INTERNATIONAL TELECOMMUNICATION UNION

**TELECOMMUNICATION STANDARDIZATION SECTOR**

STUDY PERIOD 1997 - 2000

**COM 12-7-E**
**February 1997**
**Original: English**

Questions: 11/12

**STUDY GROUP 12 – CONTRIBUTION 7**

SOURCE: KPN RESEARCH, NETHERLANDS

TITLE: OBJECTIVE MEASUREMENT OF VIDEO QUALITY [1]

**Abstract**

An objective perceptual video quality measure based on a simple perceptual improvement of the ANSI video quality measure is presented. The correlation between predicted objective and observed subjective scores is 0.31 for the ANSI based implementation and 0.87 for perceptual improved quality measure.

## 1. Introduction

Within question 11 objective measurement methods for evaluating audio-visual quality in multimedia services is studied. Such methods can be devided into two classes, one in which the system under test is characterised using know-how of the system, and one where the quality of the output signal is characterised using a model of human perception. The first approach will be called glass-box approach because one needs a model of the system under test (Figure 1) while in the second one the system itself is not modelled and viewed as a black box (Figure 2). The advantage of the black box approach is that no information is needed about the system under test. Glass box approaches for modern telecommunication systems are extremely difficult because of the non-linear, time-varying behaviour of these systems.

[1] Contact: John G. Beerends, KPN Research, +3170-3325644,
E-mail J.G.Beerends@research.kpn.com

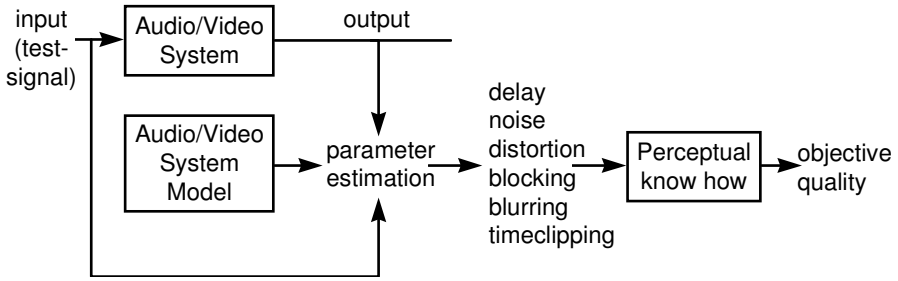


FIGURE 1

**Glass box approach towards objective measurement of the quality of an audio/video system. A model of the system is used to estimate its parameters, using a test signal that is appropriate for the device. The set of parameters is mapped to the subjective quality of the audio/video system**

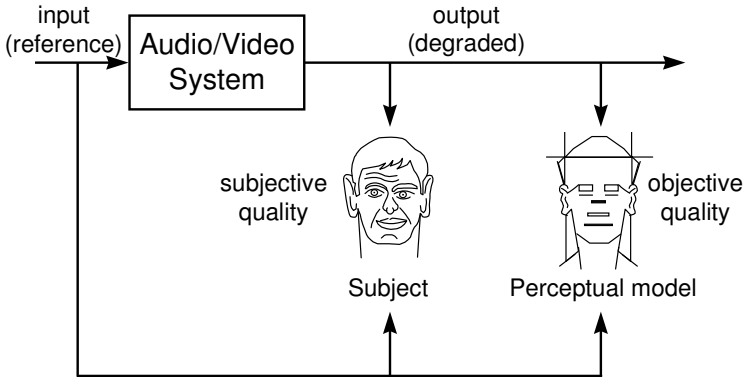


FIGURE 2

**Black box approach towards objective measurement of the quality of an audio/video system. A computer model of the subject is used to compare the output of the device under test with the input, using any audio/video signal that is appropriate for the service provided (speech, music, natural video scenes, etc).**

For speech quality measurement the glass box approach has resulted in a number of models [1], [2] while for the black box approach recently a perceptual model (PSQM, [13], [14]) was accepted for measuring telephone-band speech quality [5]. Because the glass box approach is not suited for modern non-linear, time-varying systems such as low bit rate speech codecs they are supplemented with subjective evaluations [1]. The black box approach can be used for any system as long as large sets of speech signals are used in combination with an accurate model of human auditory perception.

