Enriched in-band video : from theoretical modeling to new services for the society of knowledge

Maher Belhaj Abdallah

Ecole Doctorale EDITE

Thèse présentée pour l'obtention du diplôme de Docteur de Télécom & Management SudParis

Doctorat conjoint Télécom & Management SudParis et Université Pierre et Marie Curie

Spécialité : Télécommunications

Maher Belhaj abdallah

In-band enriched video: de la modélisation théorique aux nouveaux services pour la société des connaissances

Soutenue le 5 décembre 2011 devant le jury composé de :

Pr. Amel BENAZZA	Rapporteur
Pr. Azeddine BEGHDADI	Rapporteur
HDR. Claude DELPHA	Examinateur
Pr. Patrick GALLINARI	Examinateur
Eric MUNIER	Examinateur
Pr. Françoise PRETEUX	Directeur de thèse
HDR. Mihai MITREA	Co-directeur de thèse

Thèse n° 2011TELE0030

Table of Content

Introduction

"In the past, shoes could stink.

In the present, shoes can blink.

In the future, shoes will think."

Things That Think Consortium,

MIT (2010)

Applicative framework

One key challenge of the nowadays emerging Knowledge Society is the ubiquitous computing: the ambient world becomes digital media and should be intrinsically empowered with computing, storage and interaction capabilities, Figure Intro-1.

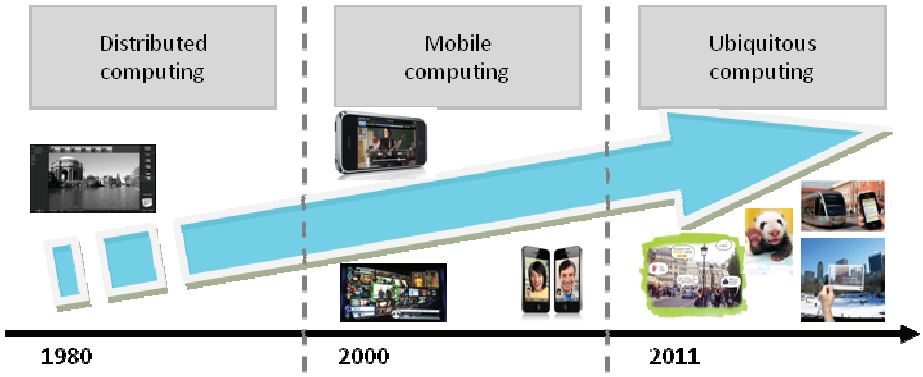


Figure Intro-1: From distributed to ubiquitous computing: the enriched man-machine or the machine-machine interaction is possible everywhere, on any device and at any time.

A significant component of this ecosystem is the digital content, considered in its widest acceptation: text, music/speech, still and animated images, 2D/3D graphics, immersive experience data ... Actually, from the user point of view, this is the brightest side of the Moon: he/she wants to solely interact with the content, while completely ignoring all the underlying technical components of the system (*i.e.* where/how his/her action is transferred to the content, where the processing actually takes place, *etc.*).

In this respect, media enrichment is a nowadays hot research topic, from both academic and industrial perspectives. The principle consists in associating to the basic data some additional information (metadata of any type, from sensorial to binary executable codes), thus making the content network/terminal/context-aware and allowing it to interact with other intelligent content/devices. Such content is gradually conquering the Internet of things [INT05], being involved in a large variety of applications, like interactive DTV, games, e-learning, and data mining.

Behind such applications, a wide range of players can be spotted out, from SMEs to large industrials and state/academic institutions. Since 2009, the Louvre Museum enhanced the visitor experience by gradually replacing the obsolete audio guides with new multimedia companions [BST11]. In August 2010, companies like BBC, HyperTV and LinkTV launched Contextually Enriched platforms that create a new www user experience by ensuring the coexistence (aggregation) of the text, video and picture; such a content is to be personalized and consumed by a client disposing of an intelligent terminal [HYP11], [LNK11]. BuilderMedia [MIT06] currently exploits indexing principles in order to establish a remote virtual connection between user and all required information about his/her content.

Research already developed within the ARTEMIS Department at Institut Télécom ; Télécom SudParis [MIT06] brought to light the new concept of *in-band enriched media*, Figure Intro-2. The principle is to insert the enrichment information in the original content itself, instead of

juxtaposing them as metadata. The challenge is to jointly observe the constraints of *imperceptibility* (*i.e.* the enrichment data should not alter the quality of the original content), of *persistency* (*i.e.* the enrichment data should serve their purpose even when the enriched content was subjected to severe alterations) and of *data payload* (*i.e.* ensuring a prescribed amount of enrichment data).

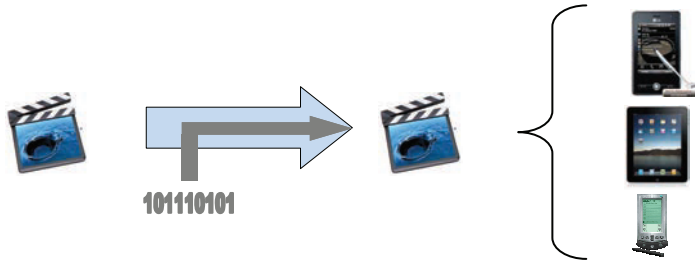


Figure Intro-2: The « *in-band enrichment* » principle: the enrichment data are imperceptibly and persistently inserted into the data to be enriched.

With respect to the traditional solutions, the *in-band enrichment* comes across with three main advantages: backward compatibility, format coherence, and virtually no network overhead.

As it can be noticed, the definition above can be considered as a general case of the digital watermarking, where a mark (some binary information) is imperceptibly and persistently inserted into some host data in order to serve the copyright applications; consequently, the mark can be considered as some enrichment data devoted to particular case of copyright protection.

This intrinsic connection between watermarking and in-band enrichment is already deployed in some pioneering industrial products.

For instance, Digimarc provided in 2010 the Discover Platform [DIG11], [DAV11] to enrich audio/video content for its further use in entertainment/e-commerce/medical assistance, Figure Intro-3. In this respect, the self-identification principle is considered: at the insertion side, the advertising posters are enriched with watermarks representing a link (a simple url) to a www page which is to be detected on the client smart phone. The client runs the detection procedure on a picture taken by his/her smart phone camera. The detection should be successful despite the errors in poster printing, poster degradation under atmospheric factors (Sun light, rain …) or the variable capturing conditions (A/D conversions, arbitrarily lighting, geometrical deformations induced by the capturing angle/distance, …). A similar mechanism is available for radio broadcasting. The *MediaSync* platform, jointly showcased by Digimarc and ABC, goes one step further and deals with audio/video enrichment for TV applications. The mobile viewers (smart phone/tablet) of the *My Generation* mockumentary[1] enjoyed video content enriched not only with basic interactivity mechanisms (quizzes, polls) but also with social media functionalities and with advertising features; this experience has been also extended to *Grey's Anatomy* soap opera.

Related products are also available from Clic2C [CLC11].

[1] A film that has the look and feel of a television documentary, but with the irreverent humor and slapstick of a comedy, designed to "mock" the documentary or subject it features.

Figure Intro-3: The *Discover Platform* promotional image [DIG11].

Theoretical framework

From the information theory point of view, any technique for information embedding can be modeled by a noisy channel [COX02], [ARN03], [KAT00], [DAV04], Figure Intro-4.

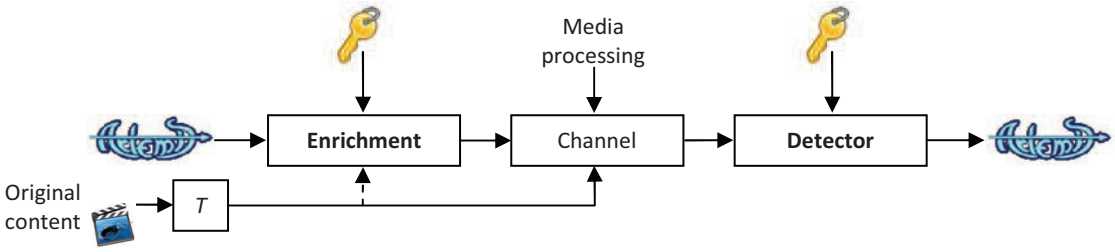


Figure Intro-4: *In-band enrichment* as a noisy channel: the enrichment information (*e.g.* the ARTEMIS logo) is encrypted with a key, then imperceptibly inserted into the original content. These enrichment data should be detected at the terminal side, despite the enriched media processing.

According to this model, the enrichment data represent a sample from the information source which is processed prior to the transmission, according to some deterministic operations given by the key and, eventually, the content to be enriched. The noise is represented by the data to be enriched and by the processing the enriched media suffers. Consequently, the proper theoretic model is noisy channel with non-causal side information at the embedder (the original content acts as a noise source completely known at the embedder but unknown at the detection) [COS83]. At the detection, the enrichment data should be retrieved from the corrupted enriched content, by using the key.

In practice, designing an in-band enrichment method means designing a transmission technique on the corresponding channel, *i.e.* finding the transmission technique able to afford the trade-off between the required data payload and the noise on the channel. Consequently, in-band enrichment comes with additional constraints on this general theoretical framework, from both conceptual and practical points of view.

In-band enrichment deployment

When considering the conceptual point of view, suspicions related to the enrichment data payload, imperceptibility, and persistency can arise.

On the one hand, the size of the enrichment data (expressed in bits) should be large enough so as to serve the targeted application; consequently, the transmission technique should allow a variable data payload (set according to each application) to be successively conveyed at the detector.

On the other hand, the *imperceptibility* and the *persistency* of the enrichment data lead to antagonist power constraints on the signal sent through the channel: generally, the lower the power of the inserted signal, the better the imperceptibility but the worse the persistency and *vice versa*. Actually, the *imperceptibility* exploits the redundancy of the original content in order to hide the enrichment data.

When considering the applicative point of view, issues connected to multimedia data representation should be dealt with. Actually, multimedia data are nowadays stored and exchanged in some widely accepted compressed formats. Consequently, accessing the media inside its file format (*e.g.* audio samples inside an mp3 file or video frames in an mp4 file) is a process requiring sophisticated (time & computation) operations which should not impair the performances of the enrichment applications. One possible solution would be to directly insert the enrichment data into the compressed stream but this would *a priori* be a fundamental contradiction: note that in a compressed stream there is no more redundancy left, so, there is no way to imperceptibly modify it.

Objectives

The present thesis aims at bridging the gap between a generic theoretical model and an effervescent applicative framework. In this respect, the applicative perimeter of the in-band enrichment is to be studied, in order to theoretically and practically address the following incremental issues, Figure Intro-5:

1. ***Viability***: is it possible for the *in-band enrichment* to be deployed for compressed video streams?
2. ***Feasibility***: how can be it demonstrated in practice?
3. ***Theoretical limits***: which are the limits of the *in-band enrichment*?

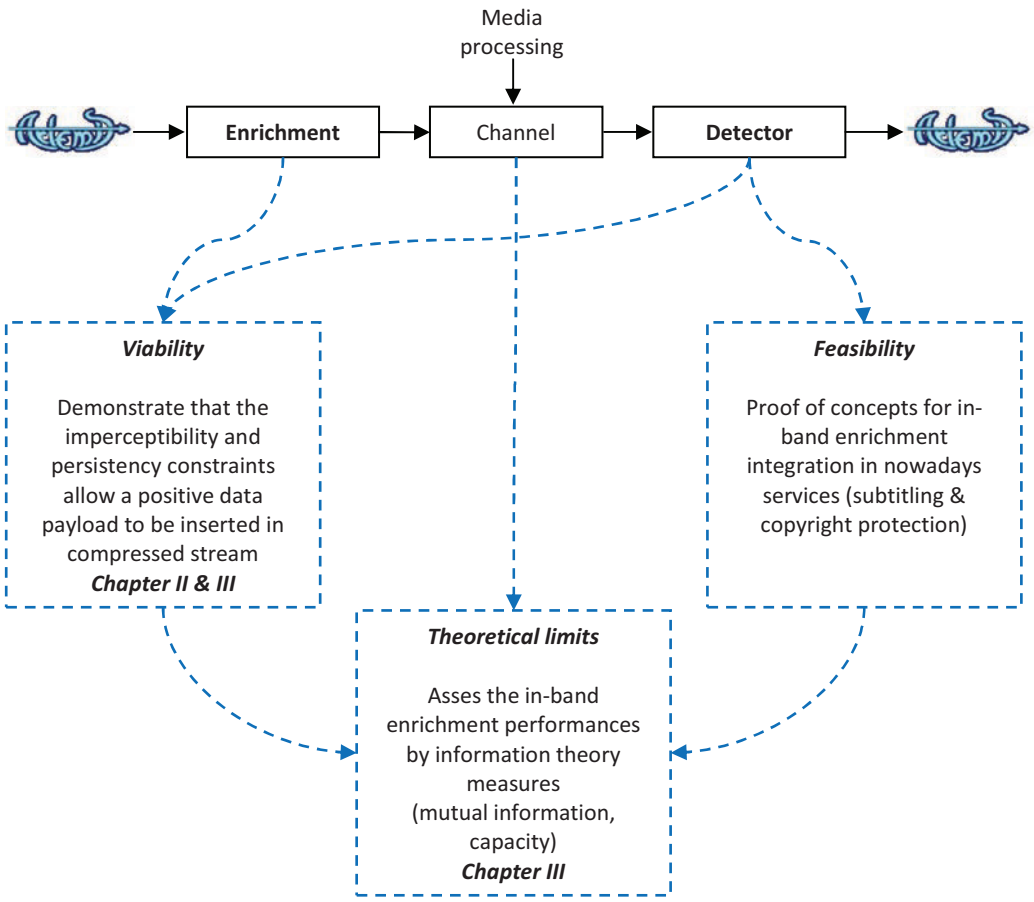


Figure Intro-5: Thesis main objectives.

Thesis structure

In order to achieve these objectives, Figure Intro-5, the manuscript is structured into four methodological chapters, preceded by this Introduction and a State-of-the-art and followed by Conclusions. Additionally, five Appendix sections will offer the relevant insights in the underlying applicative/experimental set-up.

The State-of-the-art (Chapter I) reconsiders the main concepts and results in watermarking (theoretical model, functional properties, support technologies) from the compressed domain in-band enrichment point of view.

Chapter II is devoted to the transparency concept. First, the very viability of the in-band enrichment concept is discussed. In this respect, it is first proved that the nowadays most powerful compression standard (MPEG-4 AVC, a.k.a. H.264) still allows the imperceptible modification of its syntax elements. This study requires the development of a general methodological framework, combining nine objective visual metrics, of different types (pixel-based, correlation-based and perceptual metrics). Chapter II also establishes the first perceptual masking model matched to the MPEG-4 AVC peculiarities. The related mathematical

demonstration starts from the basic Watson's model and extend it by considering a multivariate optimization procedure (under imperceptibility constraints) leading to the MPEG-4 AVC visibility thresholds, followed by an algebraic investigation of the prediction mode impact.

Chapter III aims at specifying the first MPEG-4 AVC compressed domain in-band enrichment methods. In this respect, the results are presented at three incremental levels. At the basis, the first MPEG-4 AVC method featuring robustness against noise addition, transcoding, and the StirMark (geometric) attacks is adavanced. The mark is inserted into the MPEG-4 AVC quantization indexes selected according to an energy-based selection criterion validated by information theory basic concepts. The insertion procedure combines the QIM principles and the perceptual mask obtained in this thesis. At the second level, the very QIM principles are generalized beyond the binary case, thus advancing the mQIM insertion/detection methods. Its main theoretical and practical advantage is the increase the data payload by a log(m) factor, while preserving the transparency and the robustness. Finally, the MPEG 4 AVC syntax is reconsidered as a starting point in designing a counter-attack procedure; in practice, this procedure results in BER reduction by 5% to 10%, according to the original video bit rate.

Beside this application driven presentation, Chapter III also deals with fundamental theoretical tools in information theory which allows the performances of the developed methods to be objectively compared to their theoretical limits.

Chapter IV merges the methodological results presented in Chapters II and III in order to demonstrate the real-life impact of the in-band enrichment. In this respect, applications related to subtitling and copyright protection are demonstrated.

Conclusions are drawn and perspectives are open in Chapter V.

The thesis also contains four Appendixes, devoted to an MPEG-4 AVC technical overview, to some software tolls allowing the access to the MPEG-4 syntax elements, to the video corpus considered in the experiments.

Main results

The main results of the thesis, structured according to the three main objectives previously mentioned, are synoptically displayed in Table Intro-1.

Table Intro-1: Thesis main results.

Viability	**Theoretical research**	▪ the first perceptual shaping model matched to the MPEG-4 AVC peculiarities; ▪ the first *m*QIM insertion/detection method
	Applicative research	▪ visibility limits for imperceptibly modifying the MPEG-4 AVC quantizing indexes; ▪ ComWat, the first watermarking method for the MPEG-4 AVC quantized indexes, reaching the trade-off among transparency (*e.g.* PSNR > 42dB), robustness (against transcoding and geometric attacks) and data payload (5 times larger than the DCI requirements); ▪ the first real-time MPEG-4 AVC method for MPEG-4 AVC quantized indexes featuring transparency (*e.g.* PSNR > 39dB), robustness against transcoding and data payload 5 times larger than the DCI requirements.
Feasibility	**Applicative research**	▪ prototypes demonstrating the in-band enrichment practical impact for subtitling and copyright protection.
Theoretical limits	**Theoretical research**	▪ noisy channel models (probability estimation with 95% confidence limits and relative errors lower than 5%) for compressed domain watermarking; ▪ estimating the related information theory measures (mutual information and capacity).

References

[ARN03] M. Arnold, M. Schmucker, S. Wolthusen. Techniques and Applications of Digital Watermarking and Content Protection. Artech House, 2003.

[BST11] http://www.bestofmicro.com/actualite/24309-louvre-guide-multimedia.html

[CLC11] http://www.clic2c.com/

[COS83] M. Costa. Writing on dirty paper. IEEE Trans. Inform. Theory, 29:439–441, 1983.

[COX02] I. Cox, Miller, J. Bloom. Digital Watermarking. Morgan Kaufmann Publishers, 2002.

[DAV04] F. Davoine, S. Pateux (sous la direction de). Tatouage de documents audiovisuels numériques. Lavoisier, 2004.
[DAV11] B. Davis "Signal Rich Art: Enabling the vision of Ubiquitous Computing" Din Proceedings IS&T/SPIE Electronic Imaging Science and Technology Media Watermarking, Security, and Forensics III January 2011.
[DIG11] http://www.digimarc.com/

[HYP11] http://hypertv.it/web/hypertv-corporate/home/en.

[INT05] International Telecommunication Union " internet of things" ITU internet report 2005.

[KAT00] S. Katenbeisser, F. Petitcolas. Information Hiding – Techniques for Steganography and Digital Watermarking. Artech House, 2000.

[LNK11] http://www.linktv.org/viewchange/technology.

[MIT06a] M. Mitrea, S. Duta, T. Zaharia, F. Prêteux, Ensuring multimedia content adaptability by means of data hiding techniques, Proc. SPIE, Vol. 6383, 2006, pp. 63830:1-8.
[MIT06b] M. Mitrea, S. Duta, M. Preda, F. Prêteux, In-band enriched video for interactive applications, WSEAS Trans. on Communications, Vol. 5(8), 2006, pp. 1528-1534.

Chapter I

State of the art

The future starts today, not tomorrow

Pope John Paul II

Abstract

This chapter reconsiders the main concepts and results in watermarking (theoretical model, functional properties, support technologies) from the compressed domain in-band enrichment point of view.

Contents

I.1. Overview

Since its first reference in 1992 by Andrew Tirkel and Charles Osborne [TIR92], the term *digital watermarking* has been referring to a continuously effervescent research and applicative field, bringing together scientists from information theory, communication and image processing worlds, Figure I-1.

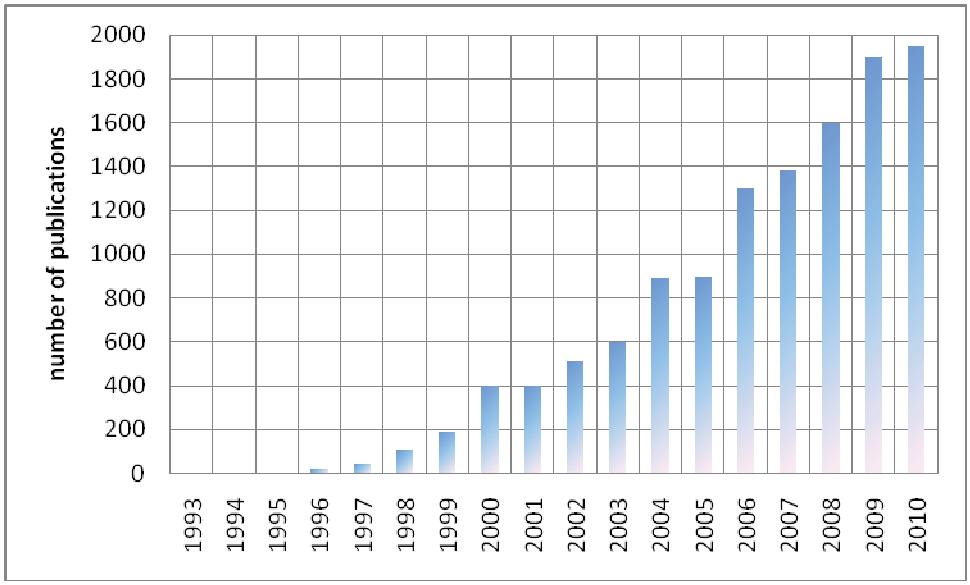


Figure I-1: Number of publications including *watermarking* as a key-word available in IEEEXplore, since 1993.

It its widest acceptation, this term regroups the applications where a mark (some binary information) is imperceptibly and persistently inserted into some host data (*e.g.* a video sequence), Figure I-2. The insertion of the watermark is always controlled by some secret information referred to as a *key*. Once watermarked, the host data can be transmitted and/or stored in a hostile environment, *i.e.* in an environment where changes may remove the watermark. The detection of the watermark provides evidence of the watermark presence and/or retrieves the watermark information bits. An oblivious watermarking algorithm does not require for the original host data to be available at the detection side.

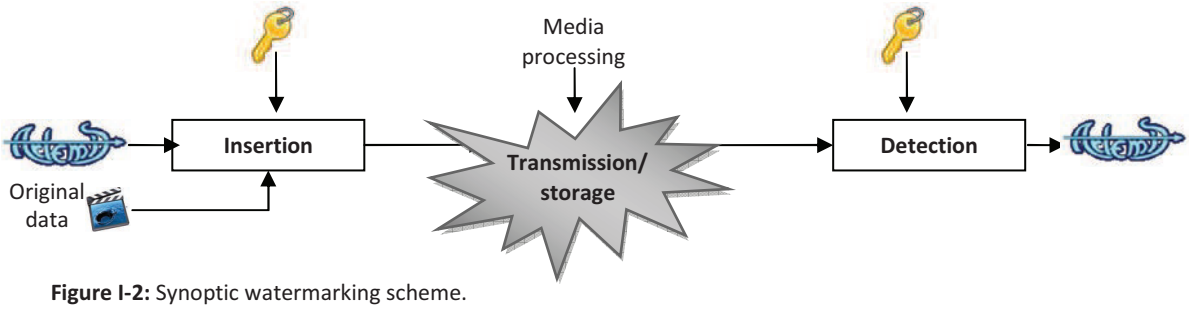


Figure I-2: Synoptic watermarking scheme.

While initially the mark size was practically restricted at some bits, the corresponding applications were bounded to the copyright world. However, recent advances pointed out that, at least from the theoretical point of view, the mark size can be big enough so as to serve in-band enrichment applications.

Under this general framework, directly inserting the mark into the compressed data is still a challenging research field, Figure I-3, with multiple-folded theoretical and applicative issues (as explained in the Introduction). Consequently, this chapter reconsiders the main concepts and results in watermarking (theoretical model, functional properties, support technologies) from the compressed domain in-band enrichment point of view.

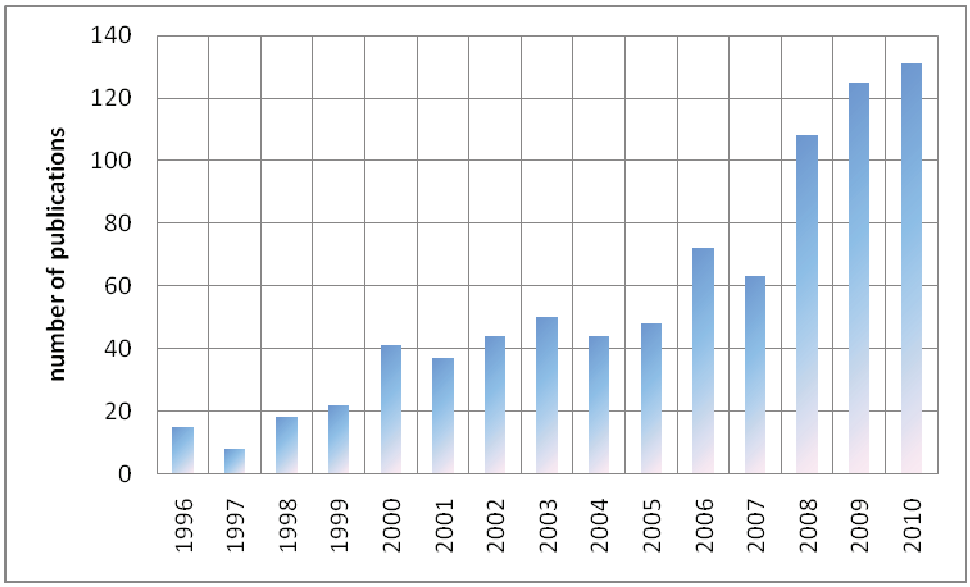


Figure I-3: Number of publications including *watermarking* and *compressed domain* as a key-words, available in IEEEXplore, since 1996.

I.2. Theoretical model

From the structural point of view, any watermarking procedure features three components: the watermark generation (*i.e.* the way in which the message to be inserted is encrypted with a secret key so as to obtain a watermarked), the watermark embedding (*i.e.* the way in which the watermark is inserted into the host document) and the watermark detection (*i.e.* the way in which the watermark is recovered).

From the theoretical point of view, any watermarking system can be presented as a noisy channel [COX08], [MIT07a], Figure I-4. Under this framework, the watermark generation is the sampling of the information source. The watermark insertion becomes the modulation technique, *i.e.* the way in which a given message is fit to the channel peculiarities. The original (host) data and all the transformations the marked data suffer emulate the channel. The watermark detection is modelled by the decision making procedure, implemented at the channel output.

From the functional point of view, the watermarking procedure is evaluated according to at least three requirements, namely the transparency, robustness and data payload, each of which comes across with antagonistic theoretical constraints. When considering transparency, the watermark insertion alters the user's multimedia experience and a high signal to noise ratio (where the host data represent the signal and the watermark is the noise) is desired (*e.g.* larger than 30 dB, in the case of still images). From the robustness point of view, the watermark detection is disturbed by the channel noise (the original content and the attacks). Consequently, a large signal to noise ratio (where, this time, the host data become the noise and the watermark is the signal to be recovered) is searched for. Note that the data payload results into a minimal value constraint on the mutual information.

The practical watermarking schemata are generally divided into two main classes, namely spread spectrum (SS) and side information (SI).

The SS methods have already been used in telecommunication applications (*e.g.* CDMA), providing a good solution for very low power signal transmission over noisy channels [COX95]. Consequently, an SS method will spread the mark across the host media, occupying a much larger bandwidth than strictly necessary. Thus, the mark becomes a very low power signal, practically undetectable in any frequency sub-band. In practice, this approach is very robust against attacks, but limited in terms of data payload [MIT07].

The SI principle [EGG03], [CHE98] states that a channel noise known at the transmitter and unknown at the receiver would not decrease the channel capacity (the maximum amount of information which can be theoretically transmitted). Thus, the original document should no longer be considered as an impediment to the watermark detection. Consequently, the side information watermarking is *a priori* optimal from the data payload point of view (under fixed transparency and robustness constraints). However, in practice, the methods following this approach allow the insertion of a very high quantity of information, but only have a very weak robustness.

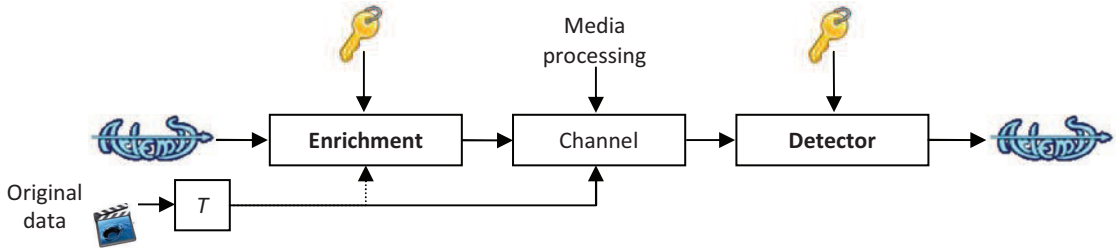


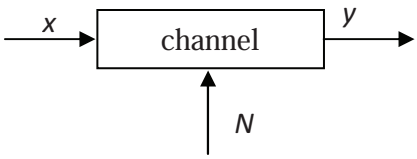
Figure I-4: *In-band enrichment* as a noisy channel: the enrichment information (*e.g.* the ARTEMIS logo) is encrypted with a key, and then imperceptibly inserted into the original data. This enrichment data should be detected at the terminal side despite the enriched media processing.

In general, to describe a noisy channel means to describe the probabilistic dependencies between the input and the output information sources, and to evaluate the average amount of information transmitted from input to output, Figure I-5. As the insertion of the watermark can be performed in different ways (additive/replacement ...) and in different domains of media representation (spatial, transformed, compressed ...), the underlying watermarking channel is of different types and of different mathematical models.

The discrete channel corresponds to a channel where the input x, the output y and the noise N information sources are discrete. The input and the output alphabets are not necessarily identical; be they denoted by $(x_1, x_2, \ldots x_n)$ and $(y_1, y_2, \ldots y_n)$, respectively. An error means to receive a symbol that does not correspond to the emitted one. In this case the channel is described by a noise matrix *i.e.* by the matrix of the conditional probabilities p(y/x).

The continuous channel is characterised by continuous input, output and noise sources. In this case, the channel is described by its noise *pdf*, *i.e.*by the noise probability density function .

Note: These definitions correspond to a basic case, the zero-memory, time invariant channel [SHA48]. More details about the viability of such a model in watermarking applications will be presented in Chapter IV, where previous results [DUT07], [DUM07], [COX95], are reconsidered from the compressed domain watermarking point of view.



Discrete

$p(y/x)$	y_1	y_2
x_1	0.08	0.92
x_2	0.89	0.11

Continuous

$$p_N(n) = \sum_{i=1}^{10} \frac{P(k_i)}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(n-\mu_i)^2}{2\sigma_i^2}\right)$$

$P(k_i)$	μ_i	σ_i
0.008 0.089	0.145.. 0.005	0.020 0.014
0.234 0.211	0.001 0.002	0.001 0.006
0.017 0.049	-0.023 -0.054	0.024 0.018
0.161 0.042	0.013 0.067	0.009 0.015

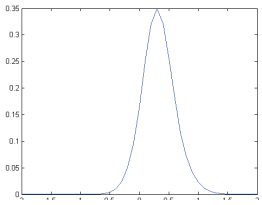


Figure I.5: Noise matrices for a discrete channel (corresponding to the transcoding of an MPEG-4 AVC watermarked video [BEL10]), left, and for a continuous channel (corresponding to the Stirmark attack applied in the largest DWT coefficient, [MIT07b], [DUM10],), right

I.3. Watermarking features

In addition to the three basic requirements already mentioned (transparency, robustness and data payload), the practical deployment of any watermarking system should also consider issues relating to the probability of false alarm, cost and key management. These six inter-dependent yet conflicting aspects, Figure I-6, will be further discussed.

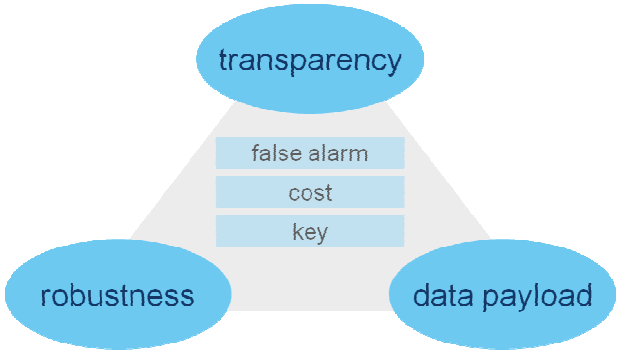


Figure I-6: Properties required for a watermark.

I.3.1. Transparency

The notion of transparency is related to perception (visual, auditory ...) of artifacts resulted from the insertion process. In literature, a distinction is done between perceptual fidelity and quality of the watermarking algorithm [Cox08], Figure I-7. A watermark is said to feature fidelity if the insertion artifacts cannot be perceived by a human observer. When the watermarked data have noticeable yet un-disturbing artifacts, they feature quality.

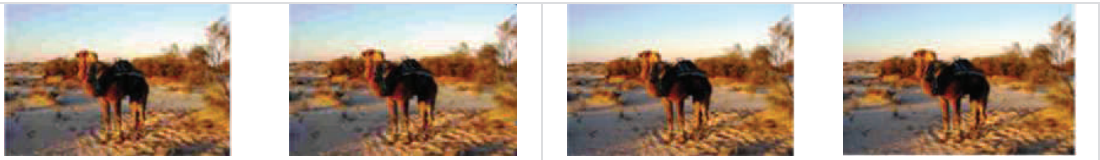


Figure I-7: The watermarking transparency: fidelity *vs.* quality.

Sometimes, the fidelity is not a strict requirement and the watermarked video has only to feature quality [COX02]. For instance, this is the case when the watermarked video is transmitted over low rate channels (*e.g* web 2.0 applications like YouTube and Metacafe, analog television, *etc.*). Conversely, in digital HDTV or DVD video, artifacts are easily perceptible; hence, the fidelity constraint should be imposed for the watermarking procedure.

The visual quality assessment of enriched data remains an important criterion for validating not only the watermarking algorithm itself, but also the complete media protection chain (compression, watermarking, transmission, *etc.*). However, it is a subjective concept that depends on various criteria: human visual system (see Chapter III.1.a), age, experience, artistic sense, observation

conditions. Thus, it is complicated to evaluate whether a watermarking method is transparent or not. Such an evaluation requires significant testing involving a wide observer's panel and many visual assessment tests [MAN74], Figure I-8. An adequate alternative is the use of objective transparency metrics related on models of human visual systems, Chapter III.1.b.

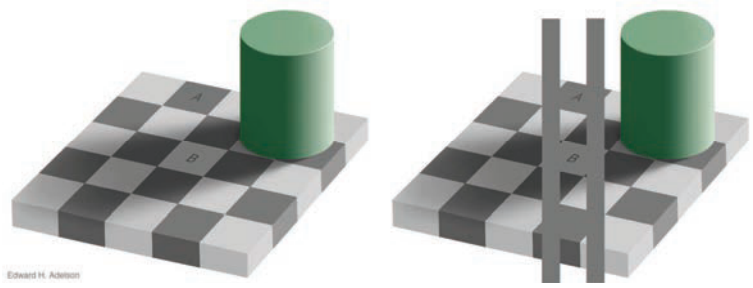


Figure I-8: Adelson's Checker-shadow illusion (by extending the similar picture in [ADE10]): although they seem different (left) the squares labeled A and B are in the same shade of gray (right).

I.3.1.a. Human visual system

The human visual system (HVS) is very complex: 80-90% of neurons in the brain are related to vision [YOU91]. The HVS depends essentially on the eye that captures light and converts it into a signal readable by the neurons. The concepts underlying the HVS functionalities are:

The contrast sensitivity

The response of the human visual system depends less on the absolute brightness of an image point than the difference in brightness of this point from its neighborhood. This property is known as Weber- Fechner law [LEG80] which measures the contrast defined by:

$$C^w = \frac{\Delta L}{L}, \qquad \text{I.1}$$

where ΔL is the minimum perceptible difference in luminance between two adjacent areas.

Neurons respond only to stimuli with a contrast greater than a specified threshold, called the contrast threshold (a.k.a inverse of the threshold contrast sensitivity). The eye is more sensitive to low spatial frequencies than to high frequency. The contrast sensitivity function is a harmonic function of spatial&temporal frequency, orientation and color.

The perception of color

In order to understand the perception of color, we should know the spectral characteristics of light. In 1666, Isaac Newton [NEW89] discovered that when white light passes through a prism, it separates into all of the visible colors and that each color is made up of a single wavelength. Further experiments demonstrated that light could be combined to form other colors. Prepared by Grassmann [GRA53], the grouping of colors satisfies homogeneity (both color produced by light of homogeneous wavelength) and overlay (both color obtained by mixing lights of different wavelengths lengths) and can be analyzed using the theory of linear systems. Hering was the first to show [HER78] that some pairs of color may form a single color sensation. For example, red is seen as yellow-orange, but no color instance is seen as red-green. Hering concluded that there are three primary psychological sensations corresponding to pairs of antagonist colors blue/yellow, green/red and white/black. These different sensations are encoded as a signal combination difference between the three primary colors. This conclusion is

inferred from experiments performed to quantify the opposite colors, called experiments cancellation of color [HUR54]. In these experiments, the observer can cancel the red of the tested light by adding the color green. In this way the color red, green, yellow and blue can be measured.

The reason we represent the color information signals in opposite colors in the human visual system is to exploit the decorrelation between colors, in order to define the best adequate color system to present the media information.

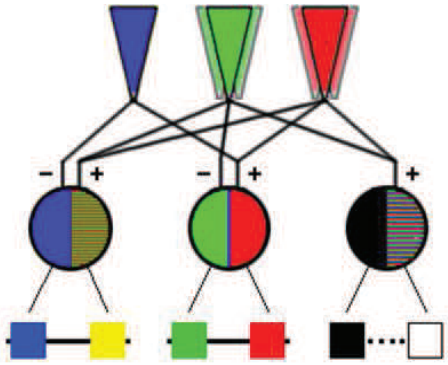


Figure I-9: Recoding opponent color [COR10].

The perceptual masking

Masking is a very important phenomenon for multimedia processing: it goes beyond the perception of individual colors and describes the interaction between multiple stimuli.

The masking effect occurs when a stimulus (spatial frequency, temporal, color, *etc*...) cannot be detected because of the presence of another. In its widest sense, a perceptual mask gives some threshold values connected to the limits under which (or above which) modification in the signal representation can be perceived.

The space masking effects are often quantified by measuring the detection threshold of a given stimulus [LEG80]. Figure I-10 shows the evolution of the logarithm of contrast sensitivity versus the spatial frequency based on Mannos Model [MAN74]: for each frequency, the maximum of spatial amplitude at which variation is perceived is presented.

