



# Performance Comparisons of Subjective Quality Assessment Methods

T. Tominaga, T. Hayashi, J. Okamoto, and A. Takahashi  
NTT Service Integration Laboratories, NTT Corporation

VQEG Boulder meeting, January 2010

# Content

---

1. Background
2. Subjective assessment methods
3. Subjective assessment test conditions
  - ▶ Test conditions
  - ▶ Video quality conditions
  - ▶ Video sources
  - ▶ Assessment environment conditions
  - ▶ Experimental design
4. Performance index
5. Test results
6. Conclusions
7. References

# 1. Background

---

- ▶ Discussions of subjective assessment methods in ITU-T.
  - ▶ ACR and ACR-HR (5pt.) were used in VQEG MM test plan (QCIF-VGA).
  - ▶ ACR and ACR-HR (5pt.) have been using in VQEG HDTV test plan.
  - ▶ ACR-HR (11pt.) is planned to be used in VQEG hybrid test plan (QCIF-HD).
  - ▶ ACR5, 9, 11pt. are candidate methods in P.NAMS qualification test (QCIF-HD).
- ▶ It is technically important to properly use the evaluation method according to technical features and differences.

⇒ Subjective quality evaluation methods' characteristics are compared in terms of consistency and sensitivity.

## 2. Subjective Assessment method

---

| Methods | Recommendations              |
|---------|------------------------------|
| DSCQS   | ITU-R BT.500-11              |
| DSIS    | ITU-R BT.500-11, ITU-T P.910 |
| ACR5    | ITU-R BT.500-11, ITU-T P.910 |
| ACR-HR  | ITU-T P.910                  |
| ACR11   | ITU-T P.910                  |
| SAMVIQ  | ITU-R BT.1788                |

# Score calculation

## ▶ Definitions

- ▶ PVS: Processed video sequence
- ▶ REF: Reference video sequence
- ▶ HREF: Hidden Reference video
- ▶  $V(\text{xxx})$ : Score of video(xxx)

| Method    | Score                                  |
|-----------|--|
| DSCQS     | $ V(\text{PVS}) - V(\text{REF}) $      |
| DSIS      | $V(\text{PVS})$                        |
| ACR-HR    | $V(\text{PVS}) - V(\text{HREF}) + 5$   |
| ACR11-HR  | $V(\text{PVS}) - V(\text{HREF}) + 10$  |
| SAMVIQ-HR | $V(\text{PVS}) - V(\text{HREF}) + 100$ |
| ACR       | $V(\text{PVS})$                        |
| ACR11     | $V(\text{PVS})$                        |
| SAMVIQ    | $V(\text{PVS})$                        |

### 3. Subjective assessment test conditions

|                     | Exp. 1  | Exp. 2  | Exp. 3  |
|---------------------|---|---|---|
| Methods             | DSCQS, DSIS, ACR, ACR-11                        | DSCQS, DSIS, SAMVIQ[1], ACR                     | DSCQS, DSIS, SAMVIQ[1], ACR, ACR-11                     |
| Total PVSs          | 56<br>(HRC 8+PVS 48)                            | 48<br>(HRC 8+PVS 40)                            | 42<br>(HRC 3+PVS 39)                                    |
| Quality Degradation | Bit rates                                       | Bit rates                                       | Bit rates<br>Frame rates<br>Freeze frames<br>Blockiness |
| Video length        | 10 sec  | 10 sec  | 10 sec  |
| No. of Viewers      | 48 non-experts<br>aged 20-39<br>males & females | 48 non-experts<br>aged 20-39<br>males & females | 48 non-experts<br>aged 19-39<br>males & females         |

# Video quality conditions

|  | Exp. 1                               |           | Exp. 2                              |           | Exp. 3                         |           |
|--|--------------------------------------|-----------|-------------------------------------|-----------|--------------------------------|-----------|
| Resolution   | HD<br>(1920 × 1080)                  |           | QVGA<br>(320 × 240)                 |           | QVGA<br>(320 × 240)            |           |
| Coding   | H. 264                               |           | H. 264                              |           | H. 264                         |           |
| Number of contents   | 8                                    | 8<br>HRC  | 8                                   | 8<br>HRC  | 3                              | 3<br>HRC  |
| Bit rates<br>30 fps  | 3, 4.3, 5.7,<br>6.9, 9.6, 15<br>Mbps | 48<br>PVS | 128, 256,<br>384, 704,<br>1500 kbps | 40<br>PVS | 128, 384, 704,<br>1500<br>kbps | 12<br>PVS |
| Frame rates (fps)<br>(1.5 Mbps)                              | –                                    |           | –                                   |           | 15, 10, 5                      | 9<br>PVS  |
| Number of block<br>distortions in 10 sec.<br>(1.5 M, 30 fps) | –                                    |           | –                                   |           | 1, 2~4, 5~<br>10               | 9<br>PVS  |
| Freeze length (frame)<br>(1.5 M, 30 fps)                     | –                                    |           | –                                   |           | 1, 5, 15                       | 9<br>PVS  |

# Video sources (ITU-R Rec. BT.1210-3)

| No.    | Exp. 1 | Exp. 2 | Exp. 3 |
|--------|--------|--------|--------|
| No. 5  | ○      | ○      |        |
| No. 7  | ○      | ○      | ○      |
| No. 10 | ○      | ○      | ○      |
| No. 12 | ○      | ○      | ○      |
| No. 16 | ○      | ○      |        |
| No. 23 | ○      | ○      |        |
| No. 25 | ○      | ○      |        |
| No. 34 | ○      | ○      |        |



No. 5



No. 16



No. 7



No. 23



No. 10



No. 25



No. 12



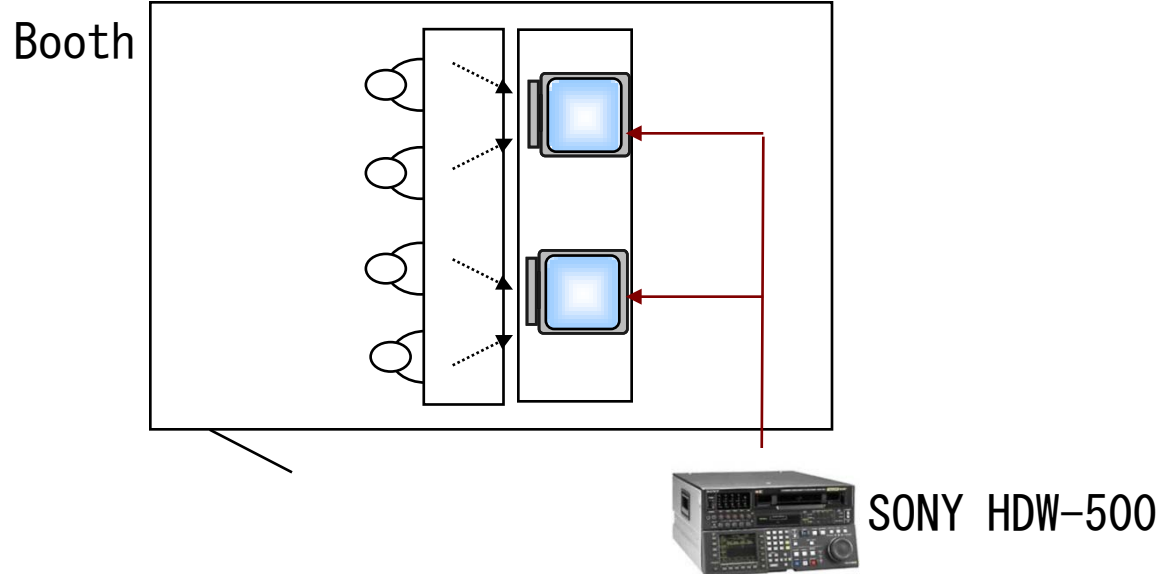
No. 34



# Assessment environment (Exp. 1)

|                  |   |
|------------------|---|
| Monitor          | 32 inch CRT Monitor<br>SONY BVM-D32E1WJ |
| Viewing distance | 3H (H: picture height) about 110 cm     |
| Illuminance      | 20lx                                    |

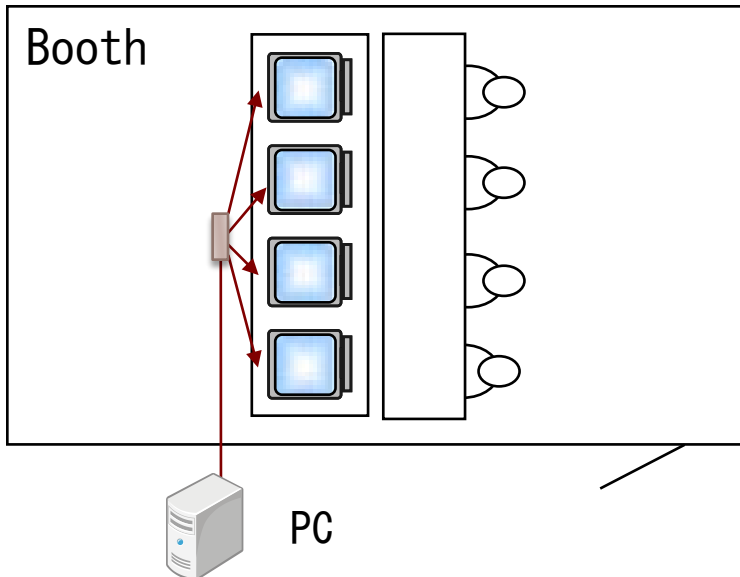
DSCQS, DSIS, ACR (-HR), ACR11 (-HR)



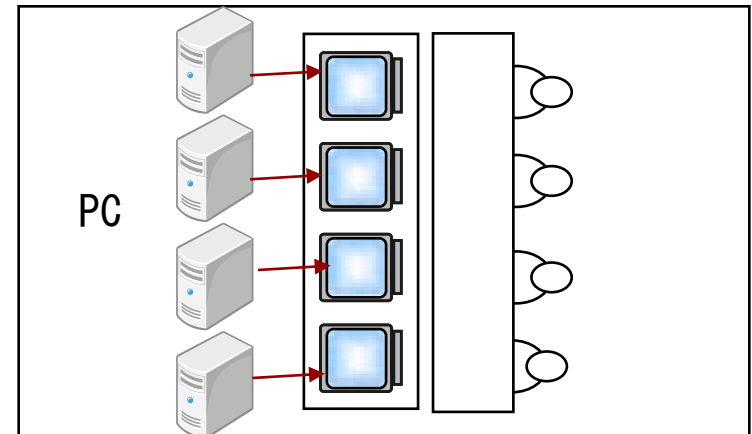
# Assessment environment (Exps. 2 and 3)

|                  |   |
|------------------|---|
| Monitor          | 17 inch LCD (EIZO FLEXSCAN M170)<br>Monitor resolution: 1280×1024 |
| Viewing distance | 8H (H:picture height) about 50 cm                                 |
| Illuminance      | 20 lx (Exp. 2), 50 lx (Exp. 3)                                    |

DSCQS, DSIS, ACR (-HR), ACR11 (-HR)



SAMVIQ (-HR)



# Experimental design

- ▶ Each participant performed all assessment methods in a day.
- ▶ Evaluation methods randomization
  - ▶ The methods were executed in a different order each day. (Exp. 1)

| order | 1 <sup>st</sup> day | 2 <sup>nd</sup> day | 3 <sup>rd</sup> day | 4 <sup>th</sup> day | 5 <sup>th</sup> day | 6 <sup>th</sup> day |
|-------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| 1     | ACR                 | DCR                 | DCR                 | ACR-11              | ACR                 | DSCQS               |
| 2     | DCR                 | ACR-11              | ACR                 | ACR                 | DSCQS               | ACR-11              |
| 3     | DSCQS               | ACR                 | ACR-11              | DSCQS               | DCR                 | DCR                 |
| 4     | ACR-11              | DSCQS               | DSCQS               | DCR                 | ACR-11              | ACR                 |

- ▶ PVSs randomization
  - ▶ Exp. 1 uses 6 randomize patterns for each method.
  - ▶ Exp. 2 uses 6 randomize patterns for each method.
  - ▶ Exp. 3 uses 12 randomize patterns for each method.