For video quality measurement ANSI accepted a standard [6] that was also submitted to ITU-T SG 12 as a proposal [7]. However, in the ANSI approach the glass box and black box approach are mixed in contrast to the recommendations on measuring speech quality. In order to keep study group 12 recommendations coherent it is to be preferred to split objective video quality recommendations into glass box and a black box approaches.

The ANSI proposal [7] starts with a number of measurements using test signals for determining model parameters like gain and offset (glass box approach). Because of the non-linear, time-varying behaviour of modern video codecs this approach only has a limited value. In a separate section a black box approach is proposed applicable to natural scenes. It allows the use of a large set of quality parameters derived from the comparison of the input and output as in Figure 2. However, no single model specification is given allowing a wide variety of different implementations. This contrasts the recommendation on measuring telephone-band speech quality [5] where a single parameter is constructed that showed a high correlation in an independent validation by the NTT [8]. In order to keep study group 12 recommendations coherent, and allow comparison of results obtained with objective measurement systems, a recommendation on objective video quality should include a single unique implementation that correlates well with the subjectively perceived video quality.

In this contribution a proposal is given for such a unique implementation. The scope of the objective perceptual video quality measure that is proposed is limited to low and medium quality video (less than television quality). The database used in this contribution resulted from a subjective test of different MPEG-4 video codec proposals [9]. This subjective test was based on ITU-R Recommendation 500 [10] using an eleven point (0 = poor quality, 10 = excellent quality) single stimulus method.

In section 2 an ANSI based video quality measure will be presented. After validation of this ANSI quality measure in section 3, the perceptual improvement is introduced and validated in section 4. An overview of the results and the conclusions are presented in the last section.

## 2. The ANSI based video quality measure

A first point of attention in implementing the ANSI method is that no space is defined in which all calculations have to be performed. Implicitly the standard only deals with the luminance signal. However, ignoring colour information is not acceptable in a video quality measure that must cope with a wide variety of distortions. Therefore, we decided to extend the recommendation to the colour domain. The best space for all calculations is the one where equal distances represent equally perceptible differences. Although literature provides many examples of such spaces the one that is most widely used for colour television is the L*u*v* space [11]. Most video representations however use a luminance signal Y and two colour difference signals derived from the RGB signals of a camera. The RGB video representations are compressed representations of the actual RGB signals because of the gamma correction used in television. In our implementation of the objective video quality measure we started to convert the CIF $YC_RC_B$ signals to RGB representations which were then expanded with a gamma of 2.8. Next the signals were transformed to the L*u*v* representation [11]. All results given in this contribution are based on calculations in this L*u*v* space.

Within the ANSI proposal a large set of quality indicators is proposed. This report starts with the evaluation of the first 10 quality indicators as given in Table A1 of [7]. These indicators

have the advantage that they can be calculated fairly easy, no large storage is required because both the input frame and output frame of the video sequence are mapped onto a single scalar. If each quality indicator is calculated in all three dimensions of the L*u*v* space we get a total number of 30 quality parameters. However, indicator 5 is only calculated on the luminance signal L* while all indicators that use the Sobel operator (edge detection, nr. 6, 7, 8 and 10) turned out to have a chrominance weighting that was about the same for both the u* and v* component. To keep the number of free parameters as low as possible the two chrominance parameters for these indicators are combined to a single parameter. This results in a total of 24 quality parameters derived from the 10 quality indicators.

Because of computational and storage limits the number of video sequences had to be limited to around 40. In the validation of the ANSI model two MPEG-4 subsets with around 40 sequences were used. A fundamental problem in the validation is that the large number of free parameters in the model will always lead to high correlations on small databases. However, the model parameters can be fitted to one set of sequences and then be used to predict the subjective scores of a second set. When the experimental contexts of both sets are the same the model should give a high correlation without having to refit the data. This procedure is carried out in the next section.