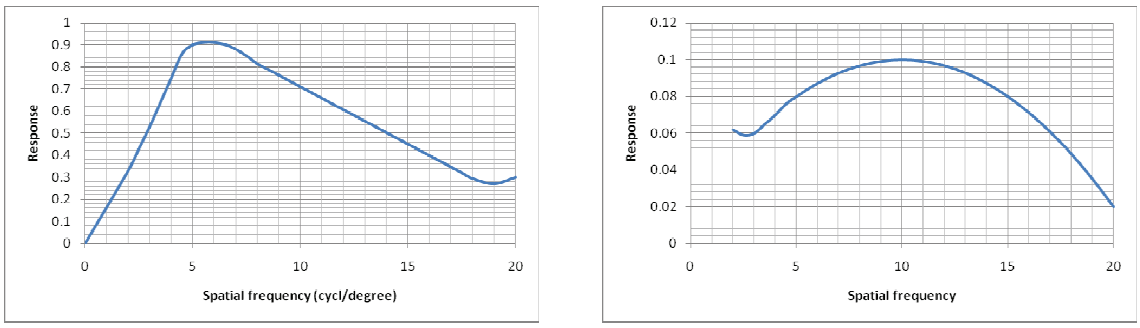


Figure I-10: Illustration of the sensitivity curve of the HVS to spatial (left) and temporal variation –right).

The temporal masking is based on a phenomenon of rising threshold of visibility from the temporal discontinuities such as the changing scenes. The first study of the temporal sensitivity function was

done by Seyler and Budrikis [SEY65] and concluded that raising the limit may take a few hundred milliseconds after a transition from darkness to light. Tam et al. [TAM95] studied the visibility of artifacts in the MPEG-2 after a change of scene and found a significant visual effect of masking only in the first frame. Carney et al. [CAR96] have shown that masking effects are strongly clear if there is a polarity (difference) between the mask and the target. The opposite of temporal masking (facilitation) can occur for the discontinuities in low contrast [GIR89].

I.3.1.b. Distortions & objective quality metrics

During the media processing, several kinds of distortions occur. These different artifacts can be classified as follows [YU98]:

- **block effects,** as a result of the quantification of individual blocks, independently with respect to their neighbors;
- **loss of spatial details and mitigating the contours**, after the suppression of high frequencies;
- **staircase effect**, induced by coarse quantization coefficients for very high frequencies, resulting in some slanted lines blurred in the reconstructed image;
- **oscillations**, due to Gibbs phenomenon occurring during image sharpening;
- **jagged motion effect**, as a consequence of the high motion prediction residual error caused by a coarse quantization;
- **mosquito noise effect**, induced by the temporal fluctuations of luminance and chrominance around the edges of high frequencies or moving objects;
- **blinking effect,** when a highly textured scene is not quantified with highly variable over time;
- **spectrum overlapping** after an un-optimal spatial sub-sampling;
- **analog/digital conversion** generally required during transmission/storage;
- **deinterlacing**, when a progressive sequence is created from an interlaced video sequence [HAA98].

In order to assess the impact of these artifacts from the human visual point of view, objective quality measures are to be defined. An objective measure is a function that takes as input some video information, calculates the distance to some reference information (be it extracted from a reference video sequence or not) and outputs a value.

While there is no conceptual limit in defining such a measure, the literature mainly considers measures which proved their ability to express quality measures related to HVS, at least for compression applications and the types of distortions mentioned above:

- ***Pixels based measures*** such as the signal to noise ratio (*PSNR*), the maximum mean square error (*PMSE*), the image fidelity (*IF*) and the average absolute difference (*AAD*).
- ***Correlation based measures*** as the normalized cross-correlation (*NCC*), the correlation quality (*CQ*), and the structural content (SC).
- ***Psychovisual measure***, namely the digital video quality (*DVQ*).

These measures will be defined in the sequel. By there S and $\hat{S}$ two video to be compared, each of

them having N_f frames, each frame of $W \times H$ pixels. Be $S_{i,j,k}$ the pixel at column *i* and row *j* of frame *k* ($1 \le i \le W, 1 \le j \le H, 1 \le k \le N_f$).

PSNR and the PMSE are defined as follows:

$$PSNR(S,\hat{S}) = \frac{1}{N_f}\sum_{k=1}^{N_f} 10\log_{10}\left(\frac{W \cdot H \cdot \max(S_{i,j,k}^2)}{\sum_{i=1}^{W}\sum_{j=1}^{H}(S_{i,j,k}-\hat{S}_{i,j,k})^2}\right), \quad \text{I.2}$$

$$PMSE(S,\hat{S}) = \frac{1}{W \cdot H \cdot N_f}\sum_{k=1}^{N_f}\frac{1}{\max(S_{i,j,k}^2)}\sum_{i=1}^{W}\sum_{j=1}^{H}(S_{i,j,k}-\hat{S}_{i,j,k})^2. \quad \text{I.3}$$

These two measures represent the difference between the two sequences. The higher the values for PSNR, the greater the similarity between images; of course, this statement should be reversed in the case of PMSE. A value of PSNR above 30 dB is generally acceptable for a good resemblance between the two tested video sequences while a value greater than 40 dB indicates an excellent likeness.

Simple and fast, PSNR and PMSE are widely used in coding schemes; however, their inner limitation is the decorrelation with the subjective visual quality. For example, it happens that the subjective quality of the image is improved by the addition of noise while the PSNR decreases. Another example is where the visibility of distortion depends on the background of the image (the property of masking), Figure I-11. The same observation applies to the PMSE.

Original image

Additive noised image
PSNR = 30 dB

Substitutive noised image
PSNR = 30 dB

Figure I-11: The same PSNR value (30 dB) may correspond to completely different visual qualities.

The *IF* is given by the following expression:

$$IF(S,\hat{S}) = \frac{1}{N_f}\sum_{k=1}^{N_f}\left(1-\frac{\sum_{i=1}^{W}\sum_{j=1}^{H}(S_{i,j,k}-\hat{S}_{i,j,k})^2}{\sum_{i=1}^{W}\sum_{j=1}^{H}(S_{i,j,k}{}^2)}\right). \quad \text{I.4}$$

As it can be seen, *IF* is a non-logarithmic version of *PSNR*, less sensitive to local contrast; the identity between the two videos tested corresponds to the value 1.

The *AAD* is expressed by the following equation:

$$AAD(S,\hat{S}) = \frac{1}{N_f}\sum_{k=1}^{N_f}\frac{\sum_{i=1}^{W}\sum_{j=1}^{H}\left|S_{i,j,k}-\hat{S}_{i,j,k}\right|}{W\cdot H}. \quad \text{I.5}$$

As it can be seen, it is also a measure of the magnitude of the error between the two videos. Compared with the PSNR, it corresponds to a linear and non-normalized difference. While a zero value always indicates the identity between two sequences, the meaning of a particular numerical value depend on the original pixel dynamic and is sometimes difficult to be interpreted. For example, for pixels coded on 8 bits, a value of 25 indicates a *AAD* average distortion of 10% (*i.e.* visible distortions); should the pixel be coded on 16 bits, the same value of 25 denotes unnoticeable artifacts.

The measures based on correlation seek a statistical matching between the tested video and reference video. They are expressed as follows:

$$NCC(S,\hat{S}) = \frac{1}{N_f}\sum_{k=1}^{N_f}\frac{\sum_{i=1}^{W}\sum_{j=1}^{H}(S_{i,j,k}\cdot\hat{S}_{i,j,k})}{\sum_{i=1}^{W}\sum_{j=1}^{H}S_{i,j,k}^{2}}, \quad \text{I.6}$$

$$CQ(S,\hat{S}) = \frac{1}{N_f}\sum_{k=1}^{N_f}\frac{\sum_{i=1}^{W}\sum_{j=1}^{H}(S_{i,j,k}\cdot\hat{S}_{i,j,k})}{\sum_{i=1}^{W}\sum_{j=1}^{H}S_{i,j,k}}, \quad \text{I.7}$$

$$SC(S,\hat{S}) = \frac{1}{N_f}\sum_{k=1}^{N_f}\frac{\sum_{i=1}^{W}\sum_{j=1}^{H}(S_{i,j,k}^{2})}{\sum_{i=1}^{W}\sum_{j=1}^{H}(^{2}\hat{S}_{i,j,k}^{2})}. \quad \text{I.8}$$

The higher the value of the *CQ*, the greater the resemblance between the two videos. As for *AAD*, this depends on the number of bits to represent pixels (*i.e.* the maximum value of the pixels). The SC and the *NIC* are close to 1 when the changes between the two videos are unnoticeable.

The *DVQ* is the most popular objective measure specifically designed for video sequences, according to the HSV peculiarities [FIS95], [WAT01]. Its computation considers dedicated blocks for contrast filtering, color perception and perceptual masking. As it is a very sophisticated measure and as it plays an important role in the present thesis, DVQ is detailed in Appendix C.

I.3.2. Robustness

Robustness is the ability of the watermark to survive after the changes undergone by the host media. These changes (be there intentional or unintentional) define the set of attacks. The various possible attacks against watermarked videos can be structured into four classes [PET98], according to the way they act: removal attacks, geometric attacks, cryptographic attacks and protocol attacks, Figure I-12.

The removal attacks try to make the watermark unreadable in light of the complexity of the watermarking detection system (computing time and processing activity). This class includes attacks

by noise addition, denoising, transcoding, quantization, modulation, ...

The geometric attacks are meant to destroy the synchronization of the watermark. After such an attack, the watermark is still present in the video, but its location is unknown at the detector. Rotations, curvatures (bending), jitter of the pixels (pixel jitter) individually considered or combined into the Stirmark attack, fall into this category [PET00].

Protocol attacks aim to make watermark unusable by creating some ambiguities concerning the mark usage. Attacks by inversion and copy belong to this class. The former creates a false key so that by applying the detection procedure, the watermark indicates a different owner for the video. The latter copies the watermark in other videos, thus decreasing the credibility of the message associated with the video.

The cryptographic attacks try to manage the watermark (detect/copy/insert a new one) without knowledge of the secret key. One solution is a brute-force search. Another, known as the oracle attack is to create an unmarked version of the signal by exploiting the responses of a detector (assuming it is available). In any event, this type of attack is very restrictive in practice because of its computational complexity.

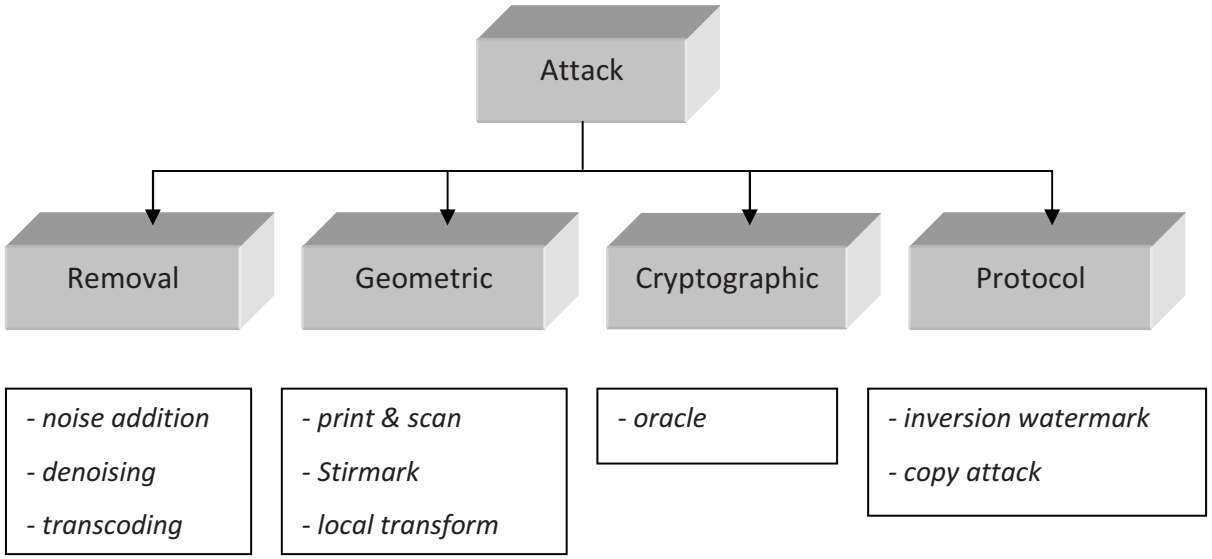


Figure I-12: Watermarking attack classification [COX08].

The literature shows that the most intensively considered attacks are the noise addition, the transcoding and the Stirmark. As these attacks are also intensively considered during the present thesis, they will be further discussed.

I.3.2.a. Additive Gaussian noise

This kind of attack has been largely addressed in communication and signal processing. It is considered (according to the Central Limit Theorem) to model the overall effect of various noise sources. Such a hypothesis is also considered in the watermarking field, where a lot of studies simply consider the Gaussian noise as a universal attack model. Although it can very effective for some real-life image/video filtering operations (*e.g.* linear filtering, JPEG compression, …), in-depth studies

proved its inaccuracy for other types of attacks (*e.g.* Stirmark, rotations, …) [MIT07a]. Note that this inaccuracy concerns the Gaussian behavior and not the additive hypothesis, which was practically each and every time validated by the experiments.

III.2.b Transcoding (lossy compression)

Lossy compression (MPEG-4 AVC, JPEG…) is an operation any multimedia content (be it marked or not) is likely to suffer before its transmission/storage.

In order to evaluate its effects, the differences between the original video and the video obtained after compression and decompression are investigated.

Note that there is a fundamental conflict between watermarking and lossy compression: a perfect compressor would virtually eliminate all the visual redundancy existing in the original video while the watermarking exploits this redundancy in order to hide the mark. Consequently, the practical challenge is to demonstrate the watermark detection is robust even against high rate compression.

I.3.2.c. Stirmark

Stirmark is a generic tool developed for robustness testing of image watermarking algorithms [PET00]. It simulates a D/A followed by an A/D conversion process, *i.e.* it introduces the same kind of errors into an image as printing it on a high quality printer and then scanning it with a low quality scanner. It also applies a minor geometric distortion: the image is slightly stretched, sheared, shifted and/or rotated by a random amount and then re-sampled using Nyquist interpolation, Figures I-13.

More distortions can be applied to a picture, Figure I-14 and Figure I-15: a global ‘bending’ and ‘random displacement’ to the image, followed by a slight deviation of each pixel (greatest at the centre of the picture and almost null at the corners) and a higher frequency displacement. Finally a transfer function that introduces a small and smoothly distributed error into all sample values is applied. In order to be most effective, a medium JPEG compression is applied after the distortions.

This attack was initially implemented by Markus G. Kuhn and enhanced and maintained by the Fabien Petitcolas [PET98].

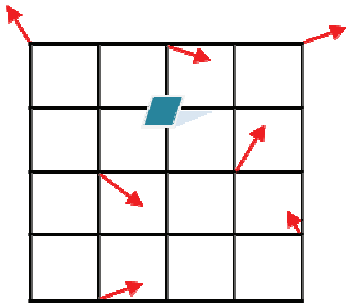
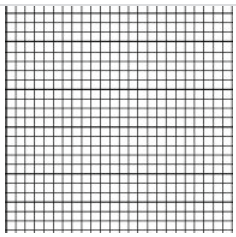


Figure I-13: Stirmark principle.

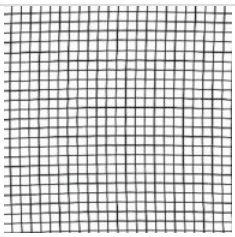
original

attacked

Figure I-14: Stirmak attack applied to natural images.



original



attacked

Figure I-15: Stirmak attack applied to test images.

I.3.3. Data payload

This is the total amount of information (in bits) inserted into the original content. According to the targeted applications, the specifications on this factor may be very different, from 64 bits per sequence for the identification of ownership up to hundreds of kilobits per frame for application of hyper-video, Table I-1:

- *copy protection*: the watermark identifies the content owner;
- *forensic tracking*: the watermark conveys a persistent ID pointing to related database;
- *broadcast and internet monitoring*: the watermark enables the broadcast or internet transmission to be monitored;
- *E-commerce*: the watermark includes the distributor identification;
- *authentification & integrity*: the watermark detects the content modification;
- *rights management*: the watermark identifies the content, includes the usage rules, billing

information of the rights manangement;

- *Hyper TV*: the watermark presents content related information (date of creation, history,related links…).

From the theoretical point of view, for prescribed transparency, the larger the data-payload, the lower the robustness.

Table I-1: In-band enrichment requirements on data-payload.

Application	*Data payload*
Copy protection	4 to 8 bits inserted in 5 min of video [COX08]
Forensic tracking	30 bit inserted in 2 min of video [PHI08]
Broadcast and internet monitoring	24 bit/s [COX08]
E-commerce	96 bit per video sequence [DIG09]
Authentification & integrety	200 bit per video sequence [XU05]
Rights management	31 bit per video sequence [DIG09]
Hyper TV	1000 bit per video sequence [MIT06a]

I.3.4. Probability of false alarm

The false alarm probability is the probability of detecting a watermark in unmarked data. This value should be arbitrarily low (*e.g.* below 10^{-6}), but in practice it can be depend on the application. For instance, a mobile service provider having 10 million subscribers should consider algorithms featuring P_f lower than 10^{-7}.

This probability depends on the watermark detection algorithm, the manner in which the detector is used, and the distribution of unwatermarked video. The problem of analyzing false detection behavior has received little attention in the watermark literature. Linnartz et al. [LIN07] provide a model to predict false positive probability in correlation-based watermarking methods and show that a non-white spectrum of a watermark causes the image content to interfere with watermark detection. Lichtenauer et al. study the false positive probability in exhaustive geometric searches [LIC08]. They show that image and key dependency in the watermark detector leads to different false positive probability for geometric searches.

Regardless the peculiarities of the watermarking scheme, from the both theoretical and practical points of view, for prescribed transparency and data-payload, the lower the P_f, the weaker the robustness.

I.3.5. Watermarking key

The key is generated randomly to control the insertion and embedding of the watermark. According to the key management, watermarking schemes can be classified into symmetric and asymmetric. A symmetric watermarking scheme uses identical keys for watermark embedding and detection [COX08]. This possesses a security problem because the knowledge of the key at the detector can be used either to remove the watermark or to maliciously watermark other images with the same key. To solve this problem, asymmetric watermarking schemes (using different keys for watermark embedding and detection) have been proposed. This makes the use of watermarking possible for public domain applications where anyone with the detection key can check the embedded watermark. One of the first asymmetric watermarking schemes was proposed by Hartung and Girod [HAR98]. It is a modification of spread spectrum watermarking in which only a part of the watermarked signal is received by the recipient. The scheme allows the recipient to check only this part of the watermark signal. However, this also enables the recipient to remove the watermark. Van Hyuk Choi et al. [HUK99] have proposed an asymmetric watermarking system by applying a linear transformation to the primitive key set to obtain the private encoding key and the public decoding key. The security of this system relies on the difficulty of estimating the private encoding key from the decoding key that is publicly available. For higher security requirements, a different transformation is required for each primitive key.

The key length is the amount of information (bits) that manages the system security. It should be long enough to avoid the exhaustive search in the key space.

I.3.6. Cost

The technical cost of the algorithm is also a significant feature of any watermarking method. From this point of view, the complexity of the algorithm is the main criterion of practical acceptance. Just as an example, the average computing time for a video watermarking system is illustrated in Figure I-15, by watermarking a 20 minutes of video in high definition resolution [RIC03].

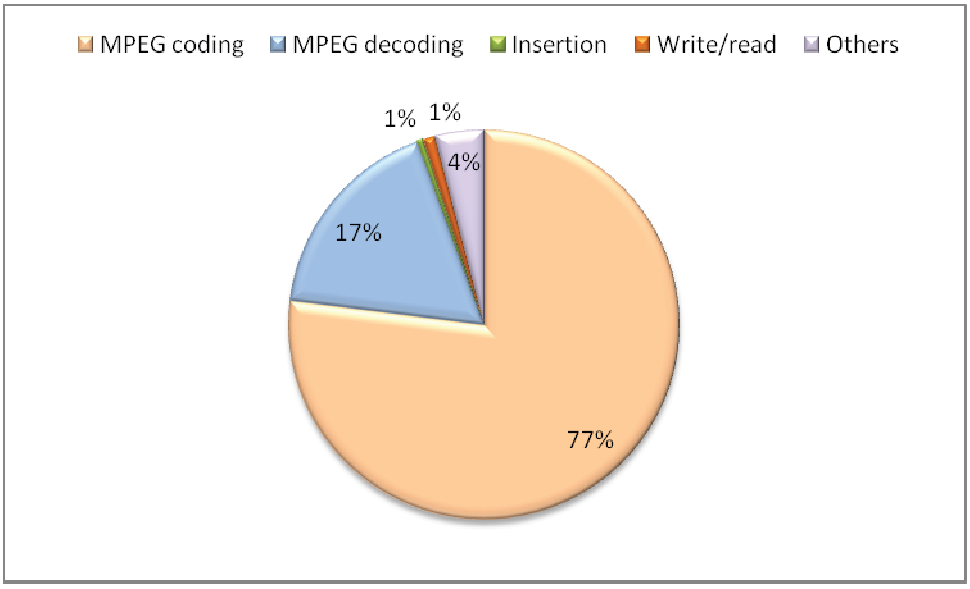


Figure I-15: Average of the various tasks performed in a watermarking system.

As depicted in Figure I-15, about 95% of total processing time is used by the coding/decoding process. Such an example brings to light that ***compressed domain watermarking*** is the only viable solution towards a practical deployment of such applications.

I.4. Watermarking in compressed domain

If the watermark has traditionally been inserted in the uncompressed media, the real-time constraints of the emerging applications (as VoD - video on demand *i*HDTV - interactive digital television) make watermarking insertion directly in compressed domain a hot research topic, Figures I-3 and I-16.

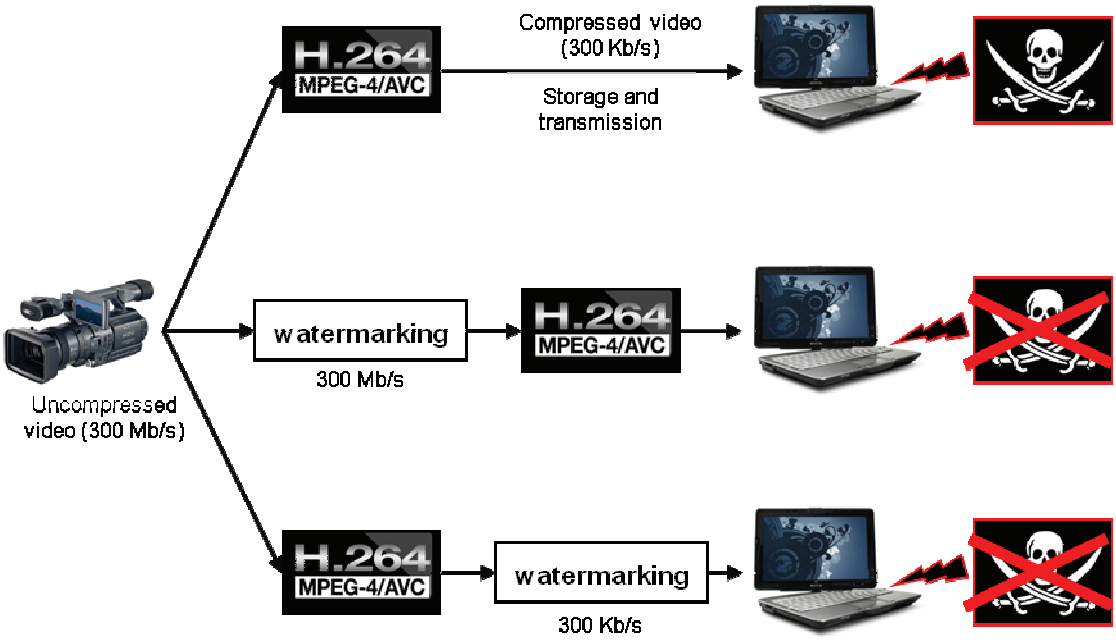


Figure I-16: Distribution chain: not secure (top), with watermarking in uncopressed domain (middle) and with watermark in the compressed domain (bottom).

In practice, video sequences are stored and distributed in compressed format. Accordingly, traditionally watermarking a video requires to decode the sequence, then to insert the watermark and, finally, to recode the video. Such an approach requires a large computing time (about 95% from the total time, *cf*. Figure I-15) to be allocated to the encoding/decoding operations. All these operations are naturally eliminated if the watermarking is applied directly to the compressed video. Currently, several research studies are conducted in the field of watermarking in the compressed domain, specifically in MPEG-4 AVC domain (which is the latest MPEG standard of compression, Appendix A).

The MPEG-4 AVC codec (coder/decoder) transforms the uncompressed data by a classic compression chain: prediction P, transformation T, quantization Q and arithmetic coding Ce, Figure I-17.

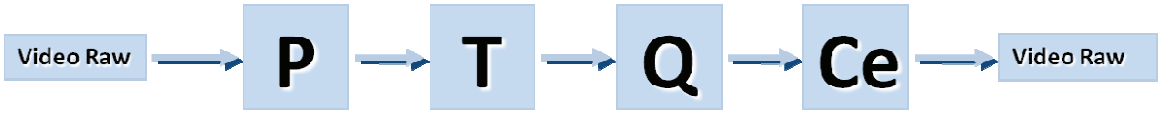


Figure I-17: MPEG-4 AVC compression chain.

Prediction is meant to eliminate the spatial (intra prediction) and temporal (inter prediction) redundancy. The transformation is applied with the view of representing the data as uncorrelated (separated into components with a minimum interdependence) and compacted (energy concentration on a small number of coefficients) information. Quantization is the phase where information is lost in the compression chain. The final phase of MPEG-4 AVC is the entropy coding (lossless).

MPEG-4 AVC is the most effective to-date lossy compression standard; consequently, any changes to the syntax of a compressed video will be *a priori* visible to the human eye, thus turning transparency into the first real challenge for watermarking in MPEG-4 AVC.

Another problem stems from the fact that the quantization indexes of coefficients of luminance are integer values, thus imposing severe constraints on the watermark robustness.

The first (and most intensive efforts) were devoted to the watermarking of the MPEG-2 compressed video streams. Langelaar et al. presented a real-time watermarking method [LAN98]. This method embedded the watermark directly by substitute the least-significant bits (LSB) of the selected suitable variable length codes (VLCs) by the watermark bit. In [HAR98], a more robust method consists in inserting the watermarking into selected high-frequency DCT coefficients of the stream. Hartung et al. proposed a spread spectrum watermarking algorithm in the MPEG-2 compressed domain Simitopoulos et al. proposed another MPEG-2 compressed-domain watermarking algorithm [SIM02]. Their method embeds the watermark generated as ramdom message to the quantized AC coefficients of intra-luminance blocks. In order to improve the transparency, the coefficients submitted of the watermarking insertion are selected according to a perceptual cost. In [WOL97], Wolfgang et al. used the Watson perceptual model improvement described in [WOL99] and to modulate the watermark by a perceptual threshold. This model consists of an IIP (image-independent perceptual) approach based on frequency, luminance sensitivity and contrast masking.

Five main studies (described below) are representative for the mark insertion in the MPEG-4 AVC domain.

Alattar et al. proposed an MPEG-4 compressed-domain video watermarking method and its performance is studied at low video bit rates [ALA03]. This approach is similar to the approach in [HAR98]; however, the authors used a synchronization template to counter-attack the geometric transformations. Their method also featured a gain control algorithm that adjusts the embedding strength of the watermark, depending on local image characteristics.

Wu et al. presented an oblivious watermarking algorithm by embedding the watermark in the MPEG-4 AVC I-frames [WU05]. Their scheme survives MPEG-4 AVC compression attacks with more than a 40:1 compression ratio. However, their scheme requires decompressing the video to embed the watermark.
In [NOO05], Noorkami et al. presented a less complex method applied at the MPEG-4 AVC coder side.

The watermark, a bipolar message, is inserted into the quantized AC coefficients, selected according to a perceptual cost. This technique, compared to existing methods, double the data payload but decreases the transparency by 0.1 dB. It is robust against linear filtering but fails the geometrical attacks. This algorithm is designed for the watermark at the encoder after the complete decompression of the stream, making it unusable for real-time applications.

In [GOL07], Golekiri et al. presented an adaptation of the popular ST-QIM method (Spread Transform - Quantization Index Modulation) applied to MPEG-4 AVC. The watermark is inserted into the perceptual projection of the selected block according to a psycho visual mask. Although it is transparent and robust transcoding, this algorithm does not meet the requirements of robustness to geometric attacks and does not meet the real time constraint.

In [ZOU08], Zou et al. tackle a real live issue, namely how the MPEG-4 AVC stream can be changed directly while respecting the video format structure and the watermarking constraints. This recent study says that the watermark direct the stream is possible within the constraint of transparency, but is very limited in terms of strength and quantity of data-payload.

Figure I-17 presents a synoptic comparison among these five methods and brings to light that no algorithm can find the optimal trade-off amongst transparency, robustness and data-payload inserted in the MPEG-4 AVC domain.

Note: All watermarking techniques presented in Figure I-17 lacks in information about the key and false positive properties.

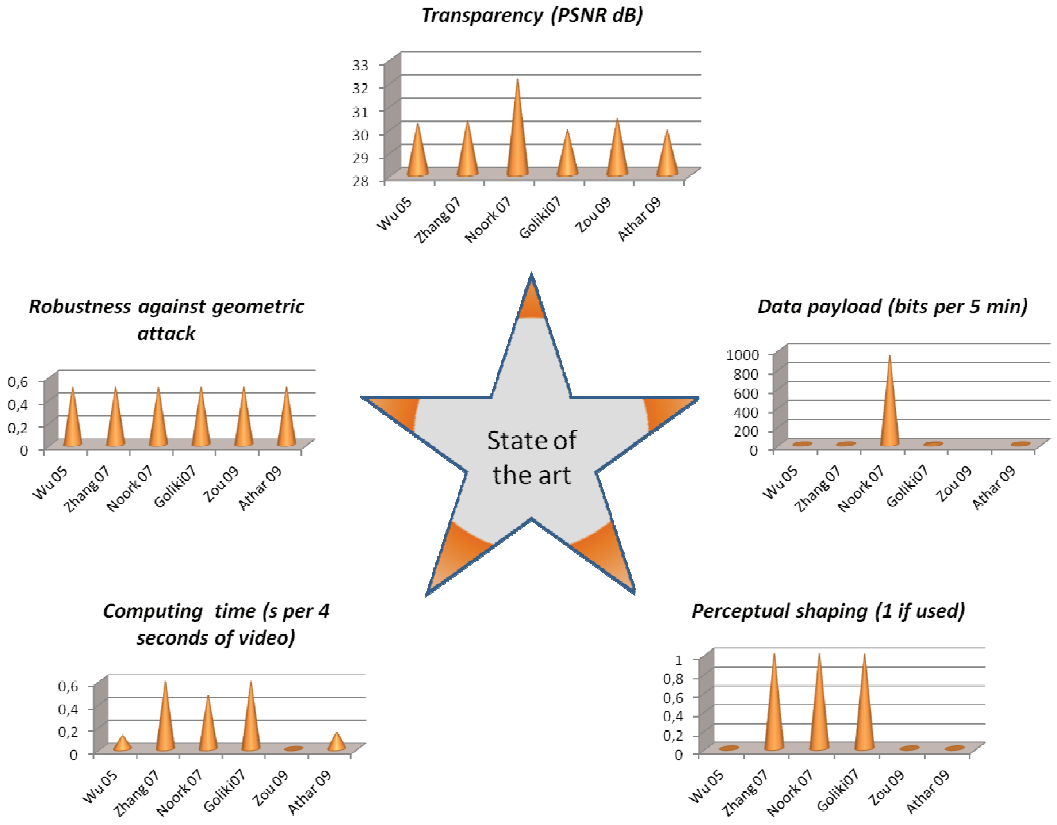


Figure I-17: A comparison of SoA watermarking methods for the MPEG-4 AVC domain.

I.5. Conclusion

By succinctly analysing the main watermarking issues (theoretical model, functional properties, support technologies), Chapter I brings to light the two main bottlenecks in compressed domain in-band enrichment: the lack of any dedicated theoretical tool and the very poor practical performances of the of-the-shelf methods.

The former bottleneck will be addressed in the present thesis at two levels. First (Chapter II), the very sophisticated and subject-dependent concept of transparency will be considered so as to establish the theoretical viability of the MPEG-4 AVC watermarking and to allow for the first MPEG-4 AVC perceptual model to be computed and validated. Secondly (Chapter V), basic tools in information theory and confidence-limit estimation are combined in order to establish the MPEG-4 AVC enrichment capacity, *i.e.* the maximal theoretical data payload for pre-established transparency and robustness constraints.

The latter bottleneck is methodologically addressed in Chapter III, where the first MPEG-4 AVC method robust against re-encoding and Stirmark attacks and in Chapter IV, where the overall results are demonstrated by two practical applications (sub-titling cartoons and copyright against camcorder recording).

References

[ALA03] A.Alattar, T. Lin,, and M. Celik, “Digital watermarking of low bit rate advanced simple profile MPEG-4 compressed video,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, pp. 787–800, August 2003.
[ADE10] http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html

[BEL10] M. Belhaj, M. Mitrea, S. Duta, and F. Preteux, “MPEG-4 AVC robust video watermarking based on QIM and perceptual masking”, IEEE International Conference on Communications, Bucharest, June 2010.
[BLD11] http://buildermedia.com/?page_id=204

[CAR96] T. Carney, S. A. Klein, and Q. Hu, “Visual masking near spatiotemporal edges”, Proc. SPIE, Vol. 2657, pp. 393-402, April 1996.
[CHE98] B. Chen, and G.W. Wornell, “Digital watermarking and information embedding using dither modulation,” in Proc. IEEE Workshop Multimedia Signal Process, Redondo Beach, CA, pp. 273–278, Dec. 1998.
[COR10] http://www.handprint.com/HP/WCL/color2.html
[COX95] I. Cox, J.Kilian, T. Leighton, and T Shamoon, “Secure spread spectrum watermarking for multimedia”, Tech Rep 95 – 10 NEC Research institute, Princeton,NJ,USA,1995
[COX08] I. Cox, M.L. Miller, J. Bloom, J. Fridrich, T. Kalker, Digital Watermarking and Steganography, Second edition, Morgan Kaufmann, Burlington, MA, 2008.

[DUM07] O. Dumitru, S. Duta, M. Mitrea, and F. Preteux, “Gaussian hypothesis for video watermarking attacks: drawbacks and limitations”, Proc. IEEE International Conference EUROCON 2007, pp. 849-855, September 2007.
[DUM10] O Dumitru, “Noise sources in robust uncompressed video watermarking”, PhD thesis, Université Paris VI - Pierre et Marie Curie, January 2011, under the supervision of F. Prêteux and M. Mitrea.
[DUT07] S. Duta, M. Mitrea, and F. Prêteux, “Compressed versus uncompressed domain video watermarking”, Proc. SPIE, Vol. 6700, p. 67000A, August 2007.

[EGG03] J.Eggers R. Bäuml R. Tzschoppe, and B. Girod, “Scalar Costa scheme for information embedding”, IEEE Trans. on Signal Processing, Vol. 51, No. 4, April 2003.

[FIS95] P. Fisher, and A. Eskicioglu, “Image quality measures and their performance”, IEEE Trans. on Communications, Vol. 43, No. 12, pp. 2959-2965, December 1995.

[GIR89] B. Girod, ”The information theoretical significance of spatial and temporal masking in video signals”, Proc. SPIE, Vol. 1077, pp. 178-187, 1989.
[GOL07] A. Golikeri. P. Nasiopoulos, and Z. Wang, “Robust digital video watermarking scheme for H.264 advanced video coding Standard”, Journal of Electronic Imaging 16 (4), 043008(Oct-Dec 2007).

[HAA98] G Haan, and E. Bellers, ”Deinterlacing: an overview”, Proc. IEEE, Vol. 86, No. 6, pp. 1839-1857, June 1998.
[HUR54] L. Hurvich, and D. Jameson, “An opponent-process theory of color vision”, Psychological Review, Vol. 64, pp. 384-404, 1954.

[LAN98] G. Langelaar, R. Lagendijk, and J. Biemond, “Real-time labeling of MPEG2 compressed video”, Journal of Visual Communication and Image Representation, Vol. 9, pp. 256–270, December 1998.
[LEG80] G. Legge, and J. Foley, “Contrast masking in human vision”, JOAM, Vol. 70, No. 12, pp 1458-1471, December 1980.
[LIC08] J. Lichtenauer, I. Setyawan, T. Kalker, and R. Lagendijk, “Exhaustive geometrical search and the false positive watermark detection probability”, in Proc. of Security and Watermarking of Multimedia Contents V, pp. 203–214, 2003.
[LIN07] J. Linnartz, T. Kalker, and G. Depovere, “Modelling the false alarm and missed detection rate for electronic watermarks,” in Proc. 2nd International Workshop on Information Hiding, pp. 329–343, 1998.

[MAN74] J. Mannos, and D. Sakrison, “The effects of a visual fidelity criterion on the encoding of images”, IEEE Trans. Information Theory, Vol. 4, pp. 525–536, 1974.
[MIT06a] M. Mitrea, S. Duta, T. Zaharia, and F. Prêteux, “Ensuring multimedia content adaptability by means of data hiding techniques”, Proc. SPIE, Vol. 6383, pp. 63830:1-8, 2006.
[MIT07a] M. Mitrea, and F. Prêteux, Tatouage robuste de contenus multimédias, La sécurité dans les réseaux sans fil et mobiles 1 : Concepts fondamentaux, H. Chaouchi and M. Laurent-Maknavicius editors, Traité IC2 - Série Réseaux et Télécoms, Hermès-Lavoisier, Paris, France, may 2007, pp. 169-224.
[MIT07b] M. Mitrea, O. Dumitru, F. Prêteux, and A. Vlad, “Zero-memory information sources approximating to video watermarking attacks”, Lecture Notes in Computer Science 4707, Vol. 3, pp. 445-459, 2007.

[NEW89] Optique de Newton. Tome 1 / , traduction nouvelle faite par M*** [Marat] sur la dernière édition originale... dédiée au roi par M. Beauzée, Auteur : Newton, Isaac (1642-1727) Éditeur : Leroy (Paris) Date d'édition : 1787 Contributeur : Marat, Jean-Paul (1743-1793). Traducteur Contributeur : Beauzée, Nicolas (1717-1789). Éditeur scientifique
[NOO05] M. Noorkami, R.M. Mersereau “Compressed-Domain Video Watermarking for H.264”, Proc. of the 2005 IEEE International Conference on Image Processing, ICIP 2005, Vol. 2, p. II-890-3, 2005

[PET98] F. Peticolas, R. Anderson, and M. Kuhn, “Attacks on copyright marking systems”, LNCS, Vol. 1525, pp. 218-238, 1998.
[PET00] F. Petitcolas, “Watermarking schemes evaluation”, IEEE Signal Processing, Vol. 17, No. 5, pp. 58-64, September 2000.
[PHI08] http://www.philips.fr/

[RIC 03] I. Richardson. H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia, John Wiley & Sons, Chichester, 2003.

[SEY65] A. Seyler, and Z. L. Budrikis, ”Detail perception after scene changes in television image presentations”, IEEE Trans. on Information Theory, Vol. 11, No. 1, pp. 31 43, 1965.
[SHA48] C.E. Shannon, “A Mathematical theory of information”, Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656, July, October 1948.
[SIM02] D. Simitopoulos, S. Tsaftaris, N. Boulgouris, and M. Strintzis, “Compressed-domain video watermarking of MPEG streams”, in Proceedings of IEEE International Conference on Multimedia and Expo (ICME), Vol. 1, pp. 569–572, 2002.

