ETSI TR 101 565 V1.1.1 (2011-09)

**Speech and multimedia Transmission Quality (STQ);
Guidelines and results of video quality analysis
in the context of Benchmark and Plugtests
for multiplay services**

Reference
DTR/STQ-00185

Keywords
analysis, quality, video

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://ipr.etsi.org).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

1 Scope

The present document presents the results obtained on technological platform where Triple Play offers are available. These results concern the quality evaluation of IPTV video streams produce by the offers available on the platform.

The determinate indicators are presented in the main part of the document. The results were obtained during a specific test campaign for SD streams analysis.

Note that determinate indicators are presented in ES 202 765-4 [i.1]. So the present document represents an implementation report for some metrics and associated methods defined in ES 202 765-4 [i.1].

The main part of the present document presents the performed indicators and charts used for results presentation.

Annex A presents the methodology implemented in a first series of tests and the results.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are necessary for the application of the present document.

Not applicable.

2.2 Informative references

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI ES 202 765-4: "Speech and multimedia Transmission Quality (STQ); QoS and network performance metrics and measurement methods; Part 4: Indicators for supervision of Multiplay services".

[i.2] ITU-T Recommendation P.505: "One-view visualization of speech quality measurement results".

[i.3] ITU-T Recommendation J.247: "Objective perceptual multimedia video quality measurement in the presence of a full reference".

[i.4] ITU-T Recommendation J.144: "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

platform: premise installed in residential environment where the accesses to different Multi Play offers proposed by ISP on the same country are made available

NOTE: This platform is generally installed in the centre of a city.

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ADSL	Asymmetric Digital Subscriber Line
DVB-T	Digital Video Broadcasting - Terrestrial
HGW	Home GateWay

NOTE: Referenced also as Residential Gateway.

IP	Internet Protocol
IPTV	IP TeleVision

NOTE: System where a digital television service is delivered using Internet Protocol.

ISP	Internet Service Provider
ITU-T	International Telecommunication Union - Telecommunication standardization sector
MOS	Mean Opinion Score
VoIP	Voice over Internet Protocol

4 Context

The deployment of multiplay offers is increasing, so it is important to measure quality performances of services proposed via these multiplay offers.

Concerning performance evaluation of multiplay offers, ES 202 765-4 [i.1] lists indicators and presents methodologies for quality characterization in a context of end-user. The present document gives practical requirements of use in the context of service verification and benchmark.

Multiplay offers developed on IP technology give access to IPTV services. These IPTV services develop in an important way in the residential context. But contrary to the VoIP services, performances of IPTV services of are not really evaluated.

Because there is a necessity of having a performance overview of the service brought to the users, the actors of the domain (operators, benchmakers) began to perform evaluations of IPTV services. In these conditions, the present document presents a benchmark of the video quality evaluated on different TV services.

The present document is intended to provide an overview of the performance of IPTV offers deployed in France (and used by customers).

The core of the report presents determined metrics and results presentation. The results presented in annex A of the present document, concern video quality of IPTV service associated to Triple Play offers implemented on the technological platform installed in a city in France.

The interest of these results is to give an overview of video quality of deployed IPTV services compared to a kind of reference which is the video quality of DVB-T (Digital Video Broadcasting - Terrestrial) service.

5 Platform presentation

The platform is a premise installed in residential environment and in which there are the accesses to different Multi Play offers proposed by ISP on the same country.

The technological platform used for performing video quality assessment is installed in a premise in the city centre of a town counting more than 210 000 residents.

During the second semester of 2010, this platform was characterized by:

- Implementation of 6 offers concerning 6 different ISPs.
- Each offer proposes 3 services: Internet access, VoIP and IPTV.
- Access to services is obtained through a Home GateWay (HGW).
- Access to the network can be ADSL or cable type depending on the ISP.
- Distance between HGW and first digital equipment is about 350 meters (Length of the ADSL line).
- A DVB-T (Digital Video Broadcasting - Terrestrial) access is also available.

So in the platform, there are 7 different TV accesses: 6 IPTV accesses and 1 DVB-T access. Figure 1 shows an overview of the architecture and a synopsis of the platform in the context of TV services.