## 3. Validation of the ANSI based video quality measure

The ANSI model was optimised on a set of 40 sequences of the MPEG-4 database (database MPEG-4a) using a multiple linear regression [12]. This optimisation resulted in a correlation of 0.89 (see Figure 3). If only the luminance signal L* is used the correlation dropped to 0.73, too low for use in practical measurements. Colour information can thus not be neglected in an objective video quality measure.

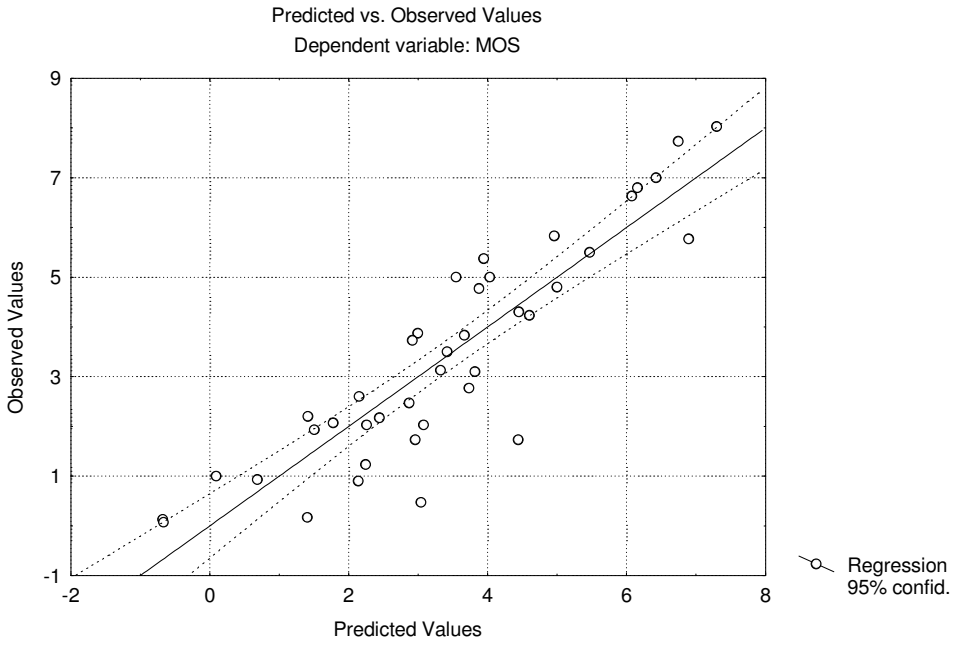


FIGURE 3

**Results of the ANSI based video quality measure on the MPEG-4a database using an optimal multiple linear regression on the subjective results of the MPEG-4a database. Correlation of this optimal fit is 0.89.**

Next the subjective scores of a second set of 36 sequences, database MPEG-4b, were predicted using the regression coefficients obtained from the optimisation of the ANSI model on the MPEG-4a database. The result was a correlation of 0.31 (see Figure 4) showing that the predictive power of the model is too low in order to be useful in practical situations. The reason for this is probably that the 24 quality parameters only have a limited perceptual interpretation. Subjects do not map solely an input frame and an output frame to a scalar and then compare these numbers, but will also compare the most relevant parts on a local space scale. Our goal is now to update the ANSI model using the best parts of it in combination with a number of new indicators that are based on know how of the human visual system.

