JRG 2

**Objective Performance Assessment: Video Quality as an Influence on Audio Perception** 4590 (L-10)

M. P. Hollier and R. Voelcker
BT Laboratories
Martlesham Heath, Ipswich IP5 7RE, UK

**Presented at the 103rd Convention 1997 September 26–29 New York**


**AES**



**AN AUDIO ENGINEERING SOCIETY PREPRINT**

# *Objective performance assessment: Video quality as an influence on audio perception*

M P Hollier, R Voelcker
MLB3/7, BT Laboratories, Martlesham Heath, IPSWICH. IP5 7RE, UK
Tel: +44 1473 644335
E.mail: mike.hollier@bt-sys.bt.co.uk

## Abstract

The objective evaluation of the audio performance of non-linear processes such as data reduction has required the development of a new generation of perception based assessment algorithms. In the future it is expected that a large proportion of communications traffic will be wideband and multi-media. In developing objective techniques to assess multi-media systems it is necessary to investigate, and model, the dependencies which exist between audio and visual perception. This paper reports an experimental investigation into combined audio/video perception and proposes an algorithmic basis for a multi-sensory perceptual model.

## 1.0 Introduction

Reliable objective assessment of perceived performance is required for optimal design, commissioning, and monitoring of quality. Non-linear processes such as data reduction have necessitated the development of a new generation of objective assessment algorithms. Conventional engineering metrics do not take adequate account of the properties of the human receiver, and new perceptually motivated techniques have started to emerge. As advanced Telepresence, virtual reality and MM (Multi-Media) applications and services become widely used a multi-sensory perceptual model will be required. This paper reviews recent advances in single sense perceptual models, reports an early experimental investigation into cross modal dependency, and proposes an algorithmic basis for a multi-sensory perceptual model.

Perceptually motivated audio assessment is relatively advanced and numerous models have been proposed for objectively assessing both high quality audio, e.g. [1] and telephone speech quality, e.g. [2,3]. Visual perceptual models which reproduce the gross psychophysics of vision are also described in the literature, usually specialising in particular aspects of visual perception, e.g. [4,5]. Models of the individual senses do not reproduce the full complexity of human perception but rather allow signals to be transformed, taking account of the dominant psychophysics, to provide a more perceptually relevant representation of the signal. In this way parts of the signal which are imperceptible are weighted down and parts of the signal which coincide with peaks in sensory sensitivity are weighted up. The value of this approach is highlighted by the success of perceptually motivated codecs which achieve data reduction by exploiting the perceptual significance of signal components. Higher level perceptual issues certainly have a bearing on opinions expressed by subjects, but the capacity to exploit basic sensory performance remains. This is our aim for multi-sensory analysis where low level cross-modal effects can be usefully measured and modelled with high level cognitive issues dealt with by constraining ourselves to particular application scenarios.

### 1.1 Auditory perceptual models

In order to determine the subjective relevance of errors in audio systems, and particularly speech systems, assessment algorithms have been developed based on models of human hearing. The prediction of audible differences between a degraded and reference signal can be thought of as the *sensory layer* of a perceptual analysis, while the subsequent categorisation of audible errors can be thought of as the *perceptual layer*. Models for assessing high quality audio, e.g. [1], have tended only to predict the probability of detection of audible errors since any audible error is deemed to be unacceptable, while early speech models have tended to predict the presence of audible errors and then employ simple distance measures to categorise their subjective importance, e.g. [2,3,8]. It has been previously shown [7] that a more sophisticated description of the audible error provides an improved correlation with subjective performance. In particular, the amount of error, distribution of error, and correlation of error with original signal have been shown to provide an improved prediction of error subjectivity.

Figure 1 shows a hypothetical fragment of an error surface. The error descriptors used to predict the subjectivity of this error are necessarily multi-dimensional, i.e. no single dimensional metric can be contrived to map between the error surface and the corresponding subjective opinion. The error descriptors, $E_d$, are in the form:

$$E_{d1} = fn_1\{e(i,j)\}$$

where $fn_1$ is a function of the error surface element values for descriptor 1. For example the error descriptor for the distribution of the error, Error-entropy (Ee), proposed in [7] was given by:

$$Ee = \sum_{i=1}^{n} \sum_{j=1}^{m} a(i,j) \ln a(i,j) \qquad \text{where } a(i,j) = |e(i,j)| / Ea$$

and Ea is the sum of $|e(i,j)|$ with respect to time and pitch.

$$\text{Opinion prediction} = fn_2\{E_{d1}, E_{d2}, ..., E_{dn}\}$$

where $fn_2$ is the mapping function between the n error descriptors and opinion scale of interest.