## 4. Performance comparison indexes

---

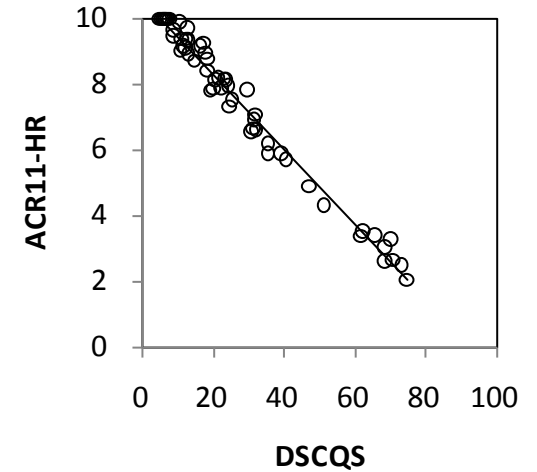
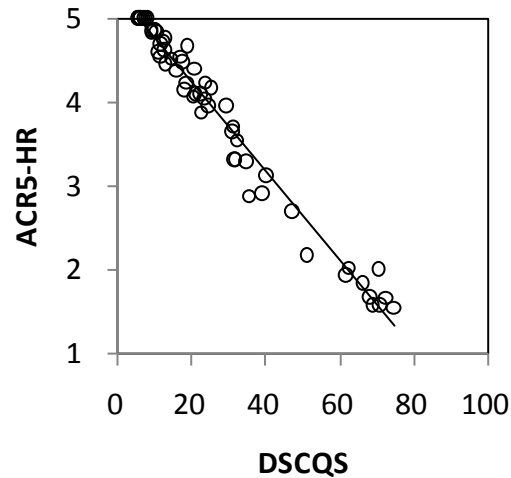
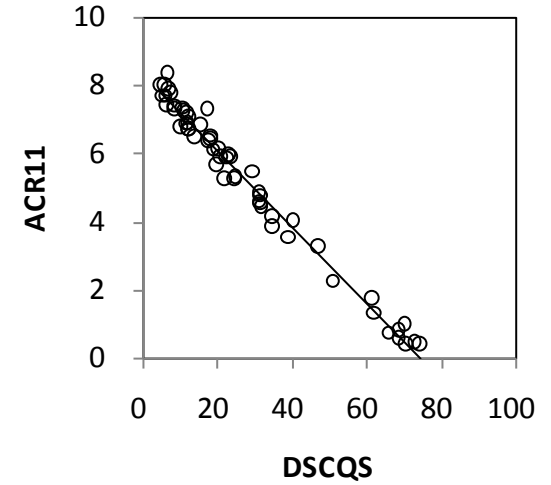
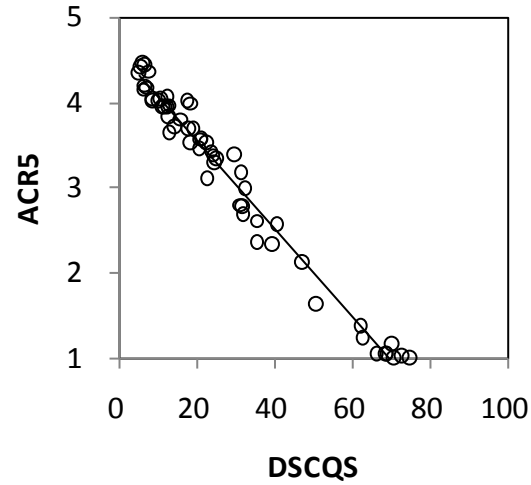
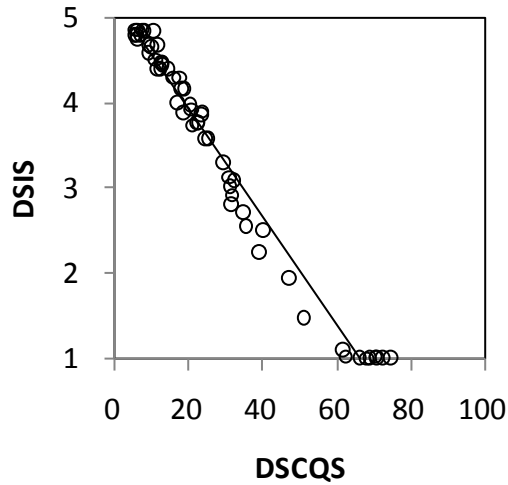
- ▶ **Correlation coefficient (R)**
  - ▶ Pearson product-moment correlation coefficient
  - ▶ Definition: Correlation coefficient between two methods' Mean Opinion Score (MOS).
- ▶ **Rank correlation coefficient (rank R)**
  - ▶ Spearman rank correlation coefficient Definition: Correlation coefficient between each method's rank.
- ▶ **Mean of 95% confidence interval – normalized (MCI<sub>norm</sub>)**
  - ▶ MCI comparison index for different scales
  - ▶ Definition:  $MCI_{norm} = MCI / (MOS \text{ range})$

---

## 5. Test results

# Correlation between MOS and DSCQS (Exp. 1)

---



## R (Exp. 1)

---

- ▶ There are no significant differences in average of Rs between each two methods with a 5% level of significance.

|          | DSCQS       | DSIS        | ACR5-HR     | ACR11-HR    | ACR5        | ACR11       |
|----------|-------------|-------------|-------------|-------------|-------------|-------------|
| DSCQS    | –           | –0.99       | –0.99       | –0.99       | –0.99       | –0.99       |
| DSIS     | –0.99       | –           | 0.99        | 0.99        | 0.99        | 0.99        |
| ACR5-HR  | –0.99       | 0.99        | –           | 0.99        | 1.00        | 0.99        |
| ACR11-HR | –0.99       | 0.99        | 0.99        | –           | 0.99        | 0.99        |
| ACR5     | –0.99       | 0.99        | 1.00        | 0.99        | –           | 0.99        |
| ACR11    | –0.99       | 0.99        | 0.99        | 0.99        | 0.99        | –           |
| Average* | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> |

\*: Average to absolute value of correlation coefficients

## rank R (Exp. 1)

---

- ▶ There are no significant differences in average of rank Rs between each two methods with a 5% level of significance.

|                | DSCQS       | DSIS        | ACR5-HR     | ACR11-HR    | ACR5        | ACR11       |
|----------------|-------------|-------------|-------------|-------------|-------------|-------------|
| DSCQS          | –           | 0.99        | 0.98        | 0.98        | 0.98        | 0.98        |
| DSIS           | 0.99        | –           | 0.98        | 0.98        | 0.97        | 0.98        |
| ACR5-HR        | 0.98        | 0.98        | –           | 0.98        | 0.99        | 0.98        |
| ACR11-HR       | 0.98        | 0.98        | 0.98        | –           | 0.98        | 0.99        |
| ACR5           | 0.98        | 0.97        | 0.99        | 0.98        | –           | 0.98        |
| ACR11          | 0.98        | 0.98        | 0.98        | 0.99        | 0.98        | –           |
| <b>Average</b> | <b>0.98</b> | <b>0.98</b> | <b>0.98</b> | <b>0.98</b> | <b>0.98</b> | <b>0.98</b> |



# MCI<sub>norm</sub> (Exp. 1)

---

- ▶ DSIS has the smallest MCI<sub>norm</sub> of the 6 methods.

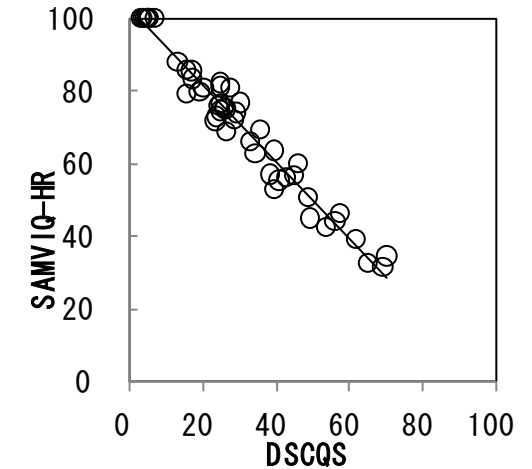
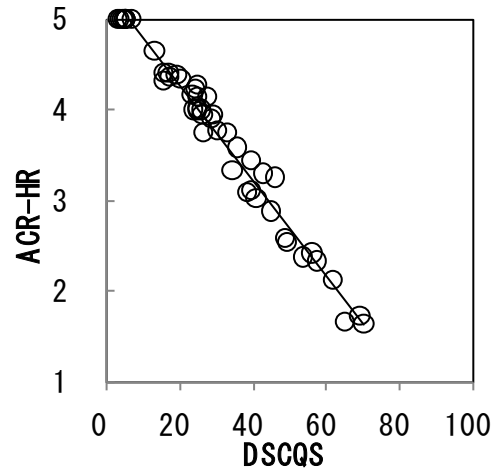
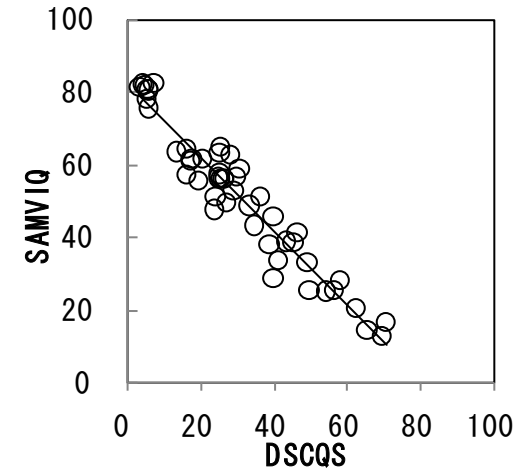
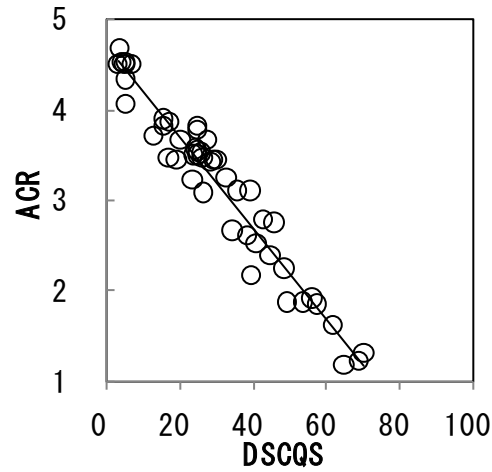
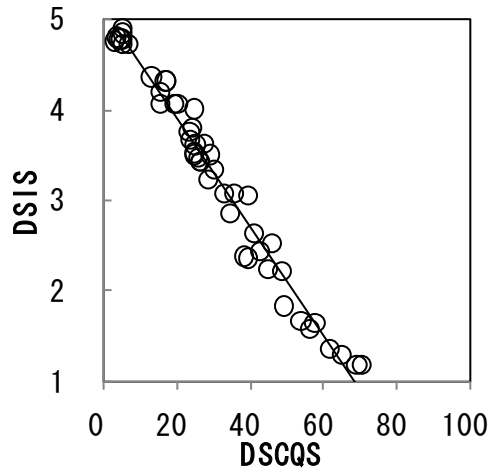
|                     | DSCQS | DSIS | ACR5-HR | ACR11-HR | ACR5 | ACR11 |
|---------------------|-------|------|---------|----------|------|-------|
| MCI                 | 4.22  | 0.17 | 0.21    | 0.44     | 0.19 | 0.42  |
| MOS range           | 69.65 | 3.88 | 3.44    | 7.94     | 3.46 | 7.94  |
| MCI <sub>norm</sub> | 0.06  | 0.04 | 0.06    | 0.06     | 0.06 | 0.05  |

