Verification of Video Quality Opinion Model for Videophone Services

Kazuhisa Yamagishi, Takanori Hayashi

NTT Service Integration Laboratories, NTT Corporation.
yamagishi.kazuhisa@lab.ntt.co.jp

Abstract

We propose a computational opinion model for estimating the video quality of videophone services. Although opinion models for speech such as the E-model have been established and widely used, little attention has been given to opinion models for video quality estimation. Our proposed opinion model is useful as a network-planning tool using several video parameters that affect the quality of videophone services. First, we established a function for estimating video quality affected by coding distortion, which expresses the quality of a video affected by coding bit rate and frame rate. Second, we established a packet loss degradation index that estimates the degree of video quality degradation due to packet loss. Third, we integrated these two functions into the opinion model for estimating video quality. We applied this model for video quality estimation of H.264 codec. The results indicated that the estimation errors of our model were less than the mean of the 99% confidence intervals for the subjective scores, which is the statistical reliability of subjective score. Finally, we conducted a verification test of our proposed model using an MPEG-4 codec and found that the form of our proposed model was valid, because video quality can be accurately estimated. Therefore, our model could be applied to effective design, implementation, and management of videophone applications and communication networks.

1. Introduction

Advances in broadband IP networks, applications, and terminal technologies have enabled multimedia communication services such as videophones, instant messaging, and teleconferencing that use speech and video to become popular as promising high-quality multimedia applications. Methodologies for evaluating the quality of such services are indispensable because quality is generally not guaranteed in an IP network. Establishing an objective quality assessment method is important for network quality planning and management.

Objective quality assessment [1] can be categorized into media-layer objective models, packet-layer objective models, and the opinion model from the viewpoint of the input information. For estimating users' perceptual quality of service (QoS), media-layer objective models use media signals [2, 3, 4, 5], packet-layer objective models use information about IP packet transmission quality [6, 7], and opinion models use network and/or application parameters that affect the quality of media [8, 9, 10, 11], where network parameters mean packet loss rate, delay, and so on and application parameters mean coding bit rate, frame rate, and so on. Media-layer objective models are highly correlated with subjective quality. However, this approach is inconvenient for network planning purposes because relationships among media quality and quality parameters were not considered. Packet-layer objective models are mainly used for in-service management and cannot be used for network planning purposes because this model cannot estimate the quality before a service is offered. The opinion model is convenient for network planning purposes because it formulates the relationships among subjective quality, network, and/or application parameters. For example, this model can be used by transmission planners to help ensure that users will be satisfied with end-to-end transmission performance and to avoid building networks that have parameters beyond the required specifications. Network, application, and terminal equipment parameters of high importance to network planners are incorporated into this model.

The opinion model, called the E-model, which has been standardized as ITU-T Recommendation G.107 [8], has been established and widely used for speech services including IP telephony. The E-model can estimate the overall communication quality using a combination of quality factors. It takes 21 parameters as inputs that represent terminal, network, and environmental quality factors. Its output is called $R-value$, which is a psychological scale that is an index of overall quality. On the other hand, few opinion models for video quality have been studied [12]. In [12], an opinion model for multimedia services was proposed, considering application parameters, but network parameters were not taken into account. In [13], the concept of an opinion model for multimedia services was proposed. This model takes into account speech quality, video quality, absolute delay, and media synchronization. In this paper, we establish an opinion model for video quality according to the concept in [13]

We establish a function for estimating video quality affected by coding distortion and a packet loss degradation index for estimating the degree of video quality degradation due to packet loss from the relationships among video quality and quality parameters. We show that our proposed model can estimate video quality of H.264 codecs and MPEG-4 codecs from video quality parameters with sufficient accuracy. We show that our model accurately estimates the quality of several codecs by changing its coefficients, which are optimized for each codec. We can integrate our model into a speech quality estimation function such as the E-model [8] and multimedia quality estimation functions [12, 13, 14, 15, 16]. An opinion model for videophones could be applied for effective design, implementation, and management of videophone applications and communication networks.

The remainder of this paper is organized as follows. An opinion model for videophone services is described in Sect. 2. A method of subjective quality experiments is described in Sect. 3. We propose a video quality estimation model using quality parameters in Sect. 4. The verification of our model is described in Sect. 5. Guidelines for quality design are described in Sect. 6. Finally, in Sect. 7, we summarize our findings and suggest possible directions for future studies.

