IJST EXPLORE International Journal of Science and Technology
CENTRE OF PROFESSIONAL RESEARCH PUBLICATIONS
ISSN 2049-7318

International Journal of Science and Technology Volume 1 No. 3, March, 2012

Markov Models of Telephone Speech Dialogues

O.I. Sheluhin, A.A. Atayero, M.O. Eyinagho, J.O. Iruemi
Department of Electrical and Information Engineering,
Covenant University, Km.10 Idiroko Road, PMB1023 Ota, Ogun State, Nigeria

ABSTRACT

Analogue speech signals are the most natural form of communication among humans. The contemporary methods adopted for the analysis of voice transmission by packet switching were designed mainly with respect to a Poisson stream of input packets, for which the probability of an active packet on each input port of the router is a constant value in time. An assumption that is not always valid, since the formation of speech packets during a dialogue is a non-stationary process, in which case mathematical modeling becomes an effective method of analysis, through which necessary estimates of a network node being designed for packet transmission of speech may be obtained. This paper presents the result of analysis of mathematical models of Markov chain based speech packet sources vis-à-vis the peculiarities of telephone dialogue models. The derived models can be employed in the design and development of methods of statistical multiplexing of packet switching network nodes.

Keywords: *Markov models, Statistical multiplexing, Speech processing, Packet switching, QoS*

1. INTRODUCTION

The main aim of statistical analysis of the characteristics of speech communication is the development of models that describe the change in the states of speech during a telephone conversation [1]. Markov processes with the necessary number of states satisfactorily describe speech signal formation mechanism, the knowledge of which is necessary for the analysis of network problems in packet switching [2, 3].

In the process of a telephone conversation between any two subscribers, each communication direction via the telephone channel is used on the average for only half the conversation time. Additionally, the active state duration of the channel is further reduced because of pauses between words and phrases. These all lead to a total active state duration of tone-frequency channel of approximately 35–40% [4].

In a bid to increase the efficiency of channel capacity usage, the pause durations of one dialogue can be used for the transmission of active state dialogues of other subscribers [5, 6]. This is achievable by means of statistical multiplexing, and systems used for its realization are known as statistical transmission systems (STS) [7]. Such systems can theoretically increase the efficiency of channel capacity usage by between 20–30% both ways.

Digital Speech Interpolation (DSI) method employed in statistical transmission systems presupposes that under a normal duplex conversation, each speaker only occupies the channel for approximately half the conversation period. Further more, pauses in between phonemes, words and phrases are added to channel inactivity (idle) time, bringing the active usage time of a normal simplex telephone channel to 25% of connection time on the average. Consequently, if that same channel can be given to other subscribers for usage during its inactive period, this will substantially increase the number of connections that can be organized in the same channel for the same time span. Subsequently the volume of traffic over the channel can be increased considerably. Suffice it to note here some characteristic advantages of statistical multiplexing [8, 9]:

i. dynamic allocation of the capacity of channel being multiplexed depending on the activity of the data transmission channel;
ii. possibility of on-demand allocation of channel capacity;

iii. possibilty of allocating priority levels for different types of traffic (i.e. introduction of QoS).

2. MARKOV MODELS OF TELEPHONE DIALOGUES

The problem of studying the dynamics of dialogue is not a new one, and it is very well covered in the literature. We consider here the most popular models in existence as at this writing. Given that there exists a means of speech transmission, in which a given interval of time is allocated for each packet, i.e. a certain device divides the time axis into segments, in the duration of which empty or occupied packets can be received from each subscriber depending on the energy level of its speech signal. The moment of appearance of active speech signal and the flow of packets are not synchronized. We shall assume hereafter that only one of the subscribers (the one initiating the dialogue) can

change the state of the speech signal within the span of a packet.

2.1 Six-state (Brody) Telephone Dialogue Model

This model consists of the following six states as its name implies:

1. subscriber A is speaking, subscriber B is silent;
2. subscriber A is preparing to pause;
3. mutual silence of both subscribers, subscriber A spoke last;
4. subscriber B is speaking, subscriber A is silent;
5. subscriber B is preparing to pause;
6. mutual silence of both subscribers, subscriber B spoke last.

This six-state model attracts a lot of interest since it excellently describes the dynamics of all possible states of a dialogue. This fact notwithstanding, however, simpler models are usually adopted provided they reflect exactly the change in duration of the active and pause states.