[TAM95] W. Tam. L. Stelmach. L. Wang. D. Lauzon, and P. Gray. ”Visual masking at video scene cuts”, Proc. SPIE, Vol. 2411, pp. 111 119, 1995.
[TIR92] A. Tirkel, G. Rankin, R. Van Schyndel, W. Ho, N. Mee, C. Osborne, “Electronic Water Mark”, DICTA 93, Macquarie University. p.666-673.

[WAT01] A. Watson, J. Hu, J.F. McGowan. “Digital Video Quality Metric Based on Human Vision”, Journal of Electronic Imaging, Vol. 10, Issue 1, pp. 20-29, 2001.
[WOL97] R.B. Wolfgang, and E.J. Delp, “A watermarking technique for digital imagery: Further studies”, in Proc. of the International Conference on Imaging Science, Systems, and Technology (CISST), vol. 1, pp. 279–287, 1997.
[WOL99] R.B. Wolfgang, C.I. Podilchuk, and E.J. Delp, “Perceptual watermarks for digital images and video”, in Proc. SPIE, Vol. 3567, pp. 40–51, 1999.
[WU05] G. Wu, Y. Wang, and W. Hsu, “Robust watermark embedding/detection algorithm for H.264”, Journal of Electronic Imaging, Vol. 14, pp. 13013–1–9, 2005.

[XU05] X. Xu, M. Tomlinson, M. Ambroze, and M. Ahmed, Techniques to provide robust and high capacity data hiding of ID badges for increased security requirement, Proc. 3rd International Conference Sciences of Electronic Technologies of Information and Telecommunications, 2005.

[YOU91] R. Young, “Oh say. Can you see? The physiology of vision”, Proc. SPIE, Vol. 1453, pp. 92-123, 1991.
[YU98] M. Yuen, and H. Wu, “A survey of hybrid mc/dpcm/dct video coding distortions”, Signal Processing, Vol. 70, No. 3, pp. 247-278, 1998.

[ZOU08] D. Zou, and J. Bloom, “H.264/AVC Stream Replacement Technique for video watermarking”, IEEE International Conference on Acoustics, Speech, and signal processing, ICASSP 2008.

Chapter II

Transparency

Let there be light

Abstract

Targeting the transparency concept, this chapter presents two studies on the use of objective assessment of the human visual perception. First, by considering eight objective quality metrics, it is proved that the MPEG-4 AVC (a.k.a. H.264) compressed stream still allows the imperceptible modification of its syntax elements. Secondly, by mathematically extending the Peterson-Watson IIP model, the first perceptual masking model matched to the MPEG-4 AVC peculiarities is computed and validated under watermarking constraints.

Contents

II.1 Introduction

As previously mentioned, when watermarking in the compressed domain, the major issue is the conceptual contradiction between watermarking and compression. On the one hand, watermarking algorithms use imperceptible features of the video to hold the watermark. On the other hand, lossy compression schemes try to remove as much as possible of the imperceptible data of a video.

The present chapter takes this challenge at two levels.

First, Chapter II.2 deals with visual quality evaluation. In this respect, the extent at which the predicted quantised luma coefficients (see Appendix A) can be altered while keeping an acceptable quality of the video is assessed. The study focuses on the *I* slice prediction errors, as they form the core of the compressed video stream.

Secondly (Chapter II.3), some quality improvement mechanisms are considered. In this respect, the first perceptual masking model for the MPEG-4 AVC compression standard is computed. This new masking model has been tested against state-of-the-art masking models, under the watermarking framework. On the one hand, when imposing a prescribed robustness and a fixed data payload, significant gains in transparency are obtained: the PSNR is increased up to 3dB while the Watson's DVQ (Digital Video Quality) is decreased by a half. On the other hand, when imposing a prescribed robustness and transparency, gains in data payload up to 50% are reached.

Chapter II.4 presents some concluding remarks.

II.2 Objective study of MPEG-4 AVC watermarking perceptual impact

The objective of this section is to establish whether the impact of the watermarking artefacts (expressed by some objective metrics) is low enough so as to allow a practical deployment of the compressed domain watermarking scheme.

Regardless the particular compressed stream syntax and watermarking insertion strategy, the human observer will always watch a video content displayed (after decoding) on some device (computer screen, TV set, video projector, ...). Consequently, any objective evaluation procedure for watermarking transparency should not be performed directly in the watermark insertion domain (*e.g.* predicted quantised luma coefficients [COX02]) but in the pixel domain, Figure II-1.

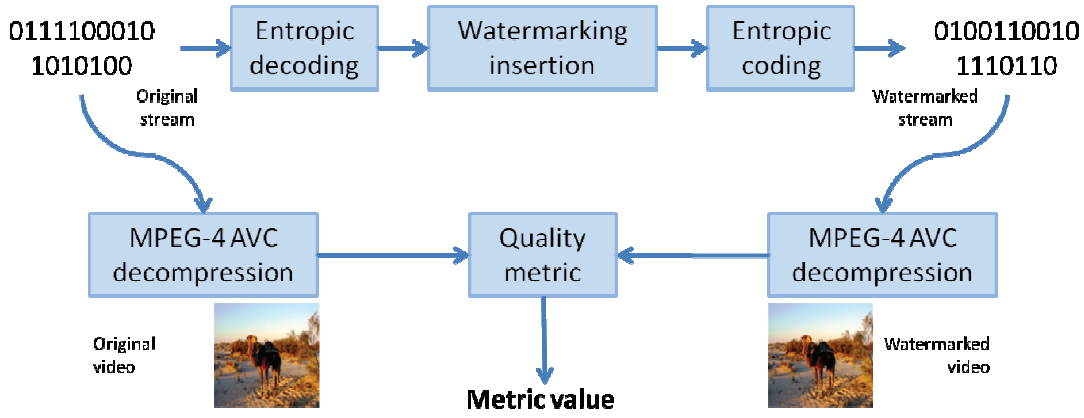


Figure II-1: The synoptic scheme of the video quality evaluation.

II.2.1. Evaluation protocol

In order to ensure generality with respect to the processed video content and effectiveness for real life applications, the evaluation protocol considered several types of original content, of the watermark insertion strategies and of the objective metrics, as follows.

Original MPEG-4 AVC content

The video test set (see Appendix D) consists of 10 sequences of 25 minutes each, coded at both SD (standard definition) and HD (high definition) resolutions.

The standard definition video main profile is encoded at two bit rates, namely 256 kbps and 64 kbps. Each GOP contains 25 frames and is structured as IPPPP (no B frames). The size of the frame is 352×576 pixels.

The high definition video sequences are also encoded at two bit rates, 5 Mbps and 10 Mbps while the frame size is 1280×720 pixels. The same GOP structure (25 frames, IPPPP) was considered.

Regardless the SD or HD profile, the video compression rate is expected to have a quite strong effect on the transparency as it acts on the quantization parameter values and on the prediction decisions.

A higher level of compression could mean larger alterations from an elementary modification of the stream.

The insertion domain

The speed constraints require for the mark to be inserted at the lowest possible levels in the compression scheme, Figure I-17. Consequently, in the present study, the predicted quantised luma coefficients domain will be considered [HAR98], [DUT07], [COX08]. Note that such a domain is not only advantageous from the speed point of view but also features some redundancy allowing the mark insertion.

The coefficients[1] corresponding to a 4×4 sub-macroblock are scanned in a zig-zag order and they will further be referred to through their index in this ordering, by c_1 to c_{15}. The DC coefficient, c_0, is not used within the experiments for two both perceptual and MPEG-4 syntactic reasons. First, the c_0 modification alters all the pixels of a 4×4 sub-macroblock with certain influences on the subsequent predicted blocks, thus amplifying the related artefacts. Second, in the 16×16 predicted macro-blocks, the DC coefficients are transformed by a Hadamard transform and then quantized, thus leaving less room for individual modifications.

In order to process these sequences (*i.e.* to modify the MPEG-4 AVC stream elements under syntax constraints), a software tool (MPEG parser and syntax analyzer and corrector) was developed starting from the MPEG-4 AVC reference software [LIN07], [HUH11], [RIC03]; Appendix B presents details in this respect.

Number of watermarked elements

In our study, two numbers of sub-macroblocks in a macroblock have been modified: 1 and 16. Concerning the number of macroblocks in a frame, different values have been considered:

- 5, 10, 50 and 100, in the SD case;
- 5, 50, 500 and 1000 in the HD case.

These parameters have a very important impact in the overall transparency because of the drift effect generated by the MPEG-4 AVC intra and inter-frame prediction. A modification made to a coefficient will propagate to the adjacent blocks and to subsequent frames through the prediction mechanisms. From the watermarking point of view, this leads to an accumulation of distortions, with disastrous effects on the transparency. In our study, the macroblocks subject to modification were randomly chosen, thus bringing the numerical values closer to the real-life application.

Note that some state of the art watermarking methods are not confronted to this problem, as they employ the encoding loop to perform the mark insertion. The encoder would then take into account the watermark when performing the prediction computations for the adjacent blocks and would naturally remove the drift effects. It should be noted that besides the speed decrease, such a system may also lead to a security fault, as an attacker can search the adjacent blocks (both in time and space) for information about the watermark signal.

[1] In the sequel, by "coefficients" we shall denote the predicted quantized luma coefficients.

The insertion strategy

Two watermarking insertion strategies are considered in our study: additive (the predilection choice of the nowadays watermarking schemes) and substitutive (mainly for comparative sakes), Figures II-2.

In the former case, the coefficients are subject to some integer value additive uniform noise. Each time, two amplitudes have been considered:

- {-1, 1} and {-2, -1, 1, 2} in the SD case;
- {-2, 2} and {-4, -3, -2, -1, 1, 2, 3, and 4} in the HD case.

Note that as the dynamic of the values in the predicted quantised luma coefficient domain are different in the SD and HD cases, we had also to consider different additive noise values, Figure II-3. This is a consequence of the fact that HD allows smaller values for the quantizing step (hence, larger values for the quantized indexes).

In the latter case, the same numerical values are considered but, this time, they are replacing the original coefficients instead of being added to them.

12	0	2	3
1	0	0	0
0	3	0	0
0	0	0	0

Original 4x4 sub-macroblock

1	-1	1	-1
1	1	-1	1
1	-1	-1	1
-1	1	-1	-1

Mark to be inserted

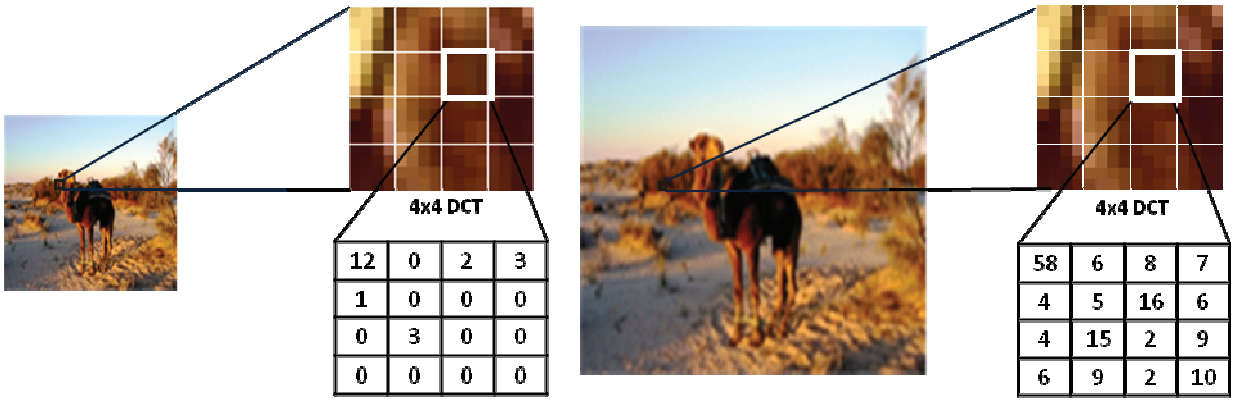
12	-1	1	2
2	1	-1	1
1	2	-1	1
-1	1	-1	-1

Additive insertion

1	-1	1	-1
1	1	-1	1
1	-1	-1	1
-1	1	-1	-1

Substitutive insertion

Figure II-2: Additive *versus* substitutive insertion (in the two cases, the lowest frequency coefficient is kept unchanged).



4×4 sub-macroblock in SD video

4×4 sub-macroblock in HD video

Figure II-3: SD *versus* HD sub-macroblock values.

Quality metrics

The following objective quality metrics have been considered (see Chapter I for their definition):

- ***Pixels based measures***: the signal to noise ratio (*PSNR*), the maximum mean square error (*PMSE*), the image fidelity (*IF*), and the average absolute difference (*AAD*).
- ***Correlation based measures***: normalized cross-correlation (*NCC*), the correlation quality (*CQ*), and the structural content (SC).
- ***Psychovisual measure***: digital video quality (*DVQ*).

The limits at which these different measures are considered as reflecting a good quality of the marked content are filled in Table II-1. Note that these limits were set according to values reported in the literature; however, as already discussed, according to the targeted application, the actual limits may vary. Note that no reference limit concerning CQ is met in literature; hence, we fill-in the corresponding column by NA (not available).

Table II-1: The limits corresponding to transparent artefacts and the related references.

	Pixels based				*Correlation based*			*Psychovisual*
	PSNR (dB)	**PMSE ($x10^{-6}$)**	**IF**	**AAD**	**NCC**	**CQ**	**SC**	**DVQ**
SD	**> 35** [QIA06]	**< 1.95** [QIA06]	**> 0.999** [QIA06]	**< 0.05** [QIA06]	**(0.95 ; 1.05)** [DUT07]	**NA**	**(0.95 ; 1.05)** [DUT07]	**< 0.033** [DUT07]
HD	**> 42** [WEI09]	**< 0.014** [WEI09]	**> 0.9999** [WEI09]	**< 0.05** [WEI09]	**(0.99 ; 1.01)** [CHI11]	**NA**	**(0.97 ; 1.03)** [CHI11]	**< 0.0016** [BEL11]

II.2.2. Experimental result

The experiments were run so as to compute the eight above-mentioned reference metrics for all the possible case studies (SD and HD, compression rate, number of modified elements in the MPEG-4 AVC syntax elements and insertion strategies). The overall results are reflected in 64 figures, each of which composed of 8 graphs (one graph for each objective measure) of 2 plots (corresponding to 1

sub-macroblock and 16 sub-macroblock, respectively). In order to alleviate the manuscript reading, only 8 (Figures II-4 to II-11) out of these 64 figures (corresponding to the SD video coded at 64kbps and additive insertion techniques) are presented below while the rest are included into the Appendix E, *cf*. Table II-2.

Tables II-3 and II-4 synoptically represent the results reported in these 64 figures.

The overall results of the study bring to light several conclusions concerning the objective evaluation of the MPEG-4 AVC modification artefacts:

- ***The three types of measures should be considered in order to obtain accurate results and significant comparison***. On the one hand, the correlation based measures are less sensitive to the frequency of the modified coefficient then the pixel-based measures but are highly sensitive to the number of modified macroblocks. On the other hand, although very seldom considered in the literature, the DVQ seems to reach a good trade-off between these two types of behaviours.
- ***All the parameters in the experimental set-up (the quality of the original video, the type of modification, the insertion strategy …) are significant***. First, as a general rule, the better the quality of the original content, the better the transparency for the additive techniques; although quite unexpected, such a result was confirmed by all the considered measures. When considering substitutive techniques, the result is inverted: the lower the quality of the original video, the better the transparency. Secondly, as expected, the larger the number of the modified elements and the larger the strength of the modification, the lower the transparency. However, for a given type of content (*e.g.* SD), these transparency differences are less important than when considering different content types (*e.g*. SD *vs*. HD). Finally, the substitutive insertion techniques are far more restrictive from the transparency point of view than the additive techniques.
- ***The experimental results allow us to bring to light the maximal modifications of the predicted quantised luma coefficients under transparency constraints.*** By comparing the numerical values presented in the 64 figures to the transparency limits presented in Table II-1, the following conclusion can be drawn, *cf*. Tables II-3 and II-4:
    - *Additive insertion*: one sub-macroblock per macroblock, 5 macroblocks per *I* frame and a four symbol noise $\{-2,-1,1,2\}$; these limits hold for both SD and HD encoded video;
    - *Substitutive insertion*: when considering SD encoded video, even the slightest modification can be detected by the pixel based measures which conflict to limits in Table II-1; hence, in this case, we cannot speak about fidelity *per-se* but rather of quality (*cf*. the numerical values in Appendix E); when considering HD encoded video, the substitutive techniques does not allow us to speak about transparency.

Table II-2: The structure of the experimental results, according to the video sequence profile (SD or HD), insertion strategy (additive or substitutive, noise dynamics), video encoding rate and number of watermarked elements.

Standard Definition	**Additive**	**{-1,1}**	**64 kbps**	Fig. II-3	Fig. II-4	Fig. II-5	Fig. II-6
			256 kbps	Fig. APP.E-1	Fig. APP.E-2	Fig. APP.E-3	Fig. APP.E-4
		{-2,-1,1,2}	**64 kbps**	Fig. II-7	Fig. II-8	Fig. II-9	Fig. II-10
			256 kbps	Fig. APP.E-5	Fig. APP.E-6	Fig. APP.E-7	Fig. APP.E-8
	Substitutive	**{-1,1}**	**64 kbps**	Fig. APP.E-9	Fig. APP.E-10	Fig. APP.E-11	Fig. APP.E-12
			256 kbps	Fig. APP.E-13	Fig. APP.E-14	Fig. APP.E-15	Fig. APP.E-16
		{-2,-1,1,2}	**64 kbps**	Fig. APP.E-17	Fig. APP.E-18	Fig. APP.E-19	Fig. APP.E-20
			256 kbps	Fig. APP.E-21	Fig. APP.E-22	Fig. APP.E-23	Fig. APP.E-24
High Definition	**Additive**	**{-2,2}**	**5 Mbps**	Fig. APP.E-25	Fig. APP.E-26	Fig. APP.E-27	Fig. APP.E-28
			10 Mbps	Fig. APP.E-29	Fig. APP.E-30	Fig. APP.E-31	Fig. APP.E-32
		{-4,-3,-2,-1,1,2,3,4}	**5 Mbps**	Fig. APP.E-33	Fig. APP.E-34	Fig. APP.E-35	Fig. APP.E-36
			10 Mbps	Fig. APP.E-37	Fig. APP.E-38	Fig. APP.E-39	Fig. APP.E-40
	Substitutive	**{-2,2}**	**5 Mbps**	Fig. APP.E-41	Fig. APP.E-42	Fig. APP.E-43	Fig. APP.E-44
			10 Mbps	Fig. APP.E-45	Fig. APP.E-46	Fig. APP.E-47	APP.E-48
		{-4,-3,-2,-1,1,2,3,4}	**5 Mbps**	Fig. APP.E-49	Fig. APP.E-50	Fig. APP.E-51	Fig. APP.E-52
			10 Mbps	Fig. APP.E-53	Fig. APP.E-54	Fig. APP.E-55	Fig. APP.E-56

Noise dynamics	Video rate	5 macroblocks	10 macroblocks	50 macroblocks	100 macroblocks
		5 macroblocks	50 macroblocks	500 macroblocks	1000 macroblocks

Table II-3: Synopsis of the results concerning transparency, obtained when a single sub-macroblock is altered in each processed macroblock. Each cell contains 8 rectangles (one rectangle for each objective metrics), colored as follows: "*green*" ■ – the values of the corresponding metric averaged over all the 15 investigated coefficients belong to the transparency limits (*cf*. Table II-1); "*red*" ■ – the values of the corresponding metric averaged over all the 15 investigated coefficients do not belong to the transparency limits (*cf*. Table II-1); "*blue*" ■ – for CQ, no transparency limit can be found in the literature.

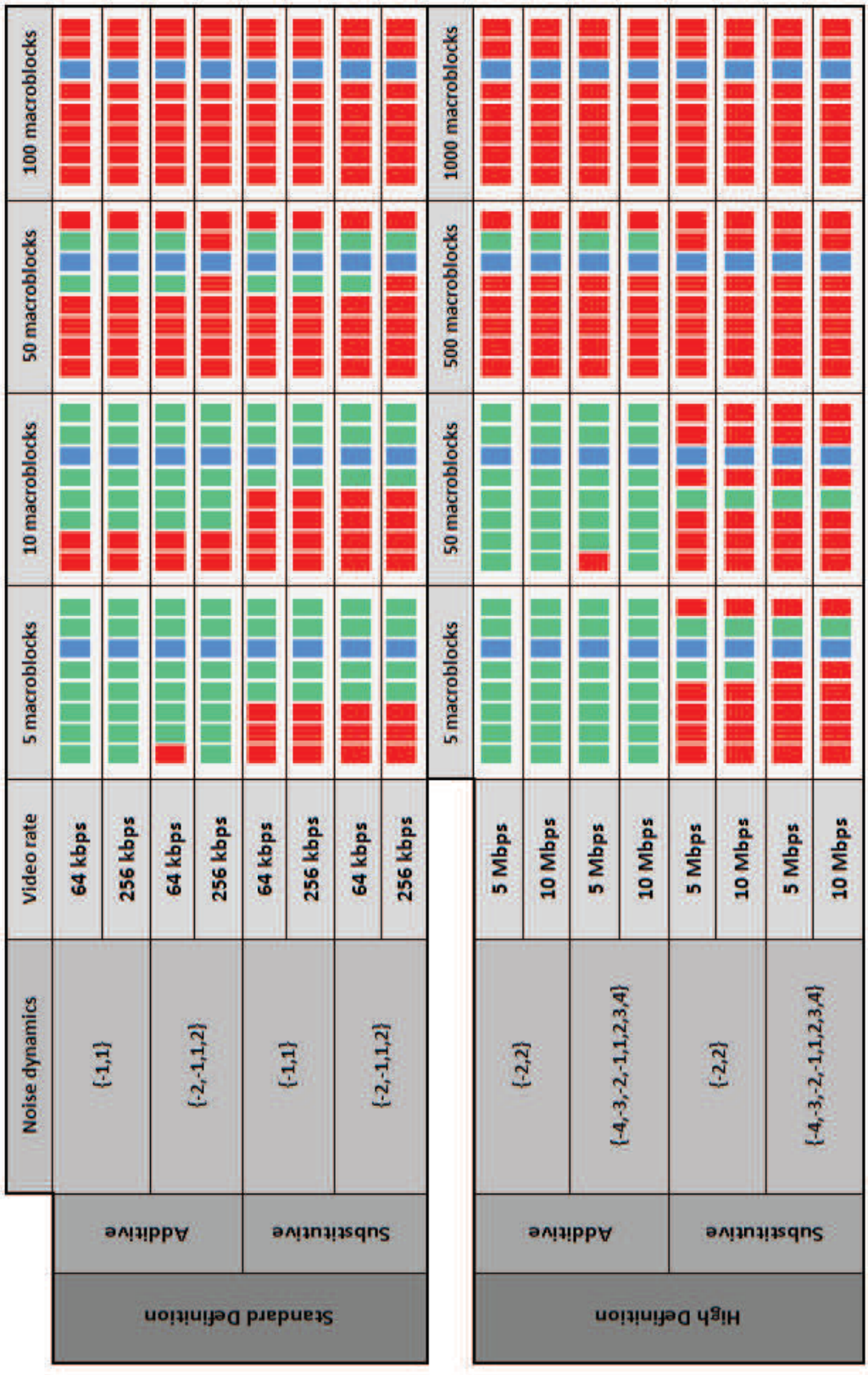
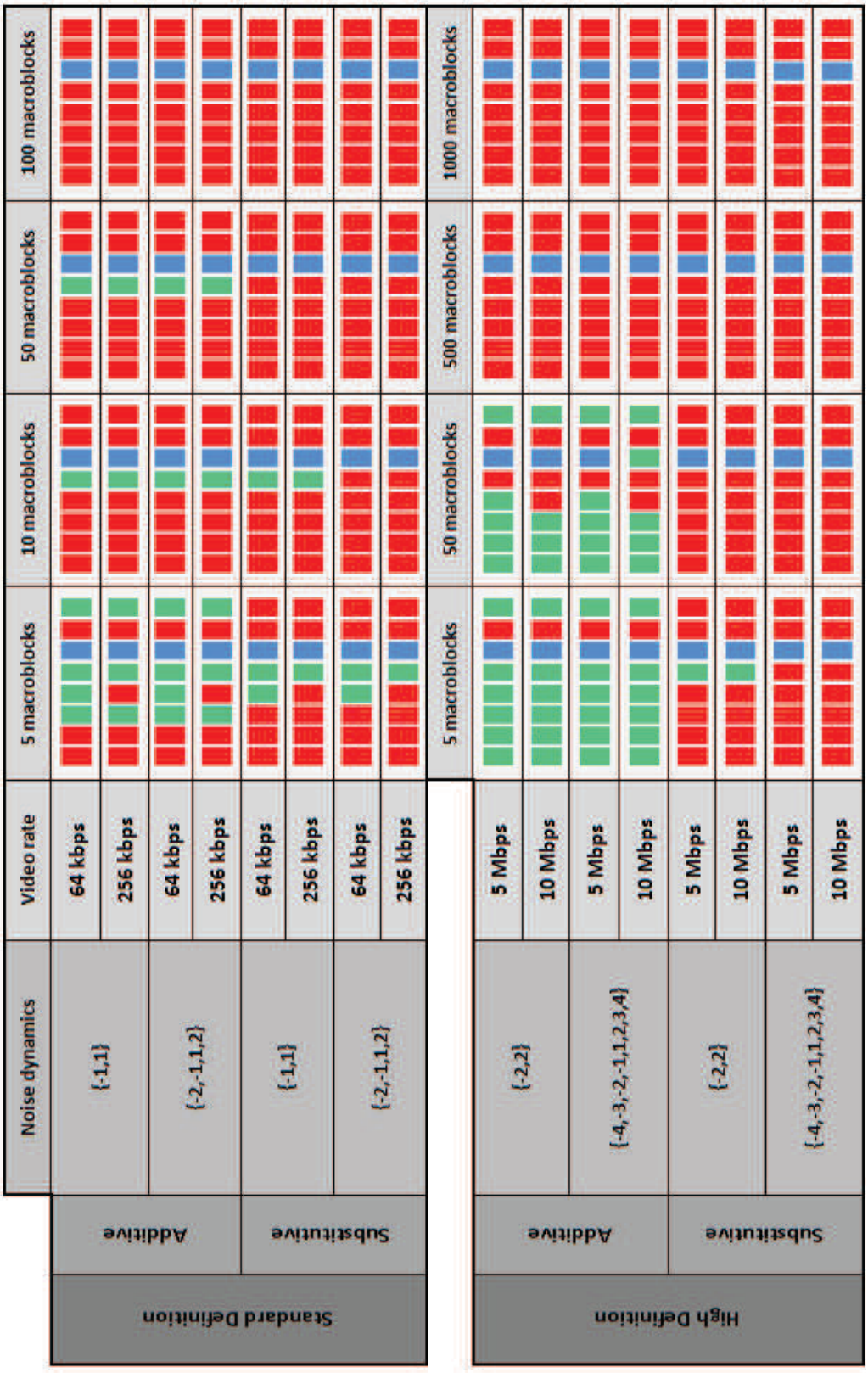


Table II-4: Synopsis of the results concerning transparency, obtained when all the 16 sub-macroblock are altered in each processed macroblock. Each cell contains 8 rectangles (one rectangle for each objective metrics), colored as follows: "*green*" ■ – the values of the corresponding metric averaged over all the 15 investigated coefficients belong to the transparency limits (*cf*. Table II-1); "*red*" ■ – the values of the corresponding metric averaged over all the 15 investigated coefficients do not belong to the transparency limits (*cf*. Table II-1); "*blue*" ■ – for CQ, no transparency limit can be found in the literature.



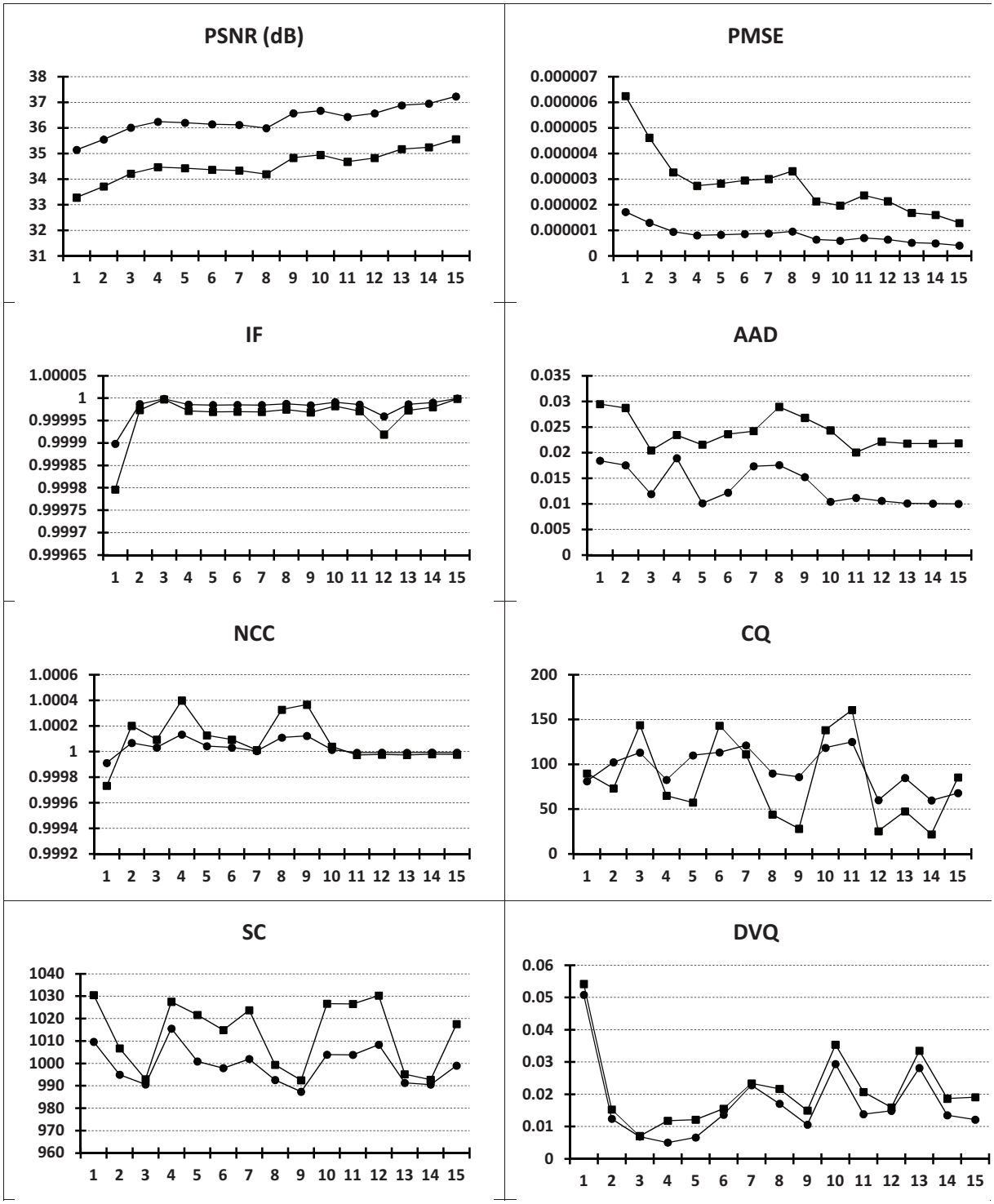


Figure II-4: Effects of {-1, 1} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

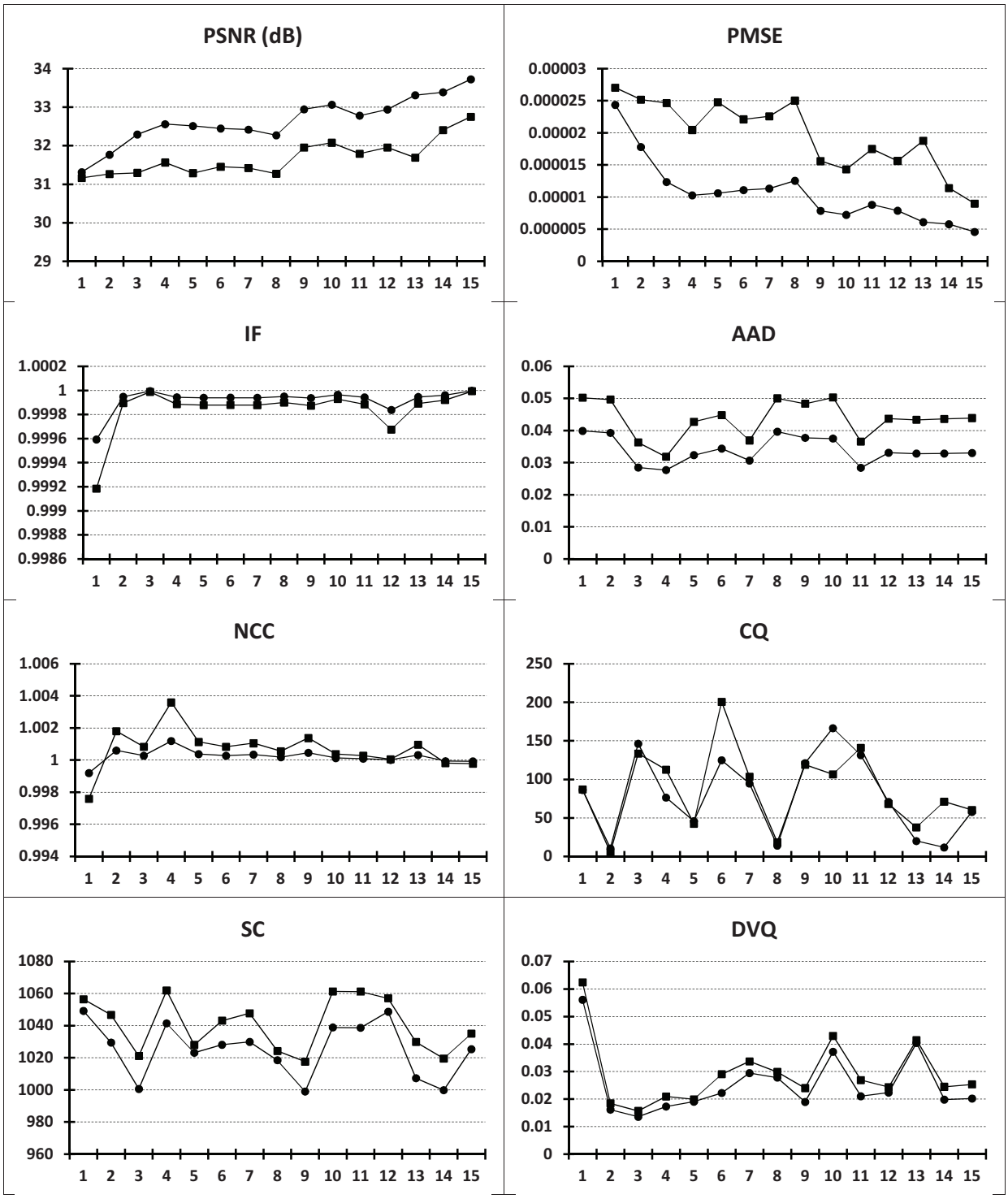


Figure II-5: Effects of {-1, 1} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 10 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

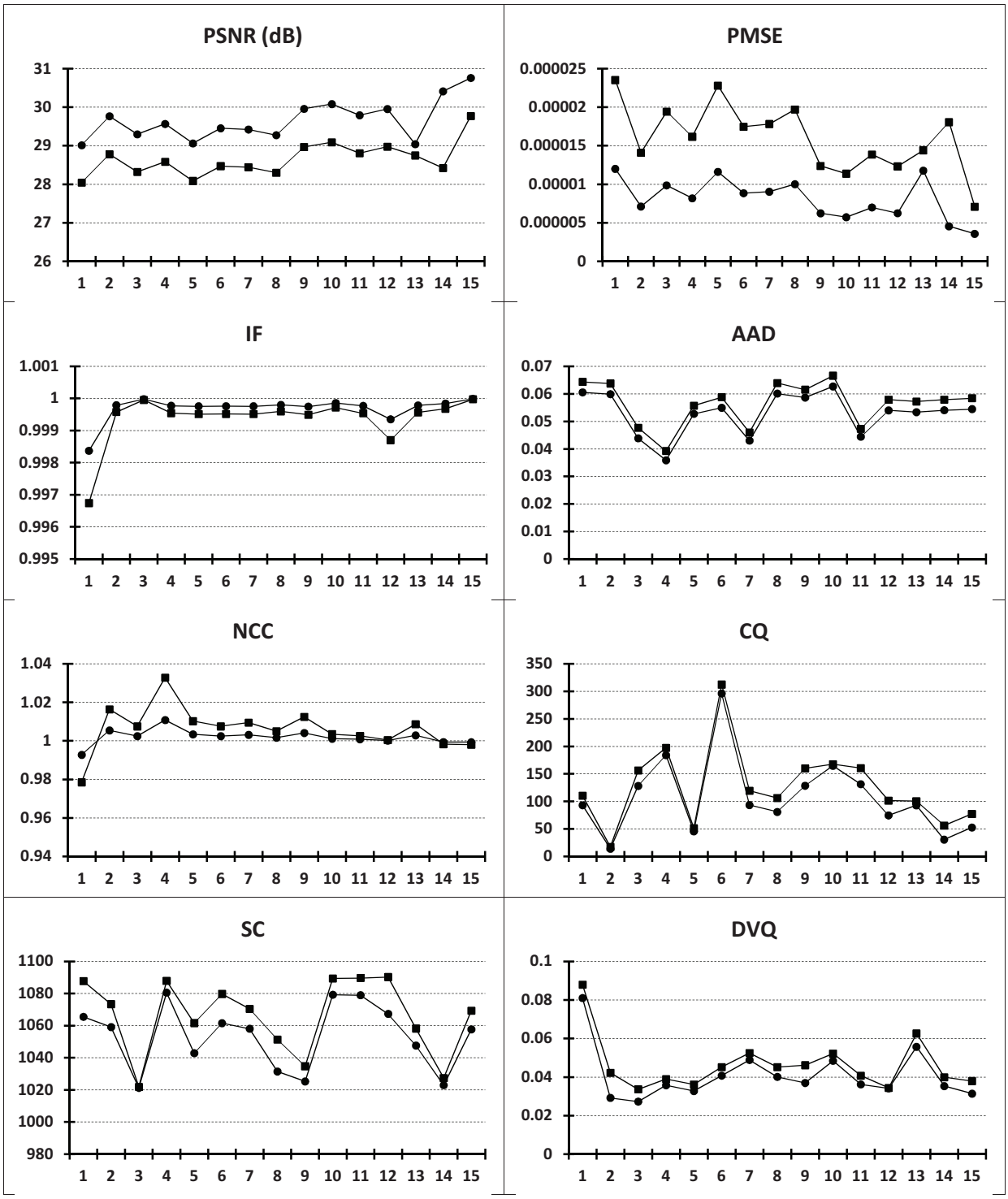


Figure II-6: Effects of {-1, 1} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