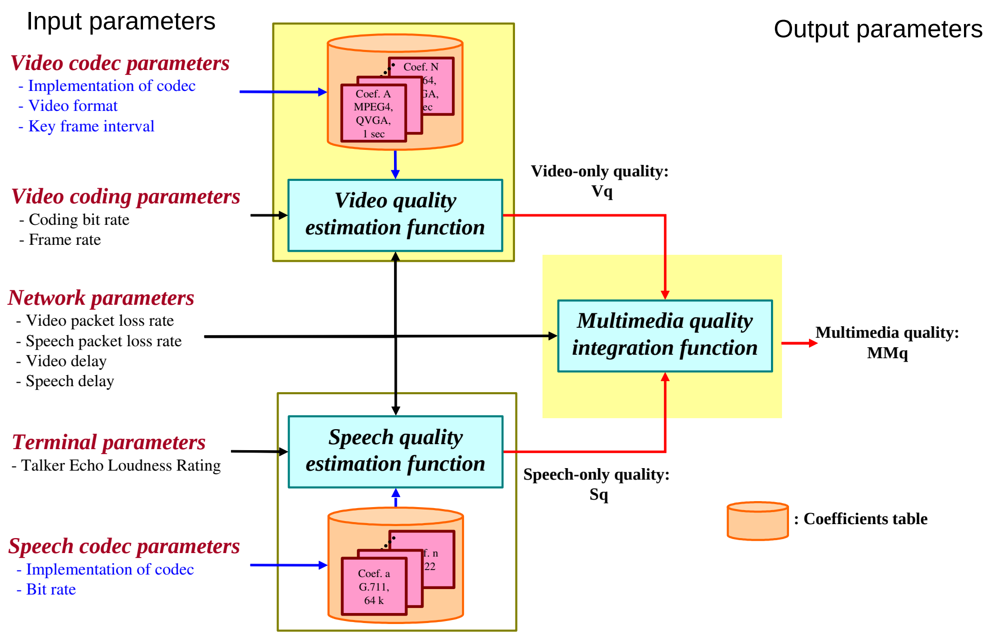

Figure 1: Opinion model for multimedia communication services.

2. Opinion Model for Multimedia Communication Services

We previously proposed the concept of an opinion model for videophone services [13]. This model can be used as a network-planning tool that estimates videophone quality based on the combined effects of variations in several video and speech parameters.

The opinion model for videophone services is shown in Fig. 1. Its input parameters are video quality parameters and speech quality parameters. This model contains three functions: a video quality estimation function, a speech quality estimation function, and a multimedia quality integration function. The output of this model are multimedia quality, which is denoted by MMq in Fig. 1. Our initial target is to establish a video quality estimation model for videophone services, as described in Sect. 1.

We consider that the effects of a codec and terminals on subjective video quality are heavily dependent on their implementation. In particular, the quality of a video codec and terminal cannot be estimated simply based on the information about the coding technology and terminal technology. For example, there are a number of different implementations for MPEG4 codecs due to variations in coding-parameter settings, variation of decoder characteristics, and the number of different types of terminals for PCs, PDAs, and mobile phones. For this reason, we have to conduct subjective quality assessment experiments to calculate the coefficients of the opinion model for each codec specification. Such coefficients should be determined based on data obtained in subjective quality experiments.

3. Subjective Quality Experiments

We built a viewing system for deriving subjective quality characteristics that are necessary for constructing the video quality estimation model. We used four video sequences that lasted 10 seconds. In each sequence, one person explains a completed figure made from multicolor, interlocking construction blocks, as shown in Fig. 2. Video movements in these sequences are larger than those of the head and shoulders that are typical in video sequences of videophones. Therefore, the test sequences were strict for quality assessment. An image whose diagonal measurement was about 4.2 inches was displayed on a 17-inch LCD monitor whose resolution was 1280 × 1024. The displayed video formats were quarter video graphics array (QVGA: 320 × 240).

Figure 2: Four video sequences.

The experimental parameters were coding bit rate, frame rate, and packet loss rate, as shown in Table 1. The number of test conditions, which is the number of combinations of parameters in Table 1, was 272.