It has been shown that a judicious choice of error descriptors can be mapped to a number of different subjective opinion scales [9]. This is an important result since the error descriptors can be mapped to different opinion scales which are dominated by different aspects of error subjectivity.

### 1.2 Visual perceptual models

Visual perceptual models are also under development and several have been proposed in the literature, e.g. [4] which proposes the use of Gabor functions to account for the inhibitory and excitatory influences of orientation between masker and maskee, and [5] in which the authors use a simple image decomposition into edges, textures and backgrounds. However, most of the published algorithms only succeed in optimising individual aspects of model behaviour; [4] provides a good model of masking, and [5] a first approximation to describing the subjective importance of errors.

An approach similar to that of the auditory perceptual model has been adopted at BT Labs. A sensory layer reproduces the gross psychophysics of the sensory mechanisms:

(i) spatio-temporal sensitivity known as the human visual filter, and
(ii) masking due to spatial frequency, orientation and temporal frequency.

The subjective testing undertaken to validate this model is the subject of a separate publication which is in preparation at the time of writing.

Following the sensory layer the image is decomposed to allow calculation of error subjectivity according to the importance of errors in relation to structures within the image, as shown in Figure 2. If the visible error coincides with a critical feature of the image, such as an edge, then it is more subjectively disturbing. The output from the perceptual layer is a set of context sensitive error descriptors which can be weighted differently to map to a variety of opinion criteria.

### 1.3 Factors contributing to subjective opinion

A key consideration for objective performance prediction in communications systems is the algorithmic prediction of the subjective consequences of perceptible errors. This "error subjectivity" is influenced by many factors including: the task undertaken, equipment utility, user expectation, and for multi-media systems the interdependencies between the senses. Factors other than audio quality which influence subjective opinion have also been discussed in order to yield practical assessment systems [9]. Recognising these factors and contriving experiments in which their influence is constrained to be representative of a particular task , e.g. a head and torso video conference task, or virtual meeting, is a key challenge in understanding and modelling multi-modal perception. Measuring and modelling the underlying inter-sensory dependencies, to which higher level task influences can be applied, relies on the design and validation of a number of sensory error descriptors which are sensitive to a variety of aspects of subjectivity. The contribution of these parameters to error subjectivity can then be predicted via a model of cross-modal effects and the influence of the task scenario. High level cognitive issues such as emotional state are not addressed, although aspects of user stressing are implicitly included via the task undertaken.

### 1.4 Multi-modal interaction

In the future it is expected that an increasing proportion of communications will be wideband and MM. In particular, video conferencing, virtual meeting spaces and networked immersive environments are expected to become commonplace. The interaction between the senses in these situations is complex and the significance of transmission errors and choice of bandwidth utilisation is correspondingly difficult to determine. This difficulty highlights the need for objective measures of the perceived performance of MM systems. Fortunately, to produce useful engineering tools, it is not necessary to model the full extent of human perception and cognition, but rather to establish and model the underlying (low level) inter-sensory dependencies.

It is possible to provide familiar examples of inter-sensory dependency, and these are useful as a starting point for discussion, despite the more sophisticated examples which soon emerge. Strong multi-sensory rules are already known and exploited by content providers, especially film makers. Consistent trajectories between scene cuts, and the constructive benefit of combined audio and video cues are obvious examples. Exploitation of this type of multi-modal relationship for human computer interface design is discussed in [10]. Less familiar examples include the mis-perception of speech when audio and video cues are mismatched, e.g. [11], and modification of error subjectivity with sequencing effects in the other modality, e.g. [12].

The body of the paper discusses a multi-modal subjective experiment, the results of this experiment and the implications of these results for objective quality assessment.