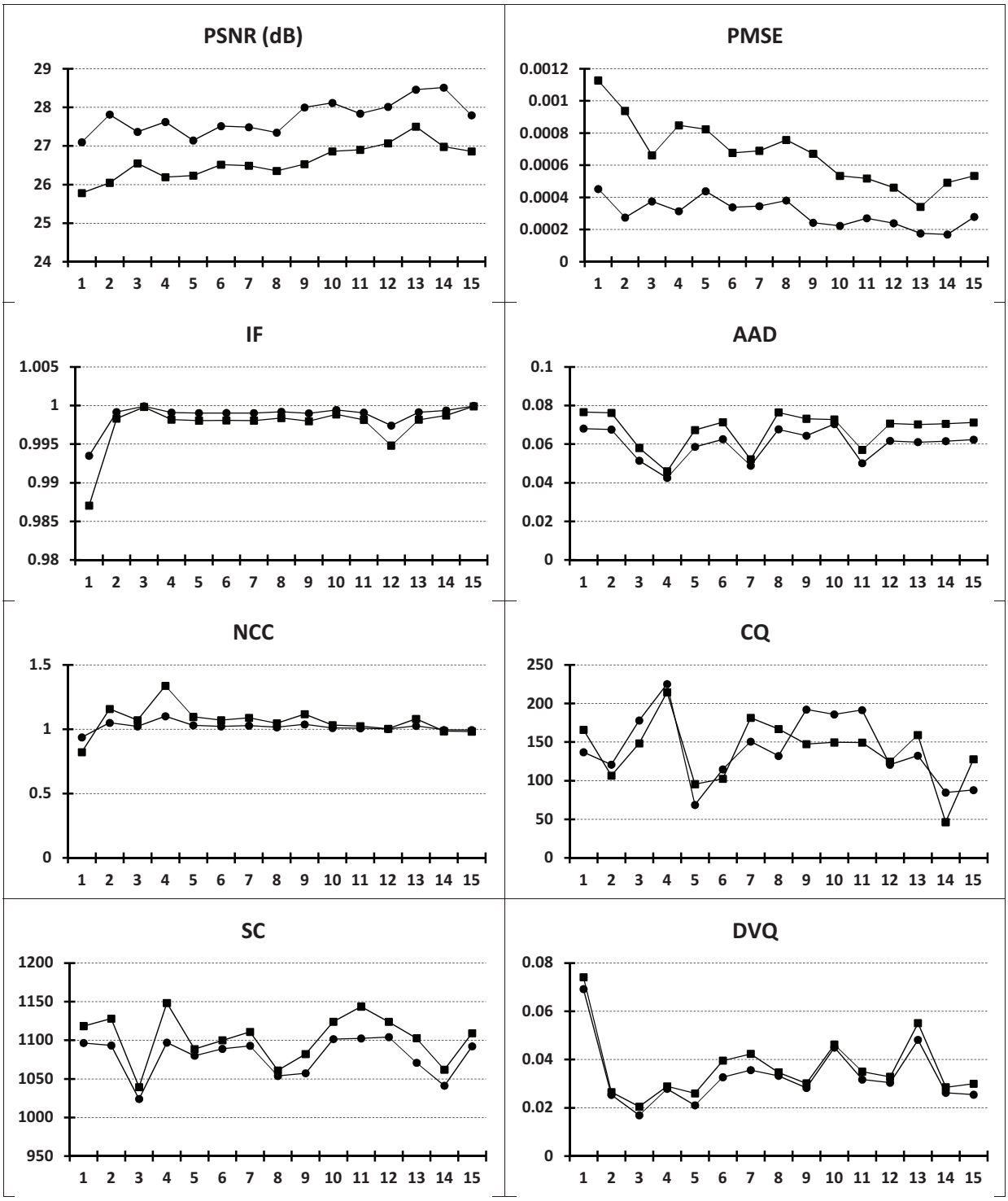


Figure II-7: Effects of {-1, 1} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 100 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

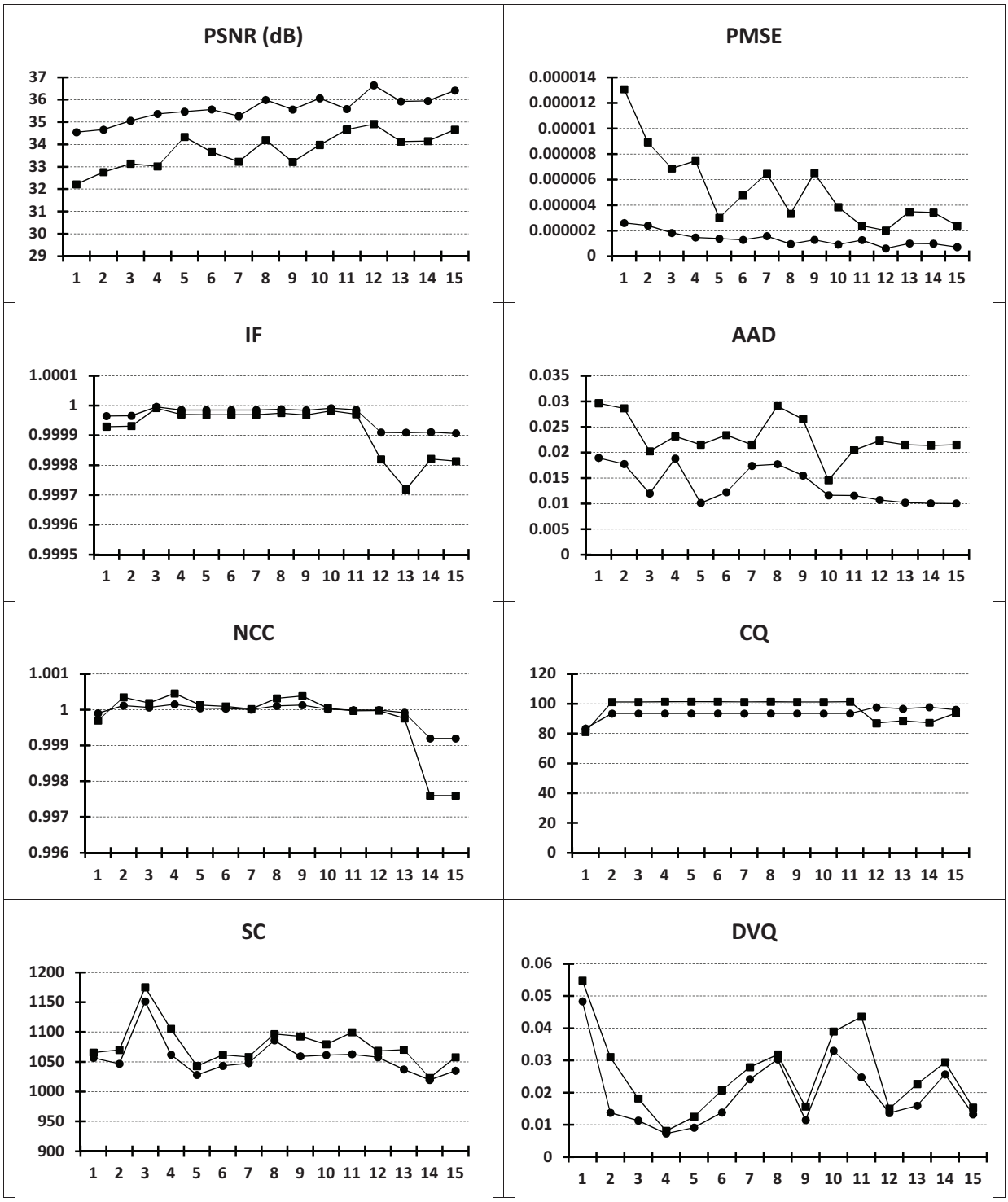


Figure II-8: Effects of {-2, -1, 1, 2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

.

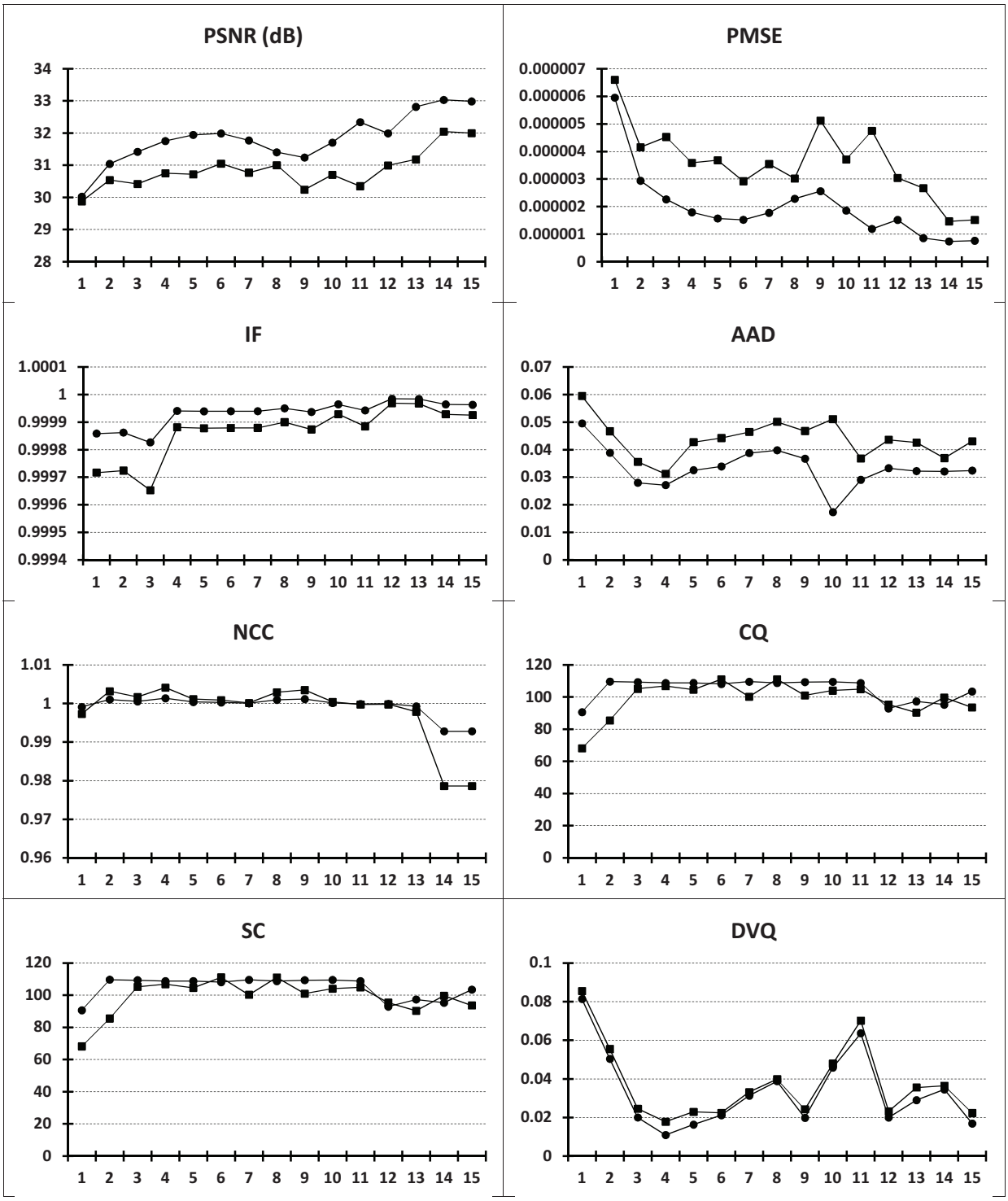


Figure II-9: Effects of {-2, -1, 1, 2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 10 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

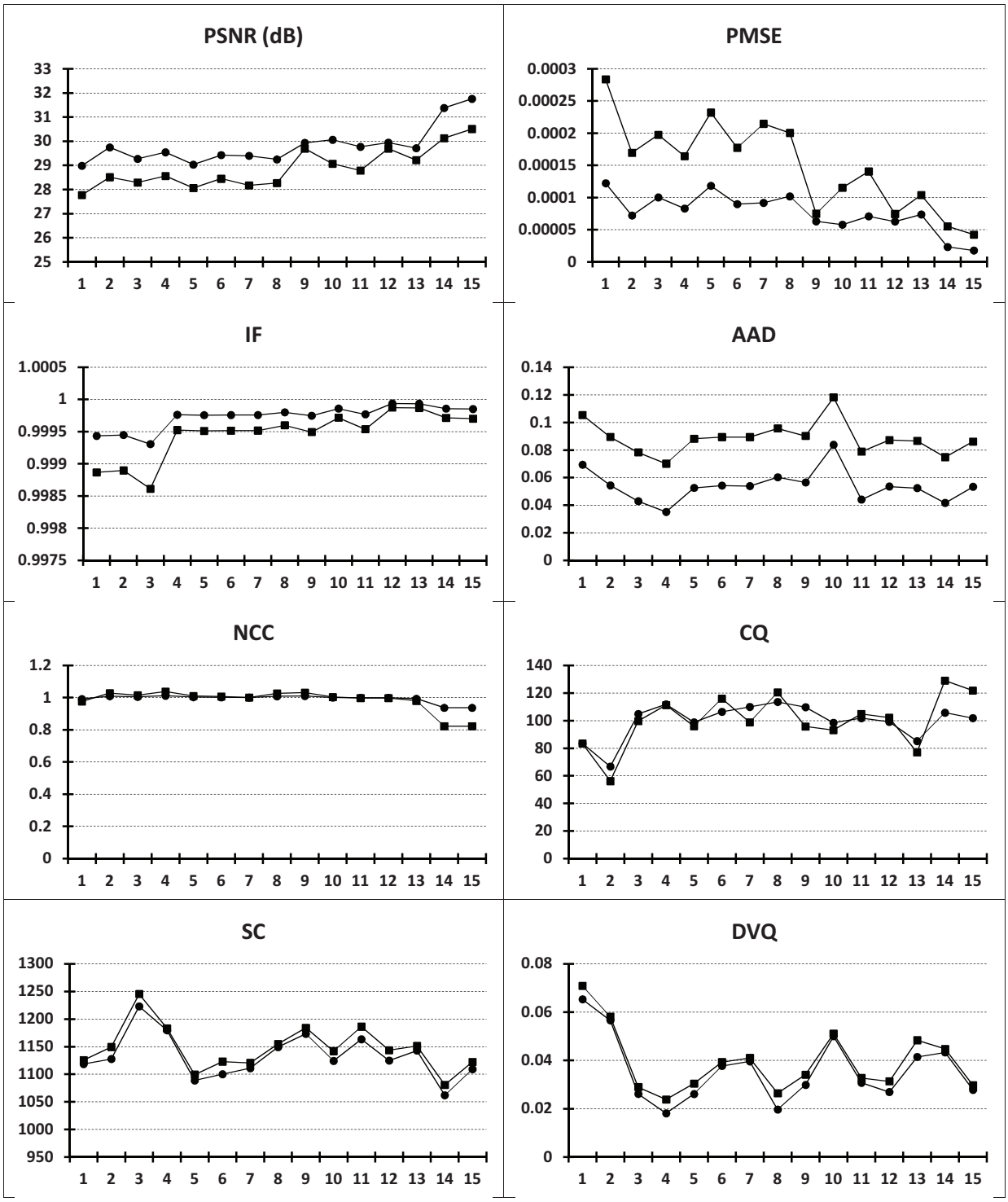


Figure II-10: Effects of {-2, -1, 1, 2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

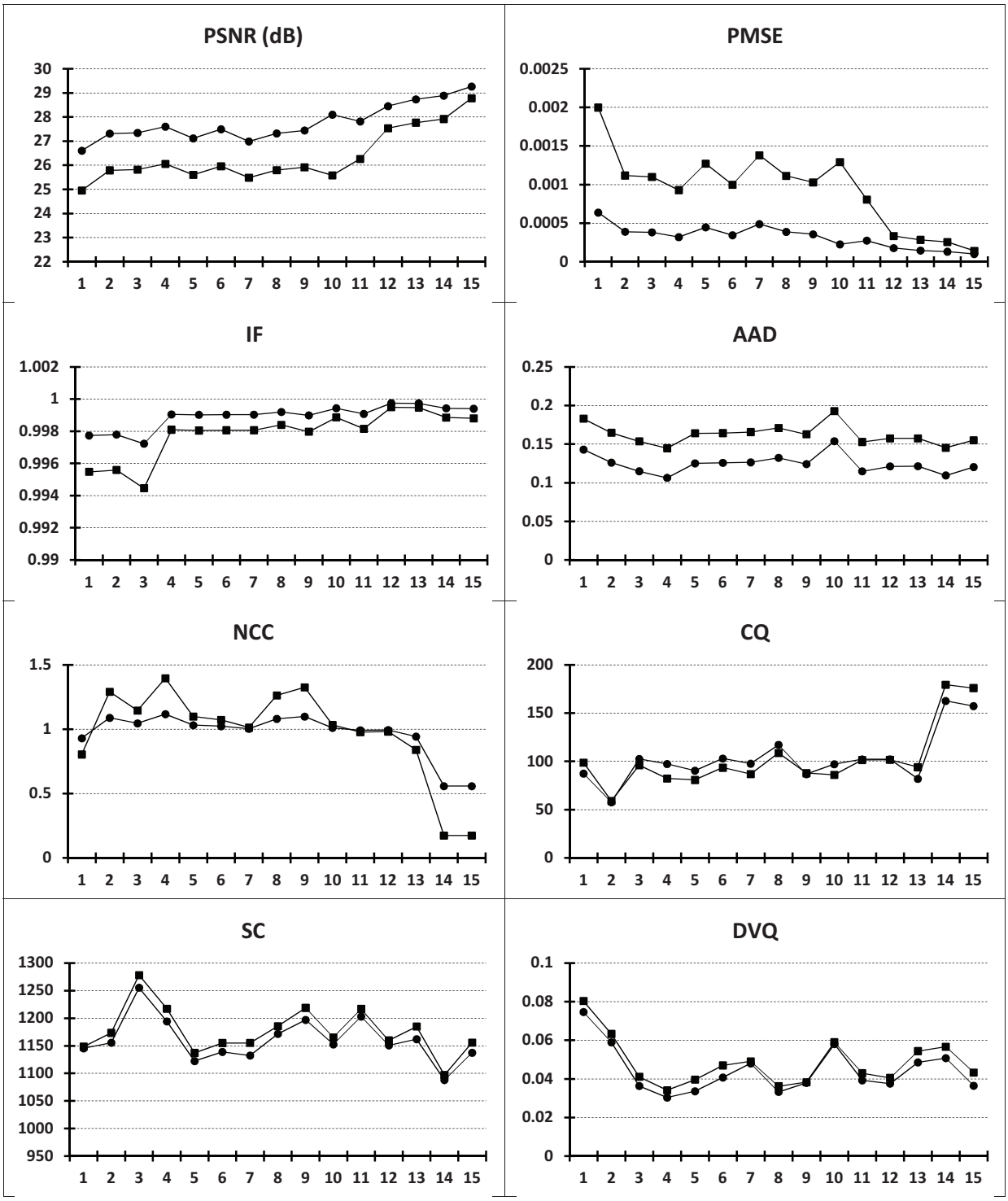


Figure II-11: Effects of {-2, -1, 1, 2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 100 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

II.3. MPEG-4 AVC perceptual masking

The present section establishes the first accurate perceptual masking model addressing the MPEG-4 AVC domain peculiarities. In other words, we computed the MPEG-4 AVC visibility thresholds, *i.e.* the maximum amount of additive noise which can imperceptibly affect the quantized values of the intra prediction errors.

In its widest acceptation, perceptual masking consists in matching the image processing algorithm to the human visual system peculiarities.

The first model of the psycho-visual system was proposed by Waston [WAT83] in 1983. On that occasion, a model was established in order to map the visual signal perceived by a human observer into an internal representation on which the observer's decision is based.

In 1990, Peterson [PET90] proposed the IIP (Image-Independent Perceptual) masking model, as an experimental estimation of the visibility threshold for images. By running psycho-visual tests, it was established that the maximum value of a perturbation which can still be imperceptibly added to the image pixels is given by a function $S(.)$ depending on the pixels position and on their neighborhood. Although this function is implicitly governing the human visual system (HVS), its mathematical form is not yet précised. In practice, $S(.)$ is projected on the DCT (Discrete Cosine Transform) domain, thus resulting in the $T_{S_{8\times 8}}$ matrix, which gives the maximum amount of the additive noise which can be imperceptibly added to the 8×8 DCT coefficients (the so-called visibility thresholds).

One year later, Albert J. Ahumada [AHU92] proved that the Peterson DCT Quantization visibility thresholds can be predicted by a simple luminance-based detection model. This model was estimated for special viewing condition (a mono color background at the luminance). However, this model somewhat lacks in practical impact, as in real-life applications the viewing conditions are unstable and depend on various parameters (display and ambient luminance, spatial frequencies and pixel spacing, viewing distances, aspect ratio, …).

Six years later, in order to solve these issues, Watson considered the contrast and the global luminance as elements in an improved IIP model [WAT98]; hence, the visibility thresholds are computed by taking into account their actual values (the so-called 8×8 Watson contrast sensitivity table).

This mask has already been used for video watermarking in both uncompressed and compressed domains, aiming at increasing the data payload and at improving the robustness, while limiting the visual distortions. However, as it was designed for classical 8×8 DCT, it can be directly exploited only in MPEG-2 applications and has no straightforward adaptation for the MPEG-4 AVC.

With respect to its predecessors, MPEG-4 AVC comes across with three new issues, Appendix A. First, the DCT is no longer applied to 8×8 blocks but to 4×4 blocks. Secondly, MPEG-4 AVC considers an integer DCT. Finally, this integer DCT is no longer applied to pixels (as it is the case in JPEG and MPEG-2) but to the intra/inter prediction errors.

When considering now the perceptual modeling from the application point of view, a wide area of approaches can be found in the last 20 year publications; in the sequel, these approaches will be illustrated for compression and watermarking applications, for both still image (JPEG) and video

(MPEG-2 and MPEG-4). Of course, the following two paragraphs will present only some examples and are not meant to cover all the related studies.

When considering the compression applications, the very first example is provided the Watson model *per-se* [TON98]: it allowed the JPEG compression ratio to be improved by 22%, for a prescribed PSNR value. In [VER96], the authors advance a novel bit allocation scheme for MPEG-2 CBR video coding using the Watson perceptual model. A measure of the macroblock activity, different from the classical MPEG-2, is subsequently defined. The underlying experiments demonstrated that for the same visual quality, the compression rate can be thus improved by 10%. A. Cabrita advanced the principle of perceptual coefficients pruning [CAB11]: all the transform coefficients which have a magnitude lower than the corresponding JND threshold can be set to zero, since these coefficients should be perceptually irrelevant. Thus, the compression rate can be improved up to 14% or, equivalently, the PSNR can be increased by 0.4 dB for rates around 8 Mbyte/s.

The perceptual masking success story can be illustrated in the watermarking field by the study of Q. Li [LI07] who advances a JPEG QIM watermarking algorithm with modified Watson perceptual mask. The adaptation provides contrast scaling invariance of the luminance mask, thus ensuring its linear relationship with the perceptual threshold values. This way, the Watson perceptual distance computed between the marked and the original JPEG images is reduced by 11.2 units; when imposing the same transparency, the new model allows a BER reduction by 10%. In [DAM06], an MPEG-2 video watermarking algorithm exploiting mark energy adaptation to the host signal values presented. This perceptual watermarking allows a 42 dB for the PSNR between the watermarked and the host video sequences to be obtained; when imposing a fixed transparency, the use of the new model allows a BER reduction by 2%.

In the sequel, the three above-mentioned main MPEG-4 AVC peculiarities (integer DCT transform applied to 4×4 blocks of predicted errors) will be incrementally considered in order to objectively establish their impact in the perceptual masking and to assess the practical impact of this new model in watermarking applications (*cf*. the low-right cell in Figure II-12).

II.3.1 Perceptual mask

II.3.1.1 The 4×4 adaptation

On the one hand, the 4×4 DCT is defined as:

$$X(n_1,n_2)=\sum_{i_2=0}^{3}\sum_{i_1=0}^{3}x(i_1,i_2)c_{i_1}c_{i_2}\cos(\frac{\pi(2i_1+1)n_1}{8})\cos(\frac{\pi(2i_2+1)n_2}{8}), \quad \text{(II.1)}$$

where $x(i_1,i_2)$ are the pixels in the 4×4 block and the c_{i_1} and c_{i_2} stand for the corresponding scaling factors. On the other hand, the 8×8 DCT is defined as:

$$X(m_1,m_2)=\sum_{j_2=0}^{7}\sum_{j_1=0}^{7}x(j_1,j_2)c_{i_1}c_{i_2}\cos(\frac{\pi(2j_1+1)m_1}{16})\cos(\frac{\pi(2j_2+1)m_2}{16}), \quad \text{(II.2)}$$

where $x(j_1,j_2)$ are the pixels in the 8×8 block and the c_{j_1} and c_{j_2} stand for the corresponding scaling factors.

According to the Noorkami study [NOO06], a 4×4 visual masking model can be obtained by sub-sampling the $T_{S_{8\times8}}$ IIP matrix and by scaling its coefficients so as to keep a normalized transform:

$$T_{S_{4\times4}}(n_1,n_2)=T_{S_{8\times8}}(2n_1,2n_2)/2=\begin{bmatrix}0.7 & 0.58 & 1.2 & 2.39\\ 0.58 & 1.12 & 1.49 & 2.3\\ 1.2 & 1.49 & 3.07 & 4.35\\ 2.39 & 2.3 & 4.35 & 7.25\end{bmatrix},\forall n_1,n_2\in\{0,1,2,3\}, \quad \text{(II.3)}$$

II.3.1.2 The integer *vs.* floating point transforms

In this step, we consider that the Peterson pixel-domain psycho-visual model $S(.)$ is true and we derived its projection on the MPEG-4 AVC DCT coefficients.

From the mathematical point of view, this means to solve the following optimisation problem:

$$t_{ij}=\arg\min_{e_{i,j}}\left(\left|dist(X,X+f(e_{i,j}))-S\right|\right), \quad \text{(II.4)}$$

where X is the original 4×4 pixel block, $S(.)$ is the Peterson's model in the pixel domain, $e_{i,j}$ is a random value added on the (i,j) MPEG-4 AVC DCT coefficient, $f(.)$ is the back projection from the MPEG-4 AVC DCT to the pixel domain, $dist(.)$ denotes the generic psycho-visual distance characterizing the HVS, and $t_{i,j}$ is the visual threshold we are interested in, $i,j\in\{0,1,2,3\}$.

As for both JPEG/MPEG-2 and MPEG-4 AVC standards the projections from the pixels to the frequency domains are governed by one-to-one differentiable linear functions, the above optimization problem can be re-written in the frequency domain as:

$$t'_{ij}=\arg\min_{e_{ij}}(dist(X+C^t(E\bullet E_b)C,X)-S), \quad \text{(II.5)}$$

where X is the original 4×4 pixel matrix, "$\bullet$" is the element-wise product, and C and E_b are provided by the standard.

The new threshold $T'=(t'_{ij})_{0<i,j<4}$ can be estimated as the value of distortion e_{ij} that minimize $dist$; thus:

$$\frac{\partial dist(X+C^t(E\bullet E_b)C,X)-S}{\partial e_{ij}}=0, \quad \text{(II.6)}$$

$$\Leftrightarrow\frac{\partial dist(X+C^t(A^t(AEA^t)A\bullet E_b)C,X)-S}{\partial e_{ij}}=0, \quad \text{(II.7)}$$

$$\Leftrightarrow\frac{\partial dist(X+A(h\circ g^{-1}(E))A^t,X)-S}{\partial e_{ij}}=0, \quad \text{(II.8)}$$

where h and g denote the two functions governing the classical and the MPEG-4 AVC DCTs, respectively.

It can be noticed that h and g are invertible, differentiable and that $h\circ g^{-1}(.)$ is stable with respect to the $e_{i,j}$ dynamic. The optimum is obtained when $E_{opt}=T'$, or when $h\circ g^{-1}(E_{opt})=T$; hence

$$E_{opt} = g \circ h(T) = T', \tag{II.9}$$

Thus, the perceptual mask for the MPEG-4 AVC DCT, denoted by $TS_{4\times4\text{int}}$ can be obtained:

$$TS_{4\times4\text{int}} = C(A^t TS_{4\times4} A) C^t \bullet E_b, \tag{II.10}$$

where C, A and E_b are the same as above.

From the intuitive point of view, the computation of $TS_{4\times4\text{int}}$ is illustrated in Figure II-12: a value in $TS_{4\times4\text{int}}$ gives the new visibility threshold, *i.e.* the maximal value of a distortion added on an integer DCT coefficient which is still transparent (imperceptible) for a human observer.

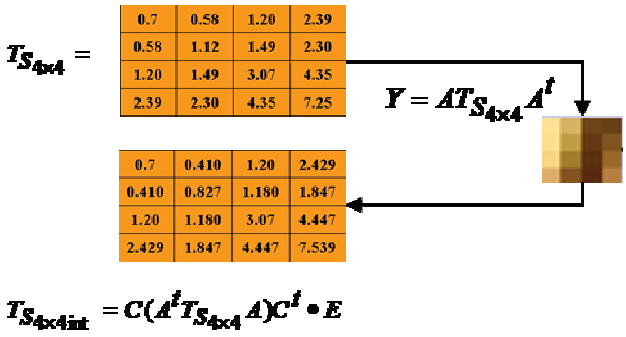


Figure II-12: The influence of the integer DCT transform in the perceptual masking.

II.3.1.3 The prediction error impact

As already mentioned, the MPEG-4 AVC stream is not composed directly by the 4×4 DCT coefficients but by the differences between these coefficients and some values which can be predicted on their neighbors. Be there C_{curr} the current 4×4 coefficient block, and be there $C_{refLeft}$, C_{refUp}, $C_{refUpRight}$, $C_{refUpLeft}$ its reference neighbors, respectively. Then, the prediction error is the E matrix, $E = C_{curr} - C_{pred}$, where C_{pred} is a function of $C_{refLeft}$, C_{refUp}, $C_{refUpRight}$ and $C_{refUpLeft}$, depending on the prediction mode.

The MPEG-4 AVC standard makes provision for 9 different prediction modes: Vertical, Horizontal, DC, Horizontal/Up, Vertical/Left, Horizontal/Down, Vertical/Right, Diagonal Down/Right, Diagonal Down/Left. Our study brought to light that despite the functional differences among these predictions modes, all the C_{pred} predicted matrices can be expressed as a combination of matrix and element-wise operations on the 4×4 integer DCT coefficients corresponding to its neighbor reference matrices:

$$\begin{aligned} C_{pred} &= P_1(C_{refLeft}, C_{refUp}, C_{refUpRight}, C_{refUpLeft}) = \\ &\sum_{i=1}^{4} L_i^{left} \bullet (R_i^t C_{refLeft} Q^t) + \sum_{i=1}^{4} L_i^{up} \bullet (QC_{refUp} R_i) + \sum_{i=1}^{4} L_i^{upRight} \bullet (QC_{refUpRight} R_i) + , \\ &+ \sum_{i=1}^{4} L_i^{upLeft} \bullet (QC_{refUpRight} R_i) \end{aligned} \tag{II.11}$$

where: P_I is a function computing the prediction mode $I=\{1,2,...,9\}$, $\bullet$ is the element-wise product, L_i^{left}, L_i^{up}, $L_i^{upRight}$, L_i^{upLeft} are 4×4 matrices depending on the I prediction mode and:

$$Q=\begin{pmatrix}0&0&0&1\\0&0&0&1\\0&0&0&1\\0&0&0&1\end{pmatrix},\quad R_i=\begin{pmatrix}0&0&0&0\\1&1&1&1\\0&0&0&0\\0&0&0&0\end{pmatrix}i^{th}line, \tag{II.12}$$

when no operation is explicitly written, the matrix product should be considered.

Although the previous equation is very sophisticated, its practical relevance is straight-forward. For instance, the $C_{refUpLeft}$ is involved in only one prediction mode (namely the Diagonal Down/Right). Moreover, several symmetry relations among the L_i^{left}, L_i^{up}, $L_i^{upRight}$, L_i^{upLeft} matrices can also be brought to light:

1. for the *Vertical* prediction mode: $L_i^{left}=L_i^{upRight}=L_i^{upLeft}=0$ and $L_i^{up}=R_i^t$, $i=\{1,2,3,4\}$.
2. for the *Horizontal* prediction mode: $L_i^{up}=L_i^{upRight}=L_i^{upLeft}=0$ and $L_i^{left}=R_i$, $i=\{1,2,3,4\}$,
3. for the *DC* prediction mode: $L_i^{upRight}=L_i^{upLeft}=0$ and:

$$L_i^{left}=L_i^{up}=\frac{1}{8}\begin{bmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{bmatrix},\ i=\{1,2,3,4\}. \tag{II.13}$$

4. for the *Horizontal/Up* prediction mode:

$$L_1^{left}=\begin{bmatrix}1&1&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\ L_2^{left}=\begin{bmatrix}1&2&1&1\\1&1&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\ L_3^{left}=\begin{bmatrix}0&1&1&2\\1&2&1&1\\1&1&0&0\\0&0&0&0\end{bmatrix},\ L_4^{left}=\begin{bmatrix}0&0&0&1\\0&1&1&3\\1&3&1&1\\1&1&1&1\end{bmatrix}, \tag{II.14}$$

$$L_I^{up}=L_j^{upRight}=L_k^{upLeft}=0;\quad i=\{1,2,3,4\},\ j=\{1,2,3,4\},\ k=\{1,2,3,4\}$$

5. for the *Vertical/Left* prediction mode:

$$L_1^{up}=\begin{bmatrix}1&0&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\ L_2^{up}=\begin{bmatrix}1&1&0&0\\2&1&0&0\\1&0&0&0\\1&0&0&0\end{bmatrix},\ L_3^{up}=\begin{bmatrix}0&1&1&1\\1&2&1&0\\1&1&0&0\\2&1&0&0\end{bmatrix},\ L_4^{up}=\begin{bmatrix}0&0&1&1\\0&1&2&1\\0&1&1&0\\1&2&1&0\end{bmatrix},$$

$$L_1^{upRight}=\begin{bmatrix}0&0&0&0\\1&0&0&0\\2&0&0&0\\1&1&0&0\end{bmatrix},\ L_2^{upRight}=\begin{bmatrix}0&0&0&0\\0&0&0&0\\1&0&0&0\\2&0&0&0\end{bmatrix},\ L_3^{upRight}=\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&1\end{bmatrix},\ L_3^{upRight}=0, \tag{II.15}$$

$$L_i^{left}=L_j^{upLeft}=0,\quad i=\{1,2,3,4\},\ j=\{1,2,3,4\}.$$

6. for the *Horizontal/Down* prediction mode:

$$L_1^{up}=\begin{bmatrix}0&1&2&1\\0&0&0&1\\0&0&0&0\\0&0&0&0\end{bmatrix},\ L_2^{up}=\begin{bmatrix}0&0&1&2\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\ L_3^{up}=\begin{bmatrix}0&0&0&1\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\ L_4^{up}=0,$$

$$L_1^{left}=\begin{bmatrix}1&1&0&0\\1&2&1&1\\0&1&1&2\\0&0&0&1\end{bmatrix},\ L_2^{left}=\begin{bmatrix}0&0&0&0\\1&1&0&0\\1&2&1&1\\0&0&1&2\end{bmatrix},\ L_3^{left}=\begin{bmatrix}0&0&0&0\\0&0&0&0\\1&1&0&0\\1&2&1&1\end{bmatrix},\ L_4^{left}=0, \tag{II.16}$$

$$L_j^{upRight}=L_i^{upLeft}=0,\ L_4^{upLeft}=\begin{bmatrix}1&2&1&0\\0&1&1&2\\0&0&0&1\\0&0&0&0\end{bmatrix},\ i=\{1,2,3\},\ j=\{1,2,3,4\}.$$

7. for the *Vertical/Right* prediction mode:

$$L_1^{up}=\begin{bmatrix}1&1&0&0\\1&2&1&0\\0&1&1&0\\0&1&2&1\end{bmatrix},\ L_2^{up}=\begin{bmatrix}0&1&1&0\\0&1&2&1\\0&0&1&1\\0&0&1&2\end{bmatrix},\ L_3^{up}=\begin{bmatrix}0&0&1&1\\0&0&1&2\\0&0&0&1\\0&0&0&1\end{bmatrix},\ L_4^{up}=\begin{bmatrix}0&0&0&1\\0&0&0&1\\0&0&0&0\\0&0&0&0\end{bmatrix},$$

$$L_1^{left}=\begin{bmatrix}0&0&0&0\\1&0&0&0\\2&0&0&0\\1&1&0&0\end{bmatrix},\ L_2^{left}=\begin{bmatrix}0&0&0&0\\0&0&0&0\\1&0&0&0\\2&0&0&0\end{bmatrix},\ L_3^{left}=\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\1&0&0&0\end{bmatrix},\ L_4^{left}=0, \tag{II.17}$$

$$L_i^{upLeft}=L_j^{upRight}=0,\ L_4^{upLeft}=\begin{bmatrix}1&0&0&0\\2&1&0&0\\1&1&0&0\\0&2&1&0\end{bmatrix},\ i=\{1,2,3\},\ j=\{1,2,3,4\}.$$

8. for the *Diagonal Down/Right* prediction mode:

$$L_1^{left}=\begin{bmatrix}0&1&2&1\\1&2&1&0\\2&1&0&0\\1&0&0&0\end{bmatrix},\ L_2^{left}=\begin{bmatrix}1&2&1&0\\2&1&0&0\\1&0&0&0\\0&0&0&0\end{bmatrix},\ L_3^{left}=\begin{bmatrix}2&1&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\ L_4^{left}=\begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},$$

$$L_1^{up}=\begin{bmatrix}0&0&0&1\\0&0&1&2\\0&1&2&1\\1&2&1&0\end{bmatrix},\ L_2^{up}=\begin{bmatrix}0&0&0&0\\0&0&0&1\\0&0&1&2\\0&1&2&1\end{bmatrix},\ L_3^{up}=\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&0&1\\0&0&1&3\end{bmatrix},\ L_4^{up}=0, \tag{II.18}$$

$$L_i^{upLeft}=L_j^{upRight}=0,\ L_4^{upLeft}=\begin{bmatrix}0&0&1&2\\0&1&2&1\\1&2&1&0\\2&1&0&0\end{bmatrix},\ i=\{1,2,3\},\ j=\{1,2,3,4\}.$$

9. for the *Diagonal Down/Left* prediction mode:

$$L_1^{up}=\begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\ L_2^{up}=\begin{bmatrix}2&1&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\ L_3^{up}=\begin{bmatrix}1&2&1&0\\2&1&0&0\\1&0&0&0\\0&0&0&0\end{bmatrix},\ L_4^{up}=\begin{bmatrix}0&1&2&1\\1&2&1&0\\2&1&0&0\\1&0&0&0\end{bmatrix}, \tag{II.19}$$

$$L_1^{upRight}=\begin{bmatrix}0&0&1&2\\0&1&2&1\\1&2&1&0\\2&1&0&0\end{bmatrix}, L_2^{upRight}=\begin{bmatrix}0&0&0&1\\0&0&1&2\\0&1&2&1\\1&2&1&0\end{bmatrix}, L_3^{upRight}=\begin{bmatrix}0&0&0&0\\0&0&0&1\\0&0&1&2\\0&1&2&1\end{bmatrix}, L_4^{upRight}=\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&0&1\\0&0&1&2\end{bmatrix},$$

$$L_j^{left}=L_i^{upLeft}=0,\ i=\{1,2,3,4\},\ j=\{1,2,3,4\}.$$

The perceptual mask in the integer DCT domain, applied to the prediction error (further denoted as T_{AVC_pred}) is defined as:

$$T_{AVC\_pred}=\max_{imperceptibility}(E_{modified}-E_{original}), \tag{II.20}$$

As the modifications are done directly on the prediction errors, the reference is kept unchanged; hence:

$$\begin{aligned}E_{modified}&=C_{currmodified}-P_l(C_{refLeft},C_{refUp},C_{refUpRight},C_{refUpLeft})\\E_{original}&=C_{currmodified}-P_k(C_{refLeft},C_{refUp},C_{refUpRight},C_{refUpLeft})\end{aligned}, \tag{II.21}$$

where l and k correspond to the original and modified prediction modes, respectively. Since the prediction mode is computed in MPEG-4 AVC out of minimizing the energies among neighbor blocks, under the imperceptibility constraints it is possible to assume that the prediction mode is not changing (this was actually the case in each and every experiment we carried out in Section II.3.2). Hence,

$$T_{AVC\_pred}=\max_{imperceptibility}(C_{currmodified}-C_{currorginal})=T_{S_{4\times4\,\mathrm{int}}}. \tag{II.22}$$

The last equation demonstrates that the perceptual masking matrix corresponding to the prediction error in the vertical mode is identical to the masking matrix in the coefficient domain.