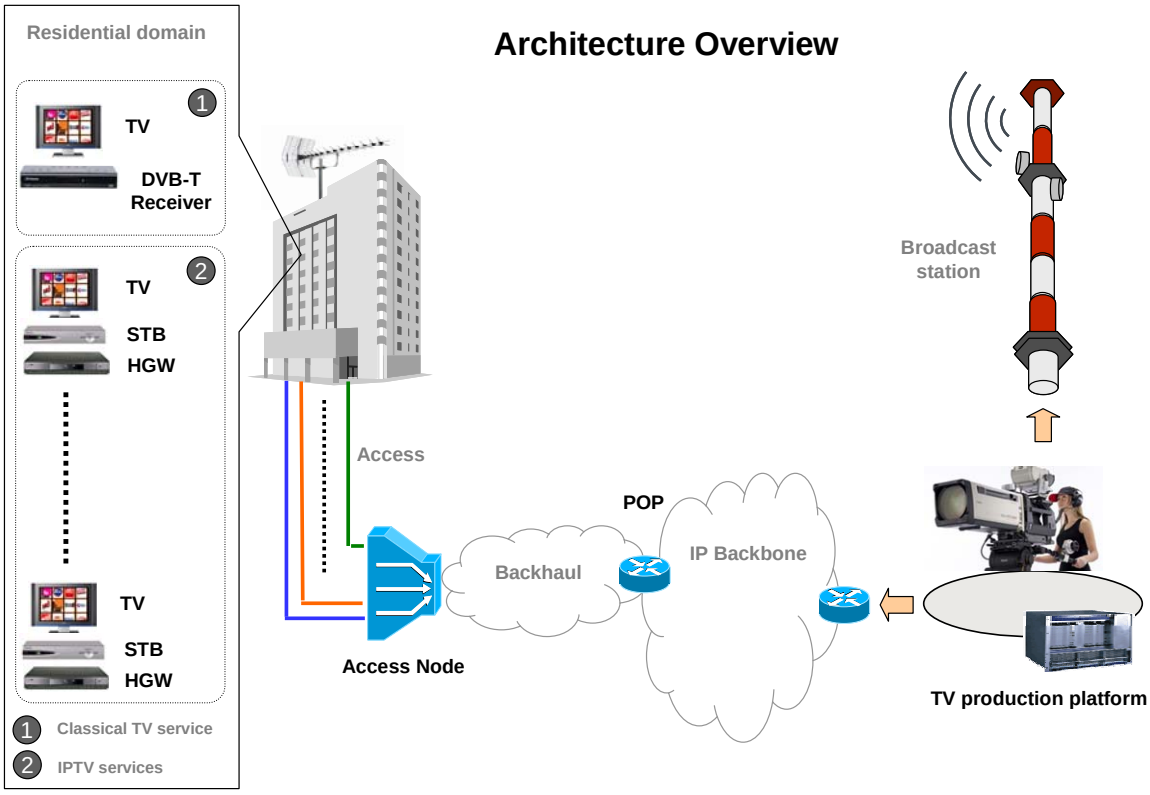


Figure 1: Architecture overview of TV service broadcasting

6 Presentation of test conditions

To assess the video quality of the different TV services, the principles are to analyze several video sequences belonging to different channels.

6.1 Indicator description

To take into account the end user context, only objective measurements models based on No Reference approach correspond to the need. The current issue is that there is no standardized algorithm in that area. The ITU-T Recommendation J.247 [i.3] is not applicable in this context because J.247 is a model functioning with reference. J.144 [i.4] is also not applicable in this context due to these restrictions (synchronisation issues, not application to MPEG4).

For the time being, it is proposed to qualify video quality by the occurrence of particular degradations like "black screen", blockiness and frozen picture.

So in this context, the indicators are the following ones.

6.1.1 "Black Screen" Occurrences

Definition	Currently, a major trouble of IPTV service is the fact of having displayed on the TV set a "black screen". Black screen can outcome from: • encoder / decoder implementation when no video stream is present • a major loss of video packet during a long period of time This metric correspond to the number of black screen sequence during a time period (24 hours, 1 week…). "Black screen" is one of the cases of "isochrominance".
Assessment method	This default is detected mainly by using robots or probes implementing objective video signal measurement algorithms that are able to detect an image fully coloured in black. Currently, this is the most suitable approach so as to perform a consistent signal-based analysis.
Unit	Number or Ratio (number of occurrence by time unit).
Standardization reference	ES 202 765-4 [i.1].
Comment	The duration of the measurement should be greater than the "inter-advertisements" duration, because sometimes the broadcasters insert "black screen" sequences not visible by the users between advertising.

6.1.2 Blockiness Occurrences

Definition	In video and image compression, a common artifact called "Blockiness" comes firstly from low-quality compression when too few bits are used. This artifact can be appeared when packet loss ratio is too high packet loss ratio on the transmission link (operator network, user installation…). Blockiness is an obviously perceptible contrast of color at the boundaries of the encoding blocks with a codec like JPEG or MPEG video. This metric correspond to the number of blockiness sequence during a time period (24 hours, 1 week…). **B-3 block** Group of pels. For example, a block of 8x8 pels is the smallest coding block used in MPEG-1 algorithms. There are 1 320 blocks in a CIF image, 44 in the horizontal direction (352 pels/8) and 36 in the vertical direction (288 lines/8). **B-4 block distortion** Distortion of the image characterized by the appearance of an underlying block encoding structure, also called *tiling*. **E-22 error blocks** A form of *block distortion* where one or more blocks in the image bear no resemblance to the current or previous scene and often contrast greatly with adjacent blocks.
Assessment method	This default is mainly detected by using robots or probes that implement objective video signal measurement algorithms that are able to detect it. Currently, this is the most suitable approach so as to perform a consistent signal-based analysis. These measurements may be done by taking into account the STB integrating error recovery.
Unit	Number or Ratio (number of occurrence by time unit).
Standardization reference	ES 202 765-4 [i.1]
Comment	It is often referred to macro blocking, this occurs when a certain amount of the IPTV streams are unavailable to the Set Top Box at playout time. This is most commonly due to packet loss at some point in the network but could be due to everything from content encoding issues to delay-variations (jitter) as packets arrive too late to the STB.

6.1.3 Frozen Picture Occurrences

Definition	The frozen picture phenomena can be expressed through some pictures appearing as stopped / freezed from time to time on the end-user screen. These freezes may be issued by the decoder or the network. There are usually very annoying for the end-user. This metric correspond to the number of freezed picture sequence during a time period (24 hours, 1 week…).
Assessment method	In practical way, this indicator can be measured by verifying on adjacent image the stability of the luminance and\or the chrominance components and this for all pixels composing an image.
Unit	Number or Ratio (number of occurrence by time unit).
Standardization reference	ES 202 765-4 [i.1].
Comment	As for "black screen" it is needed to make the measurement on a duration greater than 1 second (typically during 8 seconds or more).

6.2 Result presentations

To present results, 2 type of presentation are used in the present document:

- Sector presentation
- Pie diagram

For this first experiment, it is proposed to determine the number of impaired video sequences for each TV service and for each TV program. It is reminded that a video sequence is considered as impaired as soon as one impairment is detected. The number of different impairment types (such as blockiness, frozen and monochromatic screens) is also determined.

As an example for Channel n°2, over the 105 video sequences the following impairments re the following:

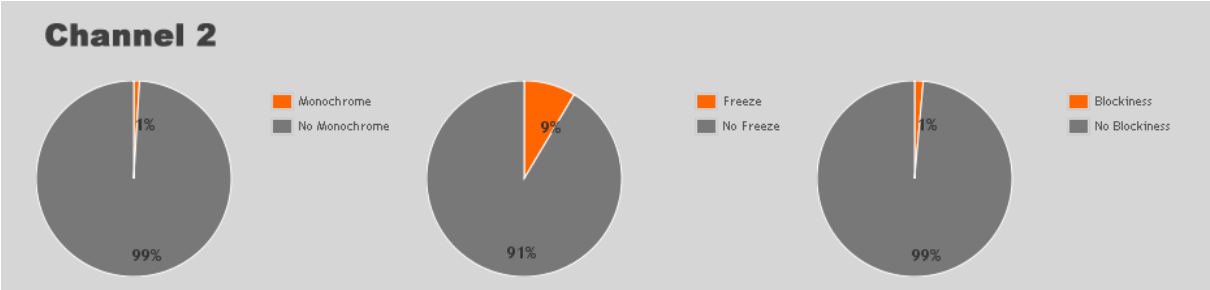
- 1 monochromatic screen
- 9 frozen screens
- 1 blockiness

As a consequence the video quality is defined as 1% of monochromatic screen, 9% of frozen screen and 1% blockiness.