## 2.0 Experimental investigation into cross-modal interaction

### 2.1 Experiment paradigm

The first stage in the development of objective techniques to measure the subjective performance of MM systems is the collation of data on how these systems are perceived, and in particular the interactions which occur between the senses. To ensure this data, which will form the basis of future multi-modal models, is statistically valid, formal subjective assessments should be performed which conform to recognised and controlled design constraints.

When characterising MM systems the potential area of investigation is vast, hence the initial scope for this work is to focus in on specific aspects such as the influence of video quality on audio perception. These specific aspects will be explored in order to determine gross inter-sensory dependencies which can be used as a foundation for the initial models and as a base for further experimentation. For this experiment, the major goal was to test



the hypothesis that video quality has an influence on the perception of audio quality, as well as obtaining general information on the interactions between the two modes in terms of overall quality.

To maximise the amount of information obtained from this experiment we chose to use a non-interactive viewing-only experiment, as using an interactive task (e.g. one involving human to human communication over a multi-media system) would have required a reduced scope to control the additional degrees of freedom. The subjects watched a set of MM presentations and were asked to give discrete opinions for each presentation on its completion.

### 2.2 Source Material

The material viewed in the experiment consisted of a series of short video clips, with supporting audio commentaries. The visual element consisted of a series of virtual reality fly-throughs of a hypothetical building. Eight different video clips were used, each lasting approximately 10 seconds. Unique audio commentaries were created for each of these clips to provide two general narratives on the themes of;

| | |
|---|---|
| Software Theme | where the methods used to create the fly-though were described, as if explaining some of the features of the packages to a potential new user. |
| Building Theme | where the virtues of the building were described from the perspective of an architect promoting a design to a potential customer. |

Digital recordings were made for each of these themes using both a male and a female talker, giving a total of 32 combinations (8 video clips x 2 audio themes x 2 talkers).

### 2.3 Degradations

A series of degradations, both audio and visual, were used to vary the quality of the presentations. These degradations will add to the variation in opinion that will occur due to the different pieces of source material.

The primary aim of the experiment was to determine if changes in the video quality affected the perceived audio quality, and so a wider range of video qualities than audio qualities was required. We therefore elected to use eight video and four audio degradations. These were used in all possible combinations, yielding a total of 32 degradations (8 video x 4 audio). This allowed the application of these degradations to the 32 source material combinations to be balanced across the experiment; described in more detail in section 2.5.

The purpose of the experiment is not the assessment of any particular multi-modal system, but rather a general investigation into the effects of mutli-modal perception. Therefore we were not constrained to any particular set of audio or video degradations and standard, well characterised degradations were chosen.

For the video degradations we used a sub-set of the degradations available in the ITU VIRIS software toolkit [13], as this provides algorithmically defined, well characterised degradations. Three video degradations were selected from the toolkit:

| | |
|---|---|
| • Edge business | a distortion concentrated at the edges of objects, characterised by temporally varying sharpness or spatially varying noise. This adds a flickering type of degradation to hard edges in the image. |
| • Blurring | makes the image blurred and unfocussed |
| • White Noise | produces a snow-storm like effect. |

Two levels of each of these degradations were selected by expert viewing. The first of these (the lowest level) was chosen so as to be noticeably present, and the second level was

chosen to be noticeably worse than the first. These six video degradations (3 types, each at two levels), together with an undegraded condition and one with no-video gave us our 8 video degradations.

The four audio conditions used were:

- No degradation
- Band-limited — telephone bandwidth (300-3500 Hz)
- MNRU (Q=20 dB) — Modulated Noise Reference [14][1]. This adds noise proportional to the amplitude of its input signal. It is a widely used degradation in audio experiments, and is designed to simulate the effect of tandemming PCM systems.
- Band-limited and MNRU — a combination of the above, where the band-limiting was performed and then the band-limited signal was passed through the MNRU.

**2.4 Experiment design**

In order to provide statistically verifiable results, a formal balanced design was used, using Graeco-Latin squares. This design methodology uses a pair of squares; one defining the order in which the source material is seen, and the other defining the degradations applied. The use of a Graeco-Latin square design ensures that over the experiment as a whole, all possible combinations of source material and degradations are used once and only once. It also ensures that each subject sees every piece of source material once only, and each degradation once only, and that for each subject the combination of source material and degradations is unique.