II.3.2 Experimental validation

The experiments were carried out on the two corpora presented in the Appendix D: 10 sequences of 25 minutes each, coded at both SD (at 64kbyte/s) and HD (5 Mbyte/s) resolutions.

The results are synoptically presented in Figures II-13 and II-14.

The watermarking method presented in [BEL10] was applied by considering four different types perceptual masks: (1) a random generated 4×4 matrix (*i.e.* no perceptual masking), (2) the Noorkami $T_{S_{4\times4}}$ mask, (3) the T_{AVC_pred} mask computed according to (II.22), and (4) eight noisy perceptual masks (denoted by T_{noise1} to T_{noise8}) obtained by adding to the T_{AVC_pred} uniform random noise of 0 mean and different variances, lower than the corresponding coefficient divided by 10.

The first experiment (Figure II-13) was devoted to the evaluation of the impact of the new masking model in the watermarking transparency. In this respect, fixed data payload (96 bits of information inserted into excerpts of 5 minutes) and robustness (against transcoding and geometric attacks – the StirMark random bending) were imposed. The corresponding transparency was evaluated according to two objective measures: the popular PSNR and the Watson's DVQ.

When considering SD encoded video, the T_{AVC_pred} masking model ensures gains in PSNR of 7 db and 3.2 dB with respect to the no-masking and the Noorkami's models are obtained, respectively. A

significantly decreased DVQ is also obtained: gains by factors of 0.52 and 0.47 are obtained with respect to the no-masking and the Noorkami's models, respectively.

The same good results are obtained when considering HD encoded video. It can be seen that the T_{AVC_pred} masking model ensures a gain of 6dB with respect to the no-masking procedure and of 2dB with respect to the Noorkami model. The gains are even more important when considering the Watson's perceptual measure: this measure is reduced by a factor of 2.58 when compared to the no-masking approach and by a factor 2.20 when compared to the Noorkami masking. It can be also seen that the T_{AVC_pred} values accurate: from the PSNR point of view, the noisy versions of the new perceptual mask lead to similar results to the Noorkami's mask; however, it significantly outperforms state-of-the art competitors from the DVQ point of view.

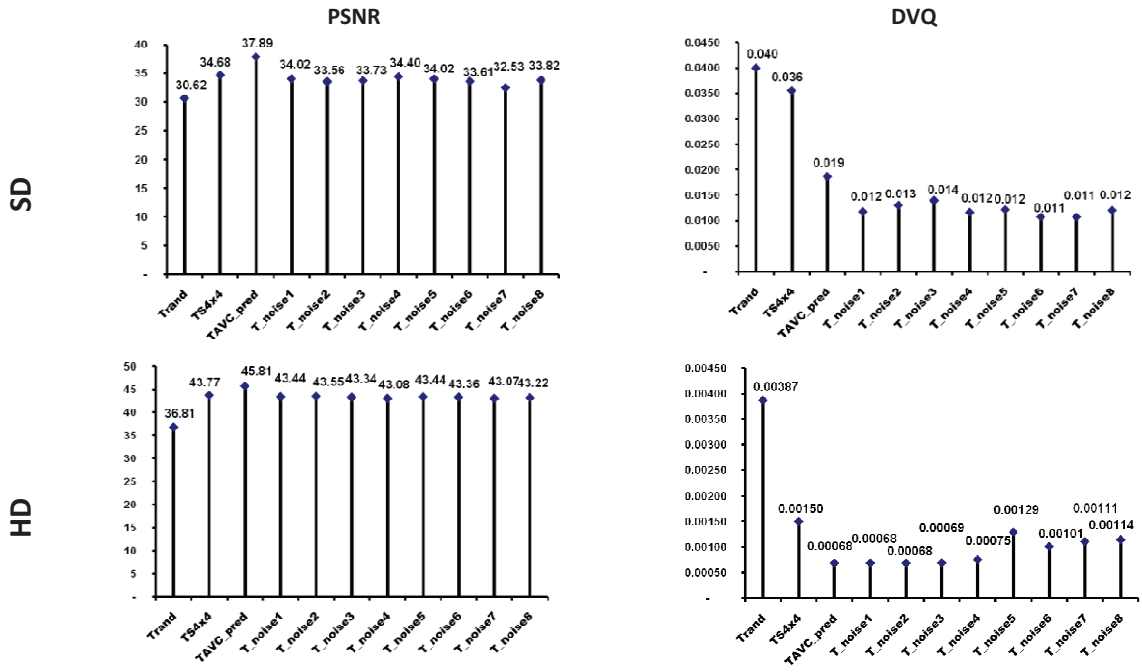


Figure II-13: The impact of the perceptual masking model in watermarking transparency for SD video encoded at 64 kbyte/s (up) and for HD video encoded at 5Mbyte/s (bottom). Two objective metrics are considered, namely the PSNR (left) and DVQ (right).

The second experiment (Figure II-13) investigates the impact of the new masking model in the data payload. In this respect, we first imposed a transparency expressed by a prescribed PSNR (35dB in the SD case and 44dB in HD case) and a robustness against transcoding the geometric attacks (SirMark random bending) and we evaluated the maximal data payload corresponding to video excerpts of 5 minutes. Secondly, we imposed a fixed DVQ (0.033 in the SD case and 0.0016 in the HD case) and we evaluated the maximal data payload corresponding to the same video excerpts of 5 minutes and to the same robustness against transcoding and geometric attacks.

When considering SD encoded video, significant gains in data payload are obtained with respect to the Noorkami's model: 20% (for a prescribed PSNR) and 54% (for a prescribed value for DVQ). Of course, the gain of the new model over the no-masking case is even larger, namely 156% (when

considering the PSNR as a transparency criterion) and 81% (when considering the DVQ as a transparency criterion).

When considering HD encoded video and the PSNR measure, it was brought to light that the data payload is increased by a factor of 15 with respect to the basic random mask and by 5% with respect to the Noorkami's masking procedure. When considering now the DVQ, gains by 30% are obtained with respect to the Noorkami's model and by factors of 2 with respect to the case in which no masking model is considered.

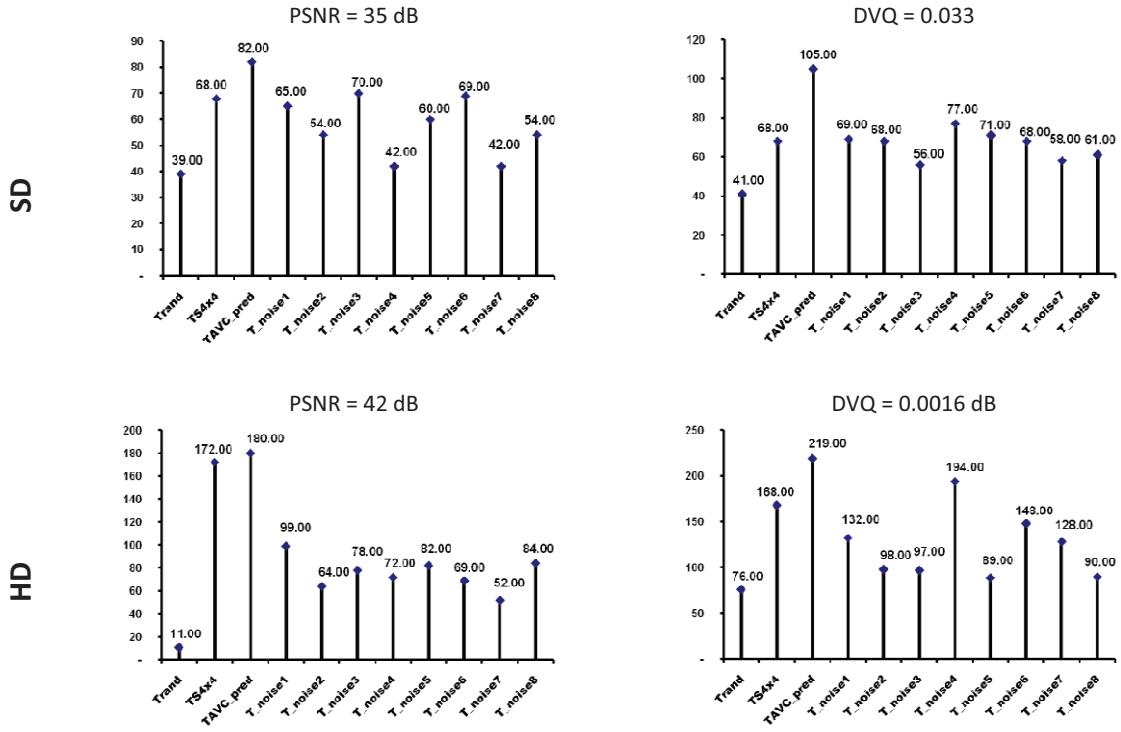


Figure II-14: The impact of the perceptual masking model in the data payload for SD video encoded at 64kbyte/s (up) and for HD video encoded at 5Mbyte/s (bottom). Two objective metrics are considered, namely the PSNR (left) and DVQ (right).

II.4 Conclusion

This chapter approaches the *in-band enriched video* paradigm from the transparency point of view. The first contribution (Chapter II.2) consists in a reference study about the possibility of modifying the MPEG-4 AVC stream elements under imperceptibility constraints. It is thus demonstrated that the redundancy still existing in the MPEG-4 AVC stream allows for several types of modifications (additive, substitutive) to be done. The related practical limits are established for both SD and HD video, by considering 8 visual quality metrics of three types (pixel based, correlation based and psychovisual based).

The second contribution consists in computing the first masking model devoted to the MPEG-4 AVC compressed stream. In this respect, the mathematical demonstration combines the algebraic modeling of the MPEG-4 AVC operations to an optimization problem solving. The practical validation

is obtained under the watermarking framework and points to significant improvement in both transparency (*e.g.* a gain of 3dB) and data payload (*e.g.* a gain of 50%) with respect to state-of-the-art masking models.

REFERENCES

[AHU92] A. J. Ahumada "Luminance-Model-Based DCT Quantization for Color Image Compression" Proc. SPIE 1666, 365 (1992).

[BEL10] M. Belhaj, M. Mitrea, S. Duta, F Preteux,"MPEG-4 AVC robust video watermarking Based based on QIM and perceptual masking", IEEE International Conference on Comminucations, Bucharest, pp. 477-480, June 2010.
[BRU96] V. Bruce, P.R. Green, and M.A. Georgeson, Visual Perception, 3rd ed.: Psychology Press, 1996.

[CAB11] A, S, Cabrita, F, P, Naccari "Perceptually driven coefficients pruning and quantization for the H.264/A VC standard". EUROCON 2011.
[CHE09] Z. Chen and C. Guillemot, "Perceptually-Friendly H264/AVC Video Coding," in 2009 IEEE International Conference On Image Processing, Cairo, Egypt, 7-10 Nov. 2009.
[CHI11] S. Chikkerur, V. Sundaram, M. Reisslein, L.J Karam "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison" Sch. of Electr., Comput., & Energy Eng., Arizona State Univ., Tempe, AZ, USA June 2011
[COX02] I. Cox, Miller, J. Bloom. Digital Watermarking. Morgan Kaufmann Publishers, 2002.
[COX08] I. Cox, M.L. Miller, J. Bloom, J. Fridrich, T. Kalker. Digital Watermarking and Steganography. Second edition, Morgan Kaufmann, Burlington, MA, 2008.

[DAM06] I. Damnjanovic, E Izquierdo "Perceptual watermarking using just noticeable difference model based on block classification" Proceeding MobiMedia '06 Proceedings of the 2nd international conference on Mobile multimedia communications , 2006.
[DUT07] S. Duta, M. Mitrea, F. Prêteux. "Compressed versus uncompressed domain video watermarking", Proc. SPIE, Vol. 6700, p. 67000A, August 2007.

[HAR98] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," Signal Process., vol. 66, no. 3, pp. 283–301, May 1998.
[HUA07] C. Huang and C. Lin, "A Novel 4-D Perceptual Quantization Modelling For H.264 Bit-Rate Control," Multimedia, IEEE Transactions on, vol. 9, no. 6, pp. 1113 - 1124, Oct. 2007.
[HUH11] K. Hühring, H.264/AVC Joint Model 8.6 (JM-8.6) Reference Software [Online]. Available: http://iphome.hhi.de/suehring/tml/

[ITU93] ITU-T Rec. H.261, "Video Codec for Audio-Visual Services at 64-1920 kbit/s," 1993.

[JVT03] JVT of ISO/IEC MPEG And ITU-T VCEG, "ITU-T Recommendation And Final Draft International Standard Of Joint Video Specification (ITU-T Rec. H.264|ISO/IEC 14496-10 AVC)," JVT-G050, 2003.

[LI04] Q. Li and E.-C. Change, , J. , Ed., “On the possibility of noninvertible watermarking schemes,” in IH 2004, Lecture notes in Computer science (LNCS). Berlin: Springer-Verlag, vol. 3200, pp. 13–2, 2004.
[LIN07] G. Lin and S. Zheng, "Perceptual Importance Analysis for H.264/AVC Bit Allocation," Journal of Zhejiang University SCIENCE A, vol. 9, no. 2, pp. 225 - 231, July 2007.
[LI07] Q. Li and I. J. Cox, "Using perceptual models to improve fidelity and provide resistance to valumetric scaling for quantization index modulation watermarking," IEEE Transactions on Information Forensics and Security, vol. 2, no. 2, pp. 127-139, 2007.

[MAI06] Z. Mai, C. Yang, K. Kuang, and L. Po, "A Novel Motion Estimation Method Based On Structural Similarity For H.264 Inter Prediction," in ICASSP 2006 Proceedings, IEEE International Conference on 93 Acoustics, Speech and Signal Processing, vol. 2, Toulouse, France, pp. 913-916, May 2006.
[MIN05] K. Minoo and T.Q. Nguyen, "Perceptual Video Coding With H.264," in Conference Record of the Thirty-Ninth Asilomar Conference, Pacific Grove, CA, USA, pp. 741-745, 2005.

[NAC10] M. Naccari and F. Pereira, "Comparing Spatial Masking Modelling In Just Noticeable Distortion Controlled H.264/AVC Video Coding," in 11th WIAMIS, Workshop on Image Analysis for Multimedia Interactive Services WIAMIS, vol. 1, Desenzano del Garda, Italy, April 2010.
[NOO05] M. Noorkami, R.M. Mersereau “Compressed-Domain Video Watermarking for H.264”, Proc. IEEE International Conference on Image Processing, ICIP, Vol. 2, p. II-890-3, Genoa, Italy, Sep. 11-14, 2005.

[PET91] Peterson H. A, Pennebaker W. B. “Quantization of color image components in the DCT domain“, Visual Processing, and digital Display II, Proc. SPIE, vol. 1453, pp. 210-222, 1991.

[QIA06] Y. Qiao, Q. Hu, G. Qian, S. Luo, and W. L. Nowinski, "Thresholding based on variance and intensity contrast," Pattern Recognition, vol. 40, no. 2, pp. 596 - 608, July 2006.

[RIC03] I.E.G. Richardson, H.264 And MPEG-4 Video Compression: Video Coding For Next-Generation Multimedia. Chichester: John Wiley & Sons, 2003.

[TAU02] D. Taubman and M.W. Marcellin, "JPEG2000: Image Compression Fundamentals," in Standards and Boston, 2002.

[VER96] O, Verscheure, A, Basso, M, El-Maliki, J.P Hubaux, "Perceptual bit allocation for MPEG-2 CBR video coding" Swiss Federal Inst. of Technol., Lausanne Image Processing, 1996. Proceedings., International Conference on Sep 1996 117 - 120 vol.2,1996.

[WAT98] A. B. Watson., “DCT Quatization Matrices Optimized for individual Images”, Proc. SPIE, Vol.1913, pp. 202-216, 1998.
[WAT83] A. B. Watson, H.B. Barlow, Robsin J. G. “What does the eyes see the best?” Marcmillan Journals Ltd 1983.
[WIE03] T. Wiegand, G.J. Sullivan, G. Bjontengaard, and A. Luthra, "Overview Of The H.264/AVC Video Coding Standard," Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560 - 576, July 2003.

[WEI09] L. Wei, O Issa, L Hong. F. Speranza, R. Renaud “Quality Assessment of Video Content for HD IPTV Applications”. ISM '09. 11th IEEE International Symposium on, pp, 517 – 522, San Diego, CA ,14-16 Dec 2009.

Chapter III

Robustness

The good tree is not growing easily.

The stronger the wind, the more robust the tree

J. W. Warriott

Abstract

By presenting COMWat, this chapter deals with the robustness issue for MPEG-4 AVC watermarking. In this respect, the results are presented at three incremental levels. At the basis, the first MPEG-4 AVC method featuring robustness against noise addition, transcoding, and the StirMark (geometric) attacks is adavanced. The mark is inserted into the MPEG-4 AVC quantization indexes selected according to an energy-based selection criterion validated by information theory basic concepts. The insertion procedure combines the QIM principles and the perceptual mask obtained in this thesis. At the second level, the very QIM principles are generalized beyond the binary case, thus advancing the mQIM insertion/detection methods. Its main theoretical and practical advantage is the increase the data payload by a log(m) factor, while preserving the transparency and the robustness. Finally, the MPEG-4 AVC syntax is reconsidered as a starting point in designing a counter-attack procedure; in practice, this procedure results in BER reduction by 5% to 10%, according to the original video bit rate.

Beside this application driven presentation, Chapter III also deals with fundamental theoretical tools in information theory which allows the performances of the developed methods to be objectively compared to their theoretical limits.

Contents

III.1 Introduction

As the Chapter II brought to light that transparent video watermarking is possible in the MPEG-4 AVC domain, the present Chapter takes the challenge of designing a practical MPEG-4 AVC watermarking technique (further referred to as on COMWat), meeting the requirements of transparency and robustness. In this respect, beyond the Introduction and Conclusion, a structure on three incremental levels is considered, see Figure III-1.

At the basis, Chapter III.2 reports on the first watermarking method in the MPEG-4 AVC domain featuring robustness against transcoding and geometric attacks and whose transparency is supported by both objective and subjective assessments. The mark is inserted into the MPEG-4 AVC quantization indexes selected according to an energy-based selection criterion, validated by information theory basic concepts. The insertion procedure combines the QIM principles and the perceptual mask obtained in this thesis.

Secondly, in Chapter III.3, in order to increase the data payload performances, an extension to multi-symbol insertion is advanced. The main novelty consists in deriving the equations underlying the multiple symbols Quantisation Index Modulation (*mQIM*) insertion rule. This method allows the size of the inserted mark (the data payload) to be increased by a factor $\log_2 m$, while keeping the same good transparency (objectively and subjectively evaluated) and robustness (against transcoding and geometric attacks).

Chapter III.4 reconsiders and extends the *m*QIM insertion technique in order to increase its overall performances (transparency and robustness against re-encoding attacks). In this respect, an objective study on the practical impact of the errors induced by the MPEG-4 AVC (iterative) decoding/re-encoding is carried out and the related counter-attack is specified. The results show significant increase of the PSNR (about 1.3dB) and decrease of the BER (by 5% to 10%).

Chapter III.5 concludes this section.

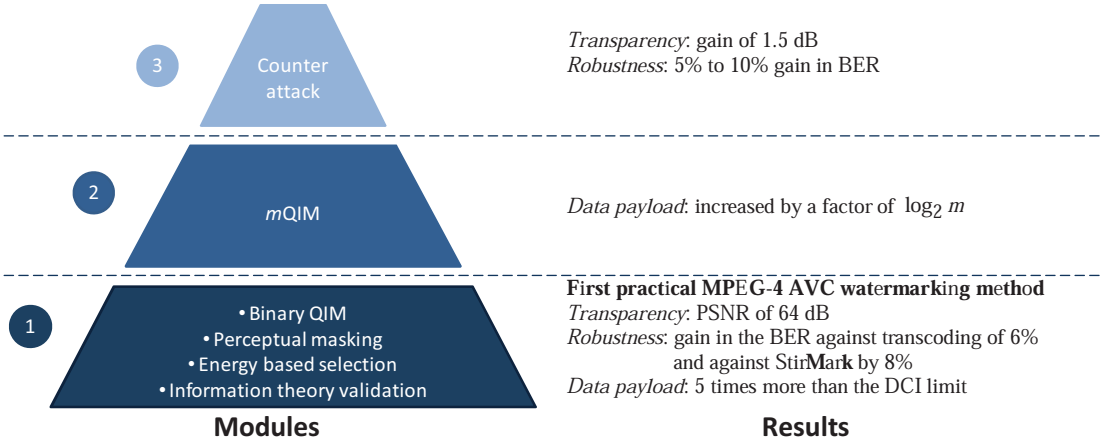


Figure III-1: Synopsis of the three layers in the COMWat method: (1) basic method based on the binary QIM, *cf*. Chapter III.2; (2) *m*QIM extension *cf*. Chapter III.3, and (3) MPEG-4 AVC based counterattacks, *cf*. Chapter III.4.

III.2 Binary QIM method for MPEG-4 AVC watermarking

At its basic level, the COMWat method reconsiders the QIM principles in order to achieve robustness against both transcoding and geometric attacks; in this respect, perceptual shaping and mark insertion after MPEG-4 AVC quantization are key enables. This new method is structured into some specific modules for insertion and detection sides (Figure III-2): perceptual masking, mark generation, embedding and block reconstruction.

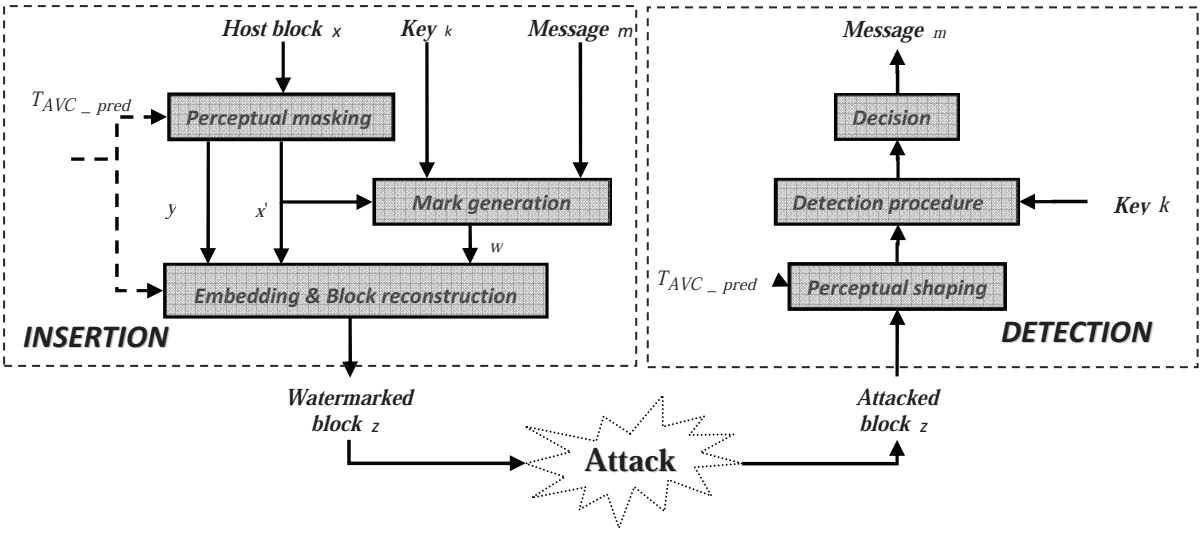


Figure III-2: Synopsis of the first layer in the COMWat method.

III.2.1 Insertion Step

The compressed domain watermarking state of art exhibits ST QIM – Spread Transform Quantisation Index Modulation methods inserting the mark in the AC coefficients [EGG03], [GOL07]. While robust against transcoding, this approach cannot withstand the geometric attacks. The method in the present chapter reconsiders the QIM principles in order to achieve robustness against both transcoding and geometric attacks.

Perceptual masking

This module was designed so as to identify the original content components which are relevant from the HVS (human visual system) point of view.

It has as input the original (unmarked) content *x* (Figure III-3) and is parameterised by the perceptual mask T_{AVC_pred} computed in Chapter II.

The method presented in this chapter deals with the AC luma quantisation indexes corresponding to the 4×4 sub-macroblocks in the *I* frames [NOO05] [NOO06] [NOO07a] [BEL10]; hence, *x* denotes a 15 component vector, obtained by zigzag scanning the quantisation indexes of such a macroblock. In order to take into consideration the peculiarities of the HSV, the input vector is projected onto a perceptual mask (denoted by T_{AVC_pred} in the matrix representation and by t_{AVC_pred} in the

vector representation), prior to the insertion. Once the T_{AVC_pred} mask matrix is computed, the t_{AVC_pred} vector is obtained by scanning in a zigzag order the values corresponding to the AC frequencies and by normalisation.

The perceptual shaping module provides $x' T_{AVC\_pred}$ and y which are the x orthogonal decomposition on the t_{AVC_pred} direction and on its normal, respectively (Figure III-4):

$$x' = x^t \cdot T_{AVC\_pred}, \quad \text{(III.1)}$$

$$y = x - x' \cdot T_{AVC\_pred}, \quad \text{(III.2)}$$

where "$\cdot$" denotes the scalar product and "-"denotes the vector difference. Note that x' stands for a scalar, while y is a vector. Assuming T_{AVC_pred} realistically models the HSV, the lowest artefacts are expected to be obtained when inserting the mark in the $x' T_{AVC\_pred}$ component.

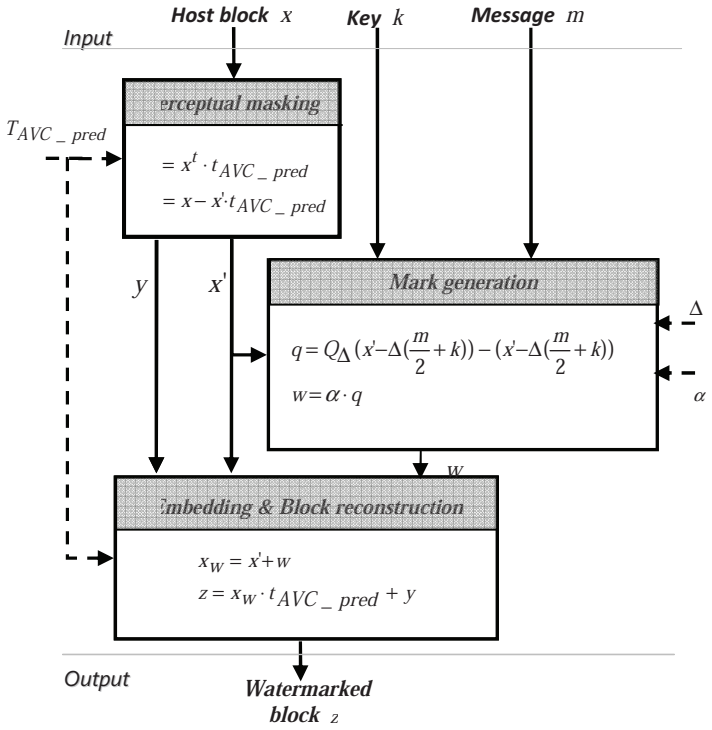


Figure III-3: The insertion synopsis: three inputs (the message m, the host x, the key k) and three parameters (the perceptual mask T_{AVC_pred}, the quantization step Δ, and the scaling factor α) are considered to compute the marked data z.

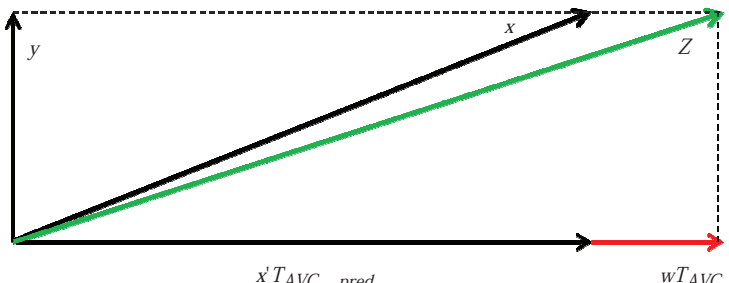


Figure III-4: Vector-based representation for the insertion: the mark is added on a direction given by the perceptual mask vector, *cf*. equations (III.1) and (III.2).

Mark generation

This module encodes the m binary message to be inserted into a quantisation error, according to the general QIM principles [EGG03][GOL07].

It has as input the m binary message, the x' projection of the original (unmarked) content on the T_{AVC_pred} perceptual mask and the k key (a value sampled from a $[0;1)$ uniform random number generator). The generation process is parameterized by the quantisation step Δ (selected in the neighbourhood of Q_{step} of the MPEG-4 AVC) and by a scaling factor α. The w watermark is generated according to a dither modulation formula:

$$q = Q_\Delta(x' - \Delta(m/2 - k)) - (x' - \Delta(m/2 - k)), \quad \text{(III.3)}$$

$$w = \alpha \cdot q, \quad \text{(III.4)}$$

where $Q_\Delta(u) = \Delta \cdot round(u/\Delta)$.

The advantages of such a technique are well known in the literature [CHE01]: independence between the watermark and the host, minimal quantisation artefacts, and improved security.

Embedding & block reconstruction

This module generates the watermarked 4×4 MPEG-4 AVC block.

It has as input the w watermark and the original content, represented by its two components $x'tAVC$ and y. In the present chapter, the insertion follows a simple additive rule and the watermarked z vector is reconstructed by summing (in the vector sense) the watermarked component on the perceptual masking direction with the initial (unmarked) component on the normal to the perceptual masking:

$$x_W = x' + w, \quad \text{(III.5)}$$

$$z = x_W \cdot T_{AVC\_pred} + y. \quad \text{(III.6)}$$

The watermarked block is obtained from the z vector by inverse zigzag scanning. The position of the 4×4 watermarked block in the 16×16 luma block is stored in order to regain synchronization at the detection side.

III.2.2 Detection procedure

The detection procedure starts by parsing the supposed watermarked 4×4 block, by scanning it in a zigzag order and by recording the corresponding AC coefficients in a vector denoted by z'. The $Y(m)$ detection variable is computed as follows:

$$Y(m) = Q_\Delta((z')^t \cdot T_{AVC\_pred} - k\Delta) - ((z')^t \cdot T_{AVC\_pred} - k\Delta), \quad \text{(III.7)}$$

where k and Δ are the key and the QIM quantisation step. Be there the ideal case in which $z' = x_W$, *i.e.* no attack occurred and the MPEG-4 AVC inner operations did not affected the watermarked block.

In this ideal case, we compute $Y(m)$ in order to get the decision's threshold value for a best detection:

$$Y(m) = Q_\Delta(x_W - k\Delta) - (x_W - k\Delta), \tag{III.8}$$

Let's $A = x' - \Delta(\frac{m}{2} + k)$ and $B = x' - \alpha q - k\Delta$ (III.9)

Thus,
$$\begin{aligned} B &= (\alpha - 1)q + Q_\Delta(A) + \Delta\frac{m}{2} \\ B'(m) &= (\alpha - 1)q + \Delta\frac{m}{2} \end{aligned} \tag{III.10}$$

$$Y(m) = Q_\Delta(B) - (B) = Q_\Delta(B'(m)) - (B'(m)) \tag{III.11}$$

$$-\frac{\Delta}{2} \le q \le \frac{\Delta}{2} \tag{III.12}$$

$$-\frac{(1-\alpha+m)\Delta}{2} \le B'(m) \le \frac{(1-\alpha+m)\Delta}{2}. \tag{III.13}$$

$Y(m)$, as any quantization error, belongs to the $\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ interval. Thus, to get a good detection, we have to divide this zone into 2 non-overlapping zones correspondently to $Y(m=0)$ and $Y(m=1)$. In this respect, the $B'(m)$ behaviour should be considered.

If $m = 0$ (III.14)

$$-\frac{(1-\alpha)\Delta}{2} \le B' \le \frac{(1-\alpha)\Delta}{2} \tag{III.15}$$

$$-\frac{(1-\alpha)\Delta}{2} \le Y(m=0) \le \frac{(1-\alpha)\Delta}{2} \tag{III.16}$$

If $m = 1$ (III.17)

$$\frac{\alpha\Delta}{2} \le B' \le \frac{(2-\alpha)\Delta}{2} \tag{III.18}$$

$$\begin{cases} -\frac{\Delta}{2} \le Y(m=1) \le -\frac{\alpha\Delta}{2} \\ \text{or} \\ \frac{\alpha\Delta}{2} \le Y(m=1) \le \frac{\Delta}{2} \end{cases}. \tag{III.19}$$

This last result (III.19) and (III.20) can be depicted in Figure III-5.

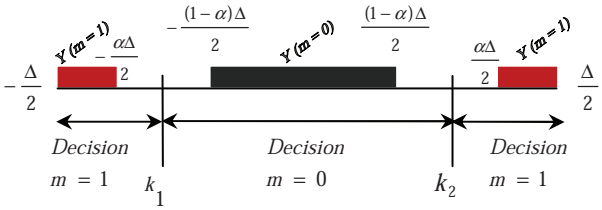


Figure III-5: Decision regions as a function of α.

The decision is done according two thresholds k_1 and k_2, resulting in the $I_1 = [-\frac{(1-\alpha)\Delta}{2} - \frac{\alpha\Delta}{2}]$ and $I_1 = [-\frac{(1-\alpha)\Delta}{2} - \frac{\alpha\Delta}{2}]$ intervals, respectively:

- If $k_1 < Y(m) < k_2 \Rightarrow m = 0$

- Else, $m = 1$.

The choice of the detection threshold is done according to a minimal error probability criterion. It can be demonstrated [BEL08], if $\alpha \geq 0.5$, then $P_{k_1,k_2}(Erreur) = 0$; otherwise (if $\alpha \geq 0.5$), then $P_{k_1,k_2}(Erreur) = \frac{1-2\alpha}{2(1-\alpha)}$. Thus we find that error probability is independent of threshold so for a low complexity we choose $|k_1| = |k_2| = \Delta/2$.

Assuming random behaviours for the original content and the key and the absence of any kind of attacks, it can be proved that $Y(m=0)$ belongs to the $(-(1-\alpha)\Delta/2\,;(1-\alpha)\Delta/2)$ interval, while $Y(m=1)$ belongs to the union of the $(-\Delta/2;-\alpha\Delta/2)$ and $(\alpha\Delta/2;\Delta/2)$ intervals, Figure III-5.

In order to ensure the method effectiveness (*i.e.* possibility of recovering the mark in the absence of attacks), the *Y*(0) and *Y*(1) variation intervals should not overlap; hence, $\alpha > 0.5$. As in practice the watermarked content is always modified by some attacks *prior* to detection, $Y(m=0)$ and $Y(m=1)$ will cover wider intervals then previously mentioned. Assuming all these attacks can be represented by random additive noise, the optimal decision rule is:

$$\begin{cases} \text{if } |Y(m)| < \Delta/4 \text{ then } m = 0 \\ \text{if } |Y(m)| > \Delta/4 \text{ then } m = 1 \end{cases}. \qquad \text{(III.20)}$$

III.2.3 EXPERIMENTAL RESULTS

Video corpus

The experiments were carried out on the corpus presented in Appendix B.

Robustness

The experimental study started by a robustness-driven analysis on the optimal way of selecting the macroblocks in which the mark is to be inserted. In this respect, a first scenario considers two non-zero macroblocks, randomly chosen in each *I* frame. The method proved its effectiveness and its robustness against noise addition (in the transformed domain). On the contrary, the mark is lost after transcoding and geometric (StirMark) attacks. From the information theory point of view, this lack of robustness is connected to a channel with very low mutual information, as illustrated in Figure III-6. The channel corresponding to the watermarking method in this study is binary: the input may be 0 or 1 (the bits composing the m mark) and the output is represented by the decision according to the rule expressed by (III-20). The corresponding noise matrices were estimated on the video corpus, for both transcoding (decoding from MPEG-4 AVC to AVI and than re-encoded to MPEG-4 AVC) and

geometric attacks (StirMark), Figure III-6. It can be directly noticed that these matrices represent channels with very strong noise: the mutual information is very close to zero. Note: the relative error in noise matrix estimation is $\hat{\varepsilon}_I \cong 5\%$.