6.2.1 "Sector" presentation

The "sector" display consists in displaying the percentage of impaired and non-impaired video sequences on one circle. The whole display is defined by one circle by impairment.

One example is provided in figure 2.



NOTE: This type of graph presents the ratio of specific degradations (Monochrome screen, Freeze and Blockiness) in video sequences.

Figure 2: Example of Sector presentation: 3 metrics determined on a TV channel

6.2.2 Pie diagram presentation

Another interesting presentation is used in the present document. It is the Pie diagram (ITU-T Recommendation P.505 [i.2]). This type of presentation offers on a single figure an overview of the performances. It is possible to present several metrics on the same graph by maintaining each indicator on its own scale. This type of presentation allows to easily displaying the strengths and weaknesses of each offer. The Pie diagram also allows to easily comparing the offer performances.

Within the framework of this platform, 3 indicators by channel are presented on a Pie Diagram (ITU-T recommendation P.505 [i.2]). These 12 indicators correspond to 3 metrics ("Black Screen" ratio, Blockiness ratio and Frozen Picture ratio) by channel and for 4 TV channels analysed.

These indicators are presented in reference to two arbitrary thresholds: 0 % and 10 %.

The 10 % threshold is represented by a red circle. The indicator value is green above and yellow below the threshold. If the indicator value is between 0 % and 10 %, the indicator is presented in light green. Dark green is used to present indicator characterizing 0 % degradation.

To facilitate the visualization of performances associated to each TV channel, a specific line separates the results by channel.

An example of this type of graph is presented on figure 3.

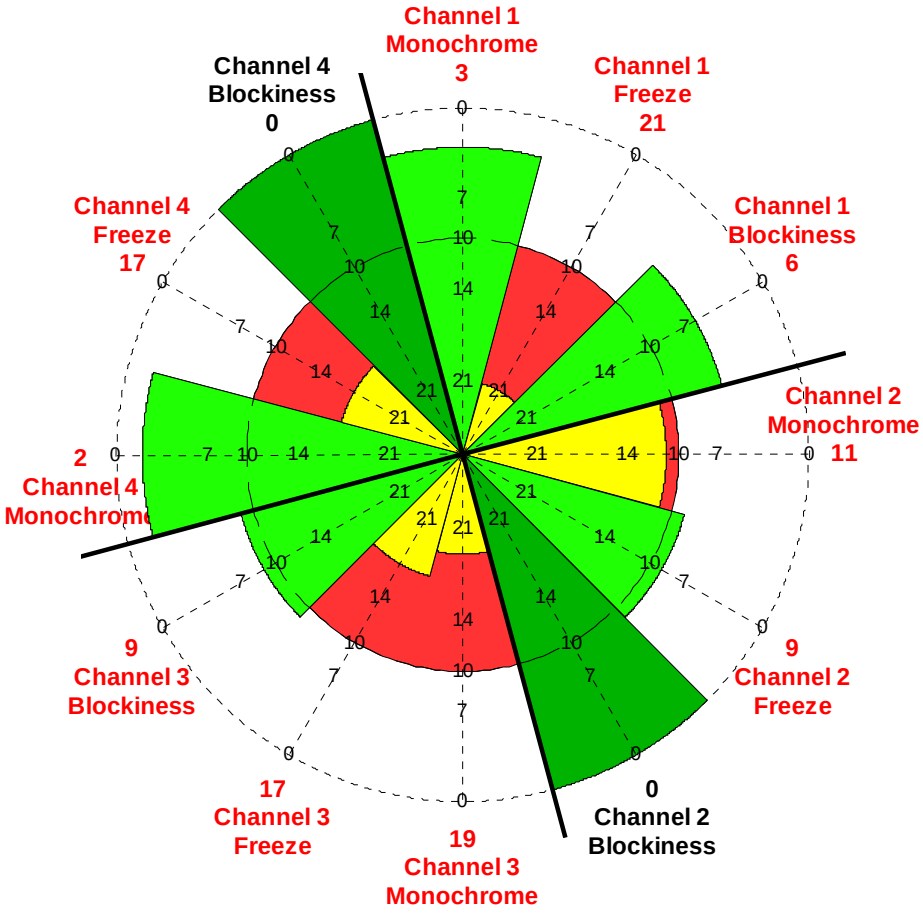


Figure 3: Example of Pie diagram with indicators determined in the context video quality evaluation (3 metrics determined on 4 IPTV channels)

Annex A:
Video quality performance evaluated on different IPTV services - Overview of results obtained at end 2010

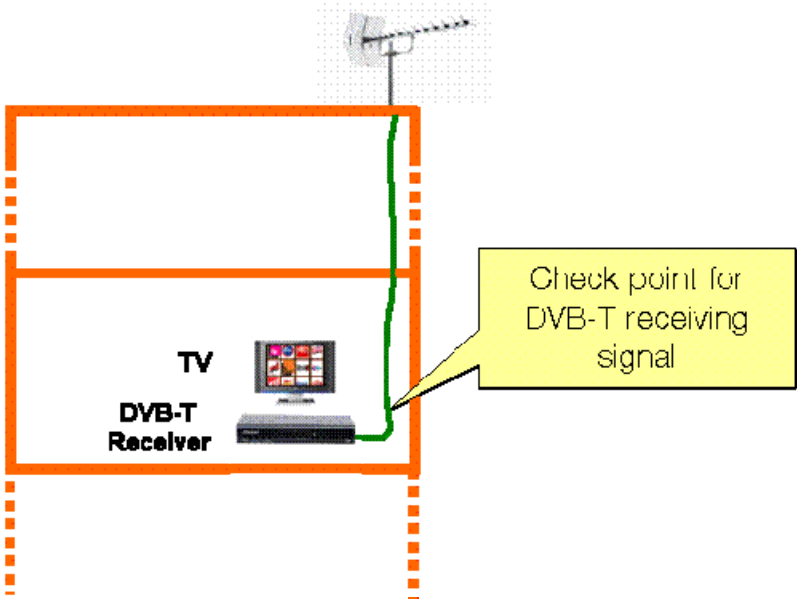
A.1 Platform presentation

On the platform, there are 7 different TV accesses: 6 IPTV accesses and 1 DVB-T access.

