Quality ranges from the highest MOS value of 3.980 at a 40 kbit/s rate to the lowest one of 3.185 at a 16 kbit/s rate. Rate switching combinations provide intermediate quality. When the RDA detects a lower MOS value at the communication node, it switches to a lower rate that brings about a reduction in network load.**

***Index Terms*— RDA, variable bit rate, G.726, bandwidth, QoS, VoIP, MOS**

I. Introduction

With the increasing availability of Voice over IP equipments the bandwidth is becoming a big concern in telecomunication designs, as much in cabled or wireless networks as in cellular communications. One way to solve these bandwidth issues is to move away from a traditional fixed-bit rate (FBR) codec scheme toward variable bit rate (VBR).

Variable bit rate (VBR) designates the bitrate used in audio or video encoding. As opposed to constant bitrate (CBR) streams, VBR streams vary the output data rate for each time segment in general. Several strategies have been proposed to adjust bitrates and studies [1] and [2] about the limitations associated with rate selection strategies attempt to integrate perceptual notions in VBR coders. In algorithms [1] and [2], speech frames are phonetically classified (e.g., as voiced, unvoiced, or onsets) according to pitch periodicity and spectral flatness measures. Indeed, shaping the quantization error according to the masking properties of the human ear is a common practice in several speech coding standards. VBR speech codecs offer improved coding efficiency, which is particularly important for multiple access systems and packet-switched networks.

Central to VBR coders is the rate-determination algorithm (RDA).

The existing RDAs have the following characteristics:

- they employ psychoacoustically-blind statistical metrics to estimate a suitable encoding bitrate;
- they have little flexibility to differentiate weak fricatives from background noise;
- they suffer from misclassification of background noise and speech, particularly, in the presence of high acoustic noise levels.

Consequently, the VBR codecs either fail to synthesize both the acoustical noise and active speech with reasonable perceptual quality or result in higher average bit rates.

Essentially, articles [1], [2], [3] and [4] are based on the signal by means of voice activity detection (VAD) and throughput measurements. This means that the RDA algorithms determine the codec rate following the characteristics of the input signal.

The main contribution of this work is an RDA that selects a codec rate according to speech quality in the network based on estimated MOS value. This work not only proposes an RDA, but it further describes a mechanism for transmitting the MOS estimate at the terminal across the network to another terminal

The whole setup in Fig. 1 was developed with open source software and consists of a network emulator, an MOS back transmitter and an ITU-T speech coder.

An MOS value ranges from 1 for an unacceptable call to 5 for an excellent call.

II. Theoretical Revision

The following describes two objective methods for estimating quality of the signal transmitted through a network and a

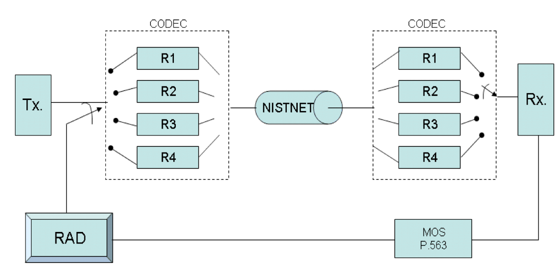

Fig. 1. A general architecture for an RDA

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) with a scholarship and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under Grants no. 2007/08288-2 and 2004/13206-7.

brief description of P.800, an RDA, the G.726 codec and the RDA notification system that is responsible for notifying the rates to be changed.

A. The rate-determination algorithm

The rate-determination algorithm (RDA) [9], [10], [11] and [12] is a mechanism that depends on some external factors to be able to switch the current rate to another one.

For the proposed RDA, the control factor is the quality of the signal at the receiver side.

B. On Recommendation P.862

ITU-T P.862 Recommendation [13] for Perceptual Evaluation of Speech Quality (PESQ) is a tool for the estimation of speech quality using two input signals: the original signal and the degraded signal.

PESQ is an objective measurement tool that predicts the results of subjective listening tests on telephone systems. PESQ uses a perceptual model to compare the original, unprocessed signal with the degraded signal from the network or network element. The resulting quality score is converted to an estimated *Mean Opinion Score* (MOS) that should actually be measured using panel tests according to ITU-T P.800. The PESQ scores are averaged over a large database of subjective tests.

The ITU-T selection process that resulted in the standardization of PESQ involved a wide range of conditions, with demanding correlation requirements set to ensure that it has good performance in assessing conventional fixed and mobile networks and packet-based transmission systems.

PESQ takes into account coding distortions, errors, packet loss, delay and variable delay, and filtering in analogue network components.