In total 32 subjects were used, each completing the experiment twice, once being asked a question on audio quality only, and once on overall quality. Half the subjects were asked the audio question first, and half were asked it second to provide balance across the experiment. All the subjects were non-expert and selected at random from the BT Laboratories subject database. People directly involved in video work were excluded. Subjects were asked to wear their glasses if they normally use them whilst watching television.

All the material for the experiment was pre-processed through each of the degradations and stored on Hard disk. This allowed a custom program to control the experiment, and accept and store the subjects' opinions.

**2.5 Experimental Procedure**

For this assessment, the subjects were seated in a chair directly in front of the screen, at a distance of 5H (5 times the vertical height of the screen). A pair of speakers, one either side of the screen were used to playback the audio commentaries.

The room in which the subjects were seated was an acoustically treated environment with an ambient noise level of less than 30 dBA, to minimise noise masking effects. The ambient light level at the viewing position (with the screen off) was 36 lux. The screen was calibrated using an intensity probe while producing a reference grey level of 57 cd/m$^2$.

All the subjects were given the same instructions at the start of the experiment. They were told that they were to watch a series of Virtual Reality Fly-Throughs and at the end of each fly-through they would be expected to give their opinion of its quality. Depending whether the session was for audio or overall quality, the wording of the question was:

[1] The algoritm used was taken from the ITU Software tools library and applied directly to the commentries sampled at 44.1kHz. Strictly speaking, this is outside the design limits of the software, which has been designed for use with speech sampled at 8 or 16kHz. However it still provides an algorithmically defined degradation which was all that was required for this experiment



*Your opinion of the overall quality of the clip*
*Your opinion of the audio quality of the clip*

For both of these questions, the permitted answers were those used in the ITU Recommendation P.800 [15] for the Quality Scale:

*Excellent*
*Good*
*Fair*
*Poor*
*Bad*

The questions and the permitted answers were displayed on the screen after each clip had finished, and the subject spoke their opinion which was then entered by the operator into the control computer (and visually confirmed to the subject). The first four presentations were preliminaries to acclimatise the subjects to the experiment and provide a constant conditioning sequence.

**2.6 Experiment results**

The first stage in reviewing the results of a subjective experiment is to review their statistical validity. As we used a balanced design methodology, the use of analysis of variance (ANOVA) is ideal, as this allows us to estimate the experimental error and determine which of the factors within the experiment (such as subjects, degradations, presentation orders, etc.) have a significant effect on the results.

Once the analysis of variance has been completed, the scores can be averaged across the factors to yield Mean Opinion Scores (MOS's) within any constraints dictated by the ANOVA process.

**2.6.1 Analysis of Variance**

An analysis of variance was performed on the results. The key points from which were:

- Subject variations were significant. This is an expected effect, due to the random sampling from a larger population. Although the fact should be noted, scores can still be averaged across subjects if they are considered, as is the case here, to be a representative sample of the larger population of interest.
- The order in which the subjects saw the material and the degradations did not have a significant bearing on the results. This was as hoped, as the order in which each subject saw the material and degradations was both unique and randomised to minimise this effect.
- The variations caused by the degradations significantly affected the results. This was expected, as it is these variations which the experiment was designed to investigate.
- The question asked (audio or video quality) had a significant effect on the resulting scores, as expected.
- The ordering of the questions (whether a subject did the audio or the overall question first) also significantly affects the results.
- The material to which the degradations were applied also significantly affected the scores. This was an expected effect, as the video material was computer generated and some of the clips contained more artefacts due to this process than others. Also, the use of the two different talkers will account for some of this variation. This effect was small compared to the variations due to the

degradations, so separate conclusions need not be drawn for each piece of source material.

### 2.6.2 Discussion of Results

For the video degradations chosen, noise was considered to be the worst of the three degradation types, with it typically scoring, at its least severe level, similar scores to the other two degradations at their most severe levels. The responses for the other two degradations were found to be broadly similar to each other. An additional point of note is that the highest level of noise yielded worse scores than when no video was present at all. This effect may be due to the fact that no specific guidance was given to the subjects on how to interpret the situation where no video was present, leading to some subjects scoring it higher than others.

For the audio degradations, the order of preference (best to worst) was undegraded, band-limited, MNRU, followed by the combination of both these degradations.