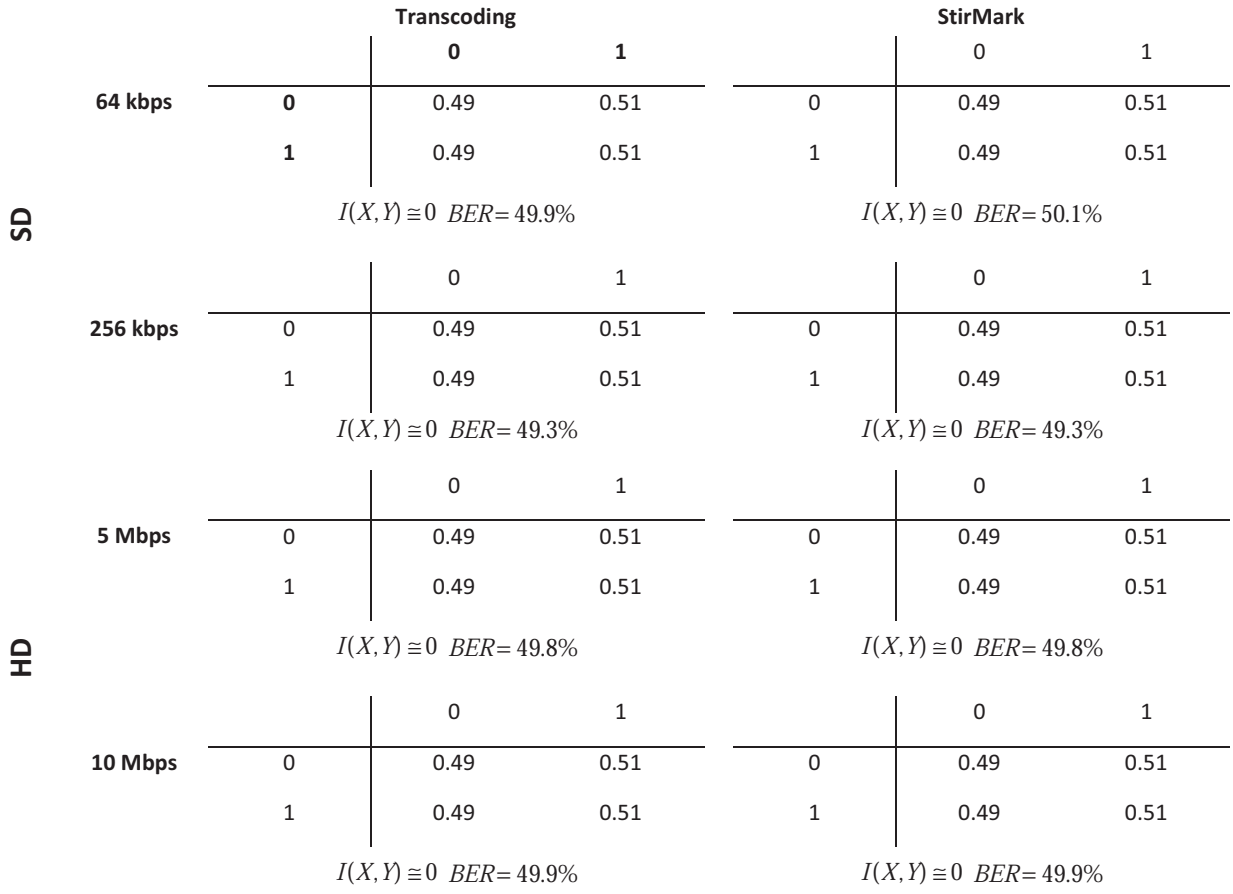


Figure III-6: Noise matrices and the crresponding mutual information values for transcoding (left) and geometric (StirMark) attacks (right), in the case the macroblocks to be watermarked are randomly chosen.

Consequently, a second scenario for macroblock selection had to be considered. This time, the mark is inserted only in the macroblocks whose energies prior to and after the insertion verify the condition:

$$\mu_X - \sigma_X/16 < 1 - \|x\|/\|z\| < \mu_X + \sigma_X/16\,, \tag{III.21}$$

where, x and z are the original and the watermarked macroblocks, $\|\ \|$ represents the energy of the block while μ_X and σ_X stand for the mean and standard deviation of energy of selected blocks. Figure III-7 illustrates the noise matrices and the related mutual information values corresponding to this situation. These values clearly point to robustness. However, even a simple visual inspection of these matrices brought into light an un-usual behaviour: in the case of the geometric attacks, the watermarking channel acts as an inversion channel (*i.e.* the inserted bits of 0 are transformed in 1 and the inserted bits of 1 are transformed in 0, with probabilities larger than 90%). Assuming the inserted mark corresponds to a logo (*e.g.* the ARTEMIS logo, Figure III-8), this means that at reception

the logo is inverted, as illustrated in Figure III 11 However, this is not a practical drawback: a simple inversion conditioned on a correlation-based threshold is able to solve the problem.

SD

64 kbps

Transcoding

	0	1
0	0.93	0.07
1	0.06	0.94

$I(X,Y)=0.653$ $BER=7.23\%$

StirMark

	0	1
0	0.10	0.90
1	0.93	0.07

$I(X,Y)=0.56$ $BER=8.11\%$

256 kbps

	0	1
0	0.93	0.07
1	0.04	0.96

$I(X,Y)=0.692$ $BER=6.51\%$

	0	1
0	0.07	0.93
1	0.92	0.08

$I(X,Y)=0.61$ $BER=8.01\%$

HD

5 Mbps

	0	1
0	0.97	0.03
1	0.09	0.9

$I(X,Y)=0.678$ $BER=7\%$

	0	1
0	0.06	0.93
1	0.95	0.05

$I(X,Y)=0.686$ $BER=5.52\%$

10 Mbps

	0	1
0	0.94	0.06
1	0.04	0.96

$I(X,Y)=0.715$ $BER=6.23\%$

	0	1
0	0.04	0.96
1	0.94	0.06

$I(X,Y)=0.705$ $BER=6.34\%$

Figure III-7: Noise matrices and the corresponding mutual information values for transcoding (left) and geometric (StirMark) attacks (right), in the case the macroblocks to be watermarked are chosen according to an energy-selection criterion.

Figure III-8: The ARTEMIS binary logo: original (left), recovered after transcoding (middle) and geometric attacks (right). The original video sequence was encoded at 10 Mbps.

Note that these good results concerning the robustness were obtained at the expense of the data payload: only 72, 102, 239 or 360 bits can be inserted per 5 minutes of video encoded at 64 kbps, 256 kbps, 5Mbps or 10Mbps, respectively. Although these values may seem quite low, they are still 2 to 10 times larger (according to the original video bit rate) than the minimal requirements set by the DCI standards [DCI08].

The overall results can be also considered from a theoretical point of view. Actually, assuming the relative error in probability estimation is low enough and that memory-less, time-invariant, independent-noise model accurate represent our watermarking method, the related capacity can be computed. The capacity value expresses the maximal theoretical limit for the data payload. The

capacity was computed in the eight above considered cases (for the four bit-rates and for the two considered attacks) and is depicted in Figure III-9, alongside with the actual mutual information values. As it can be noticed, this first method is quite close to its theoretical limits: differences lower than 8% between mutual information and capacity are obtained. Moreover, note that these differences correspond to the cases in which the content, although binary MPEG-4 AVC encoded is not practically compressed (*i.e.* content encoded at 10Mbps). For the cases in which the strongest compression is applied (*i.e.* 64 kbps), the mutual information approaches its upper theoretical limit by relative errors lower than 0.3% or by 1.3%, in the cases of the transcoding and StirMark attacks, respectively. This result demonstrates that the procedure thus obtained can be considered as optimal in the sense that maximises the data payload for prescribed transparency and robustness constraints and that such a behaviour is more effective for stronger compressions.

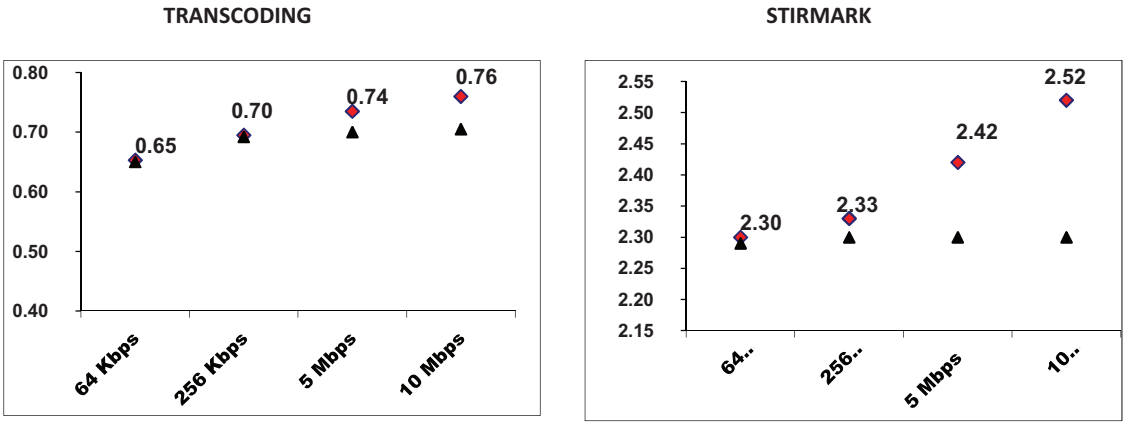


Figure III-9: Practical (mutual information, in black triangles) *versus* theoretical limits for data payload (the capacity, in diamonds), in the cases of the transcoding and geometric (StirMark) attacks on the binary QIM watermarking with energy selection criteria**.**

Transparency

The transparency of the watermarked video has been subjectively supported by a panel of 10 human observers of different ages and professional backgrounds, involved into a Two-alternative forced choice test. The objective evaluations also supported the transparency. In order to prove that, the 8 already considered metrics (see Chapter II) have been reconsidered, Table III-1.

Table III-1: Objective evaluation of the transparency for binary QIM watermarking.

		Pixels based				*Correlation based*			*Psychovisual*
		PSNR (dB)	**PMSE ($x10^{-6}$)**	**IF**	**AAD**	**NCC**	**CQ**	**SC**	**DVQ**
SD	**64 kbps**	41	0.0488	0.999	0.0017	1.00	89	0.97	0.016
	256 kbps	42.1	0.043	0.999	0.0012	1.00	90	0.999	0.023
HD	**5 Mbps**	64	0.026	0.999	0.0001	1.00	98	1.00	0.00045
	10 Mbps	63.2	0.031	0.999	0.0002	1.00	94	1.00	0.00032

III.3 mQIM method for MPEG-4 AVC watermarking

The previous section demonstrates that the advanced watermarking method is practically optimal from the quantity of inserted information under the transparency and robustness constraints for the binary case. Should we be interested in increasing the data payload, multi-symbol insertion techniques should be considered, as follows.

Be there a binary message to be inserted; instead of directly inserting it, a message d encoded into an *m*-ary alphabet $D = \{-(m-1)/2, -(m-2)/2, \ldots 0, \ldots (m-2)/2, (m-1)/2\}$ can be considered. Such a d message can be further inserted by adapting the related blocks in the basic binary QIM method (Figure III-2) and should *a priori* lead to an increase of data payload by a factor of $\log_2 m$.

III.3.1 Insertion Step

The embedding process is structured into the same three main modules (see Figure III-10 *versus* Figure III-3): *perceptual shaping*, *mark generation* and *mark embedding & block reconstruction.*

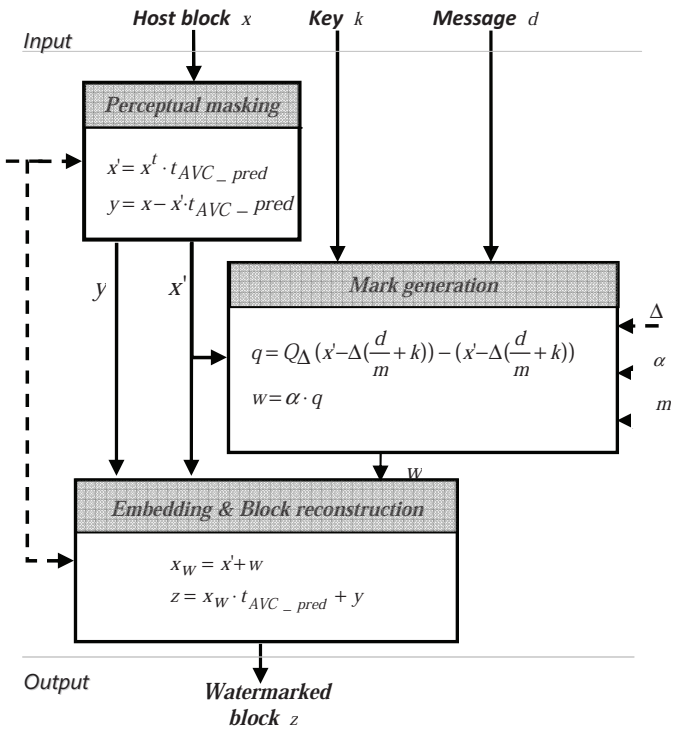


Figure III-10: The embedding synopsis: three inputs (the message d, the host x and the key k) and four parameters (the perceptual mask T_{AVC_pred}, the quantization step Δ, the scaling factor α, and the number of symbols in the alphabet m) are considered to compute the marked data z.

Perceptual masking

As previously explained, the watermark is not directly embedded into the original 4×4 block x but into the projection x' of x onto a perceptual mask T_{AVC_pred}.

Mark generation

The mark to be inserted into the host x vector depends of the m-ary message d and of the original vector x:

$$q = Q_\Delta(x^t T_{S_{AVC\_pred}} - \Delta(\frac{d}{m} + k)) - x^t T_{S_{AVC\_pred}} + \Delta(\frac{d}{m} + k)\,. \tag{III.22}$$

Embedding mark

This module generates the watermarked 4×4 MPEG-4 AVC block. It has as input the mark w, the original content and the perceptual mask T_{AVC_pred}. In the present chapter, the insertion follows a simple additive rule:

$$\begin{aligned} x_W &= x' + w \\ z &= x_W \cdot t_{AVC\_pred} + y \end{aligned}\,. \tag{III.23}$$

III.2.2 Detection process

For each supposed marked block, the detector starts by projecting the vector of 15 AC coefficients of the 4×4 sub-macroblocks in the *I* frames onto the perceptual mask T_{AVC_pred}.

Then, a decision rule optimising the probability error under the additive noise hypothesis is introduced.

Note that the decision is based on the value of $Y(d)$ which is a quantization error belonging to the $\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ interval; hence, specifying a decision rule means to divide the decision region $\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ into m nonoverlapping intervals.
$Y(d)$ can be written as follows:

$$\begin{aligned} q &= Q_\Delta(B(d)) - B(d) \\ B(d) &= (\alpha - 1)q + \Delta\frac{d}{m}, \end{aligned} \tag{III.24}$$

Given *q* a quantization error lying in the $\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ interval, then

$$\frac{\Delta((\alpha-1)m + 2d)}{2m} \le B(d) \le \frac{\Delta((1-\alpha)m + 2d)}{2m}\,. \tag{III.25}$$

Be there $I_{\sup(d)} = \frac{\Delta((1-\alpha)m + 2d)}{2m}$ and $I_{\inf(d)} = \frac{\Delta((\alpha-1)m + 2d)}{2m}$

To avoid overlapping, there will be one decision interval for each element of the alphabet; hence, $Q_\Delta(B(d)) = 0$ $Q_\Delta(B(d))$ should take eventually one value. Be there $Q_\Delta(B(d)) = 0$. Then, we have:

$$\begin{cases} I_{\sup(d)} < \dfrac{\Delta}{2} \\ \dfrac{\Delta}{2} < I_{\inf(d)} \end{cases}, \tag{III.26}$$

Hence:

$$\begin{aligned} & -I_{\sup(d)} \le B(d) \le I_{\inf(d)} \\ & \alpha > \tfrac{m-1}{m} \end{aligned}, \tag{III.27}$$

For a fixed parameter α, $I_{\sup(d)}$ and $I_{\inf(d)}$ are increasing functions of m. Hence, if each two successive symbols $(d; d+1)$ have nonoverlapping decision intervals, then we will have *m* nonoverlapping decision intervals. In this respect, $I(d)$ and $I(d+1)$ must verify the equation (III.28):

$$I_{\sup(d)} - I_{\inf(d+1)} < 0, \tag{III.28}$$

Equation (III.27) implies that: $\alpha > \dfrac{m-1}{m}$.

Therefore, the optimal value of α is $\alpha^* = \dfrac{m-1}{m}$.

We note that if $Y(m)$ is between two decision regions, we decide for the region which is closest to $Y(m)$.

Example:

For $m = 5$ we have, $\alpha^* = 0.8$. The decision regions are obtained as illustrated in Figure III-11.

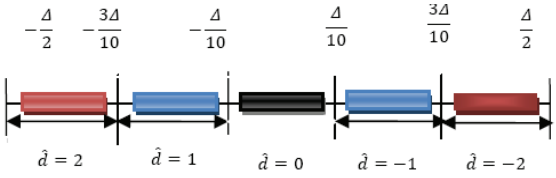


Figure III-11: Decision regions for m=5 and $\alpha \ge \alpha^*$.

III.2.3 EXPERIMENTAL RESULTS

Video corpus

The experiments were carried out on the corpus presented in Appendix B.

Data payload

As (III.21) is an equation depending on the particular video sequence to be watermarked, the data payload also depends on the particular video sequence and cannot be *a priori* predicted. The data

payload corresponding to the processed corpus averaged 136, 172, 300 or 648 bit per five minutes of video encoded at 64kbps, 256 kbps, 5Mbps or 10Mbps, respectively.

Robustness

First, the robustness against additive noise was verified by considering a bipolar white noise (-1/1), added in the 4×4 MPEG-4 AVC coefficient domain: each and every time, all the message symbols were correctly recovered.

Secondly, the robustness against transcoding is evaluated by severely compressing the watermarked files. It can be considered that *m*QIM features robustness against this attack: the BER values are presented in Figure III-12.

Finally, the robustness against geometric attacks was investigated. This time, the BER values are higher, *cf*. Figure III-12: this is not disturbing in practical applications where the inserted message is connected to some visual information, as illustrated in Figure III-13.

From the information theory point of view, this robustness is connected to channels with high mutual information, as illustrated in Figure III-12. The channel corresponding to the watermarking in this study has as input the $m=5$ symbols alphabet {-2, -1, 0, 1, 2} and the output is represented by the decision according to the above-presented rule. The noise matrices illustrated in Figure III-12 are estimated with a relative error of $\hat{\varepsilon}_I \cong 5\%$ and clearly point to robustness. When compared to the theoretical limits (Figure III-14), it can be noticed that in the case of strongly compressed content (64 kbps), the mutual information approaches its upper limit by relative errors of 0.4% and 1.7%, for the transcoding and geometric attacks, respectively. Such differences rise up to 8% and 8.7% (respectively) in the case of video sequences encoded at 10Mbps. Note that these numerical values are larger than the ones corresponding to the binary case.

SD

64 kbps

Transcoding

	-2	-1	0	1	2
-2	0.9	0.02	0.02	0.01	0.05
-1	0.06	0.87	0.05	0.01	0.01
0	0	0.06	0.89	0.04	0.01
1	0.01	0.02	0.04	0.90	0.03
2	0.02	0.01	0.11	0.04	0.82

$I(X,Y)=2.13$ $BER=4\%$

StirMark

	-2	-1	0	1	2
-2	0.8	0.03	0.05	0.03	0.09
-1	0.09	0.75	0.08	0.04	0.04
0	0.04	0.16	0.78	0.01	0.01
1	0.09	0.03	0.09	0.7	0.09
2	0.09	0.01	0.09	0.09	0.71

$I(X,Y)=1.91$ $BER=8.32\%$

256 kbps

	-2	-1	0	1	2
-2	0.92	0.01	0.03	0.02	0.02
-1	0.01	0.93	0.01	0.03	0.02
0	0.02	0.03	0.91	0.01	0.03
1	0.01	0.04	0.04	0.9	0.01
2	0.02	0.02	0.02	0.04	0.9

$I(X,Y)=2.29$ $BER=5.21\%$

	-2	-1	0	1	2
-2	0.77	0.11	0.02	0.01	0.09
-1	0.12	0.69	0.15	0.03	0.01
0	0.03	0.14	0.73	0.08	0.02
1	0.01	0.04	0.06	0.81	0.08
2	0.07	0.02	0.04	0.11	0.76

$I(X,Y)=2.014$ $BER=9.23\%$

HD

5 Mbps

	-2	-1	0	1	2
-2	0.92	0.05	0.01	0.01	0.01
-1	0.03	0.91	0.03	0.01	0.02
0	0.01	0.05	0.89	0.02	0.03
1	0.03	0.02	0.02	0.9	0.03
2	0.01	0.05	0	0.04	0.9

$I(X,Y)=2.3$ $BER=6.34\%$

	-2	-1	0	1	2
-2	0.8	0.06	0.04	0.05	0.05
-1	0.02	0.79	0.08	0.05	0.06
0	0.04	0.09	0.8	0.06	0.01
1	0.09	0.01	0.06	0.74	0.1
2	0.05	0.04	0.01	0.08	0.82

$I(X,Y)=2.09$ $BER=8.24\%$

10 Mbps

	-2	-1	0	1	2
-2	0.93	0.01	0.04	0.01	0.01
-1	0.02	0.93	0.03	0.01	0.01
0	0.01	0.01	0.95	0.02	0.01
1	0.01	0.01	0.01	0.94	0.01
2	0.01	0.01	0.01	0.01	0.96

$I(X,Y)=2.3$ $BER=5\%$

	-2	-1	0	1	2
-2	0.85	0.01	0.06	0.04	0.04
-1	0.03	0.86	0.06	0.01	0.04
0	0.04	0.05	0.84	0.04	0.03
1	0.03	0.03	0.03	0.87	0.04
2	0.01	0.05	0.05	0.05	0.84

$I(X,Y)=2.16$ $BER=9\%$

Figure III-12: Noise matrices and the corresponding mutual information values for transcoding (left) and geometric (StirMark) attacks (right), in the case the macroblocks to be watermarked are chosen according to an energy-selection criterion. These illustration correspond to the *m*QIM method.

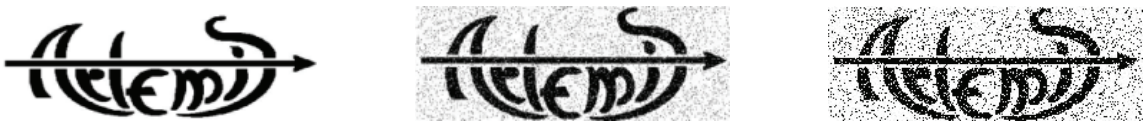


Figure III-13: The original ARTEMIS (left) binary logo recovered after transcoding (middle) and geometric attacks (right). The original video sequence was encoded at 5 Mbps.

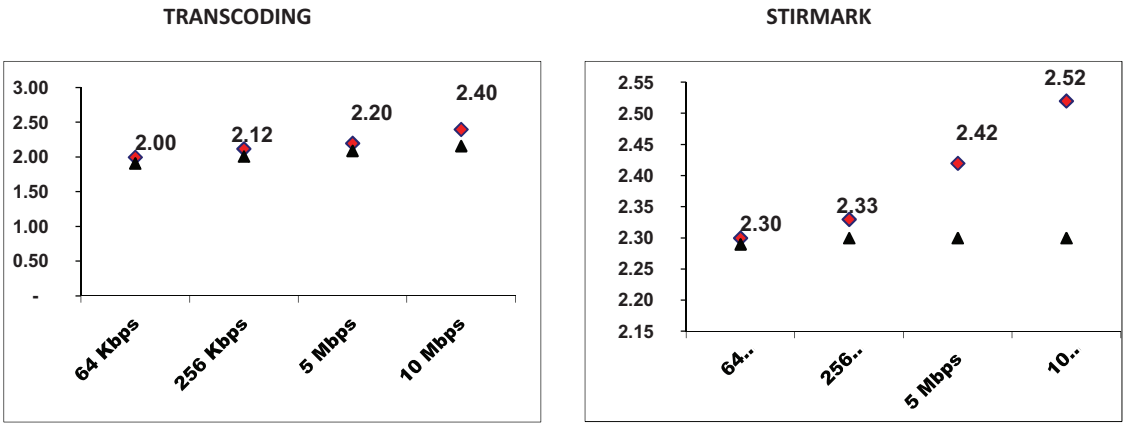


Figure III-14: Practical *versus* theoretical limits for data payload, in the cases of the transcoding and geometric (StirMark) attacks on the *m*QIM watermarking**.**

Transparency

In order to prove the transparency of the watermarked videos, the same metrics have been considered. The experimental results, filled-in in Table III-2, point to a good transparency. This conclusion was also strengthened by a panel of 10 human observers: while involved in a Two-alternative forced choice test, they agreed on the method transparency.

Table III-2: Objective evaluation of the transparency for *m*QIM watermarking.

		Pixels based				***Correlation based***			***Psychovisual***
		PSNR (dB)	**PMSE ($x10^{-6}$)**	**IF**	**AAD**	**NCC**	**CQ**	**SC**	**DVQ**
SD	**64 kbps**	39	1.05	0.9999	0.05	1.00	81	0.999	0.03
	256 kbps	38	1.65	0.9999	0.06	1.00	71	0.997	0.043
HD	**5 Mbps**	58	0.001	0.9999	0.0043	1.00	93	1.00	0.00032
	10 Mbps	62	0.0004	0.9999	0.0022	1.00	99	1.00	0.0002

III.4 MPEG-4 AVC driven counterattacks

It can also be noticed that every MPEG-4 AVC re-encoding acts as an attack: although the visual quality can be preserved, the corresponding stream is completely different from the syntax element point of view. Actually, by choosing two different sets of encoding parameters (slice structure, prediction modes, quantizing indexes, binary encoding procedure) two different streams, with practically the same visual quality and the same bit rate can be obtained. One possible counterattack procedure would be the re-encoding of the "attacked" stream by using the initial parameters. However, as the MPEG-4 AVC is a lossy compression scheme, even when keeping the same parameters, the iterative decoding/re-encoding operations may result in different syntax elements. The study reported in this section considers the JM12.2 MPEG-4 AVC reference software and objectively assesses the impact of successive encoding-decoding operations.

III.4.1 Transparency *vs*. encoding

In order to illustrate the re-encoding impact in the watermark detection, this section first considers the simplest case, in which an MPEG-4 AVC file is decoded and then re-encoded at the same rate with a subset of its initial parameters, Figure III-15. The corresponding results, obtained when decoding & re-encoding one video sequence (arbitrarily chosen from the corpus) are illustrated (for one *I* frame, arbitrarily chosen). These two images represent difference images (so, a black pixel means identity) corresponding to the Prediction values (*cf*. point ① in Figure III-15) and to the Quantized indexes (*cf*. point ② in Figure III-15), respectively. While the reference image was each time the same (the syntax elements obtained in the decoding process), three different re-encoding parameters have been considered. First, the re-encoding parameters were left to the automatic choice of the JM12.2 encoder. Secondly, the quantization parameters were the same as in the original stream and only the prediction mode was left to the encoder choice. Finally, both the quantization parameters and the prediction modes were copied from the original stream.

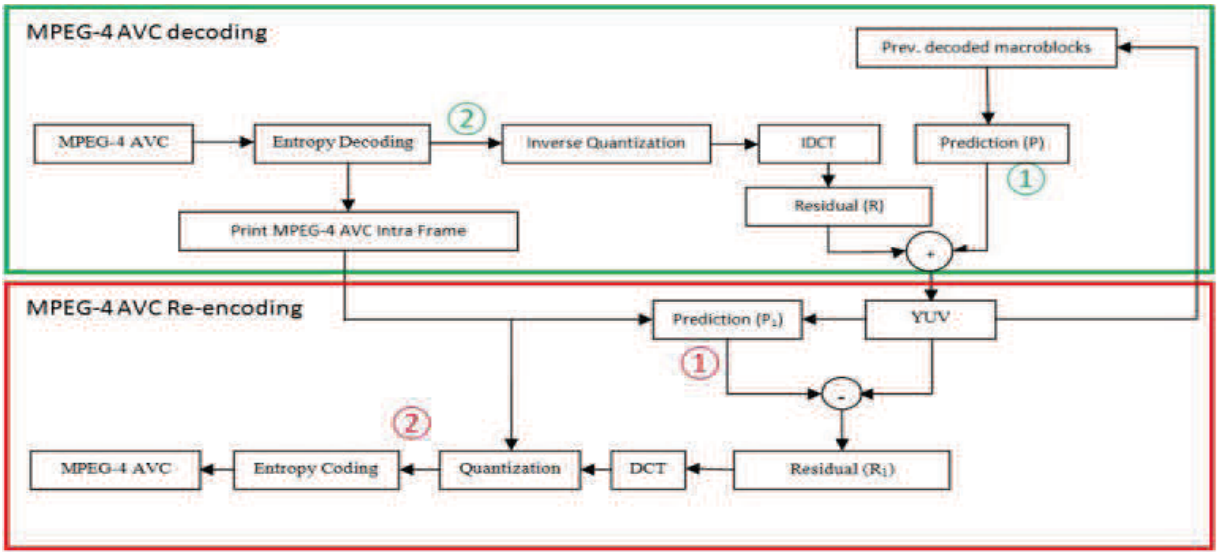
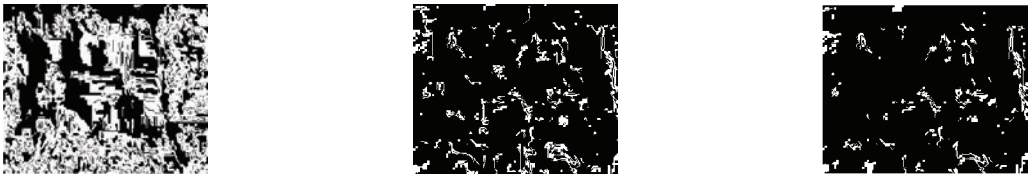


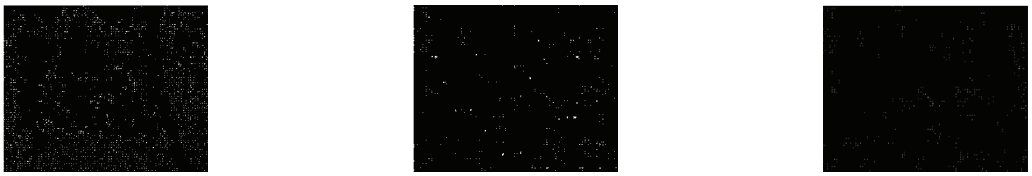
Figure III-15: MPEG-4 AVC decoding and re-encoding process.

Figures III-16 and III-17 brought to light that due to the inner MPEG-4 structure and to the way in which the related reference software is actually implemented, significant differences between the main syntax elements (like the prediction values or the quantized indexes) occurs when decoding and re-encoding the same video sequence, at the same rate (Figure III-16.a and III-17.a). However, by controlling the encoder parameters these differences can be significantly decreased (Figure III-16.c *vs*. Figure III-16.a and Figure III-17.c *vs*. Figure III-17.a).



(a) (b) (c)

Figure III-16: Difference images illustrating the MPEG-4 AVC decoding and re-encoding (to the same rate) effects in the prediction values (point ① in Figure III-15).Three different re-encoding modes have been considered: (a) automatic JM12.2, (b) automatic JM12.2 preserving the original quantization parameters and (c) JM12.2 preserving quantization parameters and prediction modes.



(a) (b) (c)

Figure III-17: Difference images illustrating the MPEG-4 AVC decoding and re-encoding (to the same rate) effects in the quantizing index values (point ② in Figure III-15).The same re-encoding modes as in Figure III-16 were considered.

To conclude with, it can be noticed that even for equivalent compression rate, the encoding parameter choice results in completely different stream syntax elements. From the watermarking point of view, this can be exploited at both transparency (Chapter III.4.1) and robustness (Chapter III.4.2) levels.

The experiments connected to the MPEG-4 AVC watermarking transparency brought to light that the main visual artefacts are induced by the drift effect. Actually, the mark insertion into a particular macroblock from one *I* frame leads to a variable and unpredictable number of modifications in other macroblocks from the same frame or from predicted frame.

One possible solution would be to restrict the drift by limiting the area in which the mark is actually inserted. To illustrate this mechanism, we considered the limit case in which each macroblock row is independently encoded and furthermore, inside each macroblock row, each of its two left/right halves are individually encoded. Then, the mQIM watermarking method was applied to the video corpus presented in Appendix B.

The practical drift reduction, for a randomly chosen *I* frame is illustrated in Figure III-18, by considering difference images. When the JM12.2 encoder is left to automatically choose the encoding parameters, the drift is significant, see Figure III-18.a. However, when an independent row encoding is imposed, the drift can be restricted even to a single macroblock, see Figure III-18.b.

The corresponding transparency was evaluated by 8 metrics (Table III-3). Table III-3 points to significant improvements in the transparency (*e.g.* a gain of 1.3dB in the PSNR) with respect to the *m*QIM (see Table III-2). However, note that this increased transparency was obtained at the expense of the compression rate: when imposing in JM12.2 independent macroblock row compression, the resulting stream is about 1% higher than the stream obtained with automatic JM12.2 compression.

(a) (b)

Figure III-18: Difference images illustrating the artifact drifts. Two MPEG-4 compression modes were considered: (a) automatic JM12.2 and (b) independent macroblock row encoding.

Table III-3: Objective evaluation of the transparency for *m*QIM watermarking *versus* controlled MPEG-4 AVC encoding.

			Pixels based				***Correlation based***			***Psychovisual***
			PSNR (dB)	**PMSE ($x10^{-6}$)**	**IF**	**AAD**	**NCC**	**CQ**	**SC**	**DVQ**
SD	**64 kbps**	basic *m*QIM	39	1.05	0.9999	0.05	1.00	81	0.999	0.03
		controlled MPEG-4	41.39	0.55	0.9999	0.04	1.00	82	0.999	0.01
	256 kbps	basic *m*QIM	38	1.65	0.9999	0.06	1.00	71	0.997	0.043
		controlled MPEG-4	40.33	1.49	0.9999	0.03	1.00	72	0.999	0.021
HD	**5 Mbps**	basic *m*QIM	58	0.001	0.9999	0.0043	1.00	93	1.00	0.00032
		controlled MPEG-4	59.78	0.0006	0.9999	0.00064	1.00	87	1.00	0.00003
	10 Mbps	basic *m*QIM	62	0.0004	0.9999	0.0022	1.00	99	1.00	0.0002
		controlled MPEG-4	63.68	0.0003	0.9999	0.0004	1.00	92	1.00	0.00001

III.4.2 Re-encoding counter-attack

The present study brought us to the specification of a basic MPEG-4 AVC re-encoding counterattack.

Assume a pirate will transcode the watermarked stream, in a different format (*e.g.* avi), or at a different bitrate (*e.g.* from 10Mbps to 64kbps).

In order to decrease the watermark detection BER, a re-encoding with the initial parameters can be performed, Figure III-19.

The corresponding experimental results are presented in Table III-4. They correspond to video corpus presented in Appendix B.

The watermarked video sequence was initially encoded at 10 Mbps. The pirate tried to transcode it at different compression rates: 64kbps, 256 kbps, 5 Mbps and 10Mbps. The three columns in Table III-4 correspond to the detection according to the *m*QIM watermarking rule, applied into three cases:

- directly on the attacked sequence;
- after a re-encoding (correction) of the attacked sequence by imposing the same quantization parameters as in the watermarked sequence,
- after a re-encoding of the attacked video sequence by imposing the same quantization parameters and predictions as in the watermarked sequence.

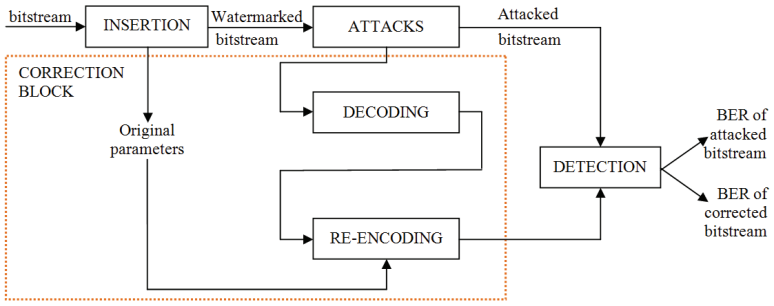


Figure III-19: Block structure of watermark insertion, attacking, correction and detection in the MPEG-4 AVC bitstream.

Table III-4 brings to light significant BER reduction, mainly in the most disturbing cases (strong attacks, like the compression from 10 Mbps to 64 kbps).

Table III-4: Bit Error Rate (BER) dependency on the re-encoding.

	***m*QIM**	***m*QIM & QP correction**	***m*QIM & QP & Prediction Correction**
64 kbps	17.6	2	1.4
256 kbps	10.1	1.6	1.6
5 Mbps	4.6	0.9	0.9
10 Mbps	0	0	0

Note that this gain in BER does not induce any collateral weakness in the watermarking scheme: the encoding parameters of the watermarked sequence are publically available (so, no pitfall in the security) and the transparency is not damaged (actually, this operation is applied only prior to detection and no human observer is supposed to watch the resulted sequence).

III.5 Conclusion

This third chapter of the thesis is devoted to the notion of robustness and presents a three-stage watermarking method referred to as COMWat. To our best knowledge, COMWat is the first watermarking method devoted to the MPEG-4 AVC compressed stream which features at the same time transparency and robustness against transcoding and geometric attacks.

At the first COMWat level, the insertion is achieved by combining basic binary QIM principles to perceptual masking (*cf*. Chapter II) and an energy-driven selection criterion. By using information theory concept and tools, it is proved that this insertion method practically optimizes the data payload under transparency and robustness constraints: for high compressed video sequences (*e.g.* 64kbps), the difference between the data payload and its theoretical optimal limit is lower than 1%.

At the second COMWat level, the data payload is to be increased by mathematically deriving the multi symbol QIM insertion/detection rules. Although relevant from the practical point of view (the data payload is increased by factors of 1.25 to 1.8 with respect to the binary QIM case), it can be noticed that in this case there is still some room for further improvements: theoretically speaking, the *m*QIM is supposed to allow an increase of data payload by factors of $\log_2 m$ (*i.e.* in the illustrations of this chapter, $\log_2 5 \cong 2.32$).

Finally, the impact of the MPEG-4 AVC (iterative) decoding/re-encoding in the watermarking applications is objectively assed. In this respect it is first brought to light that the current reference software JM12.2lead to different stream syntax elements even when the re-encoding is performed with the same parameter as the original. Then, in order to increase transparency by about 1.3 dB per frame at the expense of compression efficiency of 1% on an independent macroblock row compression is considered. Finally the re-encoding of an attacked video sequence is considered as an efficient counter attack tool: BER reduction of more than 10% is achieved without losing either in security, in transparency or in data payload.

REFERENCES

[ABR05] A. Abrardo, and M. Barni, "Informed Watermarking by Means of Orthogonal and Quasi-Orthogonal Dirty Paper Coding,"IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 53, NO. 2, FEBRUARY 2005.

[ALA03] A. Alattar, E Lin, and M Celik, “Digital watermarking of low bit rate advanced simple profile MPEG-4 compressed video,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, pp. 787–800, August 2003.

[ARN03] M. Arnold, M. Schmucker and Wolthusen., Techniques and Applications of Digital Watermarking and Content Protection, Artech House (2003).

[BEL08] M. Belhaj, M.Mitrea and F. Preteux ,“ Tatouage transparent et robuste du flux MPEG 4 AVC”, Rapport de Projet mastère d’études Mastère en Télécommunications, Option :Service des communications des réseaux multimédia, Sup’com Tunis Novembre 2010.

[BEL10] M. Belhaj, M.Mitrea, S. Duta, and F. Preteux ,“MPEG-4 AVC robust video watermarking Based based on QIM and perceptual masking”, International Conference on comminucation, Bucharest, June 2010.

[BEN95] W Bender, D Gruhl, and N Morimoto, “Techniques for data hiding,” in Proceedings of the SPIE - The International Society for Optical Engineering, vol. 2420, (San Jose, CA, USA), pp. 164–173, February 1995.

[BIR95] K Birney and T Fischer “On the modeling of DCT and subband image data for compression,” IEEE Transactions on Image Processing, vol. 4, pp. 186–193, February 1995.