This powerful test tool can be deployed in many different areas of a business, on any speech carrier technology:

- In the research laboratory; providing rapid feedback on promising areas of signal processing development, validation of design implementation, ranking alternative design solutions, providing a higher degree of confidence before submission to subjective testing;
- In network equipment evaluation; comparing different vendor offerings and determining their impact on network performance;
- In network installation and rollout of new technologies; ensuring that the desired speech quality is being delivered as the network complexity and loading increases;
- In troubleshooting network and customer problems; determining the scale of the problem, the effectiveness of the solution.

C. On Recommendation P.563

The model for the P.563 Recommendation [14] is a non-intrusive one based on signals at a node and taking measurements at the listening side, without requiring any reference signal.

The P.563 Recommendation is based on the same fundamental principles of human hearing used in intrusive methods such as PSQM, the PAMS and PESQ.

For the estimation of distortions and parameter extraction an analysis of the vocal tract is performed, the signal and noise statistics are analyzed by segment of speech, and the signal interruptions and periods of silence are detected.

D. On Recommendation P.800

The ITU-T P.800 [21] describes several methods and procedures for conducting subjective evaluations of transmission quality. The most commonly used method is Absolute Category Rating (ACR) test which gives the Mean Opinion Score (MOS). Degradation Category Rating (DCR) is also used in some occasions, which gives Degradation Mean Opinion Score (DMOS).

This work only describes the Absolute Category Rating (ACR).

1) Absolute Category Rating (ACR): For Absolute Category Rating (ACR) listening test, subjects (untrained listeners) are asked to rate the overall quality of a speech utterance being tested without being able to listen to the original reference. The rating of quality is based on an opinion scale as shown in Table I The average of opinion scores of the subjects gives the Mean Opinion Score (MOS).

TABLE I
OPINION SCALE FOR ACR TEST

Category	Speech Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

E. On the G.726 speech coder

The original adaptive differential PCM (ADPCM) algorithm standardized by the ITU in 1984 included only the 32 kbit/s bit rate, and met the goal of coding speech at half of the bit rate necessary for G.711 [5] while maintaining the same speech quality. This version, known as the Red Book ADPCM, was found to have a flaw in transmitting voice-band data signals modulated using Frequency Shift Keying (FSK), and was replaced by another version published in 1988 [6]. At the same time, 24 and 40 kbit/s extensions were also approved for the algorithm and published in Rec. G.723 [7]. In 1990, the two Blue Book recommendations were merged into G.726 [8] with the addition of the 16 kbit/s bit rate. The G.726 algorithm encodes the input 8 bit, A or u-law, PCM samples into 2-5 bit ADPCM samples. The ADPCM algorithm exploits the predictability of the speech signals. It uses an adaptive predictor to predict the present signal sample based on the past expanded input log-PCM samples. The difference between the predicted and actual sample is quantized by an adaptive quantizer using the number of bits allowed by the current bit rate. The quantizer bits are sent to the decoder where they are packed to recover the level codes that point to the quantizer levels. The difference signal is added to the adaptive predictor output at the decoder.

F. RDA notification system

We propose a notification system for the RDA rates based on the MOS estimated at the receiving network node.

According to the change in MOS index it sends notifications every 20 milliseconds. This transmission period was selected because it has been used in [16] and [17], via socket (a port between the implementation process and the transport end-to-end protocol).

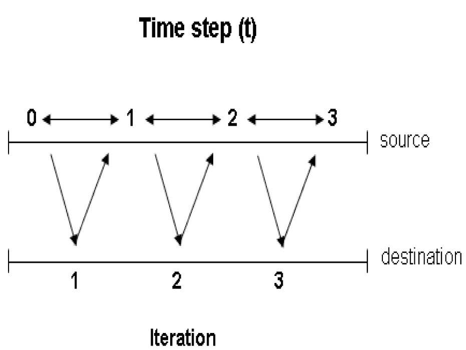

Fig. 2. Message exchange between source and destination

Fig. 2 shows the UDP (User Datagram Protocol) communication exchanged between source and destination.

The operating principles are based on [18] and further explained below.

At the receiving node the signal quality is estimated using ITU-T P.563 recommendation. This MOS estimate is stored in a variable X and then it is sent to the source point by a UDP communication, where it is input to the RDA which selects among the four rates of 16, 24, 32 or 40 kbit/s for the communication between source and destination.

This algorithm was implemented in C and is based on ITU G.191 [19] Recommendation.

III. Simulation Scenario

The experimental setup is shown in Fig. 1, where the encoded input signal is transmitted to the receiving end point through the network. The communication between the source and destination terminal computers is mediated by the Nistnet software, which has been selected to be the network emulator. It is a general purpose freeware that is widely accepted by the network research community. It injects a traffic source which may be a stream generated from a data file or an interactive audio stream originating at a microphone. For further detail the reader is referred to [15].

