Proposal of common basic objective measurement model for picture quality for FR

VQEG meeting, May 2001
KDDI R&D Laboratories Inc.*

Many excellent schemes have been proposed for FR in VQEG, but any one of them has not been distinctly superior to the others.

Therefore at the ITU-T SG9 meeting in April 1999, we have proposed the following: Instead of choosing only one scheme, we should extract a basic model that is common to the proposed schemes, and the extracted model should be standardized. Each proposed elemental technology in the model should be described as a practical example.

This proposal, however, was not fully discussed in the SG9 meeting as it was considered in the scope of VQEG.

Therefore, we request the re-consideration of this proposal (attached) at this VQEG meeting.

Attached: ITU-T SG9 (1997-2000) COM-D.84

* Contact: Ryoichi Kawada E-mail: kawada@kddlabs.co.jp

UIT - Secteur de la normalisation des télécommunications
ITU - Telecommunication Standardization Sector
UIT - Sector de Normalización de las Telecomunicaciones

Study Period 1997-2000

Commission d'études ;Study Group;Comisión de Estudio} **9**

Contribution tardive;Delayed Contribution ;Contribución tardía} **D.84**

Boston, 26-30 April 1999

Texte disponible seulement en ;Text available only in;Texto disponible solamente en} **E**

QUESTION: 22/9

SOURCE: JAPAN

TITLE: PROPOSAL OF COMMON BASIC OBJECTIVE MEASUREMENT MODEL FOR PICTURE QUALITY FOR J.ovq

1. Introduction

The present draft new recommendation J.ovq "Objective Measurement of Perceptual Video Quality for Digital Cable Television" includes some appendixes describing currently available methods of objective measurements of the subjective video quality of TV signals. In spite of VQEG (Video Quality Experts Group) efforts to intend to choose the best measurement method through the contest, it is considered very difficult to compare these proposed methods in terms of reliability. Because their measurement performances are strongly depended on video content itself to be evaluated, bit rates to be coded and so on. However, these methods seem to be integrated into a unified common measurement model which can reflect their own measurement procedures with their appropriate measurement parameter set. In this document, the common basic model on objective picture quality measurement is proposed for the draft new recommendation J.ovq. This model consists of 3 measurement layers each of which have some specific measurement parameters to be defined for each objective measurement method. All of the methods so far proposed can be expressed with their appropriate parameters in this model. These parameters are not necessarily included in a normative part but in an informative part for examples in the recommendation.

* **Contact:** S. Matsumoto
Tel: +81 492 78 7420
Fax: +81 492 78 7510
E-mail: su-matsumoto@kdd.co.jp

2. Common Basic Objective Measurement Model

2.1 General Description

Fig.1 shows the overview of objective measurement, where noise signal (n) is obtained by subtracting a test signal (t) from a reference signal (r). Objective Measurement Model is applied to the noise signal and scores are given utilizing reference signal.

In the Basic Objective Measurement Model (Fig.2), A score is determined for a video clip in a video sequence. A video clip contains several frames (z = 1, ---, Z) (Clip layer). Each frame is divided into small regions called as a "spot" (Spot layer). At pixel base, video signals r, t and n are addressed by three dimensional parameter (x, y, z) (Pixel layer). For each layer, **C**, **S** and **P** are defined as group of pixels which determine the clip, spot and pixel-base weighting functions respectively.

2.2 **Basic Concept of Three layers**

(1) Clip layer: Noise perception depends on the movement of human gaze points. Its degree of perception can be predicted by objects construction in a video clip consisted by reference signal r(x, y, z) $\in$ **C** being expressed as a noise weighting function for n(x,y,z) $\in$ **C**.

(2) Spot layer: Noise perception also depends on the characteristics of a spot such as activity and luminance level, on which human gaze points are stopping. Characteristics of a spot can be measured by a reference signal r(x, y, z) $\in$ **S** and degree of noise perception can be predicted by a weighting function for n(x, y, z) $\in$ **S**.

(3) Pixel layer: Noise is differently perceived by its frequency (Horizontal, Vertical and Temporal). Then, n(x, y, z) $\in$ **P** is converted into frequency domain and weighted.