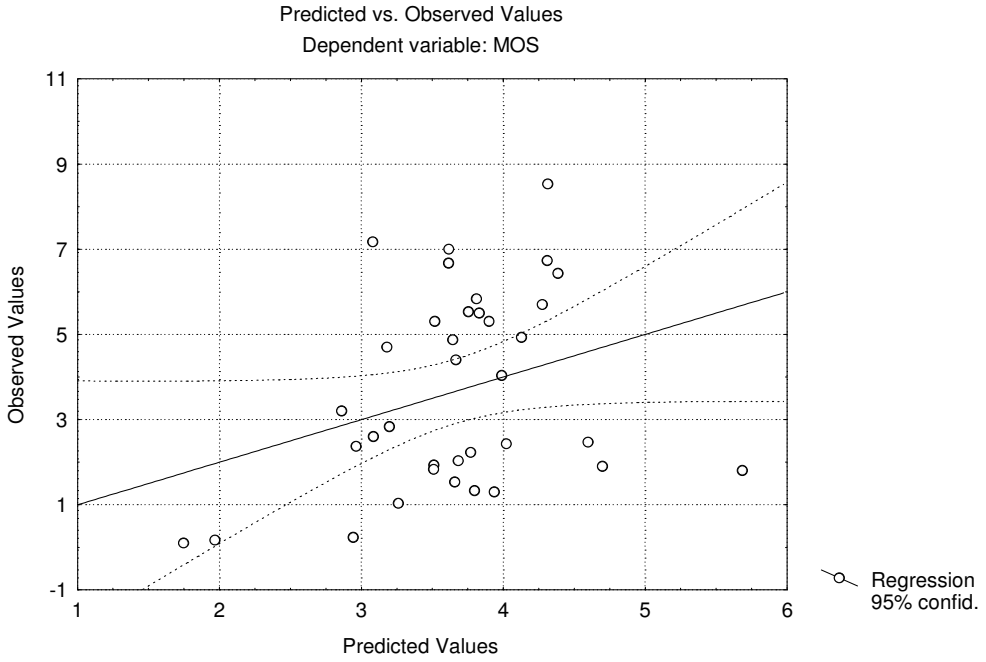


FIGURE 4

**Results of the prediction of the subjective scores of the MPEG-4b database with the ANSI based video quality measure. Correlation between predicted and observed scores is 0.31.**

## 4. A perceptual video quality measure based on the ANSI approach

Experience in the development of objective audio quality measures have led to the insight that exact modelling of the peripheral processes found in human perception has only limited value. Inclusion of simple cognitive aspects of perception are needed in order to get high correlations between objective and subjective results. A basic idea that proved to be very successful in audio quality measurements is the asymmetry between distortions caused by introducing new artefacts versus distortions caused by leaving out parts of the original signal [13], [14]. This effect is also included in the ANSI video quality measure by using separate quality indicators for added versus lost motion energy, and added versus lost edge energy (see [7]).

The rational behind the asymmetry effect is that when something is left out of the original the remaining signal is still one coherent percept while the introduction of an artefact causes a disintegrated additional percept leading to a more objectionable distortion.

The key idea to improve the ANSI quality measure is to use the idea of asymmetry on a pixel matrix basis. In the ANSI document three different levels of quality features are defined, scalar, vector and matrix features. In the previous section the implemented quality measure was derived only from the scalar features. With this implementation one needs large databases because of the great number of freedoms that are used in the multiple regression. Furthermore it was shown in the previous section that predicting subjective scores for a database using a model optimised on a slightly different database, is not feasible. Extending the model with

vector and matrix features would give an even larger number of free parameters, thus increasing the problem even more.

Therefore a more basic perceptual approach was taken to develop a quality measure that uses less parameters and has good predictive power. In literature there are many models that deal with the lower levels of auditory perception (see e.g. [15]). However, as stated, the higher levels of cognitive processing dominate the perception of audio and video quality. A combination of perceptual and cognitive modelling for measuring video quality could not be found in the literature.