2.2 Four-State Telephone Dialogue Model

This model consists of the following states:

1. subscriber A is speaking, subscriber B is silent;
2. subscriber A is speaking, subscriber B is speaking;
3. subscriber A is silent, subscriber B is speaking;
4. subscriber A is silent, subscriber B is silent.

This model clearly reflects the distribution of pause and speech activity durations, but it does not correspond with real event flow when both subscribers exhibit simultaneous speech activity.

2.3 Three-State Telephone Dialogue

This model consists of the following states:

1. subscriber A is speaking, subscriber B is silent;
2. subscriber A is silent, subscriber B is silent;
3. subscriber A is silent, subscriber B is speaking.

This model also reflects the distribution of active and pause durations efficiently. In its case, the duration of the active state is geometrically distributed, while that of the pause state differs, which corresponds to the real characteristic of speech signal [2].

2.4 Two-State Telephone Dialogue

Further simplification of the model gives the simplest model with two states:

1. subscriber A is speaking, subscriber B is silent;
2. subscriber A is silent, subscriber B is speaking.

Such model as described above corresponds to experimental data on the distribution of the active speech phase, but does not reflect the peculiarities of the distribution of pause duration. It excludes the possibility of both subscribers being in the pause state simultaneously. Thus, the dialogue activity seems to be switched between subscriber A and B, while both the active and pause states have a geometric distribution.

Table 1. Audio Files Analysis Result

Conversation №	№ of LD-CELP format samples	№ of active samples	№ of passive samples	Quantity of useful information in conversation, % of total volume
01.	62720	18317	44403	29
02.	781622	233763	547859	29
03.	147030	50458	96572	34
04.	529078	168459	360619	31
05.	292284	155586	136698	53
06.	132233	62691	69542	47
07.	917907	254628	663279	27
08.	200608	83336	117272	41
09.	25225	5987	19238	23
10.	1054297	322948	731349	30
Total №	4143004	1356173	2786831	32

All the models considered above cannot serve as solution of the stated research problem – i.e. *modeling of the flux of speech packets from a single subscriber during a telephone dialogue*. Stemming from this requirement, we will endeavor to create our own model, taking into consideration the assumption that in the course of a telephone conversation, a subscriber is either initiating the dialogue or responding in reaction to information obtained from another subscriber. In this scenario, the model will consist essentially of two main states (active- and passive-dialogue), which can be extended by adding separate *sub-states*, which allow for intra-state transitions e.g. transition from pause to active dialogue and transition to passive dialogue from active in a short span of time.

3. EXPERIMENT

The following experiment was conducted with a view to studying the properties of speech signal during a natural dialogue between two subscribers.
With the aid of an experimental setup consisting of a personal computer and an attached telephone set, telephone conversations were recorded using a special software package. Only the telephone conversation of one subscriber was recorded. In a bid to guarantee authenticity of collated data, the subscribers were made oblivious of the experiment being conducted. The recorded speech was saved as *.wav* speech files in *PCM* format with parameters 8 kHz and 8 bit, which is a standard format for the digitization of speech with minimal quality requirements [4, 10, p.675, 11, p.100]. The experiment was conducted to study the periods of subscriber speech activity and pause in the telephone dialogue, as well as their duration. The results of analysis of the files are as given in Table I. A total of ten dialogues were analyzed.

A comparison of the data gotten from the analysis of the telephone dialogue and those given in [3] leads us to conclude that the obtained statistical data gives a realistic representation of the components of telephone conversation in percentage terms. As such, a model of the source of speech signal can be developed on the basis of the obtained data vis-à-vis the model of telephone dialogue for subsequent investigation of telephone traffic formation.

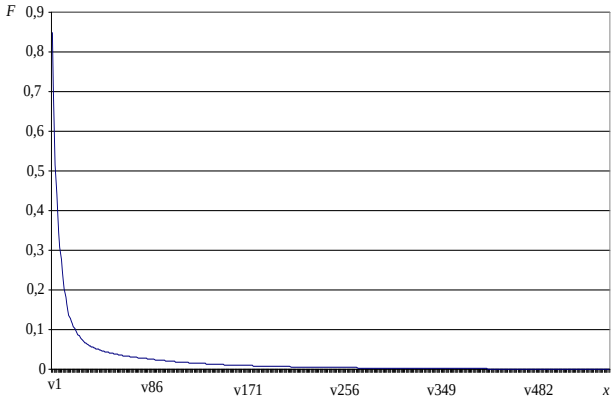

Figure 1. PDF of the duration of *active packets* for all conversations

In order to develop the model of speech packet source, it is necessary to select the appropriate analytical expression for describing real processes. At the initial modeling stage, we assume that two states can be gotten from the subscriber's speech: active state *A* and passive state *P*. The relationship $F(X > x)$ is depicted on Fig. 1. and Fig. 2. corresponding to the probability distributions of active and passive packets respectively.