Table 1: Experimental settings.

Video codec	H.264
Video format	QVGA
Key frame rate (s)	1
Video coding bit rate (kbps)	64 - 1024
Video frame rate (fps)	1 - 30
Video packet loss rate (%)	0.0 - 10.0

For subjective quality assessment, video quality was evaluated using an ACR (absolute category rating) method [17]. The quality descriptions on the rating scale are given in Japanese.

64 subjects aged 20 - 39 participated in the experiments. They were nonexperts not directly concerned with multimedia quality as part of their work, and, therefore, not experienced assessors. The subjects viewed each video sequence at a distance of about 80 cm. Video quality was represented as an average MOS (mean opinion score) of four MOSs for each sequence.

4. Video Quality Estimation Model

First, we analyzed the relationships among video quality and quality parameters. Second, we tried to establish a function for estimating video quality affected by coding distortion, which estimates the video quality affected by frame rate and coding bit rate. Third, we tried to establish the packet loss degradation index, which indicates the degree of the video quality degradation due to packet loss. Finally, we applied this model to an H.264 codec and showed that our proposed model has sufficient accuracy.

4.1. Function for Estimating Video Quality Affected by Coding Distortion

For the case of where the packet loss rate, pl, is 0, we tried to establish a function for estimating video quality affected by coding distortion from the relationships among MOS, coding bit rate, br, and frame rate, fr. There is an optimal frame rate, ofr, that maximizes the video quality, which is $MOS(ofr, br)$, at each coding bit rate, as shown Fig. 3(a). As the coding bit rate increases, the optimal frame rate and subjective MOS increase, as shown Figs. 3(b) and (c). That is, the video quality characteristics affected by coding distortion were approximated by a Gaussian function, as follows:

$$MOS(fr, br, 0) = C(fr, br) = 1 + G(fr, br) \quad (1)$$

and

$$G(fr, br) = \alpha(br)\exp(-\frac{(\ln(fr) - \ln(ofr(br)))^2}{2\omega^2(br)}). \quad (2)$$

When $fr = ofr$, $C(ofr, br)$ indicates the optimal video quality for each br, and $\omega(br)$ indicates the degree of video quality degradation due to the frame rate.

Then, when $fr = ofr$, we found that $\alpha(br)$ increases and saturates with increasing br, as shown Fig. 3(b). That is, $\alpha(br)$ was approximated by a logistic function, as follows:

$$\alpha(br) = a - \frac{a}{1 + (\frac{br}{b})^c}, \quad 0 \le \alpha(br) \le 4, \quad (3)$$

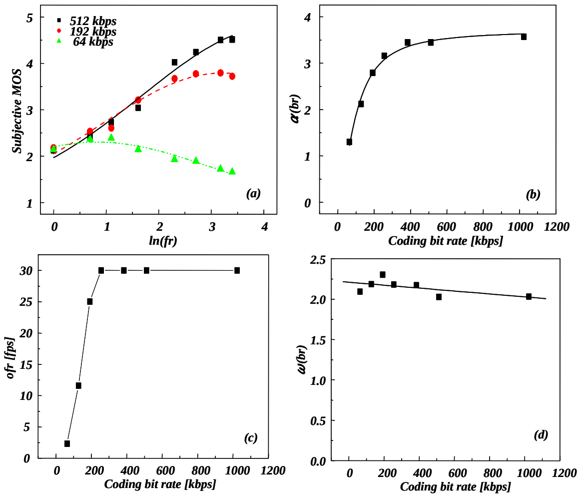

Figure 3: Quality degradation characteristics for coding distortion.

where a, b, and c are constants.

After that, Fig. 3(c) shows that the ofr increases monotonically and saturates at the maximum frame rate with increasing br, as follows:

$$ofr(br) = d + e\,br, \quad 1 \le ofr(br) \le 30, \quad (4)$$

where d and e are constants.

Finally, Fig. 3(d) shows that $\omega(br)$ is approximated by a linear function, as follows:

$$\omega(br) = f + g\,br, \quad (5)$$

where f and g are constants.