The DVB-T access characterization shows that receiving conditions are very good: only 20 km between the broadcast antenna and the platform, high level for the receiving signal into the premise, very low error bit rate…

The DVB-T access characterization is done on the received signal at the antenna output located in the test room. This characterization does not qualify the media flow!

In this condition, TV service over DVB-T access is a sort of comparison point for this benchmark.



NOTE: The characterization of DVB-T access was realized on the receiving signal before the receiver directly at the antenna output.

Figure A.1: In the premise where the technical platform is installed

A.2 Description of the methodology

The equipment needed to conduct such an experiment is rather heavy. So we have only one analyzer to lead this study.

So, the different characterizations are realized in sequence and not in parallel implying to define the whole implementation described below.

A.2.1 General Principles

As it is not possible to make all the tests in parallel, as one of the intention of the experiment is to compare the video quality of all the offers, it is needed to select carefully the time period for the test (corresponding to a rather high percentage of viewers), the most popular video programmes (for children, seniors, workers, non-active...) which have to be provided by all the ISPs and by DVB-T.

These principles are summarized in figure A.2):

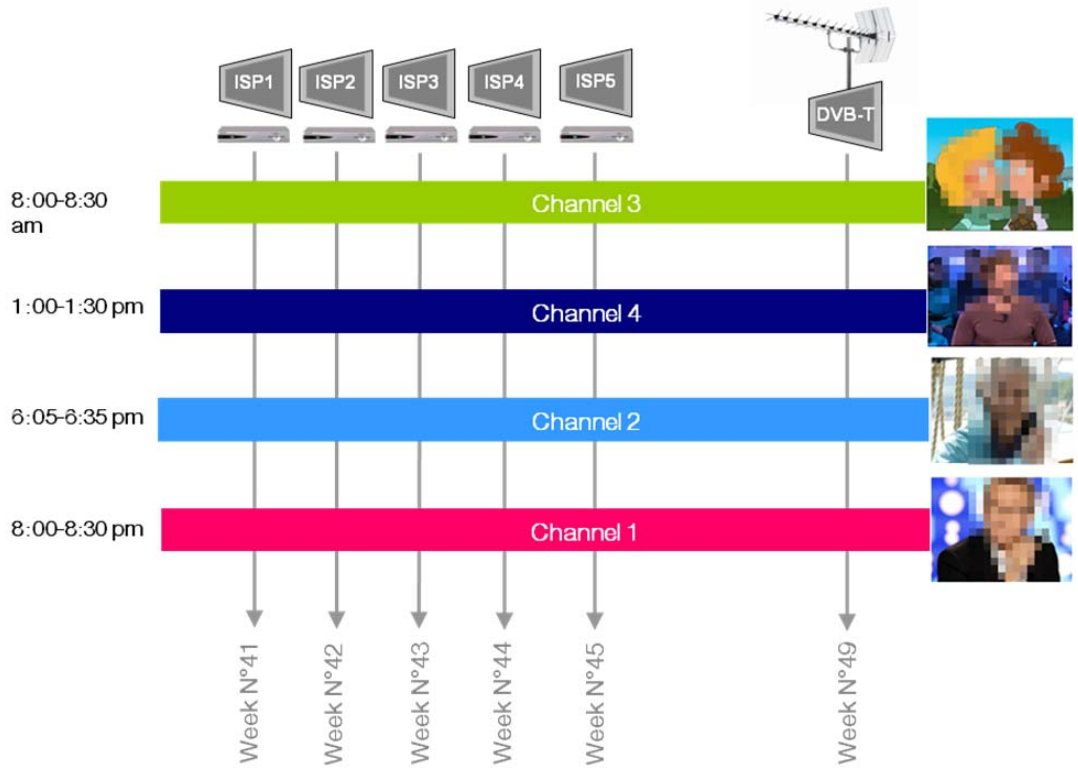


Figure A.2: Principles used to characterize video quality of TV services

Based on the principles described above the choice of the video sequences is described in figure 4.

The channel choice is based on three criteria:

- A selection of TV channels that are known and watched by a lot of people. This selection corresponds to a high representativity based on the "Audimat" percentage of viewers.
- The TV channels are viewed by different age categories such as children, seniors, workers
- The TV channels are all available in the different groups offered by the ISPs and also available in the DVB-T groups.

NOTE: In France "Audimat" is the company in charge of the monitoring of TV channel audience.

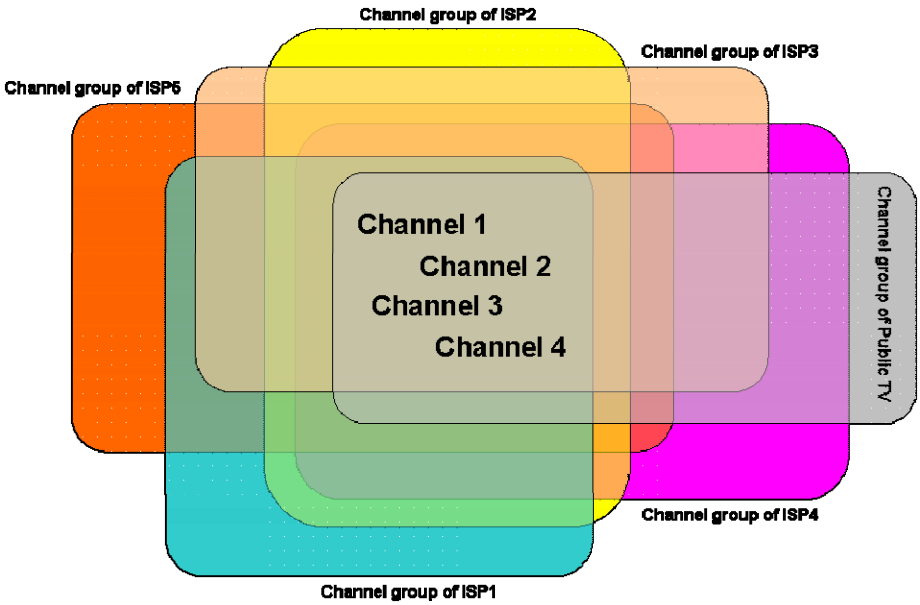


Figure A.3: The 4 channels used for video characterization are common of all IPS channel groups

The TV programs choice and consequently the timeslots choice correspond to the need to have the same kind of contents every day at the same hour. As the analysis is done sequentially for the different ISPs, it is absolutely requested to check that the different analysis is done on similar contents.