The experimental characteristics show that the model, while presupposing the presence of two states – *A* и *P*, does not reflect the dynamics of change of speech signal with enough exactitude, since the presence of one state *A* during speech period and one state *P* in the pause period suggests an exponential change of the relationship $F(X > x)$.

At the next stage, and with further analysis of the distribution function $F(X > x)$, it was observed that the distribution is accurately approximated by a function of the form:

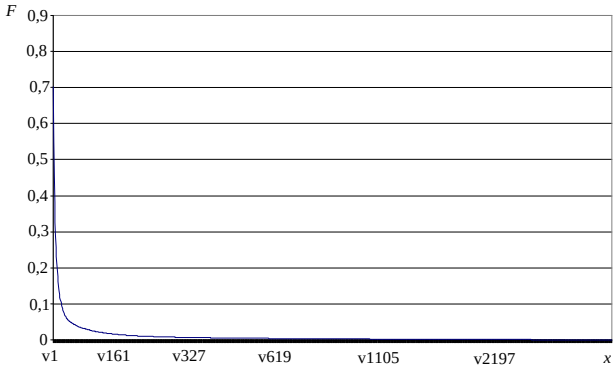

Figure 2. PDF of the duration of *passive packets* for all conversations

$$F(X > x) = S_1 e^{-\alpha_1 x} + S_2 e^{-\alpha_2 x} + S_3 e^{-\alpha_3 x} \qquad (1)$$

We write the following relationship for the distribution function of the active packet series:

$$F(X > x) = A_1 e^{-\alpha_1 x} + A_2 e^{-\alpha_2 x} + A_3 e^{-\alpha_3 x} \qquad (2)$$

In the same vein, we write the following relationship for the distribution function of the passive packet series:

$$F(X > x) = P_1 e^{-\beta_1 x} + P_2 e^{-\beta_2 x} + P_3 e^{-\beta_3 x} \qquad (3)$$

Equation (1) is the generic form of both equations (2) and (3), where S denotes state. Equation (2) describes the distribution of conditional probabilities of the occurrence of a series of state *A* given the condition that the previous state was *P*. On the other hand, equation (3) describes the distribution of conditional probabilities of the occurrence of *P* state series granted the condition that the previous state was *A*.

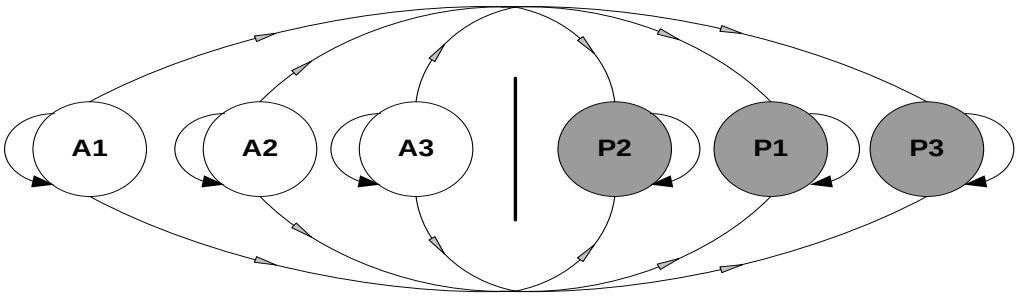

Figure 3. Simplified graph of the *six-state* model

Approximation coefficients were obtained using uniform approximation and numerical methods, for which the approximation of $F(X > x)$ was the best. Distribution function approximation coefficients thus gotten for the active and passive series of packets are given in Tables II(A) and II(B) respectively.