1.3 **Each layer specification**

(1) Pixel layer

As shown in Fig. 3, from the bottom, noise power n^2 (x, y, z) is weighted by

H($f\omega$x,$f\omega$y,$f\omega$z) on the frequency domain and gives weighted noise power n2h(x, y, z). Converting function F_P is performed on n^2(x, y, z) with (x, y, z) $\in$ **P**. This F_P includes Fourier Transform and Orthogonal Transform such as DFT, DCT, DST, WHT etc. Then, we have

$$\begin{aligned} N(\omega x, \omega y, \omega z) &= F_{\mathbf{P}}\left[n^2(x,y,z)\right] \\ n^2{}_h(x,y,z) &= F_{\mathbf{P}}{}^{-1}\left[H(\omega x,\omega y,\omega z)N(\omega x,\omega y,\omega z)\right] \end{aligned} \qquad (1)$$

Equation(1) can be expressed by filtering operation.

$$\begin{aligned} & n^2{}_h = n^2(x,y,z) * h(x,y,z) \\ & where, \\ & h(x,y,z) = F_{\mathbf{P}}{}^{-1}\left[H(\omega x,\omega y,\omega z)\right] \end{aligned} \qquad (2)$$

(2) Spot layer

In a spot, weighting function $w_{\mathbf{S}}$ is derived from reference signal belonging to **S**. That is,

$$w_{\mathbf{S}}(x,y,z) = f_{\mathbf{S}}(r(x,y,z)) \quad , where \quad (x,y,z) \in \mathbf{S}$$

(3)

(3) Clip layer

In a clip, weighting function w_C is also given by reference signal belonging to **C**. Namely,

$$w_C(x,y,z) = f_C(q(x,y,z)) \quad , where \quad (x,y,z) \in \mathbf{C} \tag{4}$$

Consequently, score for this clip is given by

$$Score = G\left[\frac{1}{XYZ}\sum_{z=1}^{Z}\sum_{y=1}^{Y}\sum_{x=1}^{X}\left\{w_C\, w_S\left(n^2 * h\right)\right\}\right] \tag{5}$$

1.4 **Parameter set to be defined**

(1) Pixel layer

F_P and H(or h) must be defined.

(2) Spot layer

f_S must be defined.

(3) Clip layer

f_C and G must be defined.

2. Example of Parameter Set

2.1 **KDD's method**

Giving five parameters to F_P H(or h), f_S, f_C and G, score can be obtained by the equation (5).

(1) Parameter set for F_P

As **P**, we employ 8 pixel(Horizontal) x 8 line block (Fig. 4) in a frame and use 8 x 8 WHT (Walsh Hadamard Transform) as F_P.

F_P **= L • L**^{T}

, where

(6)

L = $\frac{1}{\sqrt{8}}$

エラー!

+ : + 1 - : - 1

• is 8 x 8 matrix

(2) Parameter set for H

By the F_P operation, 8 x 8 matrix N(ωx,ωy) is given. H(ωx,ωy) reflects visual sensitivity on frequency domain is given in Fig.5.

(3) Parameter set for f_S (x, y, z)

Let **S** be same as **P**, and f_S is defined by considering noise masking effect using AC activity in **S**.

$$w_S = f_S(r(x,y,z)) = \frac{1}{\sqrt{2.02\log_{10} E_S}} \qquad (7)$$

$$E_S = \frac{1}{64}\sum_{x=1}^{8}\sum_{y=1}^{8}\left\{r(x,y,z) - \frac{1}{64}\sum_{x=1}^{8}\sum_{y=1}^{8} r(x,y,z)\right\}^2$$

$$, where\,(x,y,z) \in \mathbf{S}$$

(4) Parameter set for f_C (x, y, z)

f_C is given by calculating frame activity considering noise sensitivity variation by "gaze point scattering".

$$w_S = f_C(r(x,y,z)) = -0.0018A_C + 0.806$$

$$A_C = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y} E_S$$

$$, where\,(x,y,z) \in \mathbf{C} \qquad (8)$$

(5) Parameter set for G

To obtain Score as DSCQS (%: Double-Stimulus Continuous Quality-Scale), we give the following function as G.

$$G(d) = -7.32\times10^{-6} d_0 + 1.22\times10^{-3} d_0 - 7.29\times10^{-2} d_0 + 1.8 d_0 - 14.5 d_0$$
$$d_0 = 10\log\frac{255^2}{d} \quad (9)$$

2.2 **Method 2**

(1) (2) h(F_P and H) is three dimensional FIR filtering

(3) f_P is defined considering luminance energy

(4) f_C is not defined.