For this first experiment the selection, of choice for channels and TV programs are the following:

- The Channel number 1of a public TV provides all types of programs but the choice has been done on the news program at the beginning of the evening. This program is mainly viewed by workers.
- The Channel number 2 of a public TV provides all types of programs but the choice has been done on a game at the end of the afternoon, mainly viewed by seniors.
- The Channel number 3 of a private TV provides programs dedicated to cartoons and is mainly viewed in the beginning of the morning by young children.
- The Channel number 3 of a private TV provides all types of programs but the choice has been done on the news program at lunch time. This program is mainly viewed by people without professional activities and medium age.

Each TV program duration is about 30 / 35 minutes every day.

A.2.2 Technical criteria

It could be relevant to assess the quality for SD and for HD TV.

For this first experiment only the SD (Simple Definition) quality is assessed. So all the TV channels and programs are in SD quality.

To capture the video flow, a HDMI link is used between the STB (for TV services over IP) and the analyzer. We also used a HDMI link (between the DVB-T receiver and the analyzer) to analyze the video flow of the DVB-T service.

It has also been checked that image resolution at the output of the STB is the same as at the output of the DVB-T receiver whatever the ISP and the channel. This resolution is 1 920 x 1 080i for all the ISPs except one which has a smaller resolution (720x576i). This ISP has not been kept in the panel. Finally the tests are made on 5 IPTV services provided by 5 different ISPs and 1 TV service.

It should be noted that for IPTV services 4 of them are IP over an ADSL line and 1 is IP over cable.

A.2.3 Scheduling of analyses

For each TV program the video quality analysis is done on 10 seconds sequences. The test sequences are built as follows:

1) 3 successive video sequences of 10 s are recorded.

2) Each of the 3 recorded sequences is analyzed.

This process 1 and 2 is repeated all along the TV program.

The analysis duration for 3 video sequences of 10 seconds is about 5 minutes. So, every 5 minutes 3 TV sequences are recorded and analyzed.

In order to obtain a sufficient enough number of video sequences analysis, the same TV program is analyzed over several consecutive days (in practice 5 days). At the end, 105 TV sequences are recorded and analyzed for each ISP.

Globally for the 6 TV services (5 IPTV and 1 DVB-T) and for the 4 programs 2 520 video samples have been analyzed.

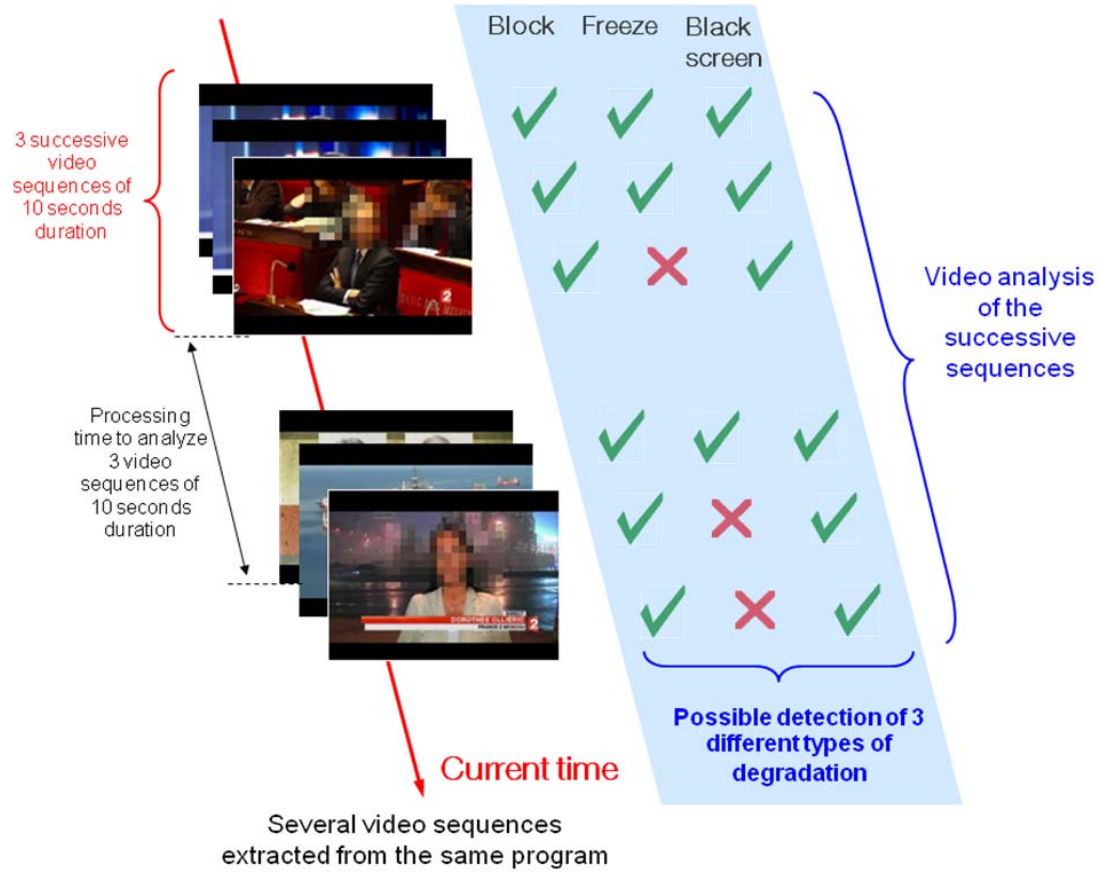


Figure A.4: Analyse scheduling and detection of video quality impairments

As shown on figure A.4 the main objective is to detect the video quality impairments (blockiness, frozen or monochromatic images) on the 10 seconds sequences.

When impairment is detected in a 10 seconds video sequence, the whole sequence is considered as impaired.

A.3 Results obtained by IPTV service of ISP1

For the two metrics "Monochrome" and "blockiness" the percentage of impaired sequences is between 0 % and 1 %, whatever the channels.

The "freeze" metric is detected on the four channels but it is extremely high for channel 3 (94 %) and rather low for channel 1 (5 %). For channels 2 and 4 the percentage of impaired sequences are respectively 12 % and 19 %.