The asymmetry idea however was used in an earlier paper by the originators of the ANSI model [16] in a more fundamental form. They implemented it by using a quality indicator derived from the positive difference between the sobel filtered output and input frames (introduced edges). Furthermore they made a distinction between still and moving parts. This idea is combined with the following basics of human perception:

1. The luminance channel is more sensitive to spatial edges then the chrominance channel [18]

2. When objects are moving spatial edges become less dominant [17]

This leads to the following sets of quality parameters:

$$Q_1 = mean_+\left\{still\left(Sobel\left(OUT_{L^*}\right)\right) - still\left(Sobel\left(IN_{L^*}\right)\right)\right\} \tag{1}$$

$$Q_2 = mean_-\left\{still\left(Sobel\left(OUT_{L^*}\right)\right) - still\left(Sobel\left(IN_{L^*}\right)\right)\right\} \tag{2}$$

$$Q_3 = mean\left|\left\{still(OUT_{u^*}) - still(IN_{u^*})\right\}\right| \quad + \quad mean\left|\left\{still(OUT_{v^*}) - still(IN_{v^*})\right\}\right| \tag{3}$$

$$Q_4 = mean\left|\left\{motion(OUT_{L^*}) - motion(IN_{L^*})\right\}\right| \tag{4}$$

$$Q_5 = mean\left|\left\{motion(OUT_{u^*}) - motion(IN_{u^*})\right\}\right| + mean\left|\left\{motion(OUT_{v^*}) - motion(IN_{v^*}\right\}\right| \tag{5}$$

The Sobel operator is again used as edge detector. The $mean_+$ and $mean_-$ operators indicate the mean over the positive parts (introduced edges, blocking) and the mean over the negative parts (left out edges, blurring) respectively. When these parameters are combined with parameters derived from the following ANSI quality indicators:

$$A_1 = RMS_{time}\left|\frac{std_{space}\left(Sobel\left(IN_n\right)\right) - std_{space}\left(Sobel\left(OUT_n\right)\right)}{std_{space}\left(Sobel\left(IN_n\right)\right)}\right|, \tag{6}$$

$$A_2 = f_{time}\left(\max\left\{0, RMS(\Delta IN) - RMS(\Delta OUT)\right\}\right), \tag{7}$$

with

$$\Delta IN = IN_n - IN_{n-1}, \qquad \Delta OUT = OUT_n - OUT_{n-1}$$
$$f_{time}\left(\{x_t\}\right) = std_{time}\left(high\ pass\ \ filter\left(\{x_t\}\right)\right)$$
$$high\ pass\ filter\ respons = (\ -1 \quad 2 \quad -1\ )$$

and

$$A_3 = \max_{time}\left\{\log_{10}\left(\frac{std_{space}(\Delta OUT)}{std_{space}(\Delta IN)}\right)\right\}, \tag{8}$$

we get a total number of 13 quality parameters. From $A_1$ two quality parameters, $Q_6$, $Q_7$ are derived, one for the luminance L* and one for the combined chrominance signal. From $A_2$ and $A_3$ six quality parameters are derived, $Q_{8,..}$ $Q_{13}$, because $A_2$ and $A_3$ operate on the L*u*v* components separately. This perceptual ANSI model thus has only about half the number of quality parameters when compared to the ANSI model presented in section 2.

When the model is optimised on the MPEG-4a database the correlation between objective and subjective results is 0.93 (see Figure 5).