It is important to note that these findings do not imply that certain types of degradation are generally considered worse than others, as selecting a different severity of each of the degradations would have produced different orderings.

When no video is present, the perceived audio quality is always worsé than if video is present.

For the audio question, the effect of whether the question was asked in the first or second session has little effect on the results. However for the overall question, the scores are higher if the overall question is asked in the first. The notable exception to this is when no video is present, where the scores are lower.

Figure 3 summarises the effect of varying the video quality on the perception of the audio quality. It is important to note when considering this that the size of this effect is influenced by the accompanying audio quality, but from the data obtained in this experiment, it was not possible to attribute this to any systematic cause. Despite this, the graph is averaged over all the audio qualities for clarity. To illustrate the effect of audio quality, the same data broken down for each of the audio degradations is shown in Figure 4 below.

The differences shown are small, but the trend in the data, as summarised in Figure 4 is consistent across the majority of the video degradations. We have therefore concluded that within the constraints of this experiment, a decrease in the visual quality had a detrimental effect on the perceived audio quality.

### 2.7 Experiment review

In the experiment reported, each opinion question (either overall or audio) was asked in separate sessions, hence the subjects knew whilst viewing the clip which aspect of performance they would be asked to judge. This was chosen since we required to determine audio perception when audio is known to be the stimulus of interest. It should be noted however that randomising the questions and not presenting them to the subjects until they had finished viewing the clips would be expected to alter the audio/video dependency measured, since the subjects would have had to pay attention to both audio and video if they did not know the nature of the question.

Using a less passive, more involving task, where the subject is not able to consciously consider the audio and video performance until they are required to vote would also be likely to alter the judgements of the subjects.

Further experimentation including questions relating to the perceived video quality are planned to enable more information on the relationships between overall perceptions and those of the individual senses to be established.



## 3.0 Multi-sensory perceptual model

An algorithmic model relating audio and video perception is required as part of the basis for a multi-sensory perceptual model. In addition, descriptions of audio and video errors are required which will allow different aspects of error subjectivity to be emphasised. This emphasis depends on cross-modal effects, the task undertaken, and high level preferences. In order to develop a practical model the individual sensory models are combined using a model of cross modal effects and a task model for a number of scenarios of interest.

### 3.1 Auditory and visual perceptual models

The concept of perceptual models consisting of sensory and perceptual layers has been introduced. The fact that the error descriptors can be sensitive to different aspects of subjectivity has been previously shown by fitting to different opinion scales. This result, together with laboratory experience, is taken to indicate that it is possible to weight a set of error descriptors to describe a range of error subjectivity since different features of the error are dominant for *quality* and *effort* opinion scales.

The gross dependencies between audio and video perception are being quantified with subjective tests, such as the one reported here, so that the gross inter-sensory effects can be included at a low level within the model.

### 3.2 Combined model architecture

Components are:

- sensory models,
- cross-modal model,
- scenario specific task model,

with an average across the target user group used for higher level perceptual issues such as personal preference and user expectation.

The main cross modal effects are:

- timing:
  - sequencing
  - synchronisation
- quality match/mismatch

Error subjectivity also depends on task

- high level cognitive preconceptions associated with task
- attention split
- degree of stress introduced by the task
- experience of user (novice vs. expert)

The advantageous architecture of the BT Labs perceptual models lends itself to multi-sensory combination due to explicit multidimensional description of error subjectivity. Figure 5 shows an outline model architecture combining individual and cross-modal model elements together with the influence of particular task scenarios.

$E_{da1}, E_{da2}, \ldots, E_{dan}$ are the audio error descriptors, and
$E_{dv1}, E_{dv2}, \ldots, E_{dvn}$ are the video error descriptors. Then, for a given task:

$fn_{aws}$ is the weighted function to calculated audio error,
$fn_{vws}$ is the weighted function to calculate video error subjectivity, and
$fn_{pm}$ is the cross-modal combining function.