Figure A.5

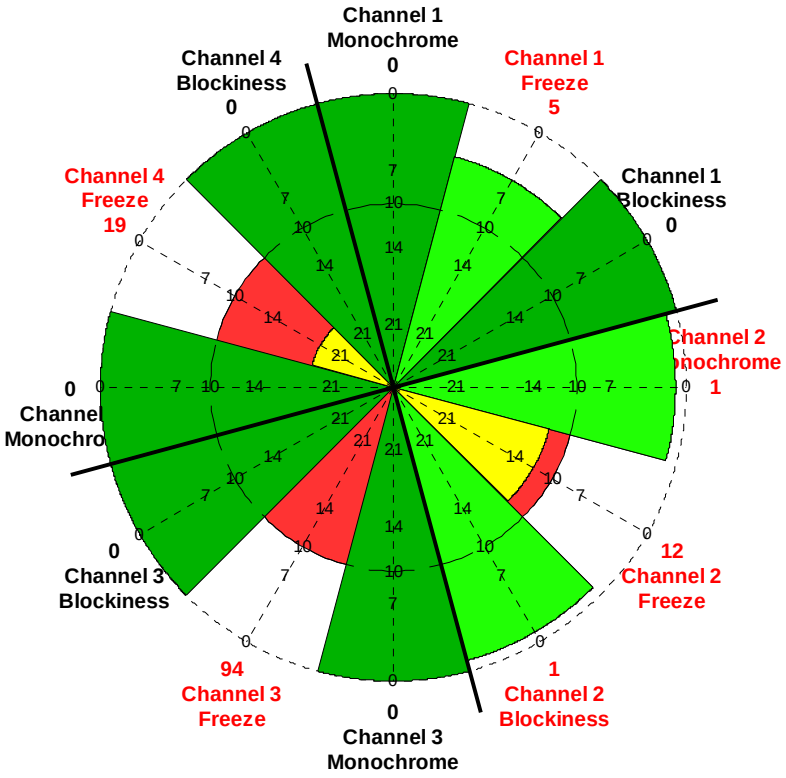


Figure A.6

A.4 Results obtained by IPTV service of ISP2

For the two metrics "Monochrome" and "blockiness" the percentage of impaired sequences is between 0 % and 2 %, whatever the channels.

The "freeze" metric is detected on the four channels but it is high (more than half of the sequences are impaired) for channel 3 (60 %). For the three other channels the percentages of impaired sequences are bigger than 10 %, the channels 2, 1 and 4 the percentage of impaired sequences are respectively 12 %, 13 % and 26 %.

Figure A.7

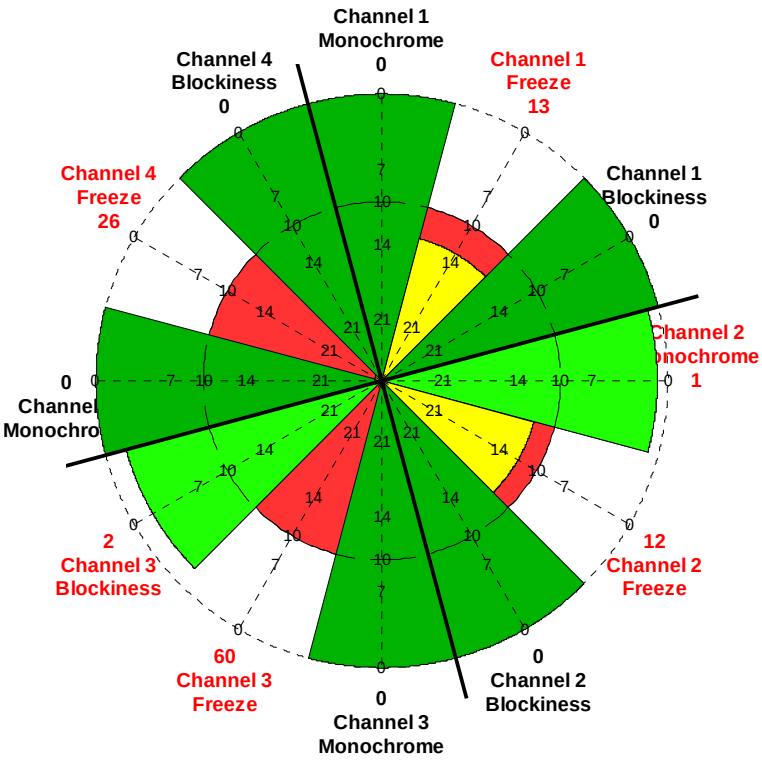


Figure A.8

A.5 Results obtained by IPTV service of ISP3

For the two metrics "Monochrome" and "blockiness" the percentage of impaired sequences is 0 % for the different channels, except for "Monochrome" on Channel 2 which reaches 3 %.

The "freeze" metric is detected on the four channels but it is very high for channel 3 (86 %) and at 10 % for channels 1 and 2 (10 %). For channel 4 the percentage of impaired sequences is 26 %.

