

HEALTH TECHNOLOGY ASSESSMENT IN DIGITAL TELE-ECHOCARDIOGRAPHY: VALIDATION OF AN AUTOMATIC TOOL FOR THE ASSESSMENT OF IMAGE QUALITY

D. Giansanti, S. Morelli, F. Martelli, G. Maccioni, A. Giordano*,
V. Macellari

Dipartimento di Tecnologie e Salute - Istituto Superiore di Sanità
Viale Regina Elena, 299 – 00161 Roma

*Fondazione S. MAUGERI IRCCS - Servizio di Bioingegneria
Via Revislate 13-28010 Veruno, Novara



INTRODUCTION

The applications of **tele-echocardiography (T-E)** are increasing due to the better availability of bandwidth, reduced costs and improvements generally in information technologies.

Two different approaches can be used, depending on the medical application:

1. **Store and forward transmission.** It can provide high fidelity video clips, but is not applicable to neonatal T E where the need for medical interaction requires real-time T E.
2. **Real-time T-E.** Serious limitations in paediatrics where high fidelity is an essential aspect. Much T E work is conducted in real time, using videoconferencing equipment to transmit video from an ultrasound machine to remote nodes for diagnosis. This process implies the compression of the original video signals and hence causes a degradation of image quality.



INTRODUCTION

The **evaluation of image quality** is fundamental to the assessment of diagnostic accuracy.

In previous works the authors introduced a complex protocol for assessing the image quality, based on:

- (1) a simulation of the \bar{T} Eprocess;
- (2) a phantom based analysis;
- (3) an investigation based on the more common quantitative parameters;
- (4) a use of the Receiver Operator Characteristic (ROC).

But the routine use of this protocol had limitations because of the high time and cost involved.