[BOL95] F Boland, J O’Ruanaidh, and C Dautzenberg, “Watermarking digital images for copyright protection,” in Proceedings of International Conference on Image Processing and its Applications, vol. 410, (Edinburgh, UK), pp. 326–330, July 1995.

[BOR96] A Bors and I Pitas “Embedding parametric digital signatures in images,” in Proceedings EUSIPCO-96, VII European Signal Processing Conference, vol. 3, (Trieste, Italy), pp. 1701–1704, September 1996.

[BUR98] S Burgett, E Koch and J Zhao, “Copyright labeling of digitized image data,” IEEE Communications Magazine, vol. 36, pp. 94–100, March 1998.

[CAR95] G Caronni, “Assuring ownership rights for digital images,” in Proceedings of Reliable IT Systems (VIS), (Germany), 1995.

[CHE98] B. Chen, and G.W. Wornell, “Digital watermarking and information embedding using dither modulation,” in Proc. IEEE Workshop Multimedia Signal Process, Redondo Beach, CA, pp. 273–278, Dec. 1998.

[CHE01] B. Chen. G. Wornell, Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. IEEE Trans. on Information Theory, Vol. 47, No. 4, May 2001, pp. 1423-1443,.

[COX97] I Cox, J Kilian, F Leighton and T Shamoon “Secure spread spectrum watermarking for multimedia,” IEEE Transactions on Image Processing, vol. 6, pp. 1673–1687, December 1997. 103

[C0X02] I.J Cox, M.L Miller and Bloom, J., “Digital Watermarking”, Morgan Kaufmann Publishers (2002).

[DEL97] J Delaigle, C Vleeschouwer, F Goffin, B Macq, and J Quisquater "Low cost watermarking based on a human visual model," in Multimedia Applications, Services and Techniques - ECMAST'97. Second European Conference Proceedings, (Milan, Italy), pp. 153–167, May 1997.
[DCI08] Digital Cinema Initiatives, DCI System Requirements and Specifications for Digital Cinema, Version 1.2, March 2008, http://www.dcimovies.com
[DOE03] G Doerr and J Dugelay "A guide tour of video watermarking," Signal Processing: Image Communication, vol. 18, pp. 263–282, April 2003.
[DUT07] S Duta, M Mitrea and F Prêteux. "Compressed versus uncompressed domain video watermarking", Proc. SPIE 6700, 67000A (2007).
[DUT08] S. Duta, M Mitrea, F Preteux, L A Riffaud, "The Watermarking Attacks in the MPEG-4 AVC Domain", Proc. SPIE 6812, 68120P (2008).
[DUT08] S. Duta, M. Mitrea, F. Preteux, M. Belhaj, The MPEG-4 AVC domain watermarking transparency, Proc. SPIE Vol. 6982, April 2008, pp. 69820F
[DUT08] S. Duta, M. Mitrea, M. Belhaj, F. Preteux, A comparative study on insertion strategies in MPEG 4 AVC watermarking, Proc. SPIE Vol. 7075, August 2008, pp. 707509.

[EGG03] J.J.Eggers R. Bäuml R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding" IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 51, NO. 4, APRIL 2003.

[GOL07] A. Golikeri. P. Nasiopoulos and Z. J. Wang, "Robust digital video watermarking scheme for H.264 advanced video coding standard", Journal of Electronic Imaging 16(4), 043008 (Oct–Dec 2007).

[HAR98] F Hartung and B Girod "Watermarking of uncompressed and compressed video," Signal Processing, vol. 66, pp. 283–301, May 1998.
[HAR99] F Hartung and M Kutter "Multimedia watermarking techniques," Proceedings of the IEEE, vol. 87, pp. 1079–1107, July 1999.
[HER00] J Hernandez, M Amado and F Perez-Gonzalez "DCT-domain watermarking techniques for still images: detector performance analysis and a new structure," IEEE Transactions on Image Processing, vol. 9, pp. 55–68, January 2000.
[HUA05] X Huang and B Zhang "Robust detection of transform domain additive watermarks," in Proceedings of Digital Watermarking 4th International Workshop (IWDW), vol. 3710, (Siena, Italy), pp. 124–138, September 2005.

[INO98] H Inoue, A Miyazaki, A Yamamoto and T Katsura "A digital watermark based on the wavelet transform and its robustness on image compression," in Proceedings of International Conference on Image Processing (ICIP), vol. 2, (Chicago, IL, USA), pp. 391–395, October 1998.

[KOC94] E Koch, J Rindfrey, and J Zhao"Copyright protection for multimedia data," in Proceedings of the International Conference on Digital Media and Electronic Publishing, (Leeds, UK), May 1994.
[KOC95] E Koch, J Rindfrey and J Zhao "Towards robust and hidden image copyright labeling," in Proceedings of IEEE Workshop on Nonlinear Signal and Image Processing, pp. 452–455, 1995.
[KUN97] D Kundur and D Hatzinakos "A robust digital image watermarking method using wavelet-based fusion," in Proceedings of International Conference on Image Processing (ICIP), vol. 1, (Santa Barbara, CA, USA), pp. 547–554, October 1997.

[KUT98] M Kutter “Watermarking resisting to translation, rotation, and scaling,”in Proceedings of the SPIE - The International Society for Optical Engineering, vol. 3528, pp. 423–431, 1998.
[KUT97] M Kutter, F Jordan and F Bossen “Digital signature of color images using amplitude modulation,” in Proceedings of the SPIE - The International Society for Optical Engineering, vol. 3022, (San Jose, CA, USA), pp. 518–526, February 1997.

[LAN98] G Langelaar, R Lagendijk and J Biemond “Real-time labeling of MPEG2 compressed video,” Journal of Visual Communication and Image Representation, vol. 9, pp. 256–270, December 1998.
[LAN96] G Langelaar, J van der Lubbe and J Biemond “Copy protection for multimedia data based on labeling techniques,” in Proceedings of 17th Symposium on Information Theory in the Benelux, (Enschede, The Netherlands), May 1996.
[LAN97] G Langelaar, J ven der Lubbe and R Lagendijk “Robust labeling methods for copy protection of images,” in Proceedings of the SPIE - The International Society for Optical Engineering, vol. 3022, (San Jose, CA, USA), pp. 298–309, February 1997.
[LIA04] Y Liang, I Ahmad and V Swaminathan “Fast priority search algorithm for block motion estimation,” in IEEE International Conference on Multimedia and Expo (ICME), vol. 1, (Taipei, Taiwan), pp. 543–546, June 2004.

[MAE98] M Maes and C Overveld “Digital watermarking by geometric wrapping,” in Proceedings of IEEE International Conference on Image Processing (ICIP), vol. 2, (Chicago, IL, USA), pp. 424–426, October 1998.
[MAL03] H Malvar, A Hallapuro, M Karczewicz, and L Kerofsky “Low-complexity transform and quantization in H.264/AVC,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, pp. 598–603, July 2003.
[MAT94] K Matsui and K Tanaka “Video-steganography,” Journal of the Interactive Multimedia Association Intellectual Property Project, vol. 1, no. 1, pp. 187–205, 1994.
[MIL99] M.L. Miller, I.J. Cox, J-P.M.G. Linnartz, T. Kalker, A Review of Watermarking Principles and Practices, Digital Signal Processing for Multimedia Systems, K. K. Parhi, T. Nishitani (eds.), Marcell Dekker, Inc. NY, pp. 461-485, 1999.
[MIT07] M. Mitrea, F. Preteux, Tatouage robuste du contenu multimedia, Chapter 5 in H. Chaouchi, M. Laurent-Maknavicius (Ed.), La sécurité dans les réseaux sans fil et mobiles 1: Concepts fondamentaux, Hermès-Lavoisier, 2007, pp. 169-224.

[NIK03] A Nikolaidis and I Pitas “Asymptotically optimal detection for additive watermarking in the DCT and DWT domains,” IEEE Transactions on Image Processing, vol. 12, pp. 563–571, May 2003.
[NIK96] A Nikolaidis and I Pitas “Copyright protection of images using robust digital signatures,” in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing Conference (ICASSP), no. 4, (Atlanta, GA, USA), pp. 2168–2171, May 1996.
[NOO05] M Noorkami and R Mersereau “Compressed-domain video watermarking for H.264,” in Proceedings of IEEE International Conference on Image Processing (ICIP), vol. 2, (Genoa, Italy), pp. 890–893, September 2005.
[NOO06] M Noorkami and R Mersereau “Towards robust compressed-domain video watermarking for H.264,” in Proceedings of SPIE - The International Society for Optical Engineering, vol. 6072, (San Jose, CA, USA), January 2006.

[NOO07a] M Noorkami and R Mersereau “Digital video watermarking in Pframes,”in Proceedings of SPIE - The International Society for Optical Engineering, vol. 6505, (San Jose, CA, USA), January 2007.
[NOO07b] M Noorkami and R Mersereau “A framework for robust watermarking of H.264-encoded video with controllable detection performance,” IEEE Transactions on Information Forensics and Security, vol. 2, pp. 14–23, March 2007.
[NOO07c]M Noorkami and R Mersereau “Video watermark detection with and without knowledge of watermark location,” in 8th ACM Multimedia and Security Workshop, (Dallas, TX, USA), September 2007.

[PIT96] I Pitas “A method for signature casting on digital images,” in Proceedings of IEEE International Conference on Image Processing (ICIP), vol. 3, (Lausanne, Switzerland), pp. 215–218, September 1996.
[PIT98] I Pitas “A method for watermark casting on digital images,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, pp. 775–780, October 1998.
[PIV97] A Piva, M Barni, F Bartolini and V Cappellini “DCT-based watermark recovering without resorting to the uncorrupted original image,” in Proceedings of International Conference on Image Processing (ICIP), vol. 1, (Santa Barbara, CA, USA), pp. 520–523, October 1997.
[POD97a] C Podilchuk and W Zeng “Perceptual watermarking of still images,”in Proceedings of Signal Processing Society Workshop on Multimedia Signal Processing, pp. 363–368, 1997.
[POD97b] C Podilchuk and W Zeng “Digital image watermarking using visual models,” in Proceedings of the SPIE - The International Society for Optical Engineering, vol. 3016, (San Jose, CA, USA), pp. 100–111, February 1997.

[QUI04] G Qiu, P Marziliano, A Ho, Q Sun “A hybrid watermarking scheme for H.264/AVC video,” in Proceedings of the 17th International Conference on Pattern Recognition, vol. 4, (Cambridge, UK), pp. 865–868, August 2004.

[RUA96] J O’Ruanaidh, W Dowling and F Boland “Phase watermarking of digital images,” in Proceedings of IEEE International Conference on Image Processing (ICIP), vol. 3, (Lausanne, Switzerland), pp. 239–242, September 1996.
[RUA97] J O’Ruanaidh and T Pun “Rotation, scale and translation invariant digital image watermarking,” in Proceedings of International Conference on Image Processing (ICIP), vol. 1, (Santa Barbara, CA, USA), pp. 536–539, October 1997.

[SAW96a] M Swanson, B Zhu and A Tewfik “Robust data hiding for images,” in Proceedings of IEEE Digital Signal Processing Workshop, (Loen, Norway), pp. 37–40, September 1996.
[SAW96b] M Swanson, B Zhu and A Tewfik “Transparent robust image watermarking,” in Proceedings of IEEE International Conference on Image Processing (ICIP), vol. 3, (Lausanne, Switzerland), pp. 211–214, September 1996.
[SAW98c] M Swanson, B Zhu and A Tewfik “Multiresolution scene-based video watermarking using perceptual models,” IEEE Journal on Selected Areas in Communications, vol. 16, pp. 540–550, May 1998.
[SIM02] D Simitopoulos, S Tsaftaris, N Boulgouris and M Strintzis “Compressed-domain video watermarking of MPEG streams,” in Proceedings of IEEE International Conference on Multimedia and Expo (ICME), vol. 1, (Lausanne, Switzerland), pp. 569–572, August 2002.

[SU05] K Su, D Kundur and D Hatzinakos “Spatially localized image dependent watermarking for statistical invisibility and collusion resistance,” IEEE Transactions on Multimedia, vol. 7, pp. 52–66, February 2005.

[TAN90] K Tanaka, Y Nakamura and K Matsui “Embedding secret information into a dithered multi-level image,” in Proceedings of IEEE Military Communications Conference, vol. 1, (Monterey, CA, USA), pp. 216–220, 1990.

[TAO97] B Tao and B Dickinson “Adaptive watermarking in the DCT domain,” in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 4, (Munich, Germany), pp. 2985–2988, April 1997.

[TRA02] W Trappe, M Wu and K Liu “Collusion-resistant fingerprinting for multimedia,” in Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 4, (Orlando, FL, USA), pp. IV3309–3312, May 2002.

[WIE05] T Wiegand, G.J Sullivan, G Bjontegaard, A Luthra “Overview of the H.264/AVC Video Coding Standard”, IEEE Trans. on Circuits and Systems for Video Technology, 13(7), 560-576(2003).

[WAN03] R.A Wannamaker, The theory of dithered quantization, thesis defended for the PhD degree in Applied Mathematics, University of Waterloo, Ontario. Canada. 2003.

[ZOU08] D. Zou, and J. A. Bloom, “ H.264/AVC Stream Replacement Technique for video watermarking”, IEEE International conference on Acoustics, Speech, and signal processing, ICASSP 2008.

Chapter IV

Demo

It is through experience that science and art both make the mankind progress.

Aristote

Abstract

This chapter presents two real-life applications of the in-band enrichment, connected to subtitling and copyright protection.

Contents

IV.1 Introduction

The previous chapters of this thesis resulted in defining COMWat, the first reliable method for the MPEG-4 AVC stream watermarking.

Within this Chapter IV, we shall demonstrate the possibility of integrating this method in the *in-band enrichment framework*. In this respect, two extreme scenarios are considered: subtitling, were the purpose is to insert a quite large data payload but were the attacks are not likely to occur and copyright protection, where although the data payload can be quite small, strong robustness constraints are imposed.

IV.2 Subtitle watermarking

IV.2.1. Experimental protocol

The aim of this section is to demonstrate that the subtitles for a video (*e.g.* Krazy Kat, 1916), available in the MPEG-4 timed-text format [GAP10], can be imperceptibly inserted into a video.

Table IV-1: Timed-text data for the Krazy Kat video.

Data in MPEG-4 Streaming text format	Inserted Message
1 00:00:22,000 --> 00:00:27,000	Data payload= 272 bit GOP = 2
I ll teach thee Bugology Ignatzes	
2 00:00:40,000 --> 00:00:43,000	Data payload= 144 bit GOP = 2
something tells me	
3 00:00:58,000 --> 00:01:04,000	Data payload= 208 bit GOP = 2
Look Ignatz a sleeping bee	
4 00:01:04,000 --> 00:01:08,000	Data payload= 144 bit GOP = 2
A dead bee y mean	
5 00:01:10,000 --> 00:01:14,000	Data payload= 120 bit GOP = 2
He must be cold	
6 00:01:22,000 --> 00:01:25,000	Data payload= 104 bit GOP = 2
Poor l il bee	
7 00:01:25,000 --> 00:01:31,000	Data payload= 280 bit GOP = 2
Ah he s happy he s in bee heaven	
8 00:01:51,000 --> 00:01:56,000	Data payload= 200 bit GOP = 2
Y See Ignatz y was wrong	
9 00:01:58,000 --> 00:02:01,000	Data payload= 40 bit GOP = 2
I WAS.	
10 00:02:45,000 --> 00:02:49,000	Data payload= 112 bit GOP = 2
Leave it to me	

The corresponding watermarking setup is synoptically represented in Figure IV-1. The Subtitle information existing in the timed-text represents the information to be inserted according to the COMWat method, using the key given by the Position information in the timed-text.

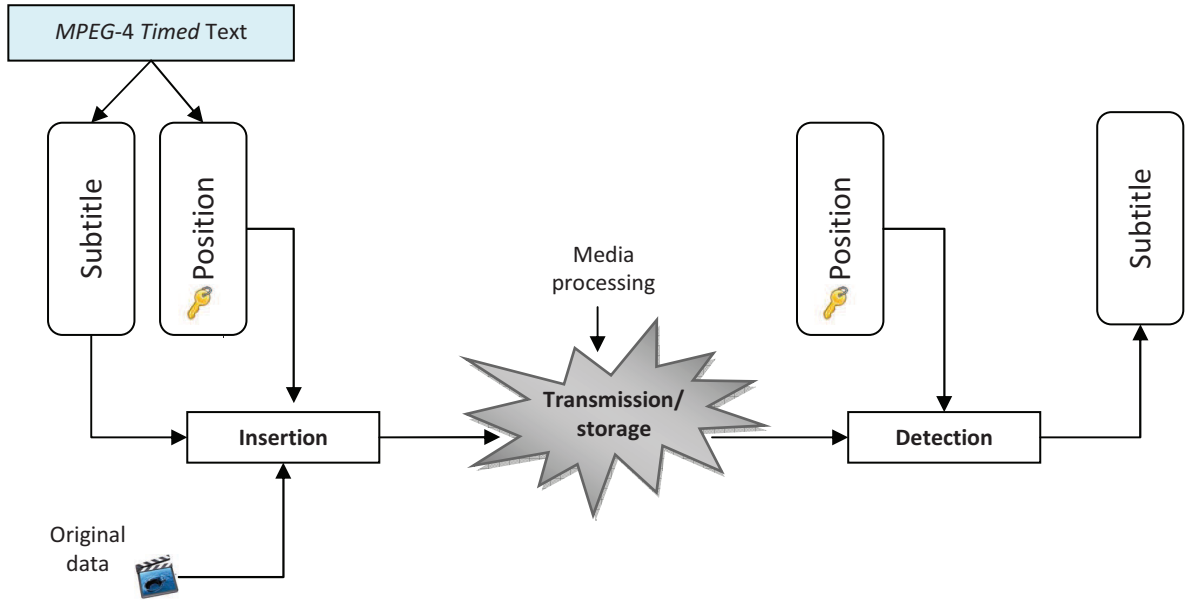


Figure IV-1: Synoptic watermarking scheme for subtitle insertion.

IV.2.2. Experimental result

Data payload

The required Subtitle information has been successfully inserted into the Krazy Kat video, according to two scenarios:

- the Subtitle was binary encoded by a simple ASCII conversion;
- the Subtitle was binary encoded by a simple ASCII conversion and each bit was repeated three times.

Robustness

The video, originally encoded at 64kbps, was subject to transcoding down to 32 kbps. Figure IV-2 presents the noise matrix and the related mutual information. Table IV-2 illustrates the effects of such errors on subtitling.

SD

64 kbps	0	1
0	0.92	0.08
1	0.06	0.94

$I(X,Y)=0.66$ $BER=7.98\%$

	0	1
0	0.98	0.02
1	0.03	0.97

$I(X,Y)=0.83$ $BER=0.42\%$

Figure IV-2: Noise matrices corresponding to the transcoding attack in the two scenarios: simple ASCII encodeing (left) and ASCII encoding and a repetition code.

Table IV-2: Subtitles recovered after the compression to 32kbps.

Inserted Data in MPEG-4 Streaming text format	Detected Data in MPEG-4 Streaming text format, no repetition	Detected Data in MPEG-4 Streaming text format, with repetition
I ll teach thee Bugology Ignatzes	Ihll tyzxh ghee Bollogy Igmatzes	I ll teach thee Bugology Ignatzes
something tells me	sobething tgllsoke	something tells me
Look Ignatz a sleeping bee	Llok Igmatz a sleeping bee	Look Ifnatz a sleeping bee
A dead bee y mean	A dead kke y meln	A dead bee y mean
He must be cold	be zust ke cold	He must be cold
Poor l il bee	Poor lgil zxe	Poor l il bee
Ah he s happy he s in bee heaven	mh he s happy hegs in bee heaven	Ah he s happy he s in bee heaven
Y See Ignatz y was wrong	P See Ignatzty was wrpng	Y See Ignatz y was wrongd
I WAS.	I GpS.	I WAS.
Leave it to me	Lerbe it to le	Leave it to me

Transparency

The transparency objective evaluations are presented in Table IV-3.

Table IV-3: Objective evaluation of the transparency for binary QIM watermarking.

		Pixels based				*Correlation based*			*Psychovisual*
		PSNR (dB)	**PMSE ($x10^{-6}$)**	**IF**	**AAD**	**NCC**	**CQ**	**SC**	**DVQ**
64 kbps	**no repetition**	37.2	0.0888	0.999	0.0027	1.00	92	0.98	0.009
	with repetition	35.8	0.0948	0.999	0.0047	0.98	81	0.99	0.0117

IV.3 Copyright protection

IV.3.1. Experimental protocol

The aim of this section is to demonstrate that the COMWat watermarking technique can protect a video content in real life application against the cam capture.

The corresponding technique is represented in Figure IV-1. The watermark is recovred even after a display of the video and cam capture

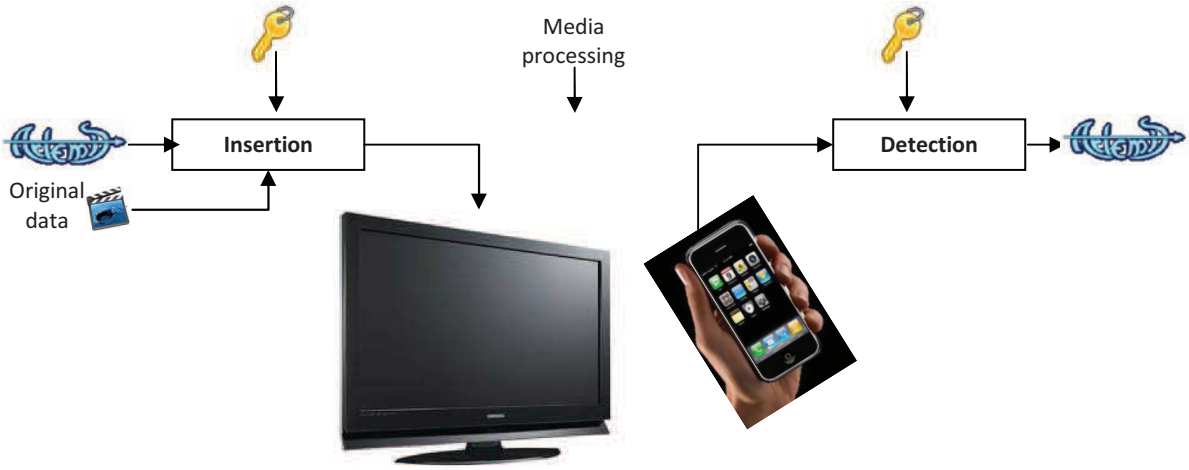


Figure IV-3: Synoptic watermarking scheme for copyright protection.

IV.3.2. Experimental result

Data payload

The required application inserts the copyright information as ARTEMIS logo.

Robustness

The video, originally encoded at 5 Mbps, was displayed on LCD Monitor, captured by an iphone's cam then encoded at 5 Mbps with the original encoder parameters. Figure IV-4 presents the noise matrix and the related mutual information. Figure IV-5 illustrates the effects of such errors on visual logo.

		StirMark	
		0	1
HD 5 Mbps	0	0.19	0.81
	1	0.83	0.17

$I(X,Y) = 0.29 \quad BER = 15.8\%$

Figure IV-4: Noise matrices and the corresponding mutual information values for capture attack, in the case the macroblocks to be watermarked are randomly chosen.

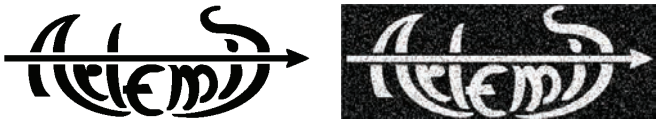


Figure IV-5: The ARTEMIS binary logo: original (left), recovered after geometric attacks (right). The original video sequence was encoded at 5 Mbps.

Transparency

The transparency objective evaluations are presented in Table IV-4.

Table IV-4: Objective evaluation of the transparency for binary QIM watermarking.

	Pixels based				*Correlation based*			*Psychovisual*
	PSNR (dB)	**PMSE ($x10^{-6}$)**	**IF**	**AAD**	**NCC**	**CQ**	**SC**	**DVQ**
5 Mbps	62	0.031	0.999	0.0006	1.00	91	1.00	0.00053

REFERENCES

[COX08] I. Cox, M.L. Miller, J. Bloom, J. Fridrich, T. Kalker. Digital Watermarking and Steganography. Second edition, Morgan Kaufmann, Burlington, MA, 2008.

[GAP10] http://gpac.wp.institut-telecom.fr/mp4box/ttxt-format-documentation/

[HUH11] K. Hühring, H.264/AVC Joint Model 8.6 (JM-8.6) Reference Software [Online]. Available: http://iphome.hhi.de/suehring/tml/

[ITU93] ITU-T Rec. H.261, "Video Codec for Audio-Visual Services at 64-1920 kbit/s," 1993.

[JVT03] JVT of ISO/IEC MPEG And ITU-T VCEG, "ITU-T Recommendation And Final Draft International Standard Of Joint Video Specification (ITU-T Rec. H.264|ISO/IEC 14496-10 AVC)," JVT-G050, 2003.

[RIC03] I.E.G. Richardson, H.264 And MPEG-4 Video Compression: Video Coding For Next-Generation Multimedia. Chichester: John Wiley & Sons, 2003.

[WEI09] L. Wei, O Issa, L Hong. F. Speranza, R. Renaud “Quality Assessment of Video Content for HD IPTV Applications”. ISM '09. 11th IEEE International Symposium on, pp, 517 – 522, San Diego, CA ,14-16 Dec 2009.

Conclusion

The important thing is not to stop questioning

A. Einstein

The present thesis follows the long way of the in-band enrichment from the very definitions of this concept, passing through its intimate relationship with watermarking, and reaching to the applications in the field of compressed video stream processing.

Despite preliminary studies already published in the literature (*cf.* Chapter I) imperceptibly inserting some extra information into a compressed stream remains a challenging research topic, mainly because of its underlying conceptual contradiction. On the one hand, compression's goal is to eliminate the redundancy from the visual content. On the other hand watermarking (and in-band enrichment) exploits this visual redundancy in order to imperceptibly hide the mark. In order to bridge the gap between these antagonistic constraints and to grant them unified theoretical and methodological basis, the present thesis provides the following main results (see Table Conclusion-1 *versus* Table Intro-1 and Fig. Intro-5):

- In absence of any objective study in the literature, the thesis presents two studies on the use of objective assessment of the human visual perception. First, by considering eight objective quality metrics, it is proved that the MPEG-4 AVC (a.k.a. H.264) compressed stream still allows the imperceptible modification of its syntax elements. Secondly, by mathematically extending the Peterson Watson IIP model, the first perceptual masking model matched to the MPEG 4 AVC peculiarities is computed and validated under watermarking constraints. This new masking model has been tested against state of the art masking models, under the watermarking framework. On the one hand, when imposing a prescribed robustness and a fixed data payload, significant gains in transparency are obtained: the PSNR is increased up to 3dB while the Watson's DVQ (Digital Video Quality) is decreased by a half. On the other hand, when imposing a prescribed robustness and transparency, gains in data payload up to 50% are reached.
- By presenting COMWat, this thesis advances the first MPEG-4 AVC watermarking method featuring robustness against noise addition, transcoding, and the StirMark (geometric) attacks. The mark is inserted into the MPEG-4 AVC quantization indexes selected according to an energy-based selection criterion validated by information theory basic concepts. The insertion procedure combines the QIM principles and the perceptual mask obtained in this thesis. Subsequently, the very QIM principles are generalized beyond the binary case, thus advancing the mQIM insertion/detection methods. Its main theoretical and practical advantage is the increase the data payload by a log(m) factor, while preserving the transparency and the robustness. Finally, the MPEG-4 AVC syntax is reconsidered as a starting point in designing a counter-attack procedure; in practice, this procedure results in BER reduction by 5% to 10%, according to the original video bit rate.
- At our best knowledge, no study on the theoretical limits of the MPEG-4 AVC watermarking was performed. Our study paves the way towards an objective assessment of the compressed watermarking: the noise channels corresponding to the transcoding and geometric attacks are estimated, the related mutual information and capacity are computed, compared among them and discussed.
- The overall results are integrated into two real-life demos, related to subtitles and camcorder recording.

Table Conclusion-1: Thesis main results and perspectives.

		Results	Perspectives
Viability	Theoretical research	• the first perceptual shaping model matched to the MPEG-4 AVC peculiarities (Chapter II.2); • the first mQIM insertion/detection method (Chapter III.3);	• exploiting this model for other fields image processing (compression, image enhancement, *etc.*); • deriving the mQIM optimal detection rule for prescribed attack probability density function;
	Applicative research	• visibility limits for imperceptibly modifying the MPEG-4 AVC quantizing indexes; • COMWat, the first watermarking method for the MPEG-4 AVC quantized indexes, reaching the trade-off among transparency (*e.g.* psnr > 42dB), robustness (against transcoding and geometric attacks) and data payload (5 times larger than the DCI requirements);	• resume this analysis for other types of insertion strategies and/or quality metrics; • exploiting the mQIM insertion/detection rules for other types of content (stereoscopic video, audio, *etc.*); • extending the watermarking to other MPEG-4 AVC syntax elements;
Feasibility	Applicative research	• prototypes demonstrating the in-band enrichment practical impact for subtitling and copyright protection.	• promoting this method within the international standardization bodies and amongst industrial partners;
Theoretical limits	Theoretical research	• noisy channel models (probability estimation with 95% confidence limits and relative errors lower than 5%) for compressed domain watermarking; • estimating the related information theory measures (mutual information and capacity).	• increase the complexity of the underlying theoretical model so as to take into account the side information model; • demonstrate and deploy the related optimal insertion rule; • take into consideration the side information constraints in the capacity evaluation;

Appendix A

MPEG-4 AVC overview

The Appendix presents some basic information about the MPEG-4 AVC (a.k.a. H.264 standard) and about the possibility of modifying some of its stream elements under syntax constraints.

Contents

A.1 MPEG-4 AVC syntax overview

Like all video codecs, the MPEG-4 AVC codec (coder/decoder), transforms the uncompressed data in a classic compression chain: prediction P, transformation T, quantization Q and arithmetic coding E [ISO10, WIE03]. A preparatory phase, preceding the video data compression chain consists in projecting the data in a *YCrCb* (respectively luma component, blue-difference and red-difference) color space that closely resembles the human perception of colors, Figure App.A-1. *YCrCb* signal is created from the source RGB (Red, Green, Blue). The values of R, G and B are added together according to their relative weight to get the signal *Y*. The latter represents the luminance of the source. The signal *Cb* is obtained by subtracting the *Y* signal original blue; similarly signal *Cr* is obtained by subtracting the signal Y:

$$\begin{aligned} Y &= 0.299 * R + 0.587 * G + 0.114 * B \\ Cr &= 0.492(B - Y) \\ Cb &= 0.877(R - Y) \end{aligned} \qquad \text{(App.1)}$$

,

Subsequently a sub-sampling is applied to the *Cb* and *Cr*, as shown in Figure App.A-1, for the *Y* component is more information than *Cb* and *Cr*.

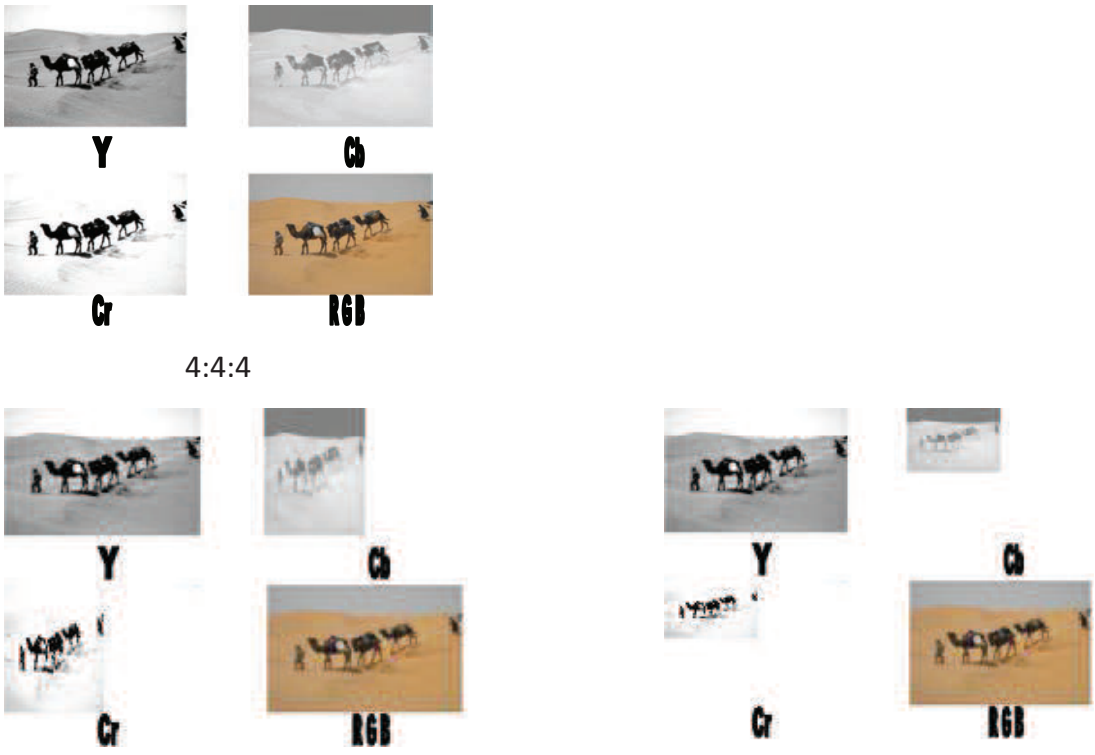


Figure App.A-1 Sub-sampling of color components.

A.1.1 Prediction

The prediction is meant to eliminate the spatial (intra prediction) and temporal (inter prediction) redundancy.

The image represented in each color component is divided into blocks of pixels 16×16 called macroblocks Figure App.A-2. For images where the number of columns/rows is not a multiple of 16, a padding (add columns/rows of pixels to obtain a multiple of 16). A particular procedure is applied to improve the resistance against errors and losses during the video reconstruction phase to gather macroblocs from the same or different images into slices. The intra prediction is made according to a prediction mode by copying the pixels in the row/column adjacent in a direction (vertical, horizontal, diagonal down/left, diagonal down/right). The prediction modes are illustrated in Figure App.A-3

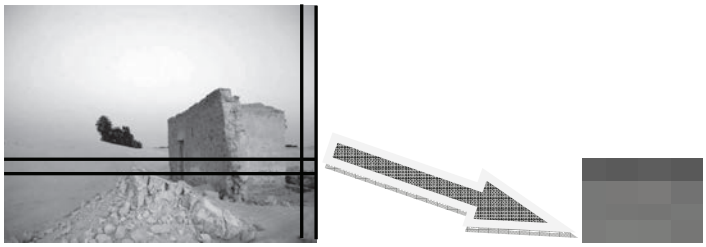


Figure App.A-2 Macroblock illustrations.

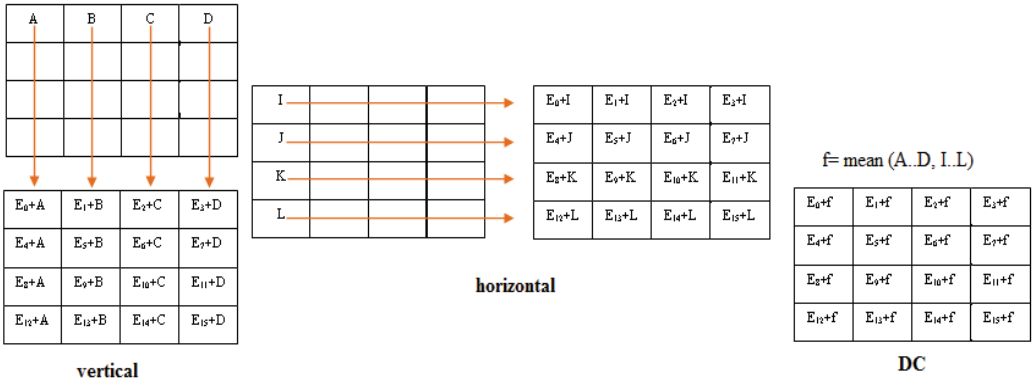


Figure App.A-3 The three most intensively considered prediction modes.

For the inter prediction, the blocks are predicted from previous or following frames, by using the spatial displacement of corresponding blocks of frames specified by a motion vector. This motion vector is estimated so that for the luminance component corresponds to an integer value. MPEG-4 AVC codec uses an estimate for the quarter pixel motion compensation, enabling very precise description of the displacement of the moving regions. Block sizes for the prediction can be 16 × 16, 16 × 8, 8 × 16, 8 × 8, 8 × 4, 4 × 8 or 4 × 4, enabling very precise segmentation of areas to predict. Note that the reference frame for inter prediction cannot be found outside of a prediction module called GOP (Group Of Picture) [WIE03].

A GOP consists of a number of images that can be 3 types, grouped according to a predetermined decoding order:

- **The I frames** correspond to a coded image independently; note that only one field I can be at the beginning of a GOP, as it serves as a starting point for coding images of two other types;

- **The P frames** are associated with an motion compensated image, predicted either from an I or from another P frame;

- **The B frames** refer to any image being double (forward and backward) motion compensated.

Regardless the type of prediction, the pixel values are subtracted from the corresponding predicted values; these differences are further transformed, quantified and binary encoded.

A.1.2 Transformation

Following the prediction, the transformation is applied with the view of representing the data as uncorrelated (separated into components with a minimum interdependence) and compacted (the energy is concentrated on a small number of values) information [HAL02]. The MPEG-4 AVC codec uses a modified version of the classical DCT (Discrete Cosine Transform) in order to work with integer coefficients and eliminate errors caused by the fact that in a conventional DCT coefficients are irrational. The transformation matrix used by MPEG-4 AVC is calculated from the matrix H by taking the rounded values of the coefficients amplified by a factor α (experimentally set to 2.5):

$$\begin{aligned} &X = H.x \\ &H = h(k,n) \\ &h(k,n) = c_k \sqrt{\frac{2}{N}} \cos\left[\left(n+\frac{1}{2}\right)\frac{k\pi}{N} \right] \end{aligned}, \qquad \text{(App.2)}$$

$$H' = round(\alpha \cdot H) \qquad \text{(App.3)}$$

$$H' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}, \qquad \text{(App.4)}$$

This transformation also has the advantage of simplicity of implementation, as shown in Figure App.A-4: its calculation is based on additions, subtractions and shifts (multiplications by 2).

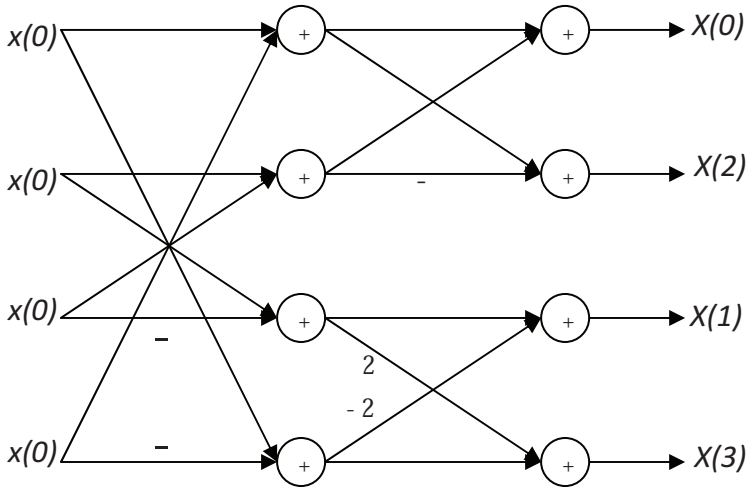


Figure App.A-4: Fast implementation of the integer DCT in MPEG-4 AVC.