The network emulator can modify the transmission parameters: PDT, PDV, BW, THRU and PLR. It is also possible to define the route to be followed by the packets.

Since the objective of this work is to evaluate the behavior of the RDA in different network scenarios and at different levels of signal quality at the destination node, the encoder and decoder are isolated in order to facilitate the tests and the network is supposed to be ideal because the loop for the retransmission of the MOS estimated is not taken into account. This avoids possible extraneous impairments to the operation of the RDA. It was done by the simulation of the estimated MOS input to the RDA so that the bit rate switches in accordance with them.

The rate switching scheme is represented in Fig.3. The RDA Input is a list of different MOS values that represent different speech quality levels at the end point in order to switch the bit rate. Additionally, these values are read at each time t. In order to evaluate the RDA performance, it was performed some tests each one with a different time t obtaining scenarios with different number of switching rate during the transmission of the same speech signal which is T seconds long.

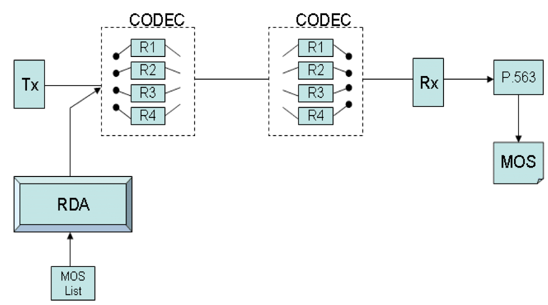

Fig. 3. Simulation Testbed

IV. Results and Discussion

The MOS is initially estimated at fixed bitrates and compared with the MOS estimated for variable bitrate combinations.

In the simulation tests the performance of the RDA was evaluated based on the P.563 tool.

Tests were conducted for quality range based on Table I and the following rate switching rules:

- $MOS \geq 3.7$ then $rate = 40$ kbps;
- $MOS \leq 3.4$ or $MOS \geq 3.2$ then $rate = 32$ kbps;
- $MOS \leq 3.2$ or $MOS \geq 3.0$ then $rate = 32$ kbps;
- $MOS \leq 3.0$ then $rate = 16$ kbps.

In all cases, the simulation has been performed 20 times in order to ensure the validity of the results. The standard deviations (SDs) are reported in Tables II, III and IV.

Table II shows the values obtained for each fixed-bitrate. It is remarked that all the signals are eight-second long.

Table III shows the MOS estimated for each test scenario conducted using the RDA that controls the rate of G.726 encoder among the bit rates of 16, 24, 32, 40 kbit/s.

The values displayed in Table III were obtained as the averages over 6 different audio files repeated 20 times each.

TABLE II

MOS ESTIMATES FOR FIXED-BITRATES.

RATE (KBPS)	AVERAGE MOS	SD
40	3.980	0.0098
32	3.906	0.0045
24	3.723	0.0029
16	3.185	0.0036

TABLE III

MOS ESTIMATES FOR DOUBLE-BITRATES.

RATE (KBPS) RATE 1 - RATE 2	MOS	SD
40 - 32	3.973	0.0022
40 - 24	3.948	0.0024
40 - 16	3.926	0.0029
32 - 24	3.822	0.0022
32 - 16	3.809	0.0027
24 - 24	3.723	0.0012
24 - 16	3.497	0.0031

These audio files are each eight-second long at a 16 bit sample resolution and an 8 kHz sampling rate.

Table IV shows the estimates obtained of MOS index for triple-bit rates.

TABLE IV

MOS ESTIMATES FOR TRIPLE-BITRATES.

RATE (KBPS) RATE 1 - RATE 2 - RATE 3	MOS	SD
40 - 32 - 24	3.909	0.0032
40 - 32 - 16	3.362	0.0029
40 - 24 - 16	3.315	0.0037
32 - 24 - 16	3.299	0.0034

The results in Tables II, III and IV are consistent because they show that when the RDA switches to a lower rate the MOS estimate decreases too.

Additionally, in an overloaded network the RDA meets the goal by switching to a lower rate to ensure a good communication between source and destination points.

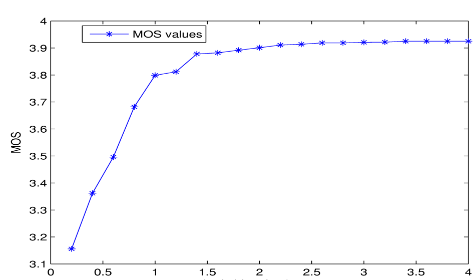

Fig. 4. MOS values for different switch time

The MOS estimates obtained by switching to different coding rates are shown in Fig. 3. For this case the double bit-rate of 40 and 32 kbit/s was used. It can be observed that a decrease in switching period entails a decrease in quality as well. In other words, when the number of rate switchings increase, the quality of the signal decreases. This trend in the results agree with [20].