We calculated the coefficient tables of the opinion model for H.264 codec in Table 1. Then, using these coefficients, we estimated the subjective video qualities. The results are shown in Fig. 4. Multiple correlation coefficients (R^2), the mean of the 99% confidence interval (MCI) for the subjective MOS, and the root mean square error (RMSE) are also shown in Fig. 4, where MCI means the statistical reliabilities of the subjective scores. We used $R^2 \ge 0.9$ and RMSE$\le$ MCI as criteria for which the accuracy of the model is sufficient. In this experiment, R^2 value was larger than 0.9, and RMSE was less than the MCI. Therefore, we concluded that the video coding quality can be formulated by Eqs. 1 to 5

4.2. Packet Loss Degradation Index

We tried to establish the packet loss degradation index, which indicates the degree of video quality degradation due to packet loss. $MOS(fr, br, pl)$ decreases exponentially with increasing the packet loss rate, pl, as shown Fig. 5 (a). That is, $MOS(fr, br, pl)$ was approximated by an exponential function as follows:

$$MOS(fr, br, pl) = 1 + G(fr, br)\exp(-\frac{pl}{\tau(fr, br)}), \quad (6)$$

where $1 + G(fr, br)$ indicates the video quality when the packet loss rate is 0, and $\tau(fr, br)$ indicates the degree of video quality degradation due to the packet loss rate.

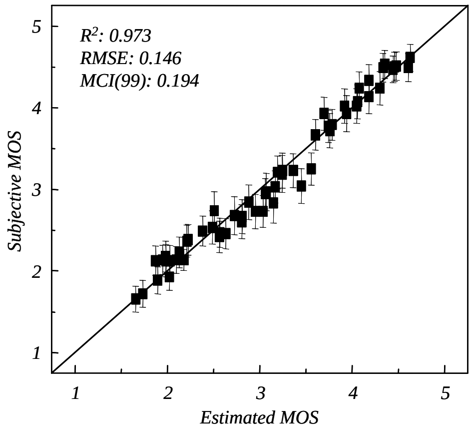

Figure 4: Relationship between subjective MOS and estimated MOS.

When fr is a constant, we define $\tau_1(br)$ as follows:

$$\tau_1(br) \overset{def}{\Longleftrightarrow} \tau(fr, br)\Big|_{fr=const}.$$

When br is a constant, we define $\tau_2(fr)$ as follows:

$$\tau_2(fr) \overset{def}{\Longleftrightarrow} \tau(fr, br)\Big|_{br=const}.$$

Figure 5(b) shows that $\tau_1(br)$ increases monotonically with decreasing br, as follows:

$$\tau_1(br) = w + x\ \exp(-br/j), \quad (7)$$

and Fig. 5(c) shows that $\tau_2(fr)$ increases monotonically with decreasing fr, as follows:

$$\tau_2(fr) = y + z\ \exp(-fr/l), \quad (8)$$

where w, x, j, y, z, and l are constants.

We integrated Eqs. 7 and 8 using regression analysis and obtained:

$$\tau(fr, br) = h + i\ \exp(-br/j) + k\ \exp(-fr/l), \quad (9)$$

where $\exp(-br/j)$ and $\exp(-fr/l)$ are independent variables, and $\tau(fr, br)$ is a dependent variable, and h, i, j, k, and l are constants.

We calculated the coefficient tables of the opinion model for H.264 codec in Table 1. Then, using these coefficients, we estimated the subjective video qualities. These results are shown in Fig. 6. R^2, MCI, and RMSE are also shown in Fig. 6.

In this experiment, R^2 values was larger than 0.9, and RMSE was less than the MCI. That is, our proposed model could be applied to effective design, implementation, and management of videophone applications and communication networks.

5. Verification

We tried to verify the generality of the opinion model using an MPEG-4 codec. The forms (Eqs. 1 to 9) of the model were the same as described in Sect. 4. Table 2 shows the verification test