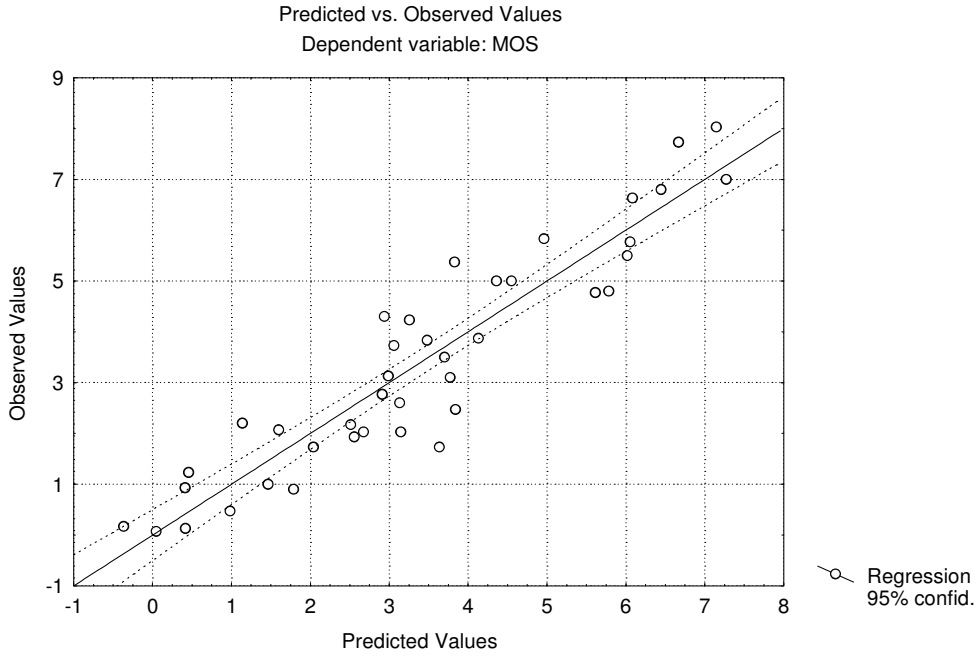


FIGURE 5

**Results of the perceptual video quality measure on the MPEG-4a database using an optimal multiple linear regression on the subjective results of the MPEG-4a database. Correlation of this optimal fit is 0.93.**

Although this is only marginally better that the ANSI approach presented in the previous section (see Figure 3), the proof of the power of this perceptual model lies in the prediction of the subjective scores of database MPEG-4b. The correlation between predicted and measured scores is 0.87 (see Figure 6) which significantly better then the 0.31 of the ANSI approach (see Figure 4).

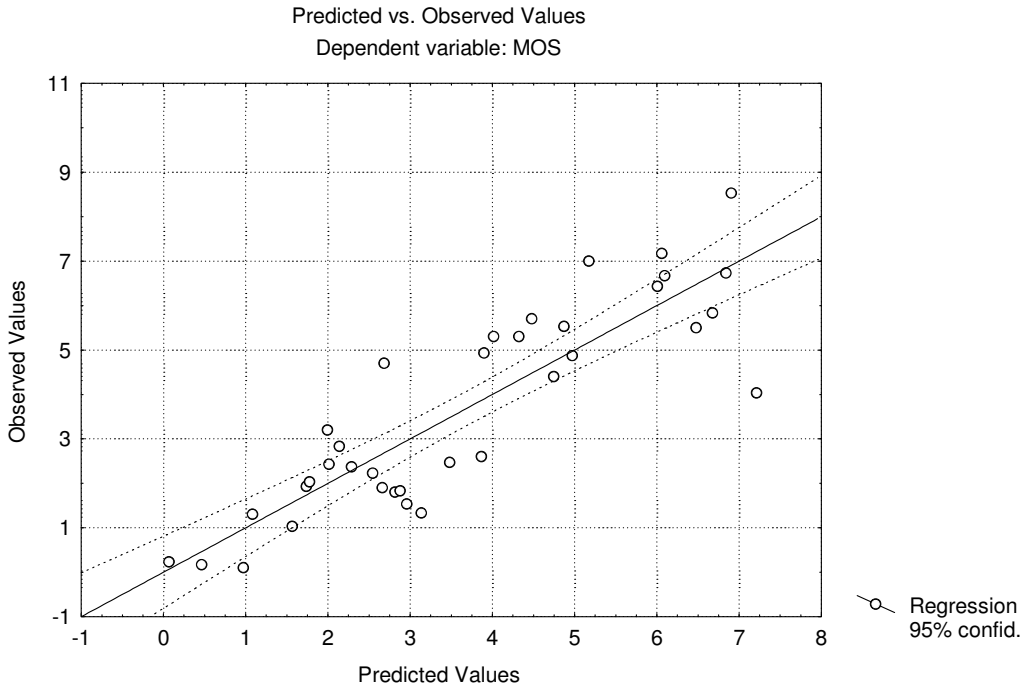


FIGURE 6

**Results of the prediction of the subjective scores of the MPEG-4b database with the perceptual video quality measure. Correlation between predicted and observed scores is 0.87.**

## 5. Results and conclusions

Table 1 gives an overview of the results. It is clear that the ANSI based video quality measure derived from [7], does not give reliable predictions of the subjective video quality. If the ANSI based quality measure would be extended to include more distortions, the number of free parameters would become too big. The perceptual video quality measure based on the ANSI approach requires only half of the number of free model parameters, while the correlation between predicted subjective scores and observed subjective scores increases from 0.31 to 0.87. Although this correlation is still below the correlation required for accurate predictions we think the model is a good basis for a newly to be developed ITU-T recommendation on the objective measurement of video quality.