The type of experimental characteristic $F(X > x)$ and the expressions that approximate them allow for the assumption of the presence of three states corresponding to active speech − A_1, A_2 and A_3, as well as the presence of three states corresponding to pause in speech − P_1, P_2 and P_3, with different values of expectation for the length of active and passive states. We note here that the graph of the prescribed model can be presented as shown in Fig. 3. For such a model, the matrix of the transition probabilities S_{ij} from state to state has the form given in equation (4).

$$S_{ij} = \begin{pmatrix} S_{A_1A_1} & 0 & 0 & S_{A_1P_1} & S_{A_1P_2} & S_{A_1P_3} \\ 0 & S_{A_2A_2} & 0 & S_{A_2P_1} & S_{A_2P_2} & S_{A_2P_3} \\ 0 & 0 & S_{A_3A_3} & S_{A_3P_1} & S_{A_3P_2} & S_{A_3P_3} \\ S_{P_1A_1} & S_{P_1A_2} & S_{PA} & S_{P_1P_1} & 0 & 0 \\ S_{P_2A_1} & S_{P_2A_2} & S_{PA} & 0 & S_{P_2P_2} & 0 \\ S_{P_3A_1} & S_{P_3A_3} & S_{PA} & 0 & 0 & S_{P_3P_3} \end{pmatrix} \quad (4)$$

The definition of the parameters of such a model has to be done under the conditions of absence of complete statistical information. Analyzing the distribution function of the length of a series of packets obtained from experimental data, it is impossible to determine to which of the three states belongs a particular packet or series of packets, making up a given period of speech activity. The same holds for states P_1, P_2 or P_3 − it is impossible to determine to which of the states belongs a particular packet or series of packets, making up a given period of pause.

Subsequently, it becomes necessary to find a compromise in the process of dividing one state into several *sub-states*. We consider a likely resolution of this problem emanating from the condition of preservation of the final probabilities of signal division into *pause* and *active speech* states. The matrix shown in equation (5) is introduced for ease of expression:

$$S_{ij} = Q_{ij} = \begin{pmatrix} q_{11} & 0 & 0 & q_{14} & q_{15} & q_{16} \\ 0 & q_{22} & 0 & q_{24} & q_{25} & q_{26} \\ 0 & 0 & q_{33} & q_{34} & q_{35} & q_{36} \\ q_{41} & q_{42} & q_{43} & q_{44} & 0 & 0 \\ q_{51} & q_{52} & q_{53} & 0 & q_{55} & 0 \\ q_{61} & q_{62} & q_{63} & 0 & 0 & q_{66} \end{pmatrix} \quad (5)$$

Matix Q_{ij} defines the final probabilities.

$$Q_i = \begin{bmatrix} Q_1 & Q_2 & Q_3 & Q_4 & Q_5 & Q_6 \end{bmatrix}^{\mathrm{T}} \quad (6)$$

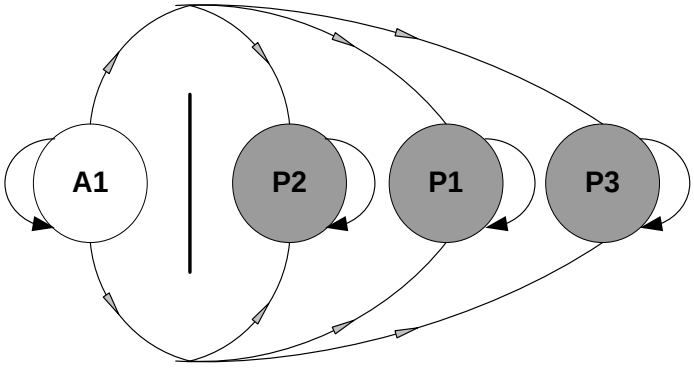

Figure 4. Graph of the *four-state* model

Similarly, the graph of the *six-state model* shown in Fig. 3 can be reduced to that of the *four-state model* as shown in Fig. 4.

For such a model, the transition probabilities matrix from state to state P_{ij} is given as follows:

$$P_{ij} = \begin{array}{c} \begin{matrix} A & P_1 & P_2 & P_3 \end{matrix} \\ \begin{pmatrix} P_{AA} & P_{AP_1} & P_{AP_2} & P_{AP_3} \\ P_{P_1A} & P_{P_1P_1} & 0 & 0 \\ P_{P_2A} & 0 & P_{P_2P_2} & 0 \\ P_{P_3A} & 0 & 0 & P_{P_3P_3} \end{pmatrix} \end{array} \qquad (7)$$

We introduce the matrix in equation (8) for ease of expression:

$$P_{ij} = R_{ij} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & 0 & 0 \\ r_{31} & 0 & r_{33} & 0 \\ r_{41} & 0 & 0 & r_{44} \end{pmatrix} \qquad (8)$$