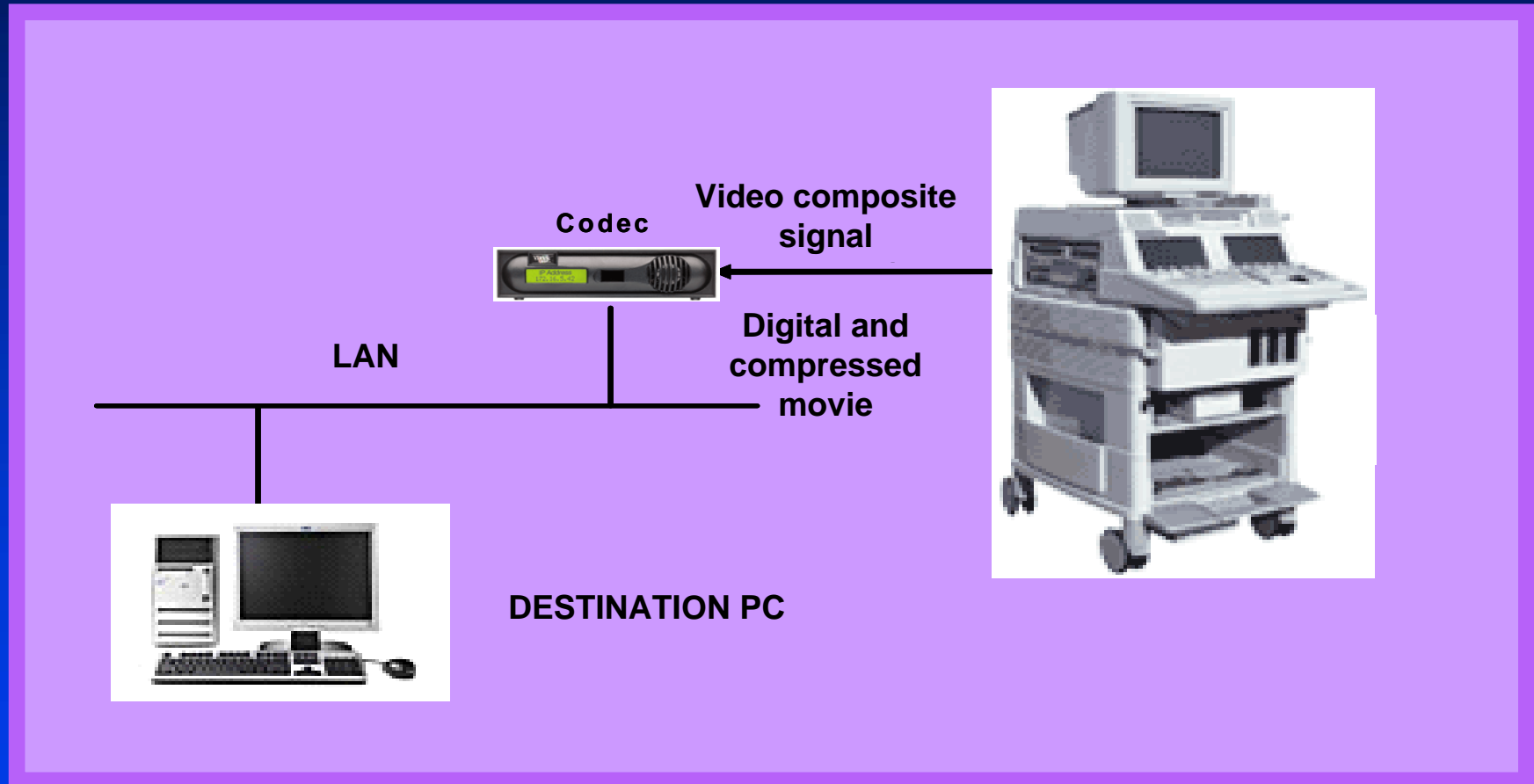
AIM OF THE WORK

- ✓ **The next step was the investigation of lower cost and faster methods of assessing image quality.**
 - ✓ The Institute of Telecommunication Sciences has developed a tool that provides an objective measurement of image quality that can be applied to digital video systems: the NTIA Video Quality Metric (VQM) tool.
 - ✓ It has been adopted as an ANSI standard T1.801.03 2003 and has been extensively tested on a wide range of video systems and bandwidths. The NTIA VQM has been validated by subjective assessments of multimedia applications using non expert viewers judging scenes from typical television programs.
- ✓ **The aim of the present study was to validate the use of the NTIA VQM for assessing the degradation in T-E.**



METHODS

Different video recordings from an echocardiographic examination were transmitted using commercial equipment (Vbrick codec, Vbrick systems Inc, USA) over a local area network (LAN).



METHODS

Different compression schemes from MPEG 1 to MPEG 4, and different bit rates from 1000 to 4500 kbit/s.

The video streams had a CIF format of 352x288 pixels and were transmitted and received on remote stations connected to the LAN.

Clip name	Type	Bit rate (kbit/s)	Duration (s)
MP1_1000_X.mp	MPEG-1	1000	72.89
MP1_1500_X.mp	MPEG-1	1500	69.97
MP2_4500_X.mp	MPEG-2	4500	13
MP4_1000_X.mp	MPEG-4	1000	111.08
MP4_600_X.mp	MPEG-4	600	78.32



METHODS

Three methodologies were used to compare the transmitted and received video sequences to evaluate the **video quality**.

- 1. Peak signal noise (PSNR):** a well accepted method of measuring picture degradation for the objective evaluation of the video quality.
- 2. Double Stimulus Impairment Scale (DSIS):** defined by the ITU for the subjective evaluation of the video quality.
- 3. NTIA General Model:** a model proposed by the National Telecommunications and Information Administration and implemented in a software tool mimicking the human perception for the automatic measurement of the video quality.



METHODS

Peak signal noise (PSNR)

The PSNR was calculated to assess the quality of the reconstruction between *transmitted video clip K* and the *received clip H*:

$$\text{PSNR} = 20 \log_{10} \frac{\max(K)}{\sqrt{\text{MSE}}}$$

$$\text{MSE} = \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n [H(i, j) - K(i, j)]^2$$

METHODS

Double Stimulus Impairment Scale (DSIS)

This method for evaluating the video quality was performed using 15 untrained observers, who graded the impairment perceived using a five-point scale (5=imperceptible to 1=very annoying). The grades were averaged over the observers to get a single subjective score for each clip.

The following five-grade impairment scale has been used:

- 5 – imperceptible
- 4 - perceptible, but not annoying
- 3 - slightly annoying
- 2 - annoying
- 1 - very annoying



METHODS

NTIA General Model

- The “**NTIA General Model**” was one of the five models implemented by the Video Quality Measurement software tool. The “NTIA VQM” is intended as the NTIA Video Quality Measurement software tool.
- The NTIA General Model was applied to obtain an *unsupervised quality parameter*.
- The NTIA VQM has been developed to achieve **maximum correlation with subjective scores** while using objective parameters for measuring the perceptual effects of a wide range of impairments (blurring, block distortion, unnatural motion, noise, and error blocks).