The task related perceived performance metric, TRPM, is then:

$$TRPM = fn_{pm} \{ fn_{aws} \{ E_{da1}, E_{da2}, \ldots, E_{dan} \}, fn_{vws} \{ E_{dv1}, E_{dv2}, \ldots, E_{dvn} \} \}$$

## 4.0 Discussion and Conclusions

This paper has introduced the requirement for a multi-sensory perceptual analysis for the objective assessment of MM systems. Part of the requirement for such an analysis is knowledge of the dependencies which exist between audio and video components. The important influence of higher level perceptual factors such as task and utility is also recognised.

Existing work on single sense perceptual analysis is summarised and attention is drawn to the advantageous nature of the BT Labs perceptual models which provide error descriptors which may be mapped to a variety of opinion scales. Knowledge of the relationship between audio and video perception, as well as specific task related influence on error subjectivity is required to complete the multi-sensory model. A subjective experiment to investigate aspects of inter-sensory dependency is reported and results shown which reveal a falling perception of audio quality with deteriorating video quality in certain circumstances.

The size and complexity of the overall modelling task is discussed and a strategy proposed to render it tractable. Gross inter-sensory dependency can be modelled and included in a multi-modal model while aspects of higher level task dependency are included as scenario specific weightings. An outline model structure incorporating these concepts is proposed.

In developing objective techniques to assess multi-media systems it is necessary to investigate and model the dependencies which exist between audio and visual perception. These complicated interactions can be rendered more manageable by constraining the scope of the investigation to a small number of relevant application scenarios. As with audio only models there are gross effects which can be measured modelled and exploited ahead of a complete description. At the time of writing a further series of experiments on talking heads is being conducted. This will allow the A/V relationships with visual speech to be compared with those for non-talking head material (within the experimental paradigm chosen).

Conclusions which can be drawn from the experiment reported include the deterioration of perceived audio performance with deteriorating video and the large influence of video performance on the subjects overall opinion. The relatively small influence of video quality on perceived audio performance in certain circumstances is a useful result since it provides an indication of where predictions from the audio only model will continue to provide an adequate prediction.

The architecture of BT Lab.'s perceptual models lends itself to multi-sensory combination due to explicit multidimensional description of error subjectivity. An outline of a multi-sensory model is proposed and further development is planned. What is apparent is that for a majority of applications both in the communications and entertainment industry separate evaluation of audio or video quality is likely to become of limited value.

## 5.0 Acknowledgements

The authors would like to thank their colleagues at BT Labs for their encouragement and support. Particular thanks are due to Alex Bourret for his work on the visual perceptual model, and Ana Belen Sanz-Duque for her work on multi-modal subjective experiments.



## 6.0 References

[1] Paillard B, Mabilleau P, Morissette S, Soumagne J, "PERCEVAL: Perceptual Evaluation of the Quality of Audio Systems.", J. Audio Eng. Soc., Vol.40, No.1/2, Jan/Feb 1992.

[2] Hollier M P, Hawksford M O, Guard D R, "Characterisation of Communications Systems Using a Speech-Like Test Stimulus", J. Audio Eng. Soc., Vol.41, No.12, December 1993.

[3] Beerends J, Stemerdink J, "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation", J. Audio Eng. Soc., Vol.40, No.12, December 1992.

[4] Watson A B, Solomon J A, "Contrast gain control model fits masking data". ARVO,. 1995.

[5] Ran X, Farvadin N, "A perceptualy motivated three-component image model- Part I: Description of the model", IEEE transactions on image processing, Vol.4, No.4 April 1995

[6] Hollier M P, Hawksford M O, Guard D R, "Objective Perceptual Analysis: Comparing the audible performance of data reduction schemes", Presented to the 96th AES Convention in Amsterdam, Preprint No.3879, February 1994.

[7] Hollier M P, Hawksford M O, Guard D R, "Error-activity and error entropy as a measure of psychoacoustic significance in the perceptual domain", IEE Proc.-Vis. Image Signal Process., Vol.141, No.3, June 1994.

[8] Wang S, Sekey A, Gersho A, "An Objective Measure for Predicting Subjective Quality of Speech Coders", IEEE J. on Selected areas in Communications, Vol.10, No.5, June 1992.

[9] Hollier M P, Sheppard P J, "Objective speech quality assessment: towards an engineering metric ", Presented at the 100th AES Convention in Copenhagen, Preprint No.4242, May 1996.

[10] May J, Barnard P, "Cinematography and interface design", in K. Norbdy et al Human Computer Interaction, Interact '95 (26-31), 1995.