# Consideration for Exp. 1

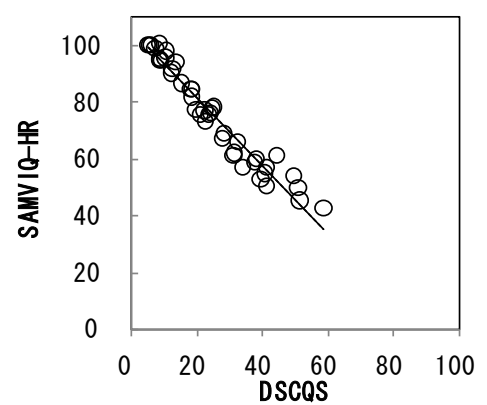
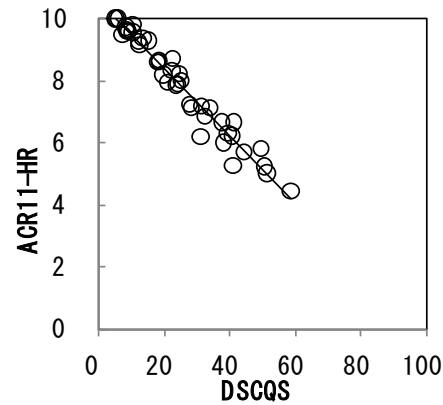
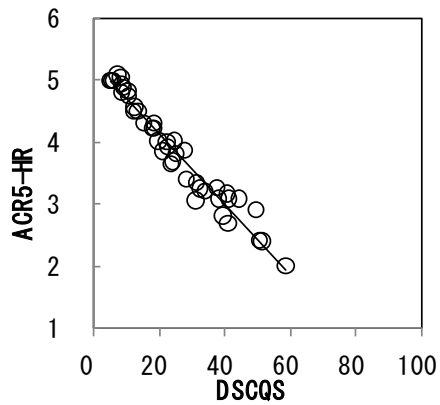
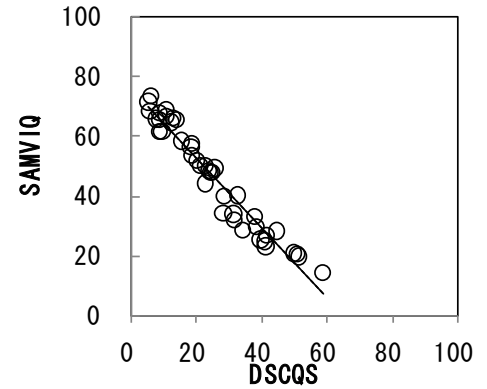
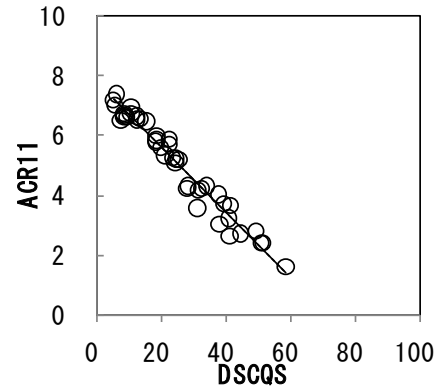
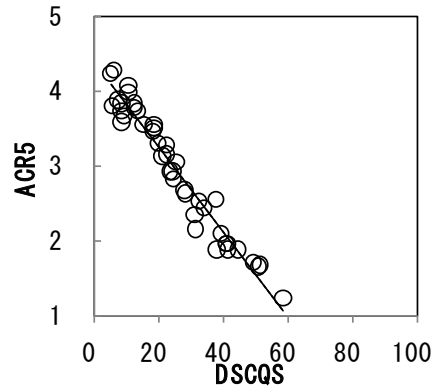
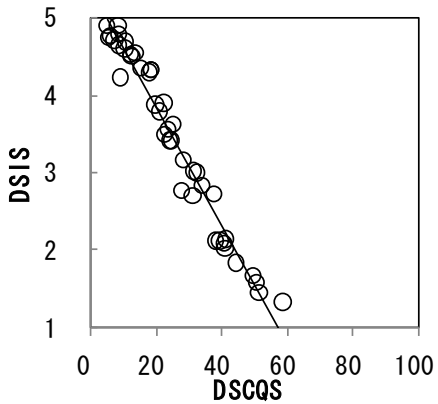
---

- ▶ For HDTV coding distortion quality assessment, combinations of all methods show high correlations and high rank correlations, there are no differences about R and rank R among methods.
- ▶ DSIS has the smallest MCI<sub>norm</sub> in 6 methods.
  - ▶ This seems it is because DSIS is a single scale.
  - ▶ Moreover, it is because DSIS method present REF before PVS, this is differ from ACR5 and ACR11.

# Correlation between MOS and DSCQS (Exp. 2)



# Correlation between MOS and DSCQS (Exp. 3)



## 6. Conclusions

---

- ▶ The assessment methods and scales in our experiments provide equivalent results with equivalent reliability.
- ▶ This conclusion holds for both low and high resolution video (Exp. 1 and Exp. 2).
- ▶ This does not change even in the evaluation of network degradation (Exp. 2 and Exp. 3).

# 7. Acknowledgments & References

---

## ▶ Acknowledgments

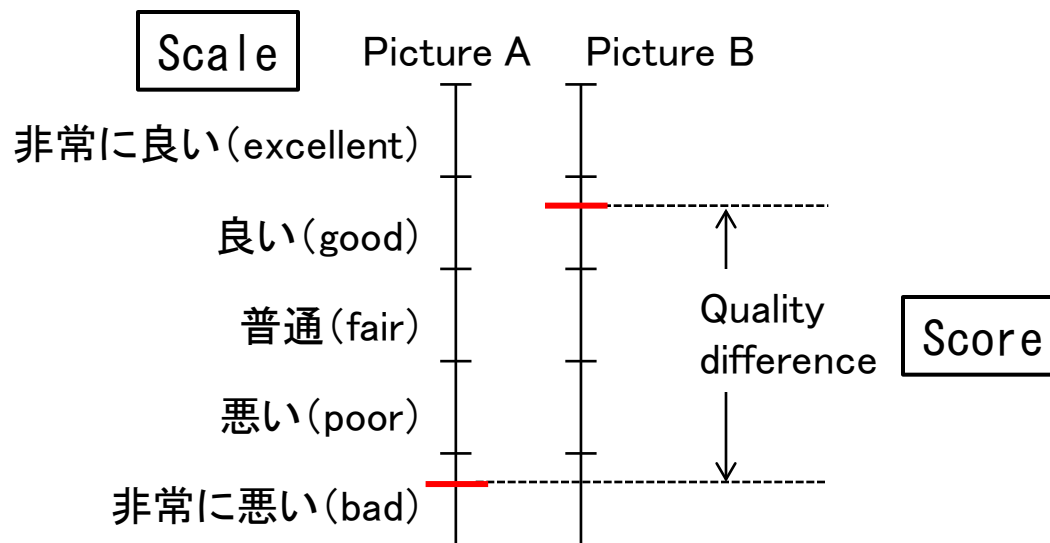
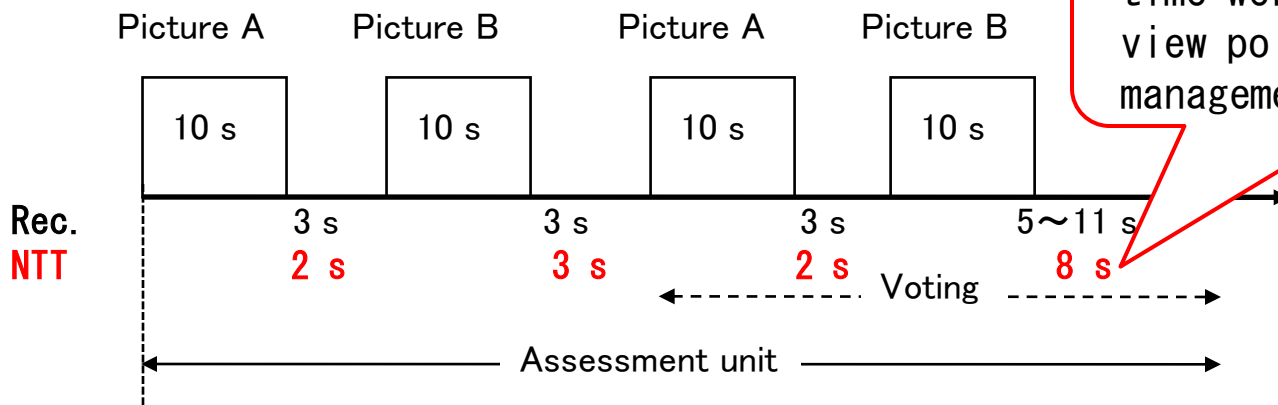
This study was supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communication of Japan (No.073103002).

## ▶ References

- [1] ITU-R Document 6Q/208, “Report on Experiment of New Subjective Video Quality Metrics SAMVIQ for Mobile Video,” May 2007.
- [2] P. Le Gallet, R. Pepion, “Comparison of subjective test methodologies,” June 2009.  
【QVGA·VGA·HD, 3, 6, 9Mbps】
- [3] Q. Huynh–Thu, A. Raake, M. N. Garcia, P. Corriveau, F. Speranza, “ Study of Rating Scales for Subjective Video Quality Assessment Using Single–Stimulus Presentation,” June 2009.  
【HD, 2–16Mbps, PLR 0.25–4% 】

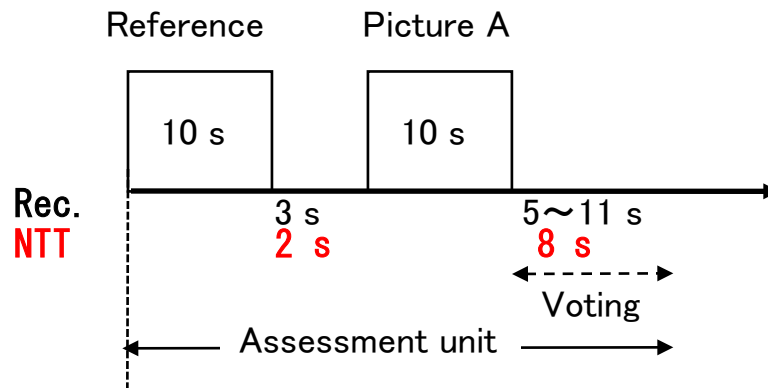
# Subjective assessment methods (detail)

# DSCQS





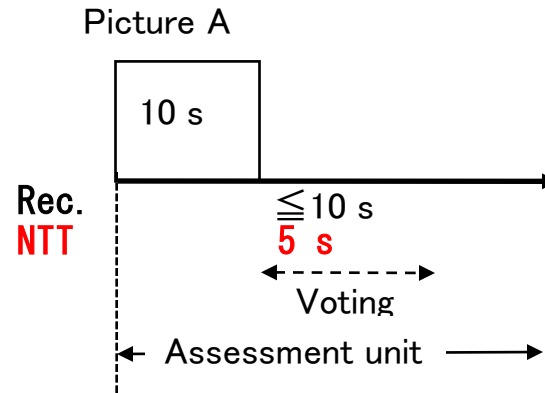
# DSIS (DCR, EBU)



| Scale  | Score |
|--|-------|
| 劣化がわからない (Imperceptible)                     | 5     |
| 劣化がわかるが気にならない (Perceptible but not annoying) | 4     |
| 劣化がわずかに気になる (Slightly annoying)              | 3     |
| 劣化が気になる (Annoying)                           | 2     |
| 劣化が非常に気になる (Very annoying)                   | 1     |