METHODS

NTIA General Model

- This model consists of a linear combination of *seven video-quality parameters* chosen to take into account
 - the decrease or loss of spatial information
 - shift of edges from horizontal and vertical orientation to diagonal orientation and vice versa
 - changes in the spread of the distribution of color samples
 - edge sharpening or enhancements
 - moving edge impairments
 - localized color impairments

- It provides a number ranging from 0 to 1:
 - *0= no degradation*
 - *1= maximum degradation*



RESULTS

From MP1 videoclip

Clip name	PSNR (db)	DSIS (1..5)	NTIA VQM
MP1_1000_A.mp	14.48	3.07	0.786
MP1_1000_B.mp	14.45	3.11	0.792
MP1_1000_C.mp	14.44	3.13	0.795
MP1_1000_D.mp	14.46	3.08	0.791
MP1_1000_E.mp	14.47	3.09	0.792
MP1_1500_A.mp	14.46	2.83	0.801
MP1_1500_B.mp	14.47	2.82	0.803
MP1_1500_C.mp	14.46	2.84	0.805
MP1_1500_D.mp	14.45	2.83	0.802
MP1_1500_E.mp	14.43	2.81	0.809



RESULTS

From MP2 videoclip

Clip name	PSNR (db)	DSIS (1..5)	NTIA VQM
MP2_4500_A.mp	14.72	4.02	0.812
MP2_4500_B.mp	14.75	4.05	0.811
MP2_4500_C.mp	14.69	4.09	0.813
MP2_4500_D.mp	14.71	4.10	0.817
MP2_4500_E.mp	14.70	4.01	0.818



RESULTS

From MP4 videoclip

Clip name	PSNR (db)	DSIS (1..5)	NTIA VQM
MP4_1000_A.mp	14.47	4.53	0.841
MP4_1000_B.mp	14.46	4.51	0.840
MP4_1000_C.mp	14.45	4.50	0.839
MP4_1000_D.mp	14.42	4.49	0.838
MP4_1000_E.mp	14.47	4.43	0.842
MP4_600_A.mp	14.49	3.81	0.848
MP4_600_B.mp	14.48	3.85	0.847
MP4_600_C.mp	14.50	3.83	0.846
MP4_600_D.mp	14.51	3.85	0.844
MP4_600_E.mp	14.52	3.82	0.849



RESULTS

The **PSNR** values (objective parameter) between the best and worse clip are too close to reflect some correlation with the subjective DSIS test.

Thus from an analytical point of a view this parameter (at least in this case of T-E) shows *a very poor usefulness*.

The more effective methodologies for the assessment of the diagnostic accuracy in T-E are those ones based on the *human eye*.

This is particularly confirmed considering that the **DSIS** shows a *very high acceptance for* the clips with a higher and better refined MPEG version (MPEG4) or transmitted with a high bandwidth.



RESULTS

We have also investigated the statistical correlation between the three different methodologies using the *Pearson index*.

The significance (assessed by ANOVA repeated measures) for each group of five sequences was very high.

Clip name	P (PSNR) %	P (DSIS) %	P (NTIA VQM) %
MP1_1000.mp (5 sequences)	0.7	0.7	0.3
MP1_1500.mp (5 sequences)	0.7	0.8	0.4
MP2_4500.mp (5 sequences)	0.8	0.9	0.5
MP4_1000.mp (5 sequences)	0.8	0.8	0.5
MP4_600.mp (5 sequences)	0.9	0.8	0.4



RESULTS

In particular a better value of significance was found for the **NTIA VQM**, confirming the point of force of the methodology: *an objective tool using subjective models which shows the better statistic significance.*

In addition we found a high correlation between the **DSIS** (subjective) and the **NTIA VQM**.

Pearson's index	PSNR	DSIS	NTIA VQM
PSNR	1.00	0.78	0.75
DSIS	0.78	1.00	0.94
NTIA VQM	0.75	0.94	1.00



DISCUSSION

- The analysis of T-E video quality demonstrated that the most grading procedures were the **subjective DSIS** ones.
- The comparison of the **NTIA VQM** used as an AT versus the **DSIS** methodology showed a **high correlation**. ANOVA tests conducted on the complete data-set showed a high significance ($P < 0.01$).



DISCUSSION

- The NTIA VQM could thus be an effective tool to evaluate telemedicine T E transmission systems in terms of image quality, avoiding the costs and times associated to the set up of a number of exercises for subjective tests in repeatable conditions.
- Furthermore this tool could be a solution to the lacking reference tool to be used to test T E applications in the phases of acceptance as a routine methodology in the National Health Care Systems NHCSs.



DISCUSSION

Significance of the research

Validation of a novel commercial tool (NTIA VQM) for assessing the image quality in T E, with two important characteristics:

the tool is automatic, objective and furnishes quantitative data
easy methodology that does not require subjects to perform tests, and thus saves cost and time

the tool embeds subjective human models
like subjective response to tests



DISCUSSION

Future work

- The T E is first of all a complex and heterogeneous telemedicine system.
- The assessment of image quality is a basic matter, even if obviously it is not the only one to be considered.

The next phase of research will consider the insertion of this test in the HTA methodology in telemedicine in order to simplify the steps dedicated to the tests of diagnostic accuracy of an application of T-E.