*Table 1. Overview of the results. The trained correlations are derived from an optimal fit of the free parameters on the subjective data. The untrained correlations are real predictions.*

| **MODEL** | **NUMBER OF FREE PARAMETERS** | **CORRELATION TRAINED** | **CORRELATION UNTRAINED** |
|---|---|---|---|
| ANSI | 24 | 0.89 | 0.31 |
| ANSI + perceptual | 13 | 0.93 | 0.87 |

## 6. References

[1] ETSI, ETR 250 (VTQM-E), *Speech communication quality from mouth to ear of 3,1 kHz handset telephony across networks*, July 1996.

[2] CCITT, Blue Books, Volume 5 - Supplement No.3, *Models for predicting transmission quality from objective measurements.*

[3] J.G. Beerends and J.A. Stemerdink, *A perceptual speech-quality measure based on a psychocoustic sound representation,* J. Audio Eng. Soc., Vol.42 no.3, pp.115-123, March 1994.

[4] J.G. Beerends, *Modelling cognitive effects that play a role in the perception of speech quality,* in Speech quality assessment, workshop papers, Bochum, pp 1-9, November 1994.

[5] ITU-T, Contribution COM 12-67, Draft recommendation P.861, *Objective quality measurement of telephone-band (300-3400 Hz) speech codecs*, study period 1993-1996.

[6] ANSI, Standard T1.801.03-1996, *Digital transport of one-way video signals - parameters for objective performance assessment.*

[7] ITU-T, Contribution COM 12-66, *Selections from the draft American national standard: - Digital transport of one-way signals - parameters for objective performance assessment*, study period 1993-1996.

[8] ITU-T, Contribution COM 12-74, *Review of validation tests for objective speech quality measures*, study period 1993-1996.

[9] T. Alpert et al. *Subjective Evaluation of MPEG-4 Video Codec Proposals: Methodological Approach and Test Procedures.* Image Communication, special issue on MPEG-4, May 1997.

[10] ITU-R, Recommendation BT.500, *Methodology for the Subjective Assessment of the Quality of Television Pictures*, November 1993

[11] CIE. *Colorimetry.* Publication 15.2, 1986.

[12] Statistica, *General conventions & statistics 1,* Chapter 12*, Multiple regression.* StatSoft USA, 1995.

[13] J.G. Beerends, *Modelling a cognitive effects that play a role in the perception of speech quality,* in Speech Quality Assessment, Workshop papers, Bochum, pp1-9, November 1994.

[14] J.G. Beerends, *Modelling a cognitive aspect in the measurement of the quality of music codecs,* AES Convention paper, preprint 3800, Amsterdam 1994.

[15] A. B. Watson (edt), *Measurement and prediction of visual quality,* in Digital Images and Human Vision, MIT press, Cambridge Massachusetts, USA, 1993.

[16] S.D. Voran, *The development of objective video quality measures that emulate human perception.* Globecom conf.publ.no.1776-1781 vol.3, 1991.

[17] L.A. Olzak, J.P. Thomas. *Seeing spatial patterns.* Handbook of Perception and Human Performance (Wiley, New York, 1986), sec. II: "Basic sensory processes I," chap. 7.

[18] G.J.C. van der Horst, C.M.M. de Weert, M.A. Bouman. *Transfer of spatial chromaticity-contrast at the threshold in the human eye*. J. Opt. Soc. Am. 57, pp 1260-1266, 1967.