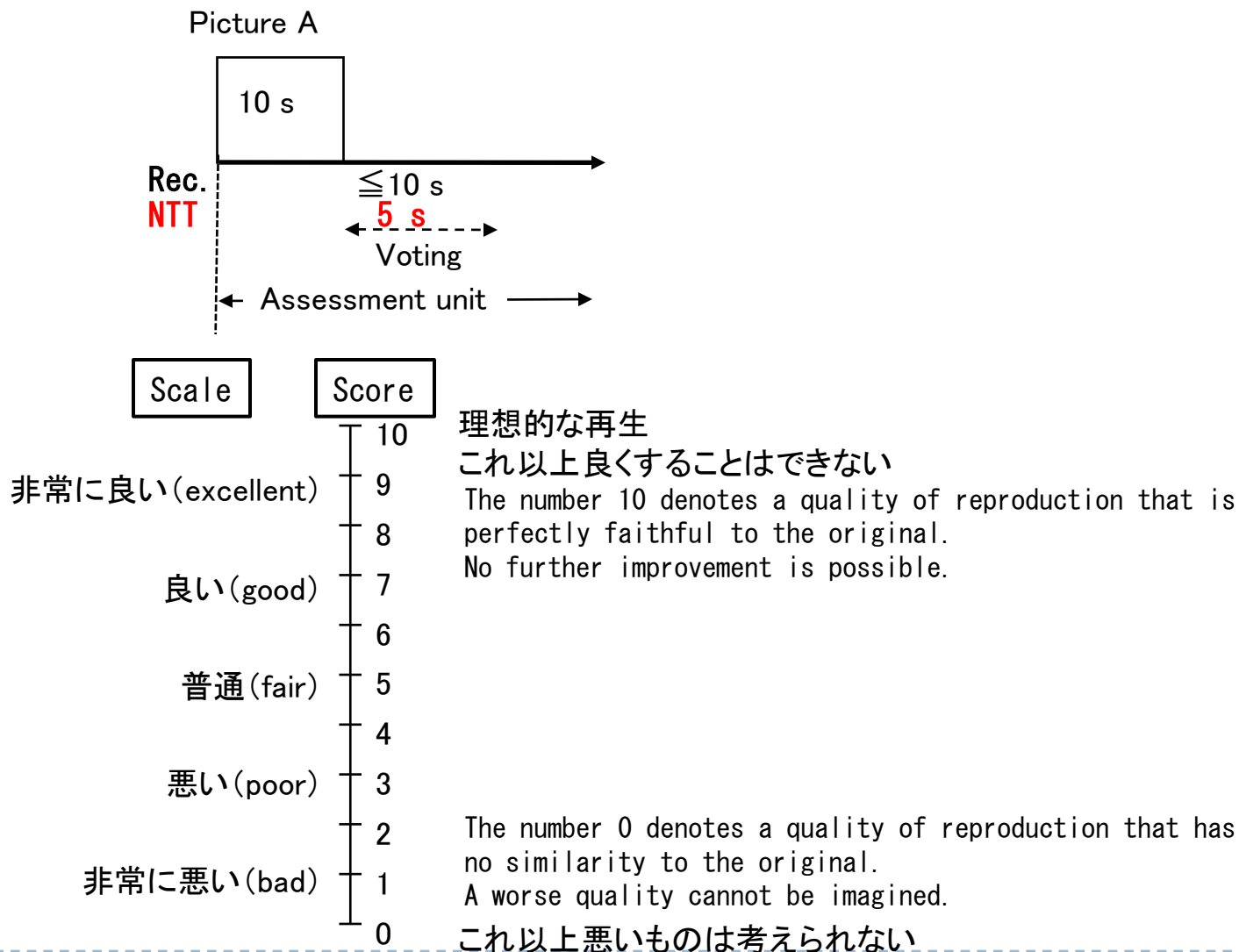
# ACR 5pt.

---



| Scale             | Score |
|-------------------|-------|
| 非常に良い (Excellent) | 5     |
| 良い (Good)         | 4     |
| 普通 (Fair)         | 3     |
| 悪い (Poor)         | 2     |
| 非常に悪い (Bad)       | 1     |

# ACR 11pt.



# SAMVIQ

The interface displays a video player with a scene of a man in red overalls and a child. A 'Voting slider' on the right shows a score of 66, with qualitative markers for 'Excellent', 'Good', 'Fair', and 'Poor'. Below the video is a 'REF' label and a sequence selection menu with buttons labeled 'Ref.', 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', and 'J'. The 'Ref.' button is selected, showing a score of 84. Buttons for 'Play', 'Stop', 'Previous sequence', 'Next sequence', and 'Voting end' are located at the bottom.

| Sequence | Score |
|----------|-------|
| Ref.     | 84    |
| A        | 66    |
| B        | X     |
| C        | X     |
| D        | X     |
| E        | X     |
| F        | X     |
| G        | X     |
| H        | X     |
| I        | X     |
| J        | X     |



# Instruction sample (Exp. 3 in Japanese)

## 評価法比較実験について(オペレータ用)

### ● 試験の概要

本試験は、複数の映像品質評価法を比較するため、同一映像に対して計5つの評価法を用いて評価していただきます。各評価方法の詳しい説明は、試験の直前にいたします。また、評価方法が変わるたびに練習を行いますので、質問や疑問は、練習の時にオペレータに訊ねてください。

すべての試験に共通していることは、観視する映像は、モバイル端末の映像の品質としてどうか、という評価をして頂くことです。モバイル端末の映像とは、世間一般のPDAや携帯電話の映像をいいます。また、すべての試験は映像品質に着目して頂きたいため、音は流れません。1回のセッションは10分～20分程度です。

では、評価内容の説明にうつる前に、一般的な注意事項を説明します。

### ● 評価中の注意事項

1. **映像の評価**を行っていただきますので、**画面を注視**して下さい。
2. 評価はあまり考え込まず、悩まないようにして下さい。
3. 他人や自分への影響を避けるため、**評価に関する話題は試験が全て終わるまでご遠慮下さい。**
4. 評価ブース内の**机**、**ケーブル類**は移動させないで下さい。また、ディスプレイから皆様の目までの距離を一定にしたいため、指定の位置に椅子を合わせ、深く腰かけ背筋を伸ばした状態で試験を実施してください。
5. 評価ブース内には、飲食物、**携帯電話**を持ち込まないで下さい。
6. 試験の途中で気分が悪くなったり、何か不都合があったりした場合には、ブース内部の音声と映像をモニタしていますので、遠慮なく申し出て下さい。
7. メガネやコンタクトレンズを使用して**矯正視力1.0以上**になる方は、必ず着用してください。
8. ブースの扉は、危険防止のためオペレータが開閉しますので、自分で**扉の開閉はしない**で下さい。(緊急時を除く)。

### ● その他の注意事項

1. 試験以外の時は休憩場所で待機して下さい。
2. 携帯電話はマナーモードにし、休憩場所での**通話をご遠慮下さい。**
3. 所持品は休憩場所に置いて頂いて結構ですが、貴重品は各自で管理して下さい。万一、紛失した場合には責任を負いかねますので、ご協力をお願いします。
4. **非常口**は本日、入ってきていただいた入口になります。
5. トイレ、喫煙等で席をはずす場合は、カードを借り、**管理簿に記入**してください。
6. 敷地内での写真撮影は固く禁じられております。携帯電話による撮影もご遠慮下さい。

全般の注意事項は以上です。

これから試験の説明をいたしますが、該当の説明ページだけをご覧頂くようお願いいたします。他の試験のページを同時にご覧頂きますと、ご自身が試験の時に混乱しますので、よろしくお願ひします。また、休憩中も他の評価方法の説明ページを読まないようにお願ひします。ただし、注意事項はいつでも読み直して頂いてけっこうです。それでは、評価**X**(その日の最初の試験方法)のページを開いてください。

# Time schedule sample (Exp. 3)

| time          | min. | Content                       | group | Conditions |
|---------------|------|-------------------------------|-------|------------|
| 9:00 - 9:20   | 0:20 | Instructions                  | A•B   | -          |
| 9:20 - 9:27   | 0:07 | ACR5 instruction & practice   | A     | 10         |
| 9:27 - 9:44   | 0:17 | ACR test                      | A     | 21+21      |
| 9:44 - 9:51   | 0:07 | ACR5 instruction & practice   | B     | 10         |
| 9:51 - 10:08  | 0:17 | ACR test                      | B     | 21+21      |
| 10:08 - 10:18 | 0:10 | Break (change settings)       |       |            |
| 10:18 - 10:28 | 0:10 | DSIS instruction & practice   | A     | 10         |
| 10:28 - 10:52 | 0:24 | DSIS test                     | A     | 21+21      |
| 10:52 - 11:02 | 0:10 | DSIS instruction & practice   | B     | 10         |
| 11:02 - 11:26 | 0:24 | DSIS test                     | B     | 21+21      |
| 11:26 - 11:28 | 0:02 | change settings               |       |            |
| 11:28 - 11:38 | 0:10 | ACR11 instruction & practice  | A     | 10         |
| 11:38 - 11:55 | 0:17 | ACR-11test                    | A     | 21+21      |
| 11:55 - 12:05 | 0:10 | ACR11 instruction & practice  | B     | 10         |
| 12:05 - 12:22 | 0:17 | ACR-11test                    | B     | 21+21      |
| 12:22 - 13:22 | 1:00 | Lunch                         | A•B   |            |
| 13:22 - 13:27 | 0:05 | change settings               |       |            |
| 13:27 - 13:42 | 0:15 | DSCQS instruction & practice  | A     | 10         |
| 13:42 - 14:05 | 0:23 | DSCQS test1                   | A     | 11+10      |
| 14:05 - 14:20 | 0:15 | DSCQS instruction & practice  | B     | 10         |
| 14:20 - 14:43 | 0:23 | DSCQS test1                   | B     | 11+10      |
| 14:43 - 15:05 | 0:22 | DSCQS test2                   | A     | 11+10      |
| 15:05 - 15:27 | 0:22 | DSCQS test2                   | B     | 11+10      |
| 15:27 - 15:37 | 0:10 | Break (change settings)       |       |            |
| 15:37 - 15:47 | 0:10 | SAMVIQ instruction & practice | A     | 1content   |
| 15:47 - 16:00 | 0:13 | SAMVIQ test1                  | A     | 3 contents |
| 16:00 - 16:10 | 0:10 | SAMVIQ instruction & practice | B     | 1content   |
| 16:10 - 16:23 | 0:13 | SAMVIQ test1                  | B     | 3 contents |
| 16:23 - 16:36 | 0:13 | SAMVIQ test2                  | A     | 3 contents |
| 16:36 - 16:49 | 0:13 | SAMVIQ test2                  | B     | 3 contents |
| 16:49 - 17:00 | 0:11 | Dara confirmation             |       |            |