For $x=1$, $x=2$, $x=3$, equation (3) becomes

$$F = \begin{cases} P_{AP}P_1 e^{-\beta_1} + \\ +P_{AP}P_2 e^{-\beta_2} + P_{AP}P_3 e^{-\beta_3} & , \quad x=1 \\ P_{AP}P_1 e^{-\beta_1} e^{-\beta_1} + \\ +P_{AP}P_2 e^{-\beta_2} e^{-\beta_2} + P_{AP}P_3 e^{-\beta_3} e^{-\beta_3} & , \quad x=2 \\ P_{AP}P_1 e^{-\beta_1} e^{-\beta_1} e^{-\beta_1} + \\ +P_{AP}P_2 e^{-\beta_2} e^{-\beta_2} e^{-\beta_2} + P_{AP}P_3 e^{-\beta_3} e^{-\beta_3} e^{-\beta_3} & , \quad x=3 \end{cases} \qquad (9)$$

Probabilities for the graph shown in Fig. 4 can be calculated from equation (8):

$$F = \begin{cases} r_{12}r_{22} + r_{13}r_{33} + r_{14}r_{44} & , \quad x=1 \\ r_{12}r_{22}r_{22} + r_{13}r_{33}r_{33} + r_{14}r_{44}r_{44} & , \quad x=2 \\ r_{12}r_{22}r_{22}r_{22} + r_{13}r_{33}r_{33}r_{33} + r_{14}r_{44}r_{44}r_{44} & , \quad x=3 \end{cases} \qquad (10)$$

From the comparison of equations (9) and (10), we may write the following:

$$r_{ij} = e^{-\beta_i}, \; r_{12} = P_{AP}P_1, \; \ldots, \; r_{14} = P_{AP}P_3 \qquad (11)$$

and consequently, the matrix P_{ij} may be rewritten in the following form:

$$P_{ij} = \begin{pmatrix} P_{AA} & P_{AP}P_1 & P_{AP}P_2 & P_{AP}P_3 \\ 1-e^{-\beta_1} & e^{-\beta_1} & 0 & 0 \\ 1-e^{-\beta_2} & 0 & e^{-\beta_2} & 0 \\ 1-e^{-\beta_3} & 0 & 0 & e^{-\beta_3} \end{pmatrix} \qquad (12)$$

Similar to equation (10), the model can be made more exact by introducing three states – A_1, A_2, A_3 – and state P.

Consequently, the transition probabilities matrix S_{ij} of the states of the model of speech packet source, considering the peculiarities of six-state dialogue will be of the form:

Table 2. Approximation coefficients for probability distribution function

(A)

Coefficient	A_1	A_2	A_3	α_1	α_2	α_3
Value	0.859	0.1348	0.0065	0.248	0.0165	0.00084

(B)

Coefficient	P_1	P_2	P_3	β_1	β_2	β_3
Value	0.859	0.1348	0.0065	0.248	0.0165	0.00084

(A) – of active packet series; (B) - of passive packet series

$$S_{ij} = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} \qquad (13)$$

$$\mathbf{A} = \begin{vmatrix} e^{-\alpha_1} & 0 & 0 \\ 0 & e^{-\alpha_2} & 0 \\ 0 & 0 & e^{-\alpha_3} \end{vmatrix} \qquad (13a)$$

$$\mathbf{B} = \begin{vmatrix} (1-e^{-\alpha_1})P_1 & (1-e^{-\alpha_1})P_2 & (1-e^{-\alpha_1})P_3 \\ (1-e^{-\alpha_2})P_1 & (1-e^{-\alpha_2})P_2 & (1-e^{-\alpha_2})P_3 \\ (1-e^{-\alpha_3})P_1 & (1-e^{-\alpha_3})P_2 & (1-e^{-\alpha_3})P_3 \end{vmatrix} \qquad (13b)$$

$$\mathbf{C} = \begin{vmatrix} (1-e^{-\beta_1})A_1 & (1-e^{-\beta_1})A_2 & (1-e^{-\beta_1})A_3 \\ (1-e^{-\beta_2})A_1 & (1-e^{-\beta_2})A_2 & (1-e^{-\beta_2})A_3 \\ (1-e^{-\beta_3})A_1 & (1-e^{-\beta_3})A_2 & (1-e^{-\beta_3})A_3 \end{vmatrix} \qquad (13c)$$

$$\mathbf{D} = \begin{vmatrix} e^{-\beta_1} & 0 & 0 \\ 0 & e^{-\beta_2} & 0 \\ 0 & 0 & e^{-\beta_3} \end{vmatrix} \qquad (13d)$$