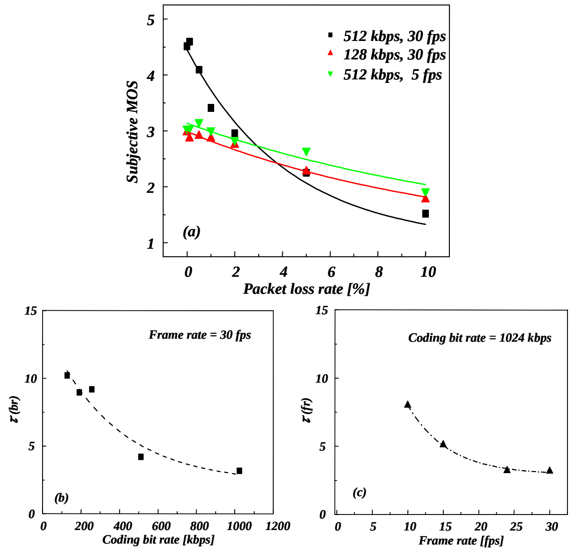

Figure 5: Quality degradation characteristics for packet loss distortion.

conditions. There were 49 test conditions, which is the number of combinations of parameters in Table 2.

The test sequences, the rating scale, and the other assessment environment conditions were the same as those used in Sect. 3 In the verification test, 32 nonexperts aged 20 - 39 participated and they were different from those in Sect. 3.

Table 2: Experimental settings.

Video codec	MPEG-4
Video format	QVGA
Key frame rate (s)	1
Video coding bit rate (kbps)	512 - 2048
Video frame rate (fps)	2 - 30
Video packet loss rate (%)	0.0 - 2.0

We calculated the coefficient tables of the opinion model for MPEG-4 codec. Then, using these coefficients, we estimated the subjective video qualities. The estimation accuracy of the model is demonstrated in Fig. 7. We found that the quality estimation accuracy of our proposed model was sufficient for the verification data.

6. Quality Design Using our Proposed Model

Using our proposed opinion model for an MPEG-4 codec, we can give some guidelines for quality design and management of videophone services. Designing an application and/or network to improve video quality for various video sequences is extremely important to avoid the over-engineering network.

First, we show an example of an application guideline. In general, a network designer can use only a limited bandwidth. For example, when bandwidth is restricted to 0.5 or 2.0 Mbps, our proposed opinion model enables us to find the optimal frame rate, which is about 15 or 30 fps, respectively, as shown in

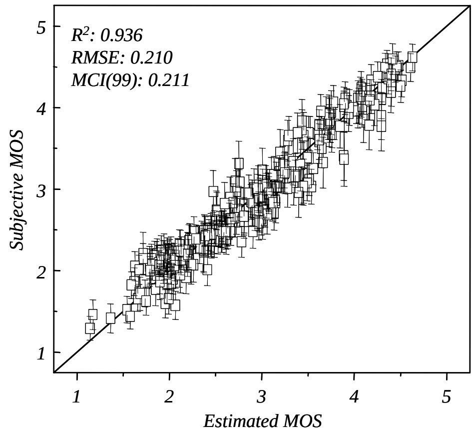

Figure 6: Relationship between subjective MOS and estimated MOS.

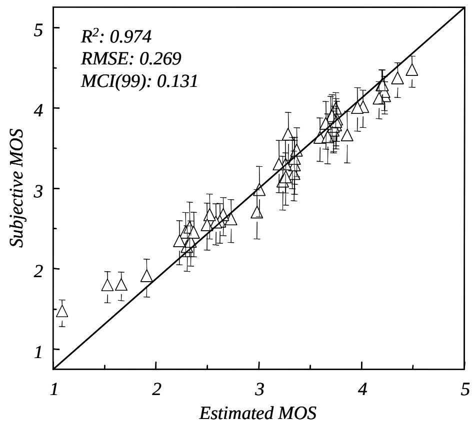

Figure 7: Relationship between subjective MOS and estimated MOS

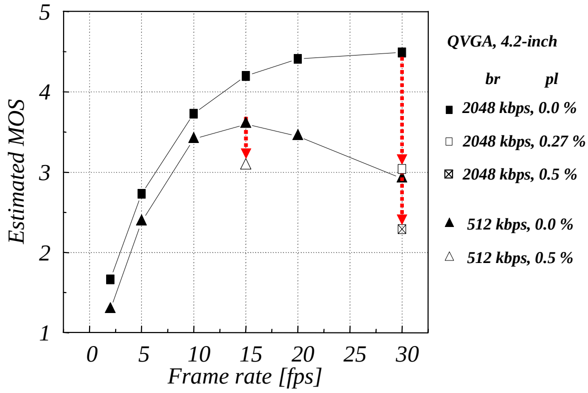

Figure 8: Relationship between subjective MOS and estimated MOS.