# Comparison indices

# Comparison of CI methods

A. MCI normalized by MOS range

$$MCI_{norm} = \frac{MCI}{MOS\ range}$$

$$MOS\ range = MAX(MOSs) - MIN(MOSs)$$

B. MCI normalized by scale range

$$MCI_{scale} = \frac{MCI}{Scale\ range}$$

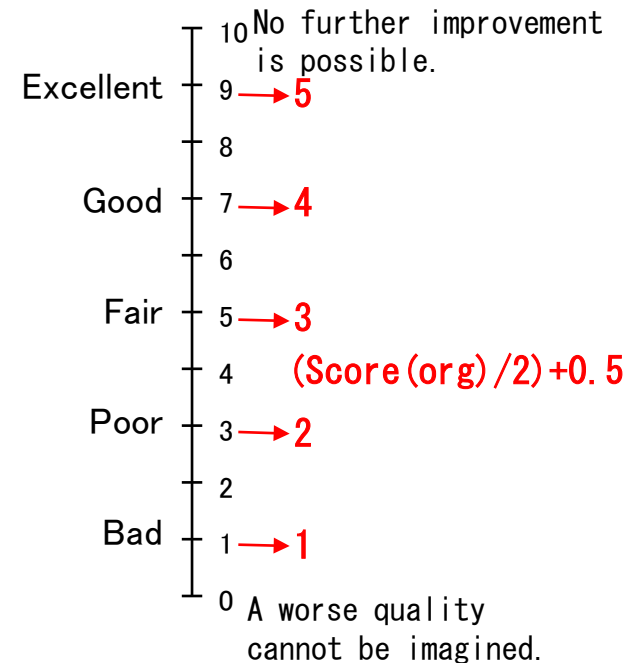
Scale range

$$ACR5: 5 - 1 = 4$$

$$ACR11: 10 - 0 = 10$$

$$SAMVIQ: 100 - 0 = 100$$

C. Score transform

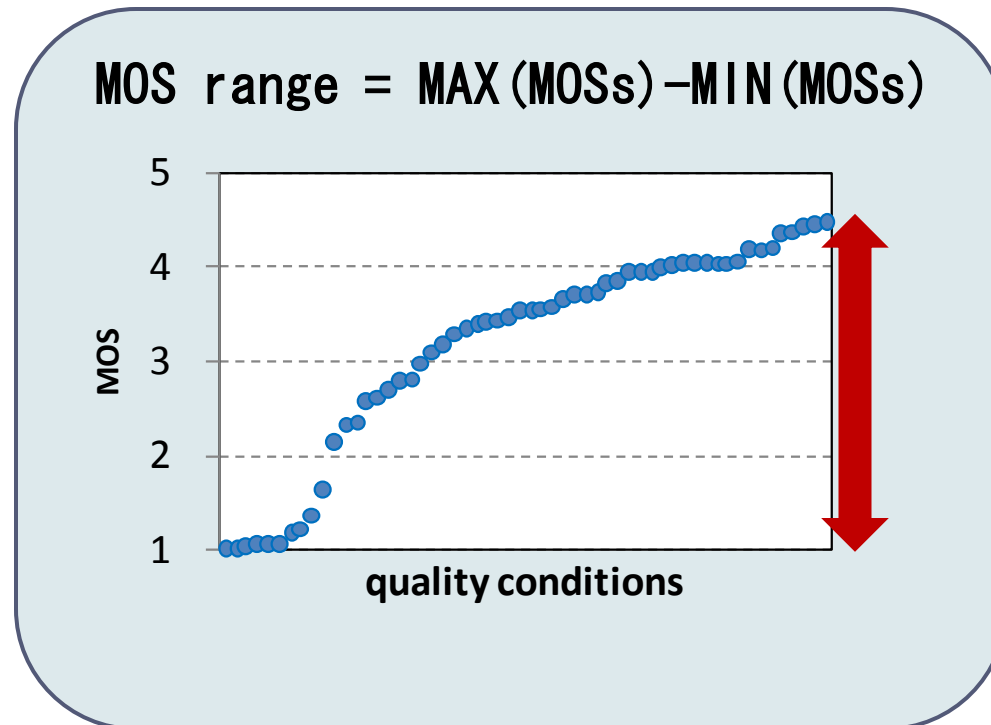


|              | Method A (proposal)  | Method B ( Ref. [2])                                    | Method C ( Ref. [3])                          |
|--------------|--|---|---|
| Advantage    | Normalization within the effective MOS range for each scale is possible. | Normalization is constant according to the scale range. | Direct comparison of MOS is available.        |
| Disadvantage | Direct comparison of MOS is not available.                               | Direct comparison of MOS is not available.              | Transform function for each method is needed. |



# MCI<sub>norm</sub>

- ▶  $MCI_{norm} = MCI / \text{MOS range}$



# CI comparison (Exp. 1)

---

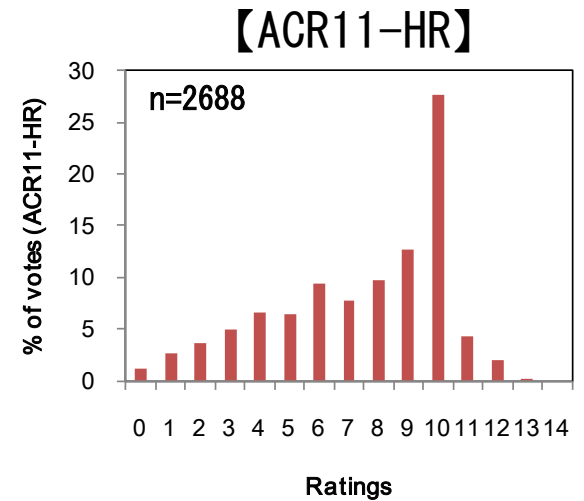
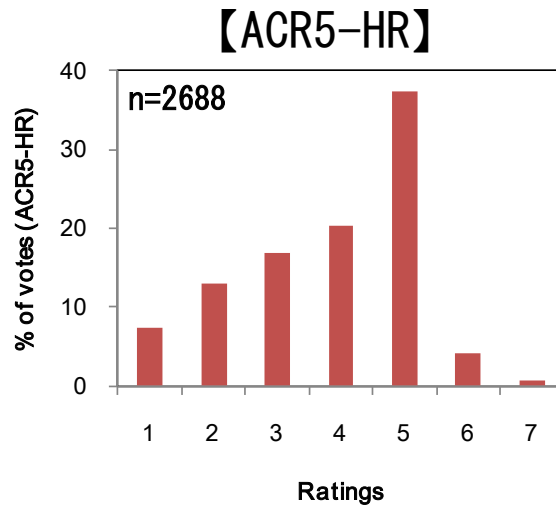
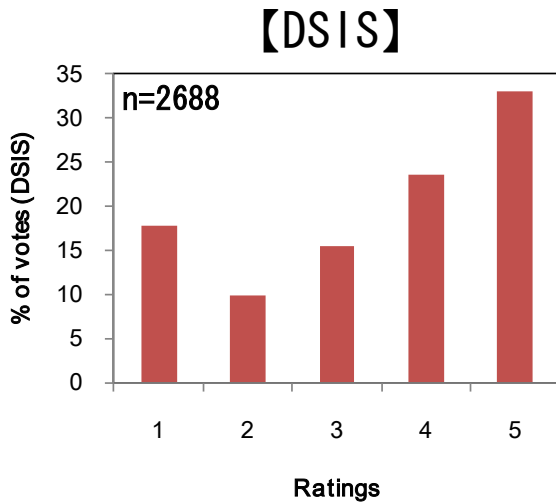
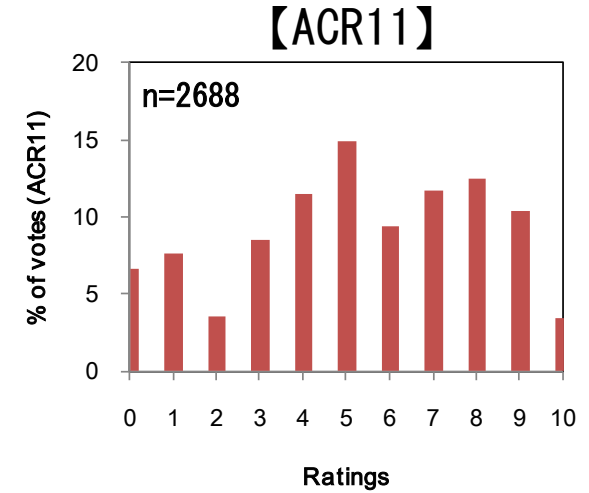
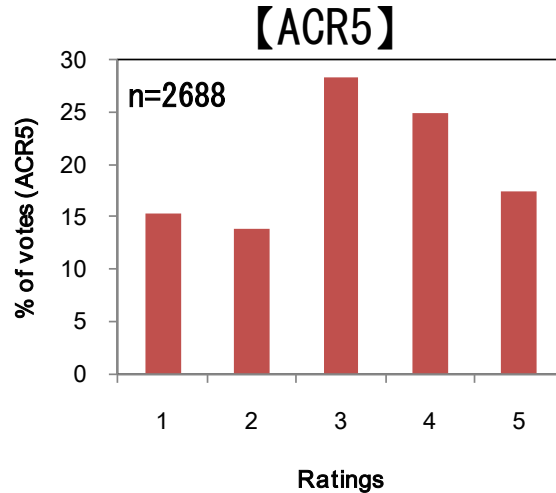
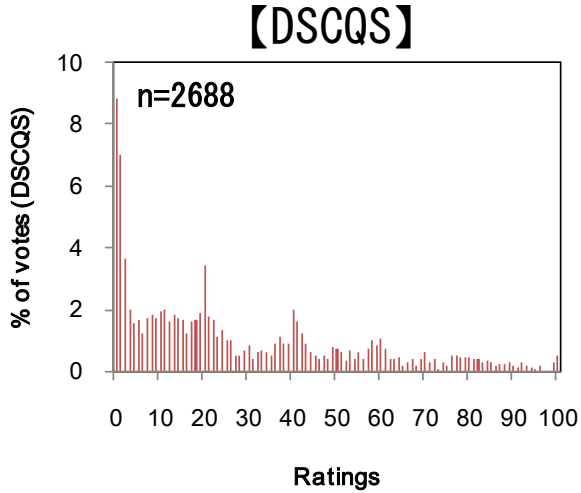
- ▶ The differences in MCI of the compared CI methods is 2% in MCInorm, and 1% in MCIscale.

|                     | DSCQS | DSIS | ACR5-HR | ACR11-HR | ACR5 | ACR11 |
|---------------------|-------|------|---------|----------|------|-------|
| MCInorm (NTT)       | 0.06  | 0.04 | 0.06    | 0.06     | 0.06 | 0.05  |
| MCIscale (Ref. [2]) | 0.04  | 0.04 | –       | –        | 0.05 | 0.04  |

# Experimental results (Exp. 1)

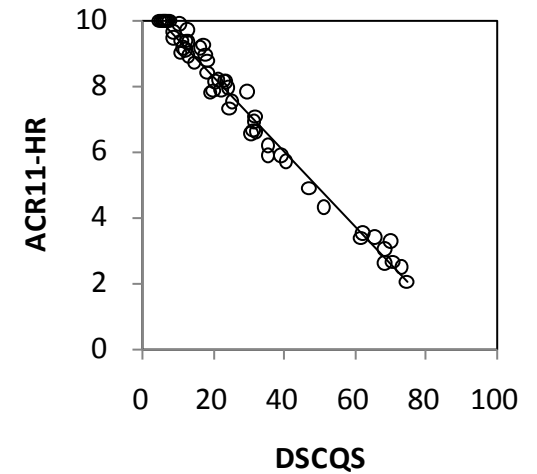
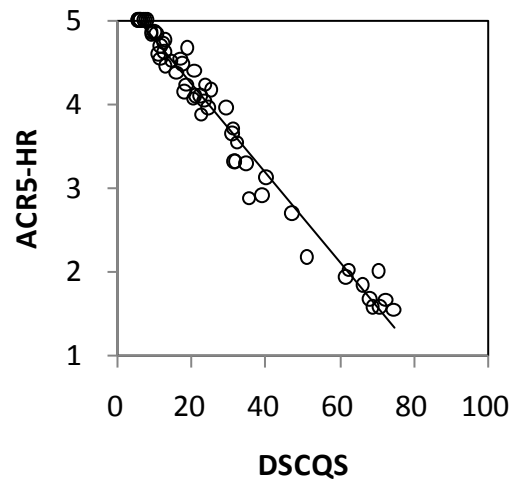
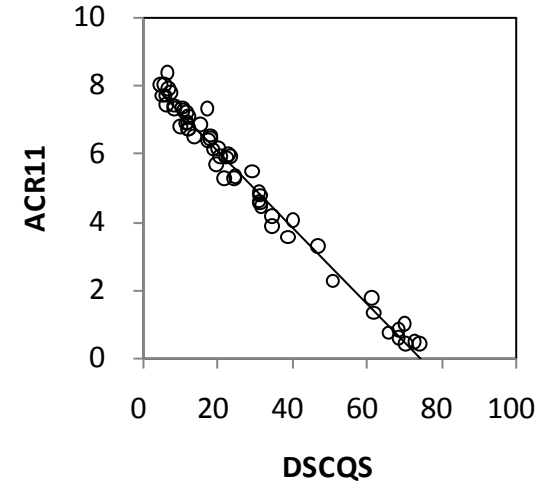
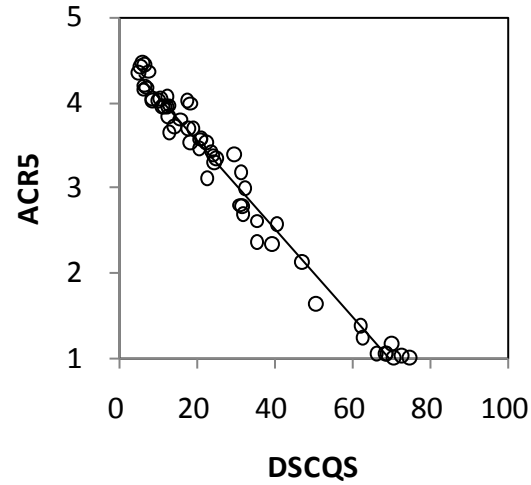
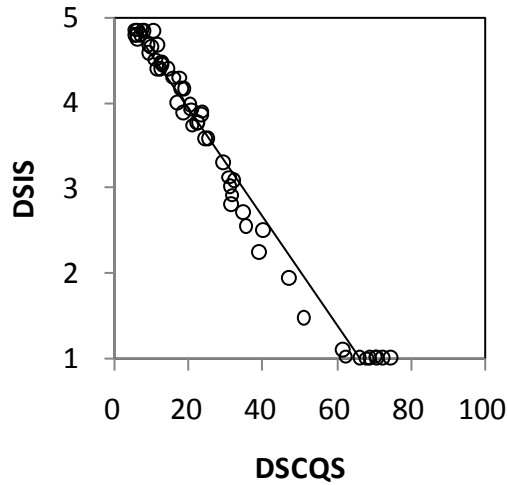
# Test results for Exp. 1