[11] McGurk H, MacDonald J, "Hearing lips and seeing voices", Nature, 264 (510-518), 1976

[12] O'Leary A, Rhodes G, "Cross-modal effects on visual and auditory perception", Perception and psychophysics, 35 (565-569), 1984.

[13] ITU P.930 "Principles of a reference impairment system for video", August 1996

[14] ITU-T Recommendation P.810 "Modulated noise reference unit (MNRU)", Feb. 1996.

[15] ITU-T Recommendation P.800 "Methods for subjective determination of transmission quality". Note that this recommendation was previously numbered as P.80. Aug. 1996

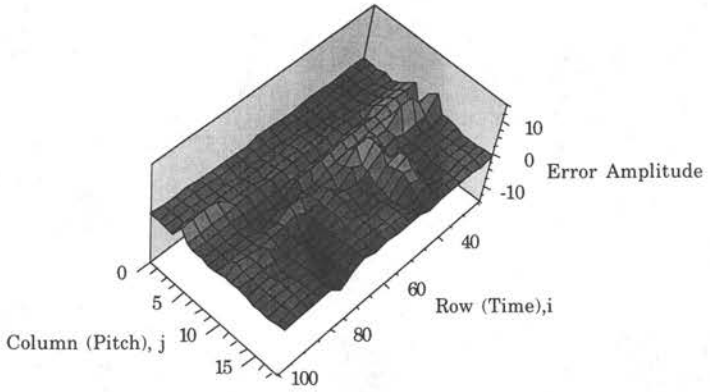


Figure 1, Fragment of audible error surface.

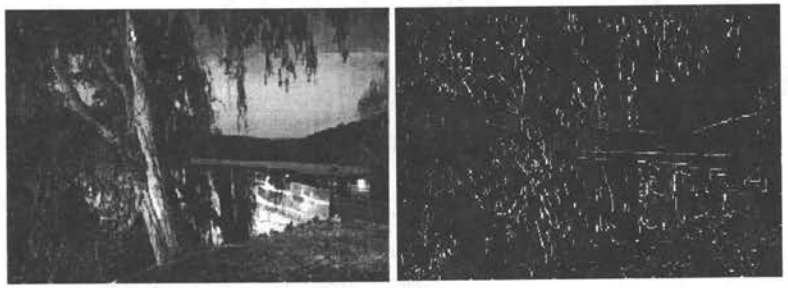

Figure 2, Image decomposition for error subjectivity prediction.

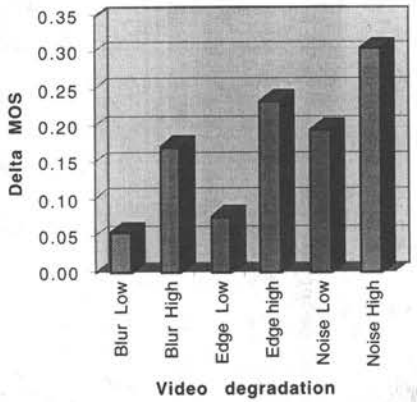


Figure 3, Differences between MOS for degraded video and for undegraded video, averaged over all audio conditions. The higher the bar, the greater this difference.

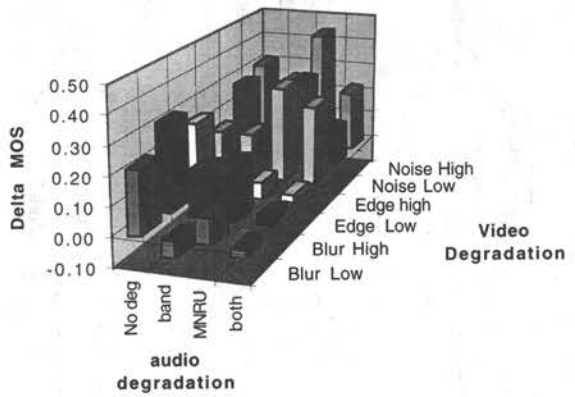


Figure 4, Differences between MOS for degraded video and for undegraded video, shown separately for each audio degradation. The higher the bar, the greater this difference.

Figure 5, Overview of proposed multi-sensory perceptual model.