V. CONCLUSION

This work shows that using an RDA based on the quality of the signal at the reception point there is an improvement in the use of the transmission channel and hence on the quality of the transmission of the signal.

The results look reliable since the higher rate corresponds to the highest quality. It is further concluded that when the number of rate switchings increase for a given speech signal, the resulting quality tends to decrease.

The bandwidth can be better used to avoid a network overload.

It is also observed that the setup can be perfectly reproduced by other researches because it consists of free tools and their interconnection makes for an easy implementation.

The continuation of this work is expected to highlight that variable rate switching results in an increase in MOS estimates and improved VoIP quality.

Experiments with network noises are planned to measure the sensitivity of the RDA additional tests will be performed more tests for evaluating how much each one of the network parameters (PDT, PDV, BW, THRU and PLR) affect the RDA performance.

REFERENCES

[1] Paksoy, E.; Gersho, A.,*Multimode and variable-rate coding of speech*, in Speech Coding and Synthesis, Eds. W. B. Kleijn and K. K. Paliwal, Elsevier, 1995.

[2] Gersho, A.; Paksoy, E.,*Variable rate speech coding for cellular networks*, in Speech and Audio Coding for Wireless and Network Applications, Eds. B. Atal, V. Cuperman, Kluwer Academic Publishers, 1993.

[3] Paulus, J. W.,*Variable bitrate wideband speech coding using perceptual motivated thresholds*, IEEE Workshop on Speech Coding for Telecommunications, pp 35-36, 1995.

[4] Hermansky, H.,*Perceptual linear predictive (PLP) analysis of speech*, J. Acoust. Soc. Amer., pp 1738-1752, 1990.

[5] ITU-T Rec. G.711, *General Aspects of Digital Transmission Systems Terminal Equipments - Pulse Code Modulation (PCM) of Voice Frequencies*, 1972. http://www.itu.int/rec/T-REC-G.711/en.

[6] CCITT R. G.721, *32kbits/s adaptative differential pulse code modulation (ADPCM)*, IXth Plenary Assembly, Melbourne, 14-25 Nov. 1988, Blue Book.

[7] ITU-T Rec. G.723.1, *General Aspects of Digital Transmission Systems - Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s*, May. 2006. http://www.itu.int/rec/T-REC-G.723.1/en.

[8] ITU-T Rec. G.726, *General Aspects of Digital Transmission Systems, Terminal Equipment - 40, 32, 24,16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)*, 1990. http://www.itu.int/rec/T-REC-G.726/en.

[9] S. El-Ramly, A. Tawfik, *A Rate-Determination Algorithm for Variable-Rate Speech Coder*, IEEE Workshop on Speech Coding for Telecommunications, pp 697-701, 2004.

[10] Atti, V. Spanias, A., *Rate determination based on perceptual loudness*, IEEE International Symposium on Circuits and Systems, Vol 2. , May 2005.

[11] Sheung. C. N., *Rate determination algorithms in IS-95 forward traffic channels*, Vehicular Technology Conference, 1998. VTC 98. 48th IEEE, Vol 13, Sept. 2005.

[12] Tammi, M. Jelinek, M. Ruoppila, V.T., *Signal Modification Method for Variable Bit Rate Wide-band Speech Coding*, IEEE Transactions on Speech and Audio Processing, Vol 13, no. 5, pp 799-810, Sep. 2005.

[13] 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*, 2005. http://www.itu.int/rec/T-REC-P.862/en.

[14] P.563, *The ITU-T Standard for Single-Ended Speech Quality Assessment*, 2005. http://www.itu.int/rec/T-REC-P.563/en.

[15] Rodriguez D. Z., Ramirez M. A., *Determinação do comportamento de codificadores em diferentes condições de rede em um cenário experimental*, XXVI Simpósio brasileiro de Telecomunicações, 2008.

[16] Kambiz, H.,*Rate adaptative speech coding for Universal Multimedia Access*, IEEE Signal Processing Magazine, 2003.

[17] Usman, M.,*Performance analysis for Adaptative source rate control algorithm for VoIP*, Proceedings of the American Control Conference, 2003.

[18] Cerello, P.,*The FINUDA Data Acquisition System as Event Monitor*, IEEE Transations on Nuclear Science, Vol. 4, 1998.

[19] G.191, *Software tools for speech and audio coding standardization*, 2005. http://www.itu.int/rec/T-REC-G.191-200509-I/en.

[20] Hoene C., Karl H., *Perceptual Quality Model Intended for Adaptive VoIP Applications*, International Journal of Communication Systems, 2005.

[21] P.800, *Methods for Subjective Determination of Transmission Quality*, ITU Recommendation P.800, August 1996.