(5) G is defined for DSCQS(%).

2.3 **Method 3**

(1) F_P is Pyramid Decomposition (Down sampling)

(2) H is three dimensional filtering

(3) f_S is defined considering activity

(4) f_C is not defined

(5) G is defined for an original score scale

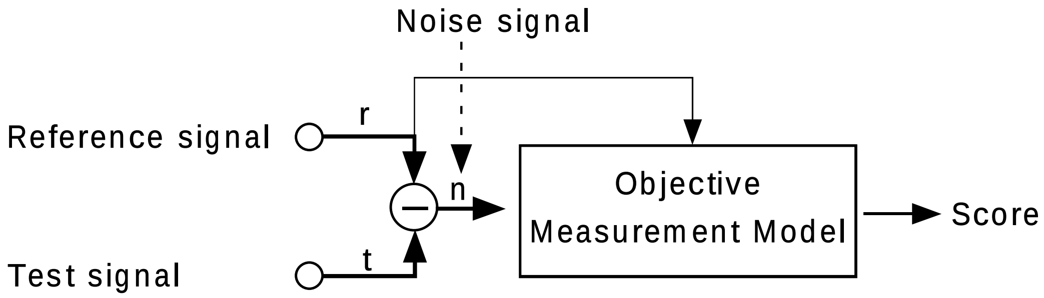

Fig.1 Overview of Objective Measurement