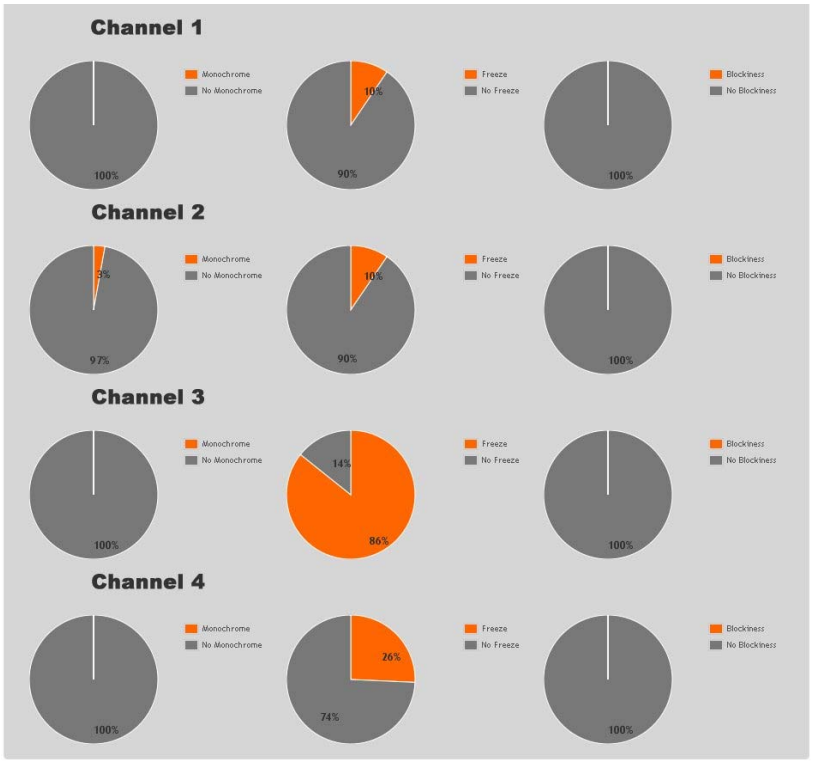


Figure A.9

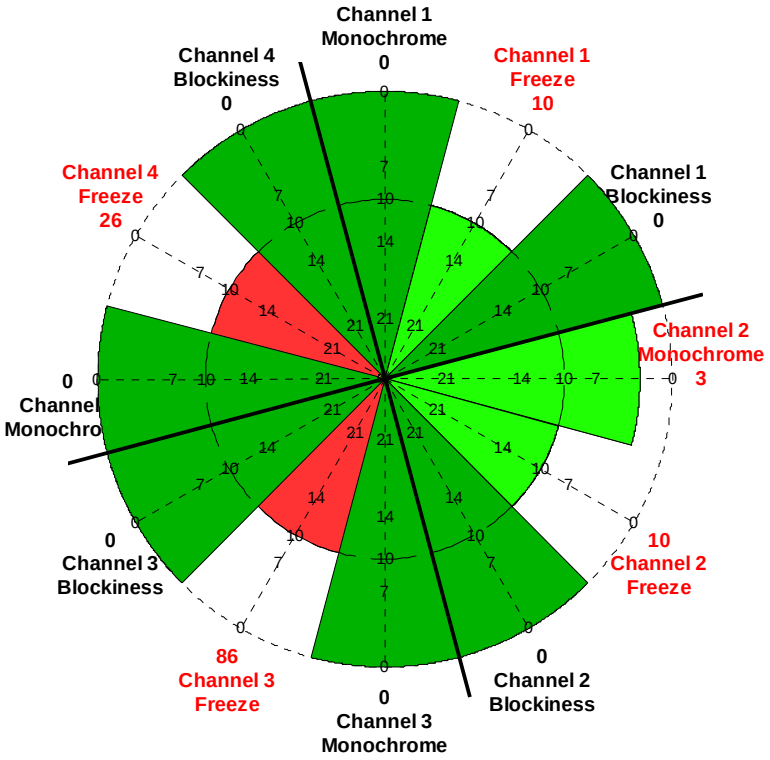


Figure A.10

A.6 Results obtained by IPTV service of ISP4

For the two metrics "Monochrome" and "blockiness" the percentage of impaired sequences is between 0 % and 1 %, whatever the channels.

The "freeze" metric is detected on the four channels but the percentage of impaired sequences is lower than for the other ISPs. For channel 3 more than ¼ of the sequences are impaired (28 %). For the three other channels the percentages of impaired sequences are respectively 4 %, 12 % and 14 % for channels 1, 2 and 4.

Figure A.11

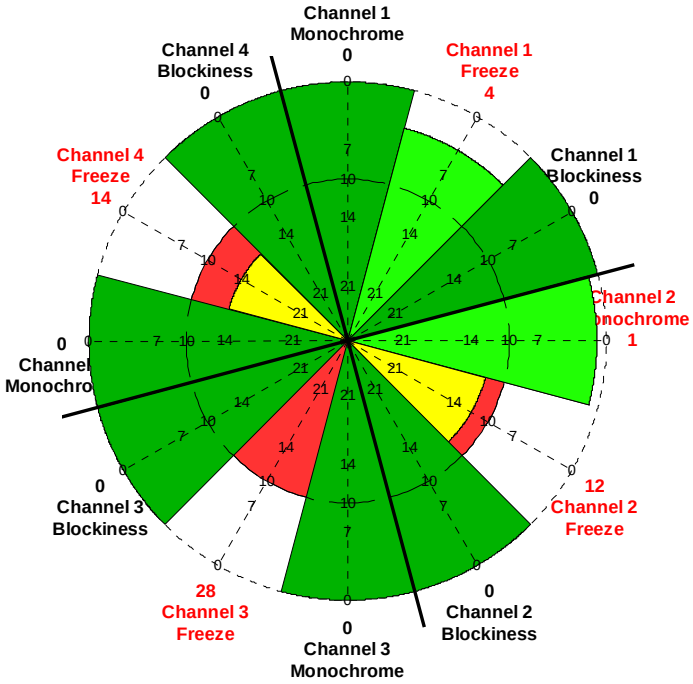


Figure A.12

A.7 Results obtained by IPTV service of ISP5

For the two metrics "Monochrome" and "blockiness" the percentage of impaired sequences is between 0% and 1%, whatever the channels.

For the three other channels the percentages of impaired sequences are respectively 13% for channel 1 and 17 % for channels 2 and 4.

Figure A.13

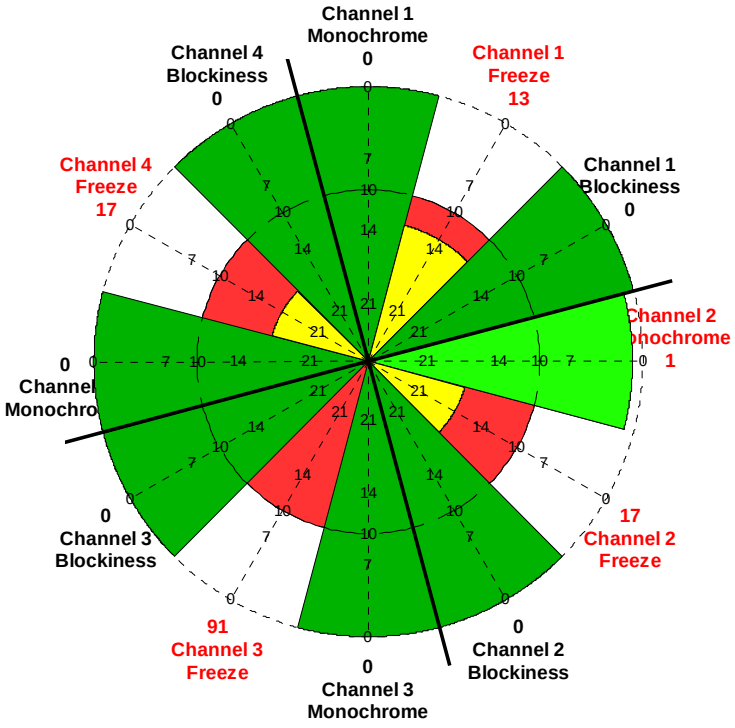


Figure A.14

A.8 Results obtained by TV service of DVB-T

Initially it was intended to use the DVB-T service as the reference for quality. However, the results shown for DVB-T indicate that the three metrics are impaired (except for "blockiness" on channels 2 and 4).