## Score distributions



# Correlation between MOS and DSCQS (Exp. 1)

---



## R (Exp. 1)

---

- ▶ There are no significant differences in average of Rs between each two methods with a 5% level of significance.

|          | DSCQS       | DSIS        | ACR5-HR     | ACR11-HR    | ACR5        | ACR11       |
|----------|-------------|-------------|-------------|-------------|-------------|-------------|
| DSCQS    | –           | –0.99       | –0.99       | –0.99       | –0.99       | –0.99       |
| DSIS     | –0.99       | –           | 0.99        | 0.99        | 0.99        | 0.99        |
| ACR5-HR  | –0.99       | 0.99        | –           | 0.99        | 1.00        | 0.99        |
| ACR11-HR | –0.99       | 0.99        | 0.99        | –           | 0.99        | 0.99        |
| ACR5     | –0.99       | 0.99        | 1.00        | 0.99        | –           | 0.99        |
| ACR11    | –0.99       | 0.99        | 0.99        | 0.99        | 0.99        | –           |
| Average* | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> |

\*: Average to absolute value of correlation coefficients

## rank R (Exp. 1)

---

- ▶ There are no significant differences in average of rank Rs between each two methods with a 5% level of significance.

|                | DSCQS       | DSIS        | ACR5-HR     | ACR11-HR    | ACR5        | ACR11       |
|----------------|-------------|-------------|-------------|-------------|-------------|-------------|
| DSCQS          | –           | 0.99        | 0.98        | 0.98        | 0.98        | 0.98        |
| DSIS           | 0.99        | –           | 0.98        | 0.98        | 0.97        | 0.98        |
| ACR5-HR        | 0.98        | 0.98        | –           | 0.98        | 0.99        | 0.98        |
| ACR11-HR       | 0.98        | 0.98        | 0.98        | –           | 0.98        | 0.99        |
| ACR5           | 0.98        | 0.97        | 0.99        | 0.98        | –           | 0.98        |
| ACR11          | 0.98        | 0.98        | 0.98        | 0.99        | 0.98        | –           |
| <b>Average</b> | <b>0.98</b> | <b>0.98</b> | <b>0.98</b> | <b>0.98</b> | <b>0.98</b> | <b>0.98</b> |

# MCI<sub>norm</sub> (Exp. 1)

---

- ▶ DSIS has the smallest MCI<sub>norm</sub> of the 6 methods.

|                     | DSCQS | DSIS | ACR5-HR | ACR11-HR | ACR5 | ACR11 |
|---------------------|-------|------|---------|----------|------|-------|
| MCI                 | 4.22  | 0.17 | 0.21    | 0.44     | 0.19 | 0.42  |
| MOS range           | 69.65 | 3.88 | 3.44    | 7.94     | 3.46 | 7.94  |
| MCI <sub>norm</sub> | 0.06  | 0.04 | 0.06    | 0.06     | 0.06 | 0.05  |



# Consideration for Exp. 1

---

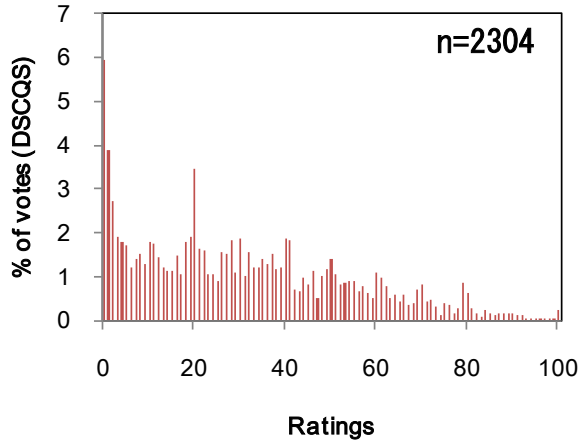
- ▶ For HDTV coding distortion quality assessment, combinations of all methods show high correlations and high rank correlations, there are no differences about R and rank R among methods.
- ▶ DSIS has the smallest MCI<sub>norm</sub> in 6 methods.
  - ▶ This seems it is because DSIS is a single scale.
  - ▶ Moreover, it is because DSIS method present REF before PVS, this is differ from ACR5 and ACR11.