Fig. 8. When the packet loss rate is 0.5%, we need to find the optimal coding bit rate because the degrees of degradation due to packet loss are different at each coding bit rate. That is, we can find the optimal coding bit rate, which is about 0.5 Mbps, as shown in Fig. 8.

Next, we explain an example of a network guideline. To satisfy the requirement that $MOS \geq 3.0$ when the bandwidth is 0.5 or 2.0 Mbps, the thresholds for the packet loss rate must be 0.50% or 0.27%, respectively, as shown in Fig. 8.

From these guidelines, we can design the application and/or network appropriately using our proposed model.

7. Conclusion

We defined video quality estimation functions for codecs and a packet loss degradation index from the relationships among video quality and quality parameters. The former function can be used to estimate the optimal video quality at each coding bit rate considering the video frame rate. The latter can be used to estimate the degree of video quality degradation due to packet loss for each application. From these functions, we can estimate video quality using video quality parameters because the RMSEs of H.264 and MPEG-4 codecs in our proposed model were less than the statistical ambiguity of the subjective scores.

We described an example of quality design guidelines for videophone services by using our proposed model. We believe that our study provides a basis for an opinion model such as the E-model for estimating video quality of videophone services.

The following issues call for further study. First, we will apply our model to multimedia quality integration model [13] and verify that the model will have sufficient accuracy. Then, we need to construct the video quality estimation model which can evaluate various displayed video size for PCs, PDAs, and mobile phones. Moreover, we will extend our model to estimate the various video sequences for video streaming services including IPTV.

8. References

[1] A. Takahashi, H. Yoshino, and N. Kitawaki, "Perceptual QoS assessment technologies for VoIP," IEEE Communications Magazine, pp. 28-34, July 2004.

[2] ITU-T Recommendation P.862, "Perceptual Evaluation of Speech Quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," Feb. 2001.

[3] Video Quality Experts Group website, http://www.vqeg.org/.

[4] J. Okamoto, T. Hayashi, A. Takahashi, and T. Kurita, "Verification of objective quality assessment method for arbitrary video sequence," IMQA, Sep. 2005.

[5] ITU-T Recommendation J.148, "Requirements for an objective perceptual multimedia quality model," May 2003.

[6] A. Clark, "Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality," IP Telephony Workshop. Mar. 2001.

[7] M. Masuda and T. Hayashi, "Non-intrusive Quality Monitoring Method of VoIP Speech Based on Network Performance Metrics," IEICE Trans. Commun. Vol. E89-B, No. 2, pp. 304-312, Feb. 2006.

[8] ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning," Mar. 2005.

[9] A. Takahashi, "Opinion model for estimating conversational quality of VoIP," Proc. IEEE ICASSP' 04, vol. III, pp. 1072-1075, May 2004.

[10] K. Yamagishi and T. Hayashi, "Analysis of psychological factors for quality assessment of interactive multimodal service," Electronic Imaging 2005, pp. 130-138, Jan. 2005.

[11] K. Yamagishi and T. Hayashi, "Opinion Model using Psychological Factors for Interactive Multimodal Services," IEICE Trans. Commun. Vol. E89-B, No.2, pp. 281-288, Feb. 2006.

[12] N. Kitawaki, Y. Arayama, and T. Yamada, "Multimedia opinion model based on media interaction of audio-visual communications," MESAQIN2005, pp. 5-10, June 2005.

[13] T. Hayashi, K. Yamagishi, and H. Yoshino, "Perceptual QoS Evaluation Model for Audiovisual Communication Services," World Telecommunications Congress 2006 (WTC' 06), May 2006.

[14] C. Jones and D.J. Atkinson, "Development of Opinion-Based Audiovisual Quality Models for Desktop Video-Teleconferencing," 6th IEEE International Workshop on Quality of Service, May 1998.

[15] T. Hayashi, "State of the Art of Multimedia Quality Assessment Methods," ITU-T Workshop on Video and Image Coding and Applications (VICA), July 2005.

[16] D. S. Hands, "A Basic Multimedia Quality Model," IEEE TRANS. ON MULTIMEDIA, VOL. 6, NO. 6, pp. 806-812, Dec. 2004.

[17] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Sep. 1999.