Substituting the numerical values in equation (13) for approximation coefficients earlier obtained (see Table II.), matrices **A**, **B**, **C**, **D** become:

$$\mathbf{A} = \begin{vmatrix} 0{,}79453 & 0 & 0 \\ 0 & 0{,}94904 & 0 \\ 0 & 0 & 0{,}99164 \end{vmatrix}; \qquad (14a)$$

$$\mathrm{B} = \begin{vmatrix} 0{,}01765 & 0{,}0277 & 0{,}00127 \\ 0{,}04377 & 0{,}00687 & 0{,}00032 \\ 0{,}00719 & 0{,}00113 & 0{,}00005 \end{vmatrix}; \qquad (14b)$$

$$\mathbf{C} = \begin{vmatrix} 0{,}16166 & 0{,}04876 & 0{,}00922 \\ 0{,}01204 & 0{,}00363 & 0{,}00069 \\ 0{,}00062 & 0{,}00019 & 0{,}00004 \end{vmatrix}; \quad (14c)$$

$$\mathbf{D} = \begin{vmatrix} 0{,}78036 & 0 & 0 \\ 0 & 0{,}98364 & 0 \\ 0 & 0 & 0{,}999916 \end{vmatrix}. \quad (14d)$$

The final probabilities matrix Q_i of the occurrence of the speech packets source in each of the states at any given moment of time can be gotten from matrices 14a–14d:

$$Q_i = \begin{pmatrix} 0{,}110555 \\ 0{,}134454 \\ 0{,}155010 \\ 0{,}120709 \\ 0{,}254354 \\ 0{,}224918 \end{pmatrix} \quad (15)$$

4. CONCLUSION

We have presented in this paper the methodology for and means of deriving mathematical models of speech sources based on Markov chains vis-à-vis the peculiarities of the model of telephone dialogue. The derived models will find useful application in the design of mathematical models of packet switching network nodes.

REFERENCES

[1] Rasool Tahmasbi and Sadegh Rezaci, "A Soft Voice Activity Detection Using GARCH Filter and Variance Gamma Distribution", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 4, pp. 1129-1134.

[2] Sheluhin O.I., Lukiantsev N.F., "Digital processing and transmission of speech" (in Russian), Radio and Communication, Moscow, 2000.

[3] Minoli D., Minoli E., "Delivering Voice over IP Networks", John Willey & Sons Inc., 1999.

[4] Atayero A.A. (2000), "Estimation of the Quality of Digitally Transmitted Analogue Signals over Corporate VSAT Networks", Ph.D Thesis (unpublished).

[5] Wuncheol Jeong, Mohsen Kavehred, Jungnam Yun (2004), "Spectral Efficiency of Dynamic Time-Division System for Asymmetric and Dynamic Multimedia Traffic, International Journal of Wireless Information Networks", Vol. 11, No. 4, pp.173~185.

[6] Sangwan, A., Chiranth, M. C., Jamadgni, H. S., Sah, R., Prasad, R.V., Gaurav, V. (2002), "VAD Techniques for Real-Time Speech Transmission on the Internet", 5th IEEE International Conference on High-Speed Networks and Multimedia Communications, pp.46~50.

[7] Ahmad Sun Yu, Dongdon, I., Zhang Li Ya-Bin (2002), "Region-based rate control and bit allocation for wireless video transmission", IEEE Transaction on Multimedia, Vol. 8, Issue 1, pp.1~10.

[8] Ajiboye Johnson and Adediran Yinusa (2010), "Performance Analysis of Statistical Time Division Multiplexing Systems", Leonardo Electronic Journal of Practices and Technologies, Issue 16, pp.151~166.

[9] Gringeri Steve, Bitar Nablir, Egorov Roman, Basch Bert, Suitor Craig, and Peng Harry (2007), "Optimizing Transport Services to Integrate TDM and Packet Services", National Fiber Optic Engineers Conference-Networking Convergence for Multi-Haul Architecture, Anaheim, CA.

[10] Tanenbaum Andrew (2006), "Computer Networks", Prentice-Hall of India Private, New Delhi, India.

[11] Anurag Kumar, Manjunath D., Joy Kuri (2004), "Communication Networking: An Analytical Approach", Morgan Kaufman Publishing, San Francisco, USA.