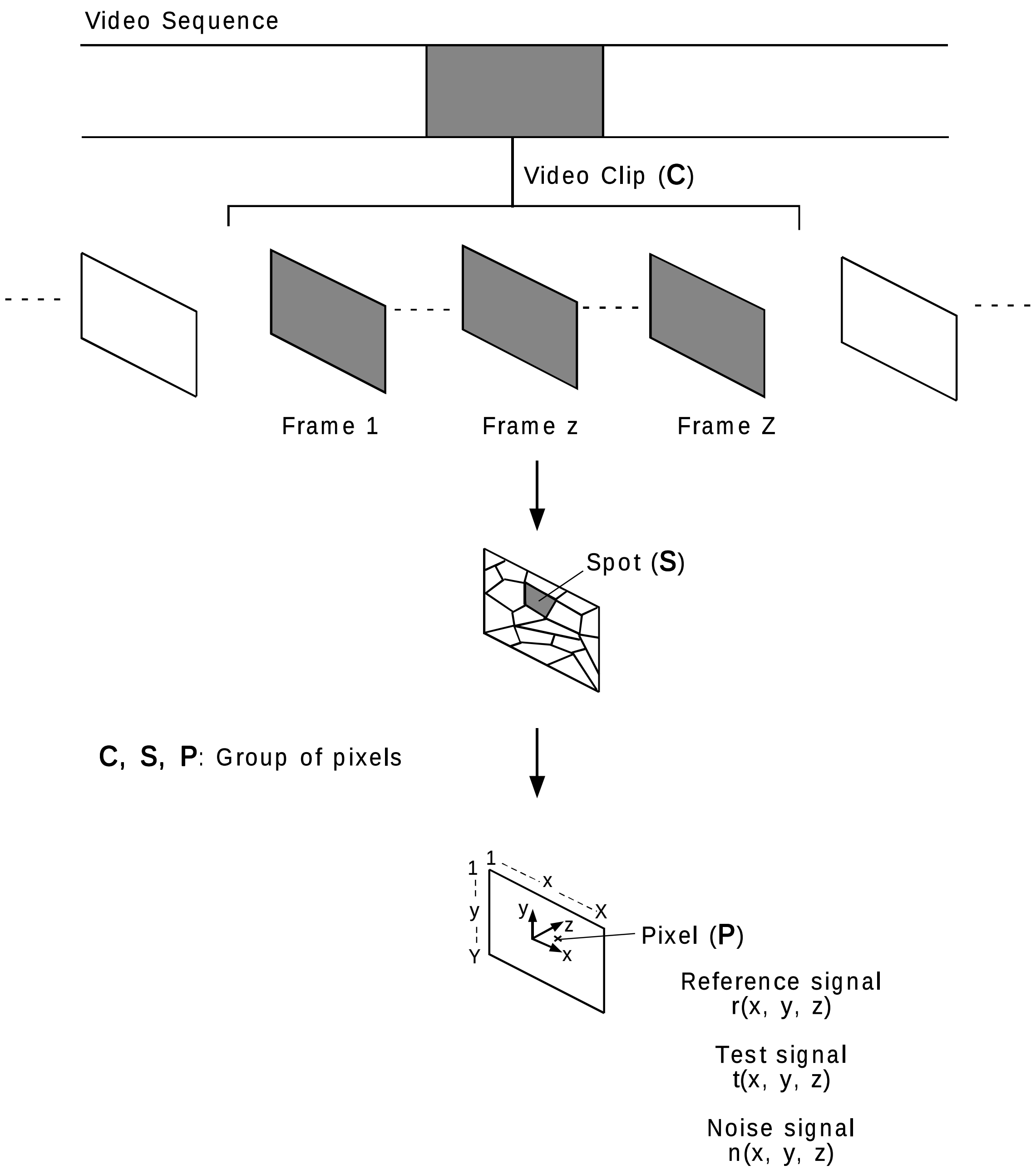

Fig.2 Basic Objective Measurement Model

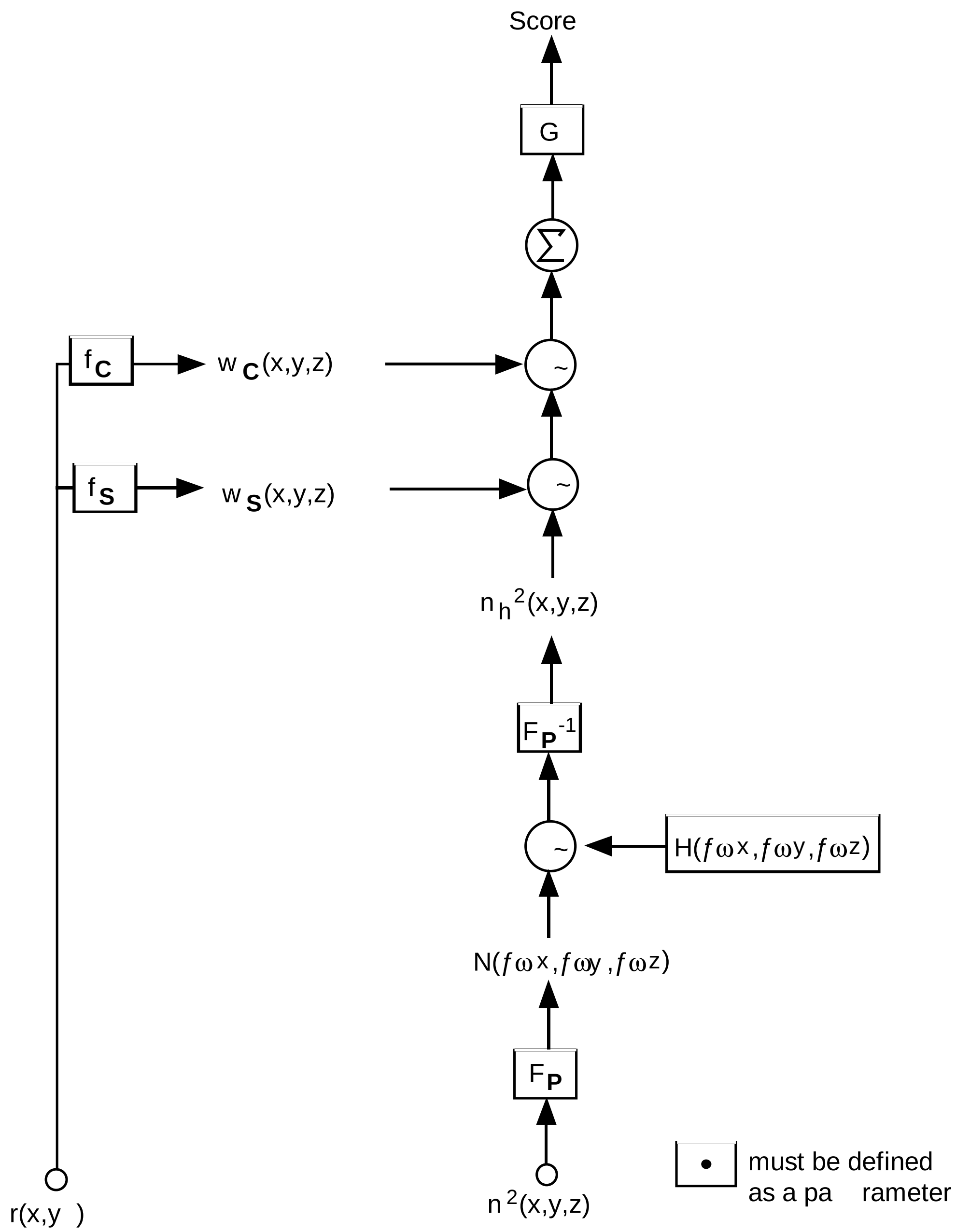

Fig.3 Signal processing

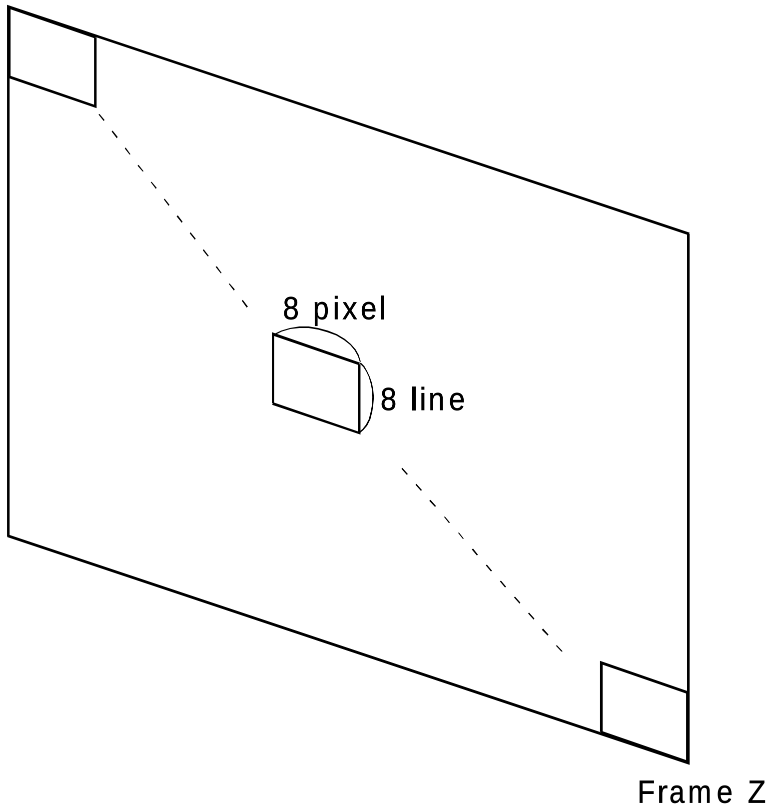

Fig.4 8 pixel x 8 line as P and S

1.00	1.00	0.75	0.60	0.50	0.50	0.40	0.40
1.00	0.75	0.60	0.40	0.30	0.30	0.30	0.15
0.75	0.60	0.40	0.60	0.75	0.30	0.15	0.15
0.60	0.40	0.60	1.00	1.00	0.75	0.15	0.15
0.50	0.30	0.30	0.30	0.15	0.15	0.15	0.15
0.50	0.30	0.30	0.15	0.15	0.15	0.15	0.15
0.75	0.50	0.15	0.15	0.15	0.15	0.15	0.15
0.75	0.75	0.50	0.50	0.15	0.15	0.15	0.15

Fig.5 Weighting function matrix H on (8×8) Hadamard Transform Domain