"Monochrome" metric is respectively observed on 1 % of sequences for channels 1 and 2, and on 7 % for channel 3. "Blockiness" metric is observed as 1 % of sequences for channel 1 and as 2 % on channel 3. Even if "freeze" metric does not seem as important as for IPSs it can be seen that the percentage of "freeze" sequences is respectively 4 % for channel 2, 9 % for channel 1, 27 % for channel 3 and 28 % for channel 4.

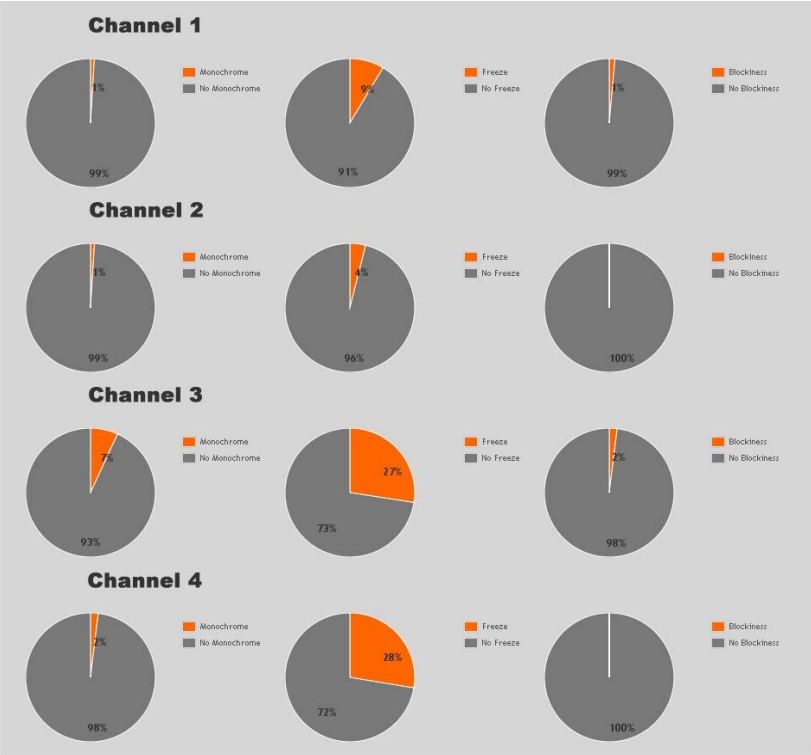


Figure A.15

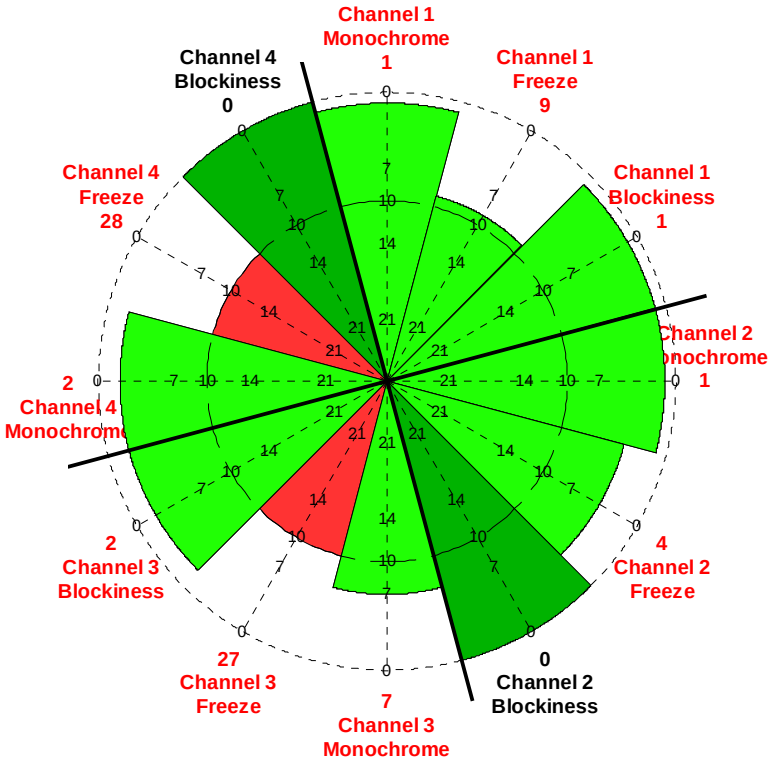


Figure A.16

A.9 Result discussion

This first test campaign of IPTV quality offered by ISPs and compared with DVB-T shows that the three parameters qualifying the video quality may be assessed in "real time" and validates the principles defined in ES 202 765-4 [i.1].

The results on the three metrics ("Monochromatic", "Freeze" and "blockiness") do not give similar results: the metric "Freeze" is the most common detected, the two other metrics being in general rather low for the most of the ISP x channel.

It is also seen that the "Freeze" metric is observed on the different channels and for all the ISPs but it seems that Channel 3 is more affected than the others. As Channel 3 proposes mainly cartoons, it should be needed to check carefully the results: the cartoons are sometimes produced by creating a limited number of pictures, compared to the number of video image. For such programs it will be perhaps needed to define specific parameters (increase the duration criteria for detection).

Globally, ISP 4 provides a lower percentage of frozen pictures.

An observation that was not expected is to discover that DTT also provides a rather large percentage of impaired video sequences. As a global result, if we consider only indicator without degradation, **DVB-T provides the worst quality**.

The different channels are not impaired in a similar way. Channels 1 and 2 provide better quality than channel 4, channel 3 being identified as the poorest quality channel.

History

Document history		
V1.1.1	September 2011	Publication