# Experimental results (Exp. 2)

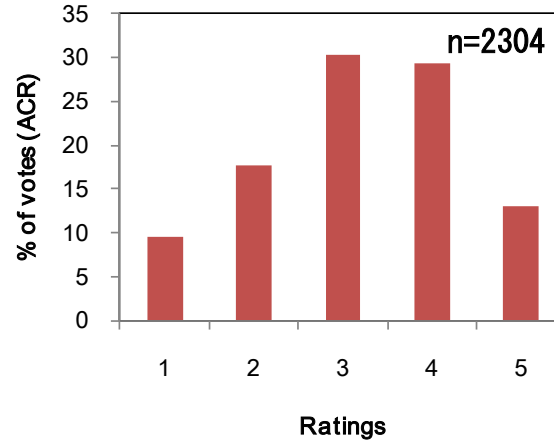
# Test results for Exp. 2

## Score distributions

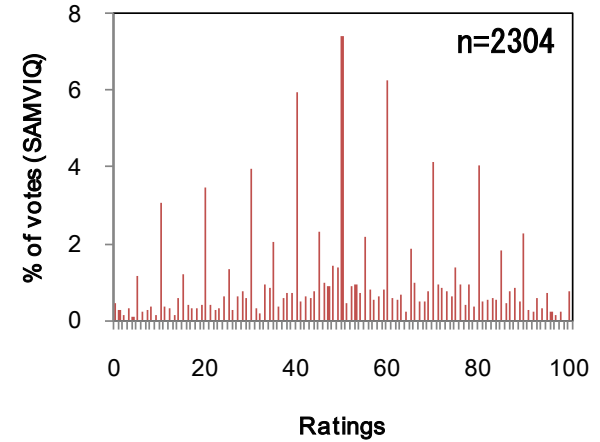
**【DSCQS】**



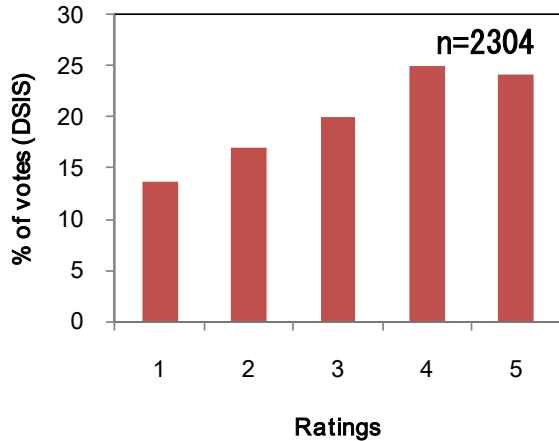
**【ACR5】**



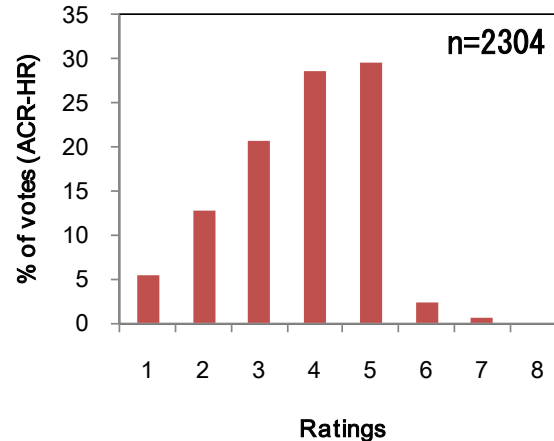
**【SAMVIQ】**



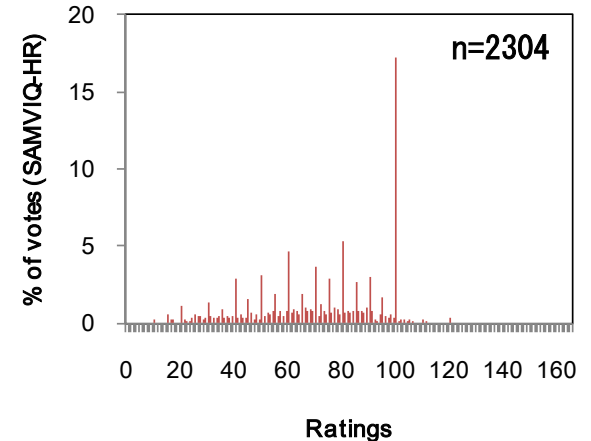
**【DSIS】**



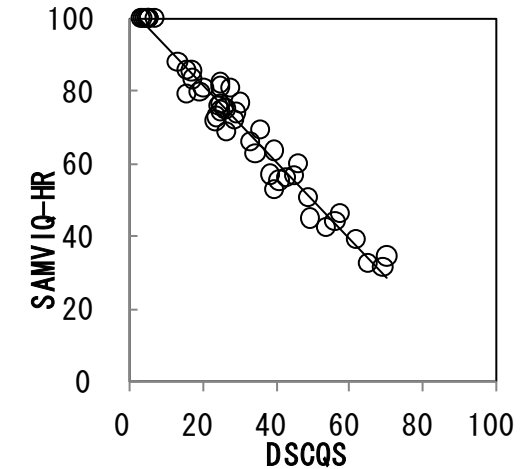
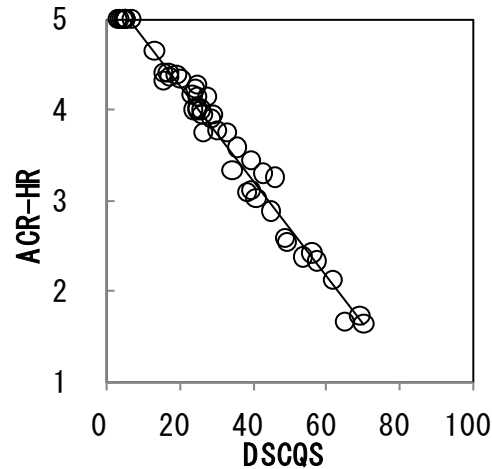
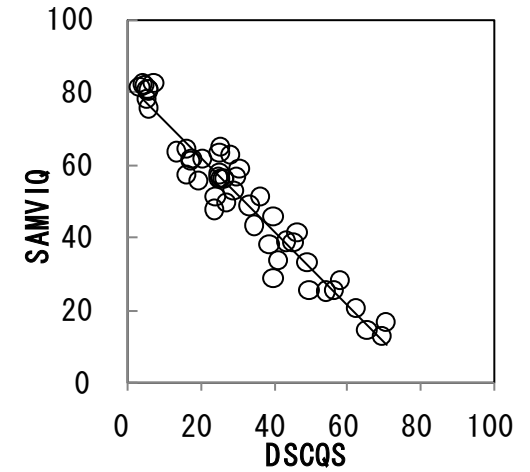
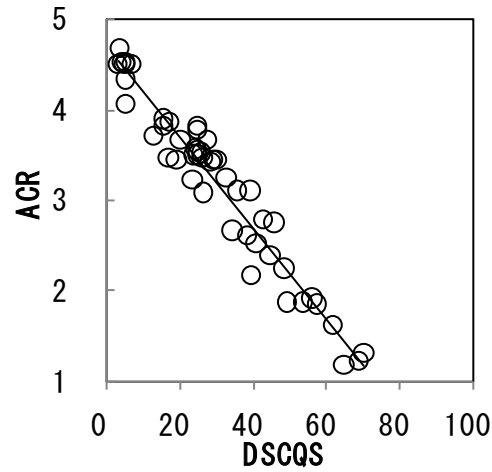
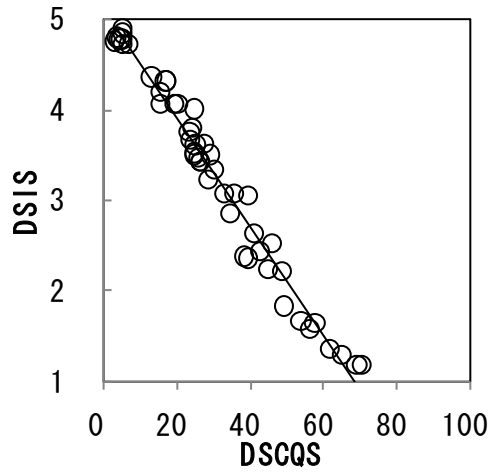
**【ACR5-HR】**



**【SAMVIQ-HR】**



# Correlation between MOS and DSCQS (Exp. 2)



## R (Exp. 2)

- ▶ There are no significant differences in average of Rs between each two methods with a 5% level of significance.

|                 | DSCQS       | DSIS        | ACR5-HR     | SAMVIQ-HR   | ACR5        | SAMVIQ      |
|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|
| DSCQS           | –           | –0.99       | –0.99       | –0.98       | –0.97       | –0.96       |
| DSIS            | –0.99       | –           | 0.99        | 0.99        | 0.97        | 0.97        |
| ACR5-HR         | –0.99       | 0.99        | –           | 0.99        | 0.98        | 0.97        |
| SAMVIQ-HR       | –0.98       | 0.99        | 0.99        | –           | 0.98        | 0.99        |
| ACR5            | –0.97       | 0.97        | 0.98        | 0.98        | –           | 0.99        |
| SAMVIQ          | –0.96       | 0.97        | 0.97        | 0.99        | 0.99        | –           |
| <b>Average*</b> | <b>0.98</b> | <b>0.98</b> | <b>0.99</b> | <b>0.99</b> | <b>0.98</b> | <b>0.98</b> |

\*: Average to absolute value of correlation coefficients

## rank R (Exp. 2)

- ▶ There are no significant differences in average of rank Rs between each two methods with a 5% level of significance.

|           | DSCQS       | DSIS        | ACR5-HR     | SAMVIQ-HR   | ACR5        | SAMVIQ      |
|-----------|-------------|-------------|-------------|-------------|-------------|-------------|
| DSCQS     | –           | 0.99        | 0.98        | 0.96        | 0.94        | 0.93        |
| DSIS      | 0.99        | –           | 0.99        | 0.97        | 0.96        | 0.95        |
| ACR5-HR   | 0.98        | 0.99        | –           | 0.98        | 0.95        | 0.95        |
| SAMVIQ-HR | 0.96        | 0.97        | 0.98        | –           | 0.97        | 0.99        |
| ACR5      | 0.94        | 0.96        | 0.95        | 0.97        | –           | 0.98        |
| SAMVIQ    | 0.93        | 0.95        | 0.95        | 0.99        | 0.98        | –           |
| Average   | <b>0.96</b> | <b>0.97</b> | <b>0.97</b> | <b>0.97</b> | <b>0.96</b> | <b>0.96</b> |

## MCI<sub>norm</sub> (Exp. 2)

---

- ▶ MCI<sub>norm</sub> of the 6 methods are almost same.

|                     | DSCQS | DSIS | ACR5-HR | SAMVIQ-HR | ACR5 | SAMVIQ |
|---------------------|-------|------|---------|-----------|------|--------|
| MCI                 | 4.07  | 0.22 | 0.21    | 4.06      | 0.20 | 4.06   |
| MOS range           | 57.21 | 3.17 | 3.00    | 56.21     | 2.71 | 51.79  |
| MCI <sub>norm</sub> | 0.07  | 0.07 | 0.07    | 0.07      | 0.07 | 0.08   |

## Consideration for Exp. 2

---

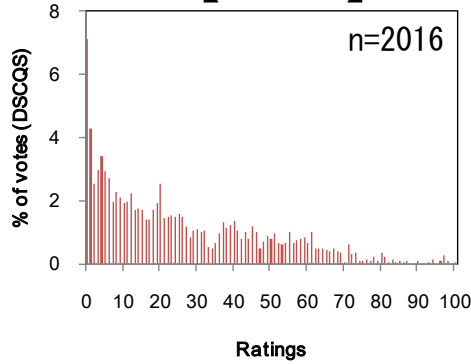
- ▶ For QVGA coding distortion quality assessment, combinations of all methods show high correlations and high rank correlations, there are no differences about R and rank R among methods.



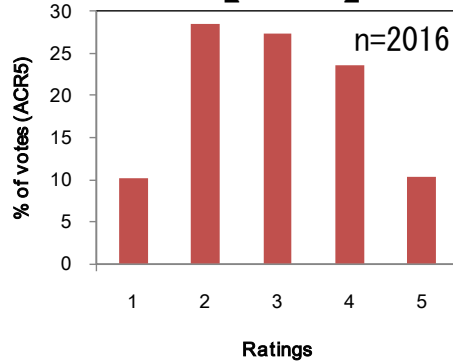
# Experimental results (Exp. 3)

# Score distributions (Exp. 3 QVGA)

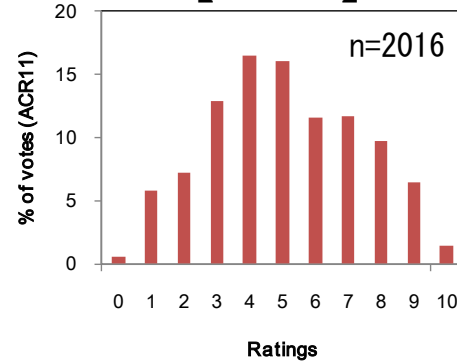
**【DSCQS】**



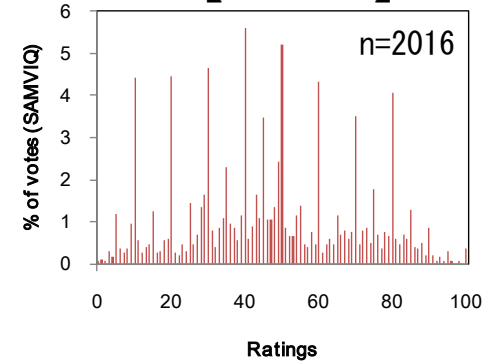
**【ACR5】**



**【ACR11】**



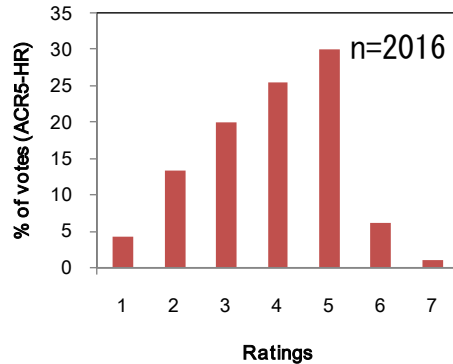
**【SAMVIQ】**



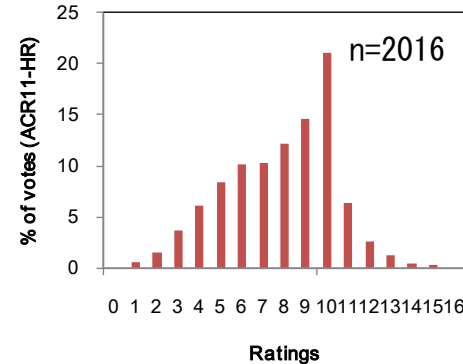
**【DSIS】**



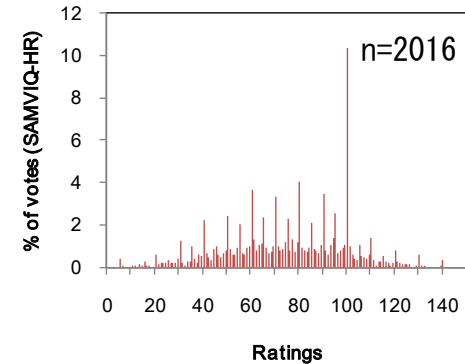
**【ACR5-HR】**



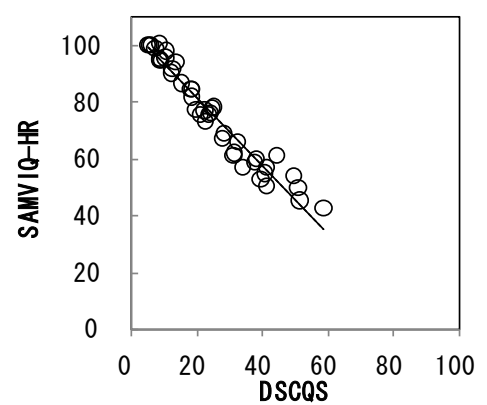
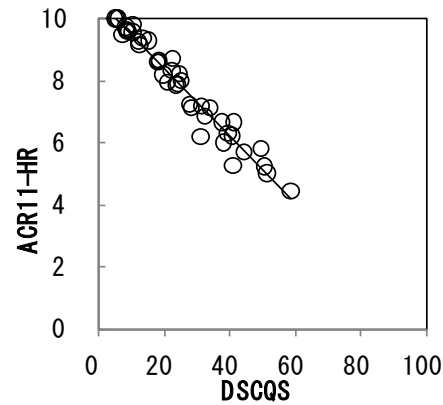
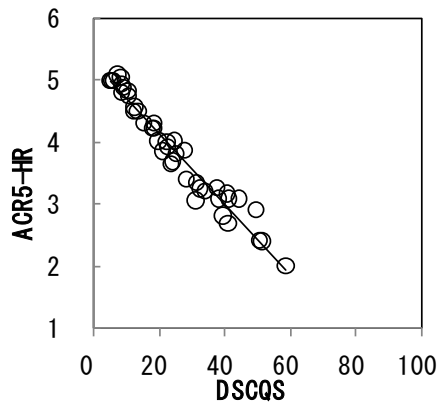
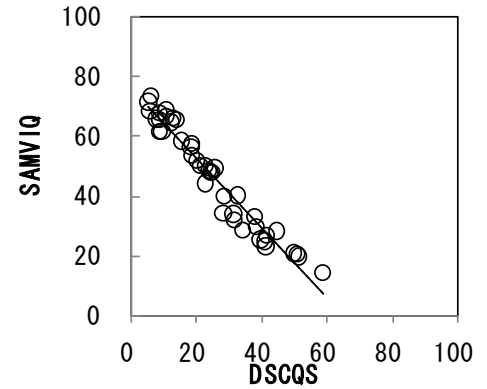
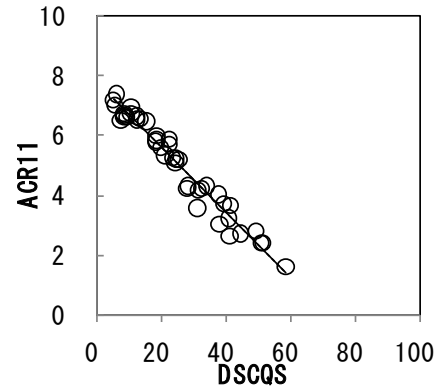
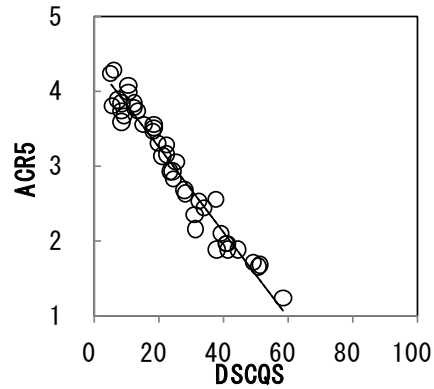
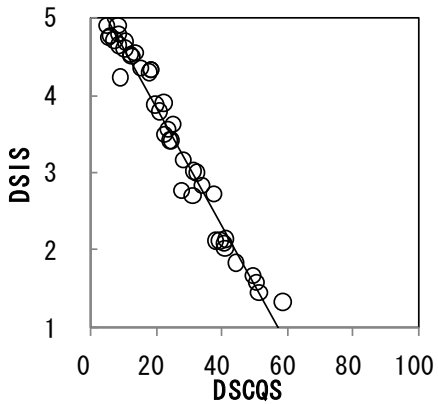
**【ACR11-HR】**



**【SAMVIQ-HR】**



# Correlation between MOS and DSCQS (Exp. 3)



## R (Exp. 3)

- ▶ There are no significant differences in average of Rs between each two methods with a 5% level of significance.

|                 | DSCQS       | DSIS        | ACR5<br>-HR | ACR11<br>-HR | SAMVIQ<br>-HR | ACR5        | ACR11       | SAMVIQ      |
|-----------------|-------------|-------------|-------------|--------------|---------------|-------------|-------------|-------------|
| DSCQS           | –           | –0.98       | –0.98       | –0.98        | –0.97         | –0.98       | –0.98       | –0.98       |
| DSIS            | –0.98       | –           | 0.96        | 0.98         | 0.97          | 0.98        | 0.98        | 0.98        |
| ACR–HR          | –0.98       | 0.96        | –           | 0.98         | 0.98          | 0.97        | 0.97        | 0.97        |
| ACR11–HR        | –0.98       | 0.98        | 0.98        | –            | 0.97          | 0.97        | 1.00        | 0.97        |
| SAMVIQ–HR       | –0.97       | 0.97        | 0.98        | 0.97         | –             | 0.97        | 0.97        | 0.99        |
| ACR             | –0.98       | 0.98        | 0.97        | 0.97         | 0.97          | –           | 0.98        | 0.98        |
| ACR–11          | –0.98       | 0.98        | 0.97        | 1.00         | 0.97          | 0.98        | –           | 0.98        |
| SAMVIQ          | –0.98       | 0.98        | 0.97        | 0.97         | 0.99          | 0.98        | 0.98        | –           |
| <b>Average*</b> | <b>0.98</b> | <b>0.98</b> | <b>0.97</b> | <b>0.98</b>  | <b>0.98</b>   | <b>0.98</b> | <b>0.98</b> | <b>0.98</b> |

\*: Average to absolute value of correlation coefficients

## rank R (Exp. 3)

- ▶ There are no significant differences in average of rank Rs between each two methods with a 5% level of significance.

|           | DSCQS       | DSIS        | ACR5<br>-HR | ACR11<br>-HR | SAMVIQ<br>-HR | ACR5        | ACR11       | SAMVIQ      |
|-----------|-------------|-------------|-------------|--------------|---------------|-------------|-------------|-------------|
| DSCQS     | –           | 0.98        | 0.98        | 0.97         | 0.97          | 0.97        | 0.97        | 0.98        |
| DSIS      | 0.98        | –           | 0.97        | 0.97         | 0.97          | 0.97        | 0.97        | 0.98        |
| ACR5-HR   | 0.98        | 0.97        | –           | 0.98         | 0.98          | 0.97        | 0.96        | 0.96        |
| ACR11-HR  | 0.97        | 0.97        | 0.98        | –            | 0.97          | 0.97        | 0.99        | 0.96        |
| SAMVIQ-HR | 0.97        | 0.97        | 0.98        | 0.97         | –             | 0.97        | 0.96        | 0.98        |
| ACR5      | 0.97        | 0.97        | 0.97        | 0.97         | 0.97          | –           | 0.98        | 0.99        |
| ACR11     | 0.97        | 0.97        | 0.96        | 0.99         | 0.96          | 0.98        | –           | 0.97        |
| SAMVIQ    | 0.98        | 0.98        | 0.96        | 0.96         | 0.98          | 0.99        | 0.97        | –           |
| Average   | <b>0.98</b> | <b>0.97</b> | <b>0.97</b> | <b>0.97</b>  | <b>0.97</b>   | <b>0.97</b> | <b>0.97</b> | <b>0.98</b> |

## MCInorm (Exp. 3)

- ▶ DSIS, ACR5, and SAMVIQ has the smallest MCInorm of the 8 methods.
- ▶ MCInorm of ACR5–HR, ACR11–HR, and SAMVIQ–HR are larger than that of ACR5, ACR11, and SAMVIQ.
  - ▶ It is thought that this caused by the differences in MCInorm values between ACR5–HR/ACR5, ACR11–HR/ACR11, and SAMVIQ–HR/SAMVIQ so that HR may use and calculate two values of PVS and REF.

|           | DSCQS | DSIS | ACR5–HR | ACR11–HR | SAMVIQ–HR | ACR5 | ACR11 | SAMVIQ |
|-----------|-------|------|---------|----------|-----------|------|-------|--------|
| MCI       | 4.58  | 0.25 | 0.28    | 0.56     | 4.87      | 0.23 | 0.47  | 4.26   |
| MOS range | 53.54 | 3.56 | 3.08    | 5.56     | 58.13     | 3.04 | 5.77  | 59.21  |
| MCInorm   | 0.09  | 0.07 | 0.09    | 0.10     | 0.08      | 0.07 | 0.08  | 0.07   |

## Consideration for Exp. 3

---

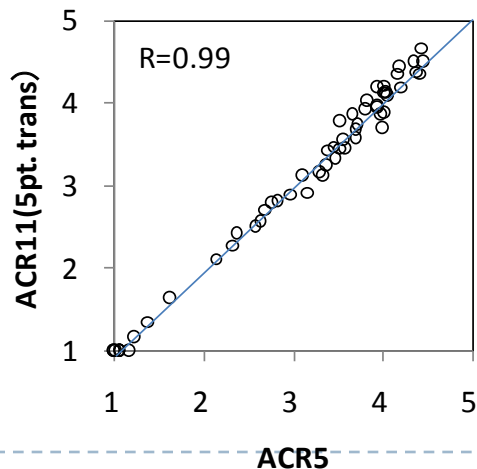
- ▶ Combinations of all methods show high correlations and high rank correlations, there are no differences about R and rank R among methods.
- ▶ DSIS, ACR5, and SAMVIQ have the smallest MCInorm of the 8 methods.
  - ▶ Because HR uses two assessment values (i.e., PVS and REF) by definition, MCInorm tends to be larger for HR methods than non-HR ones.

# Additional analysis



# Results compared to those of Ref. [3] (Exp. 1 HD)

- ▶ Scores of ACR11 transform to that of ACR5.
  - ① Correlation coefficient of MOSs is 0.99.  
(Ref. [3]: 0.99)
  - ② No statistical differences between mean quality.  
(Ref. [3]: No statistical differences)
  - ③ There is a statistical difference between MCIs.  
(Ref. [3]: No statistical differences)

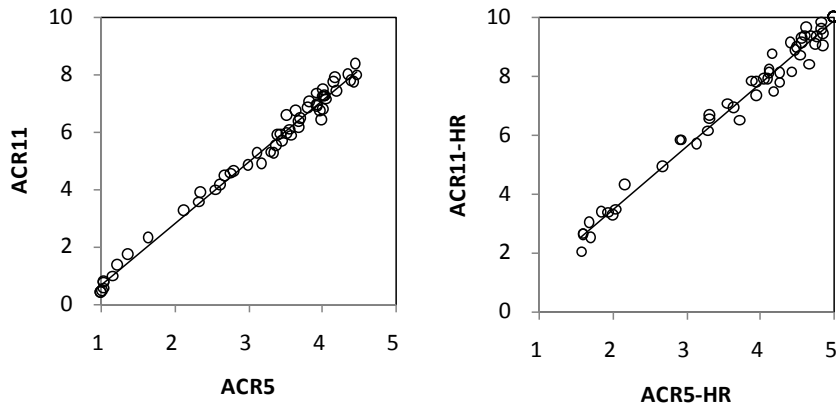


|          | ACR5 | ACR11<br>(5pt. transform) |
|----------|------|---------------------------|
| Mean MOS | 3.15 | 3.16                      |
| MCI      | 0.19 | 0.21                      |

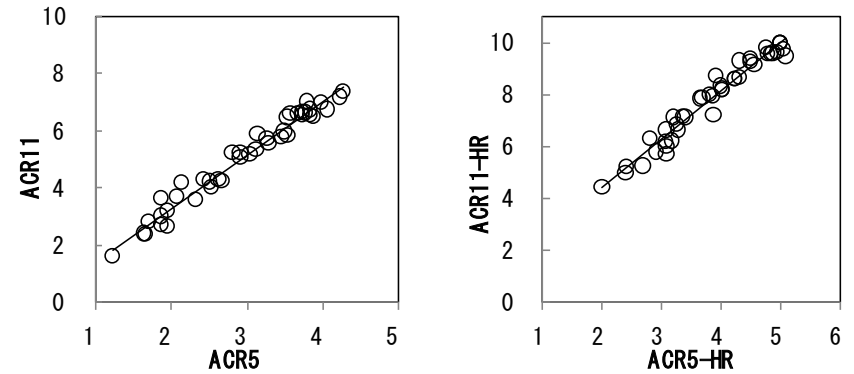
# ACR5 vs. ACR11

- ▶ ACR5 and ACR11 have almost same assessment performance.

Exp. 1 (HD)



Exp. 3 (QVGA)



|        | ACR5 : ACR11 | ACR5-HR : ACR11-HR |
|--------|--------------|--------------------|
| R      | 0.99         | 0.98               |
| rank R | 0.98         | 0.98               |

|                     | ACR5 | ACR11 | ACR5-HR | ACR11-HR |
|---------------------|------|-------|---------|----------|
| MCI <sub>norm</sub> | 0.06 | 0.05  | 0.06    | 0.06     |

|        | ACR5 : ACR11 | ACR5-HR : ACR11-HR |
|--------|--------------|--------------------|
| R      | 0.98         | 0.98               |
| rank R | 0.98         | 0.98               |

|                     | ACR5 | ACR11 | ACR5-HR | ACR11-HR |
|---------------------|------|-------|---------|----------|
| MCI <sub>norm</sub> | 0.07 | 0.08  | 0.09    | 0.10     |