A.1.3 Quantizing

Quantizing is the phase where information is lost in the compression chain [HAL02]. For a given quantization step (always an integer), the quantized value can be calculated as:

$$X_q(i,j) = sng\{X(i,j)\}\frac{X(i,j) + f(Q_s)}{Q_s}, \quad \text{(App.5)}$$

where *i* and *j* are the indices of rows and columns and $f(Q_s)$ controls the quantization values near the origin.

For de-quantization, the reconstructed information $X_r(i,j)$ is calculated as

$$X_r(i,j) = Q_s \cdot X_r(i,j), \quad \text{(App.6)}$$

To circumvent the disadvantages of an entire division, the MPEG-4 AVC offers another form of quantification, this time involving a right shift:

$$\begin{aligned} X_q(i,j) &= sng\{X(i,j)\}[|X(i,j)|A(Q) + f2^L >> L] \\ X_r(i,j) &= X_q(i,j)B(Q) \end{aligned}, \quad \text{(App.7)}$$

where f, $Q(A)$ and $Q(B)$ are association of the quantization parameter, L is the bit length parameter for the encoding process.

A.1.4 Entropy encoding

The final phase of MPEG-4 AVC is the entropy coding (lossless) which takes place in three stages. First, the quantized coefficients are scanned in a zigzag order, Figure App.A-5, in order to gather the maximum of trailing zero and trailing one coefficient. Secondly, each quantized coefficient is RL (Run-Length) encoded so as to increase the compression rate. The construction of the bitstream is achieved according to two advanced methods of entropy coding [GOL66, BJO02, MAR03]:

- CAVLC (Context-Adaptive Variable Length Coding) can be used for all encoding profiles.
- CABAC (Context-Adaptive Binary Arithmetic Coding) that can be used alternately with CAVLC only for main profile.

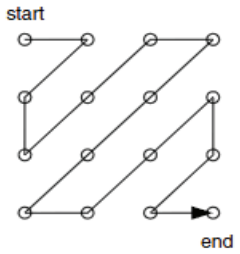


Figure App.A-5: Zigzag scan for 4x4 blocks.

A.2 MPEG-4 AVC parser

In order to allow the modification of individual elements in an MPEG-4 AVC compressed stream while observing to the related syntax constraints, the following parser has been implemented, Figure App.A-5.

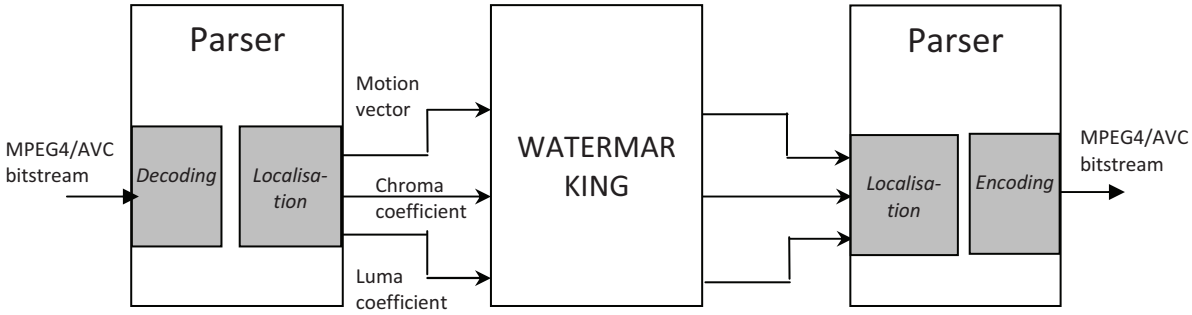


Figure App.A-5: Parser of MPEG-4/AVC bitstream.

Our parser will exploit the layer architecture of the Codec to separate the parsing and location process of MPEG-4 AVC syntax. As shown in Figure App.A-6, two main layers exist: the video coding layer describes the chain compression while the abstraction layer network controls the package of the bistream in NALU (Network Abstraction Layer Unit). After this second layer, the parser analyzes the stream and recovers the syntax elements according to the hierarchy shown in Figure App.A-6.

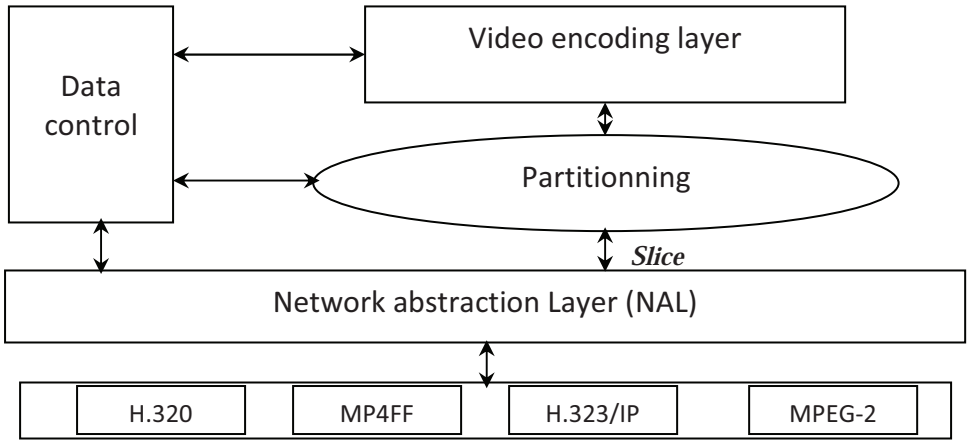


Figure App.A-6: CODEC MPEG-4 AVC structure.

Our parser, Figure App.A-7, performs a partial decoding of the AVC streams read from a file, changes its syntax elements according to the watermarking procedure, recodes it into a new stream and then writes into another file. These operations are also made into a two levels architecture.

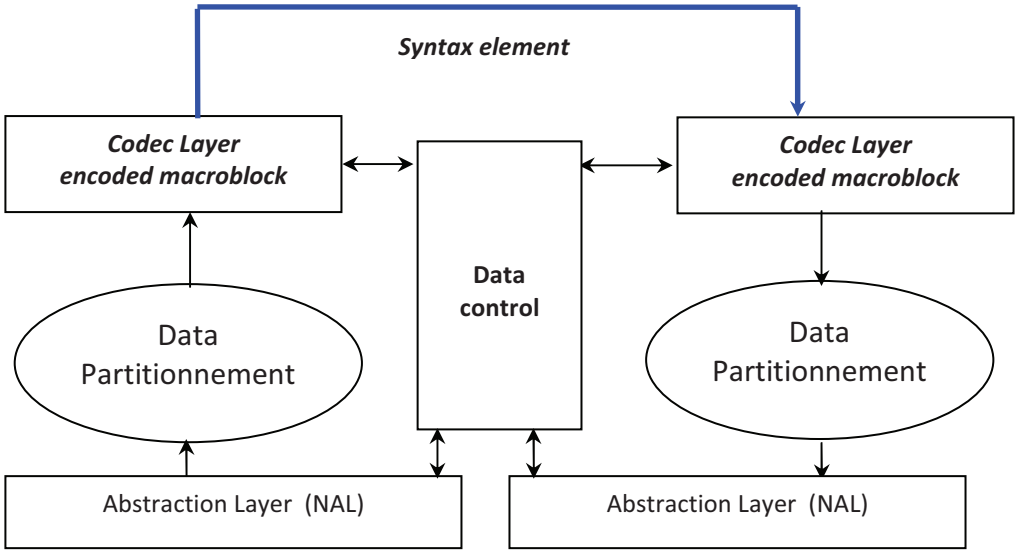


Figure App.A-7: Layer structure MPEG-4 AVC Decoder/Parser/Encoder.

Reference

[ISO10] ISO/IEC 14496-10 and ITU-T Rec. H.264, Advanced Video Coding.

[WIE03] T. Wiegand, G. Sullivan, G. Bjontegaard and A. Luthra, "Overviewof the H.264 /AVCVideo Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, 2003.

[HAL02] A. Hallapuro, M. Karczewicz and H. Malvar, "Low Complexity Transform and Quantization – Part I: Basic Implementation," JVT document JVT-B038, Geneva, February 2002.

[REF03] H.264 Reference Software Version JM6.1d, http://bs.hhi.de/~suehring/tml/, March 2003.

[GOL66] S. W. Golomb, "Run-length encoding," *IEEE Trans. on Inf. Theory*, IT-12, pp. 399–401, 1966.

[BJO02] G. Bjontegaard and K. Lillevold, "Context-adaptive VLC coding of coefficients," JVT document JVT-C028, Fairfax, May 2002.

[MAR03] D. Marpe, H. Schwarz and T. Wiegand, "Context-Based Adaptive Binary Arithmetic Coding in the H.264 / AVC Video Compression Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, 2003.

Appendix B

The video corpus

The Appendix presents a description of the experimental corpus

B.1 MPEG-4 AVC encoding parameter

In generally the standard includes two different configuration of codec called profiles, each targeting a specific class of applications:

- *Baseline Profile* (BP): Primarily for lower-cost applications that use fewer resources, this profile is widely used in mobile applications and videoconferencing.
- *Main Profile* (MP): This profile is used for standard-definition digital TV broadcasts that use the MPEG-4 format as defined in the DVB standard.

In the sequel, Table App.B-1 will provide details about the profile parameters considered in the corpus while Table App.B-2 provides the level parameter descriptions.

Table App.B-1: MPEG-4 AVC profile parameters of our experimental corpus.

Compression Process	*Configuration*	*Baseline*	*Main*
Pre -processing	**4:2:0 format**	**Yes**	**Yes**
	4:0:0 format	**No**	**No**
	4:2:2 format	**No**	**No**
	4:4:4 format	**No**	**No**
	Deblocking filter	**Yes**	**Yes**
Prediction	**Tranches I et P**	**Yes**	**Yes**
	Tranches B	**No**	**Yes**
	Tranches SI et SP	**No**	**No**
	Multiple Reference	**Yes**	**Yes**
	Redundant Slices (RS)	**Yes**	**No**
Quantization	**Quantization Matrix**	**No**	**No**
Encoding	**CAVLC**	**Yes**	**Yes**
	CABAC	**No**	**Yes**
	8 Bit per pixel	**Yes**	**Yes**

Table App.B-2: MPEG-4 AVC level parameters of our experimental corpus.

Level	*Maximum bloc number per frame*	*Maximum rate for Baseline & Main Profils*	*Resolution & fps*
1	**99**	**64 kbit/s**	**128×96/30.9**
2.2	**1620**	**256 kbit/s**	**352×576/25.6**
3	**1620**	**5 Mbit/s**	**720×576/25.0**
4	**8192**	**10 Mbit/s**	**1920×1080/30.1**

B.2 Corpus

The experiments are carried out on large number of heterogeneous video sequences, as illustrated bellow. The sizes and the types of the various contents processed in the present thesis are presented in Table App.B-3.

Video natural 64 kbps

• Baseline profile 128×96@30.9

Video natural 256 kbps

• Baseline profile 352×576/25.6

Video natural 5 Mbps

• Main profile 720×576/25.0

Video natural 10 Mbps

• Main profile 1920×1080/30.1

Video cartoons 64 kbps

• Baseline profile 128×96@30.9

Video cartoons 256 kbps

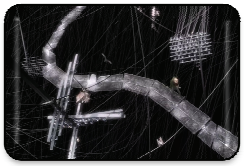
• Baseline profile 352×576/25.6

Video cartoons 5 Mbps

• Main profile 720×576/25.0

Video cartoons 10 Mbps

• Main profile 1920×1080/30.1

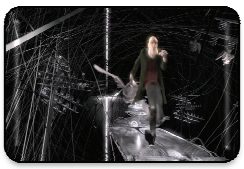


Video computer generated 64 kbps

• Baseline profile 128×96@30.9

Video computer generated 256 kbps

• Baseline profile 352×576/25.6



Video computer generated 5 Mbps

• Main profile 720×576/25.0

Video computer generated 10 Mbps

• Main profile 1920×1080/30.1

Table App.B-3: Content of our experimental corpus.

Video category	*Number of sequence*	*Total Duration*	*Related R&D project*
Natural	**17**	**2h 30 min**	**MEDIEVALS, TAMUSO, HD3D-IIO**
Cartoons	**3**	**27 min**	**MEDIEVALS**
Computer generated	**7**	**45 min**	**MEDIEVALS**

Appendix C

Digital Video Quality

The Appendix presents the DVQ as a sophisticated measure, taking into account both the filtering and the masking properties of the human visual system.

The DVQ is a sophisticated measure, taking into account both the filtering and the masking properties of the human visual system, Figure App.C-1. First, both the reference and the tested videos are spatially and temporally filtered. Then their difference is computed and the reference video is passed through a second filter so as to be used to compute the locations where the artefacts would be masked by the original data themselves, Figure App.C-2.

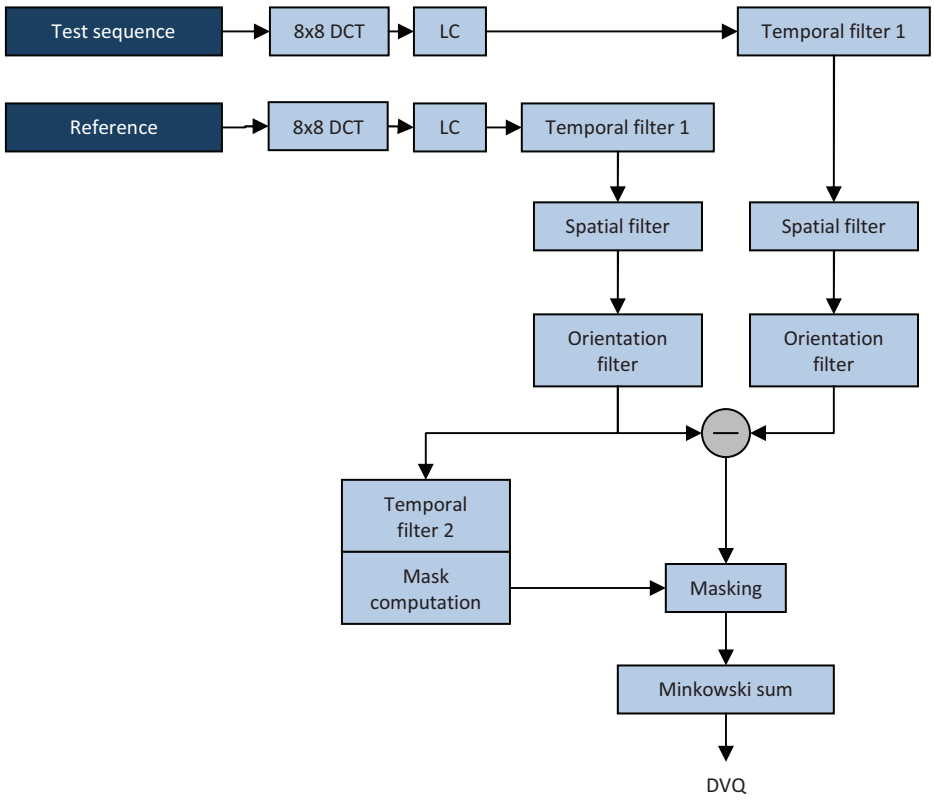
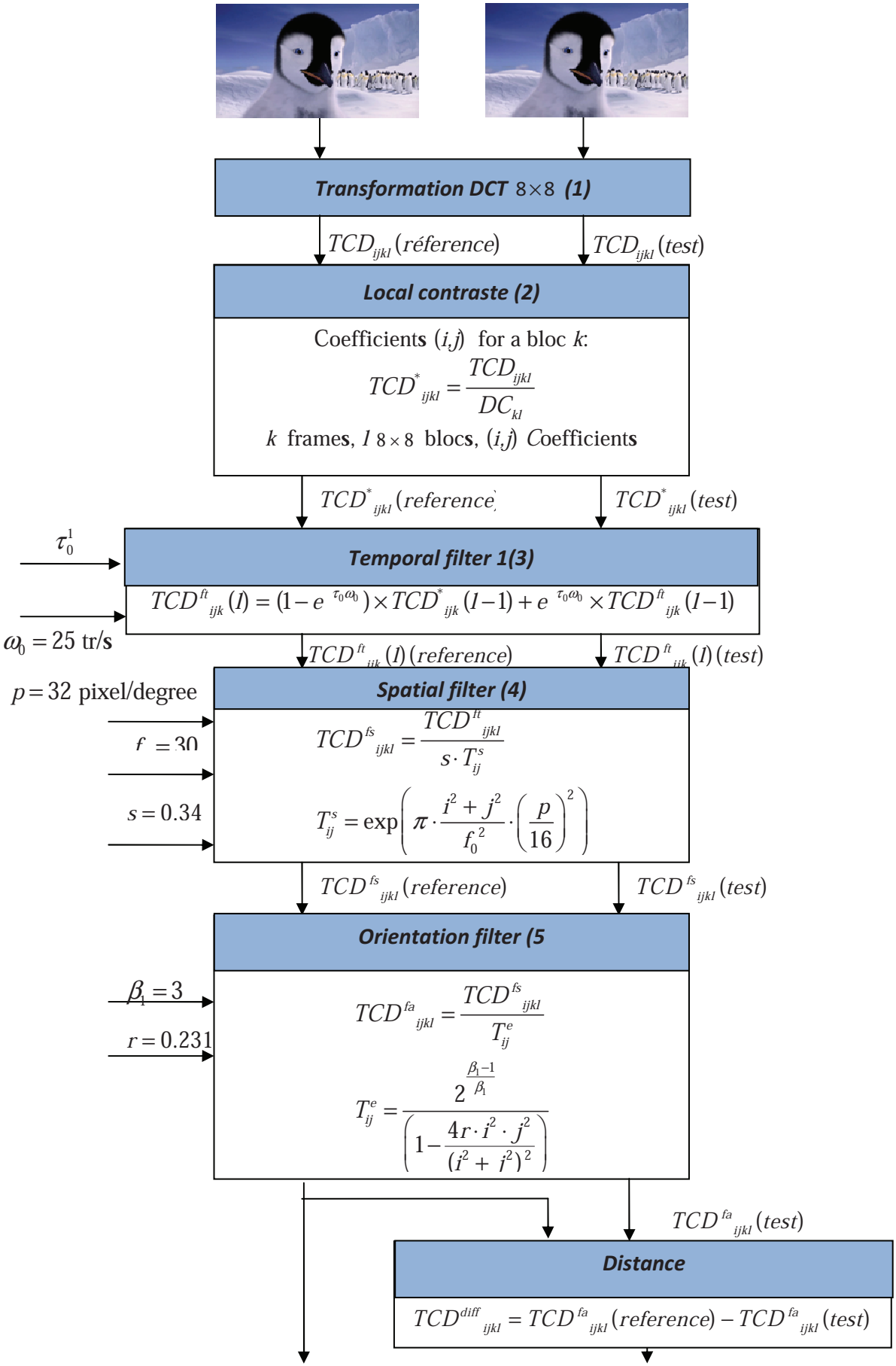


Figure App.C-1: Synopsis of DVQ


Transformation DCT 8×8 (1)
$TCD_{ijkl}(réference)$
$TCD_{ijkl}(test)$
Local contraste (2)
Coefficients (i,j) for a bloc k:
$TCD^{*}_{ijkl} = \frac{TCD_{ijkl}}{DC_{kl}}$
k frames, l 8×8 blocs, (i,j) Coefficients
$TCD^{*}_{ijkl}(reference)$
$TCD^{*}_{ijkl}(test)$
τ_0^1
$\omega_0 = 25$ tr/s
Temporal filter 1(3)
$TCD^{ft}_{ijk}(l) = (1-e^{-\tau_0\omega_0}) \times TCD^{*}_{ijk}(l-1) + e^{-\tau_0\omega_0} \times TCD^{ft}_{ijk}(l-1)$
$TCD^{ft}_{ijk}(l)(reference)$
$TCD^{ft}_{ijk}(l)(test)$
$p = 32$ pixel/degree
$f = 30$
$s = 0.34$
Spatial filter (4)
$TCD^{fs}_{ijkl} = \frac{TCD^{ft}_{ijkl}}{s \cdot T^{s}_{ij}}$
$T^{s}_{ij} = \exp\left(\pi \cdot \frac{i^2 + j^2}{f_0^2} \cdot \left(\frac{p}{16}\right)^2\right)$
$TCD^{fs}_{ijkl}(reference)$
$TCD^{fs}_{ijkl}(test)$
Orientation filter (5
$\beta_1 = 3$
$r = 0.231$
$TCD^{fa}_{ijkl} = \frac{TCD^{fs}_{ijkl}}{T^{e}_{ij}}$
$T^{e}_{ij} = \frac{2^{\frac{\beta_1 - 1}{\beta_1}}}{\left(1 - \frac{4r \cdot i^2 \cdot j^2}{(i^2 + j^2)^2}\right)}$
$TCD^{fa}_{ijkl}(test)$
Distance
$TCD^{diff}_{ijkl} = TCD^{fa}_{ijkl}(reference) - TCD^{fa}_{ijkl}(test)$

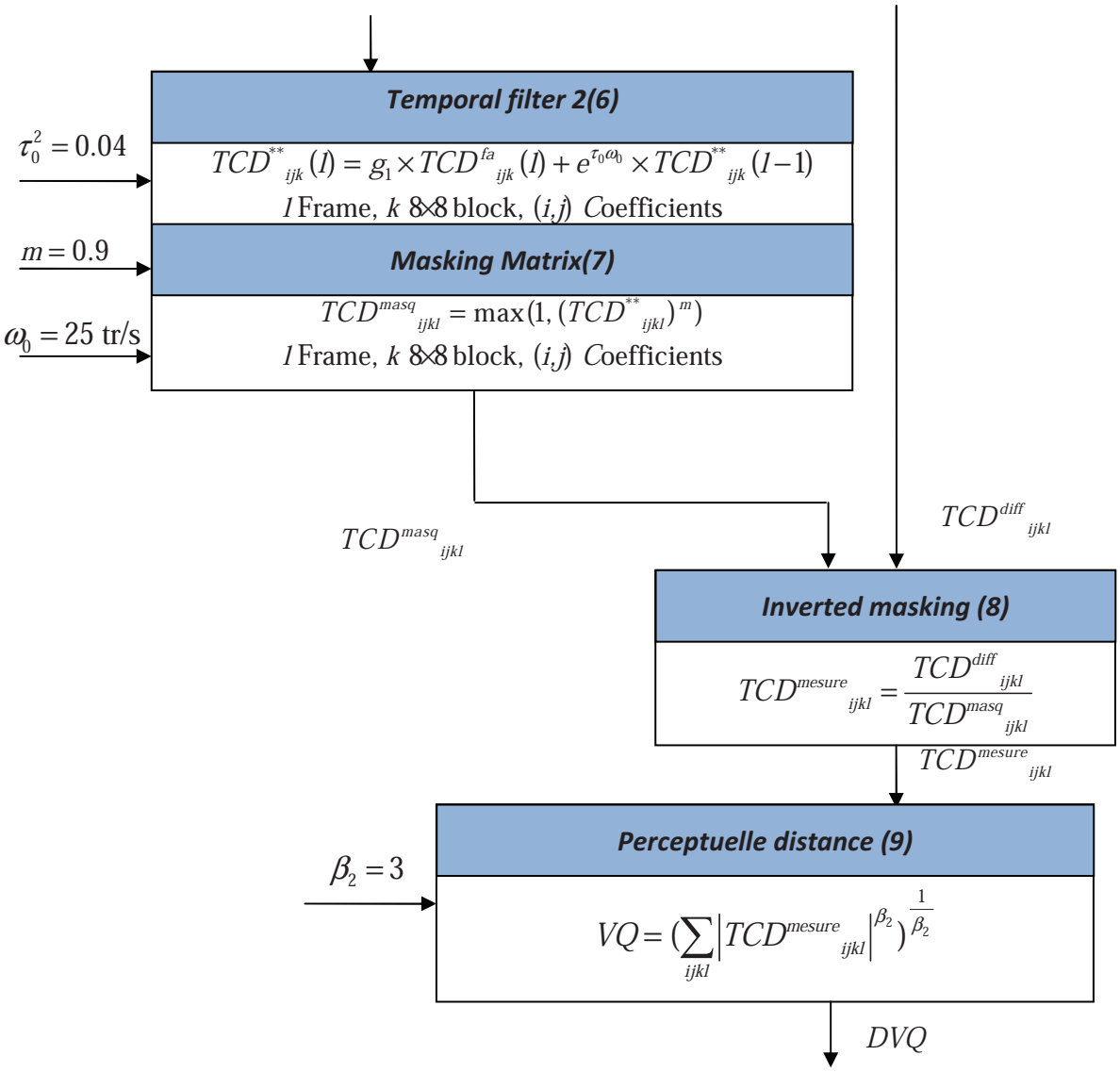


Figure App.C-2: DVQ flowchart.

(1) DCT blocs 8×8

The perceptual model of Watson was already defined and implemented for MPEG-2 uses a DCT transform on 8×8 blocks. We kept the work environment by calculating the DVQ respect to such transformation.

(2) Local contrast

The eye is less sensitive to absolute luminance value of a point in the image that unlike the luminance value from the local neighborhood. The local contrast is to translate this property of HVS: to divide each DCT coefficient by the DC component DC (total energy on a neighborhood 8×8 block).

(3) Temporal filter 1

The temporal filter is to attenuate high temporal frequencies that are less perceived by the visual system than low frequencies [WIN99]. This filter is a low-pass infinite impulse response of the first order, attacked the input by the coefficient of a block of the frame (it will be applied individually to each coefficient). As this is a recursive filter, the filtered output frame $TCD_{ijk}^{ft}(1)$ is expressed in terms of the previous filtered frame and the previous $TCD_{ijk}^{ft}(1-1)$ and original $TCD_{ijk}^{*}(1-1)$ frame.

(4) Spatial filter

This filter is a Gaussian low pass that eliminates high spatial frequencies less perceived by the visual system [WIN99].

(5) Orientation filter

At this stage, we filter the frequencies that are less oblique perceived by the visual system [WIN99].

(6) Temporal filter2

This filter is designed to prepare the mask generated from the original video. The low pass filter has the coefficients most frequently detected compared to the masking effect. This filter applies to the frame to provide the filtered output frame $TCD_{ijk}^{**}(1)$ expressed in terms of the previous filtered frame $TCD_{ijk}^{**}(1-1)$ and the original frame $TCD_{ijk}^{*}(1)$, in the same manner as the first time filter.

(7) Masking matrix

Masking is to generate the mask matrix from the filtered frame. The matrix will mask the size of the frame and this generated as follows: the image coefficients for which there is not a masking phenomenon (ratios less than 1) will be replaced by 1. the image coefficients for which there is a masking phenomenon (coefficients greater than 1) will be fitted by a power m and kept in this matrix masking. This reflects the formula $TCD_{ijk}^{masq}(1) = max(1, (TCD_{ijk}^{**})^{m})$.

(8) Inverted making

This step involves dividing the DCT coefficients of the difference between the original image and test already filtered by the matrix coefficients of masking. This division removes the DCT differences that are masked (ie forgone) in the image to keep only the perceived differences.

(9) Perceptual distance

The measurement of distance is completed by calculating the differences are between the two test images / original as measured by Minkowski with the order β=3.

Appendix D

Transparency evaluation

The Appendix exhibits reference results concerning the compressed domain stream modification visibility, thus providing some limits for compressed stream watermark insertion. Finally, the model thus obtained will be used to devise a watermarking method approaching this capacity (i.e. optimal watermarking methods).

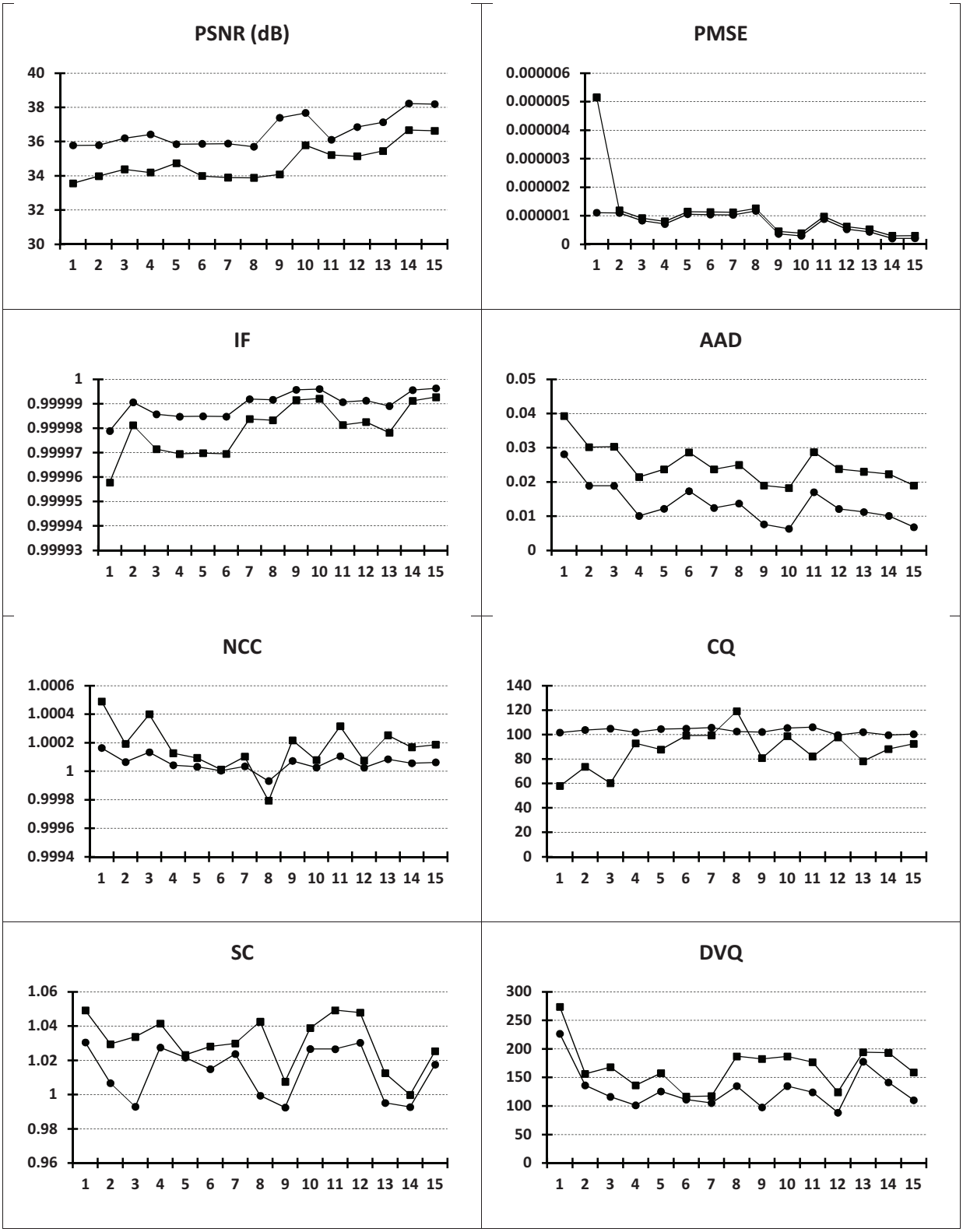


Figure.App.D-1 Effects of {-1, 1} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

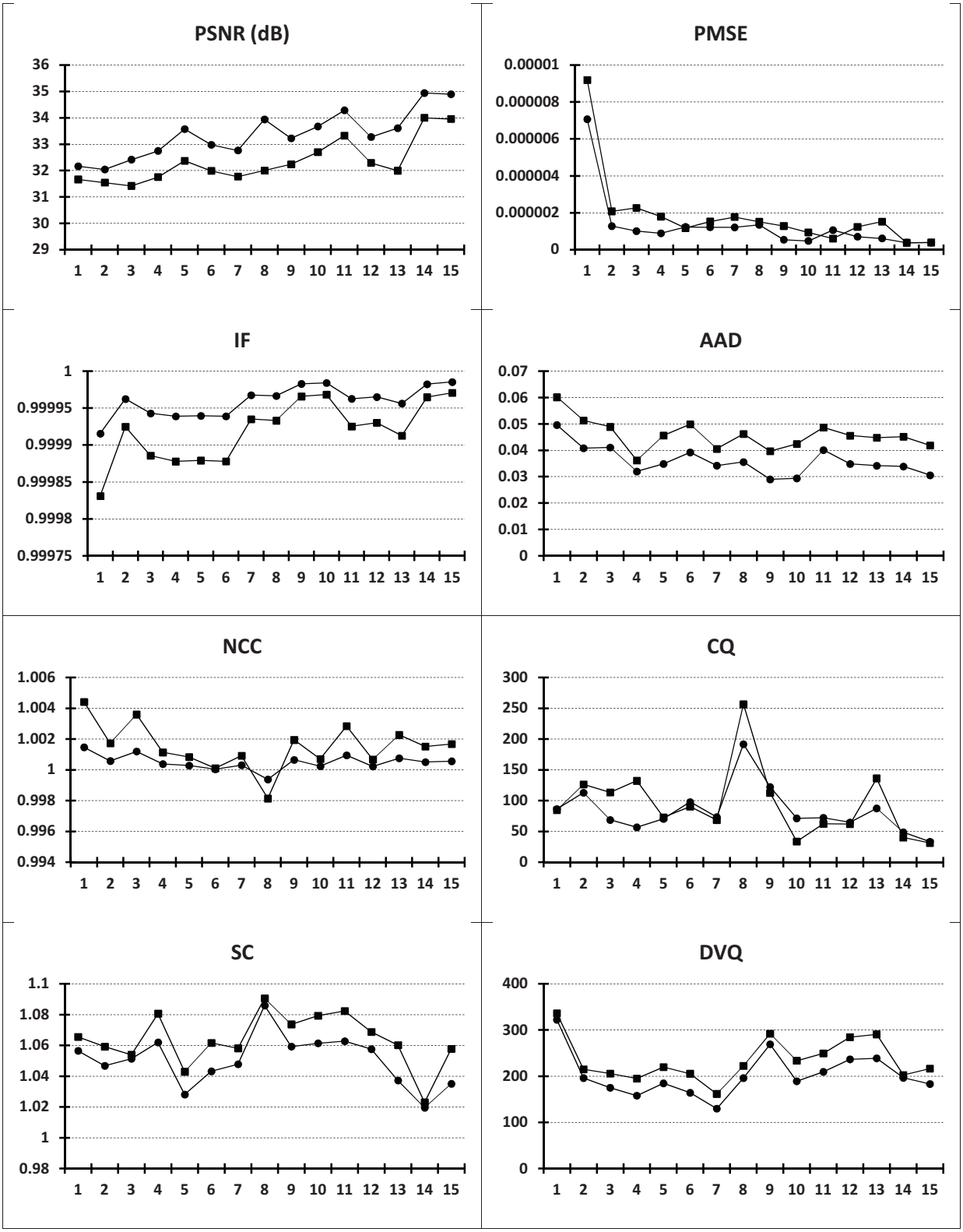


Figure.App.D-2 Effects of {-1, 1} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 10 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

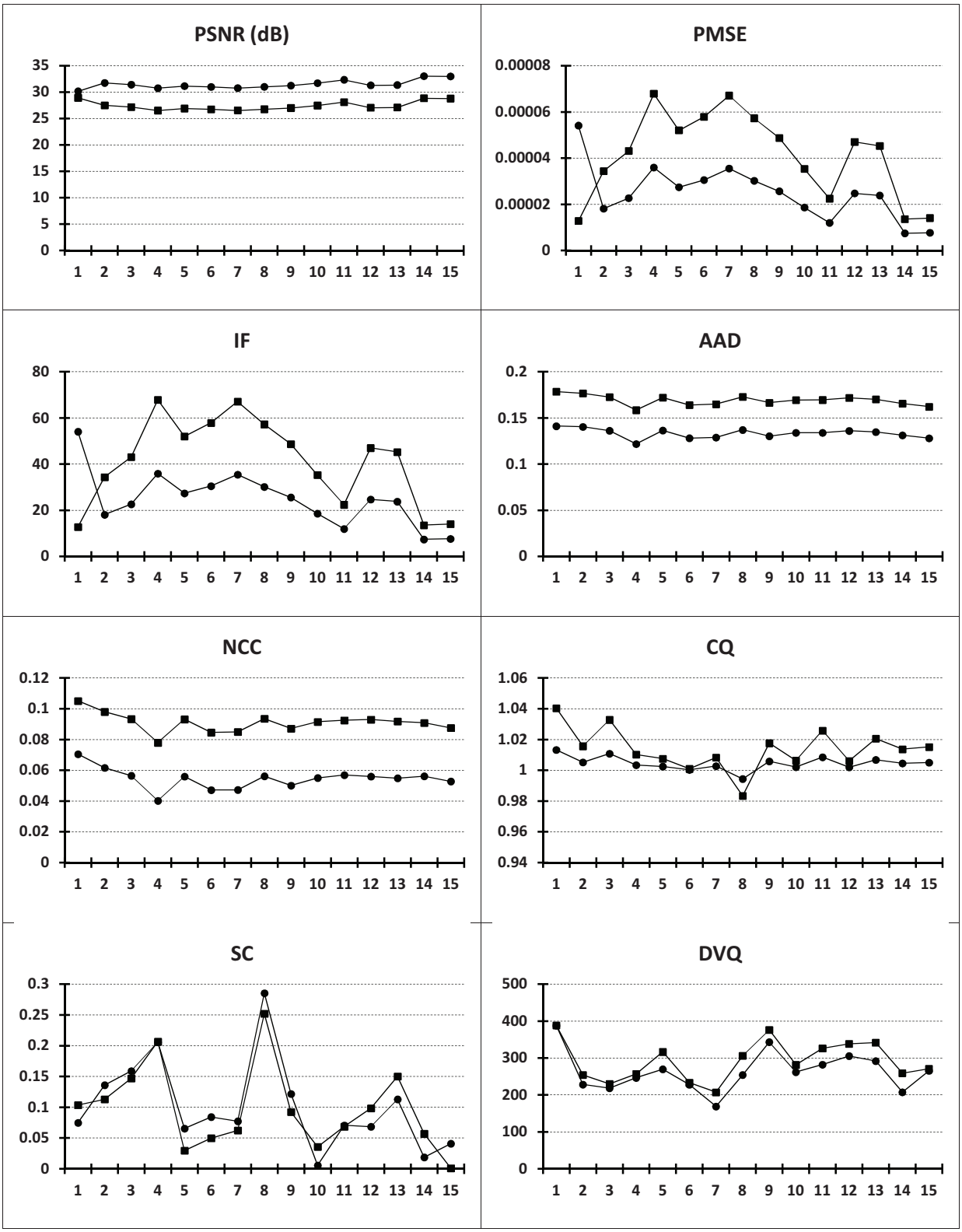


Figure.App.D-3 Effects of {-1, 1} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

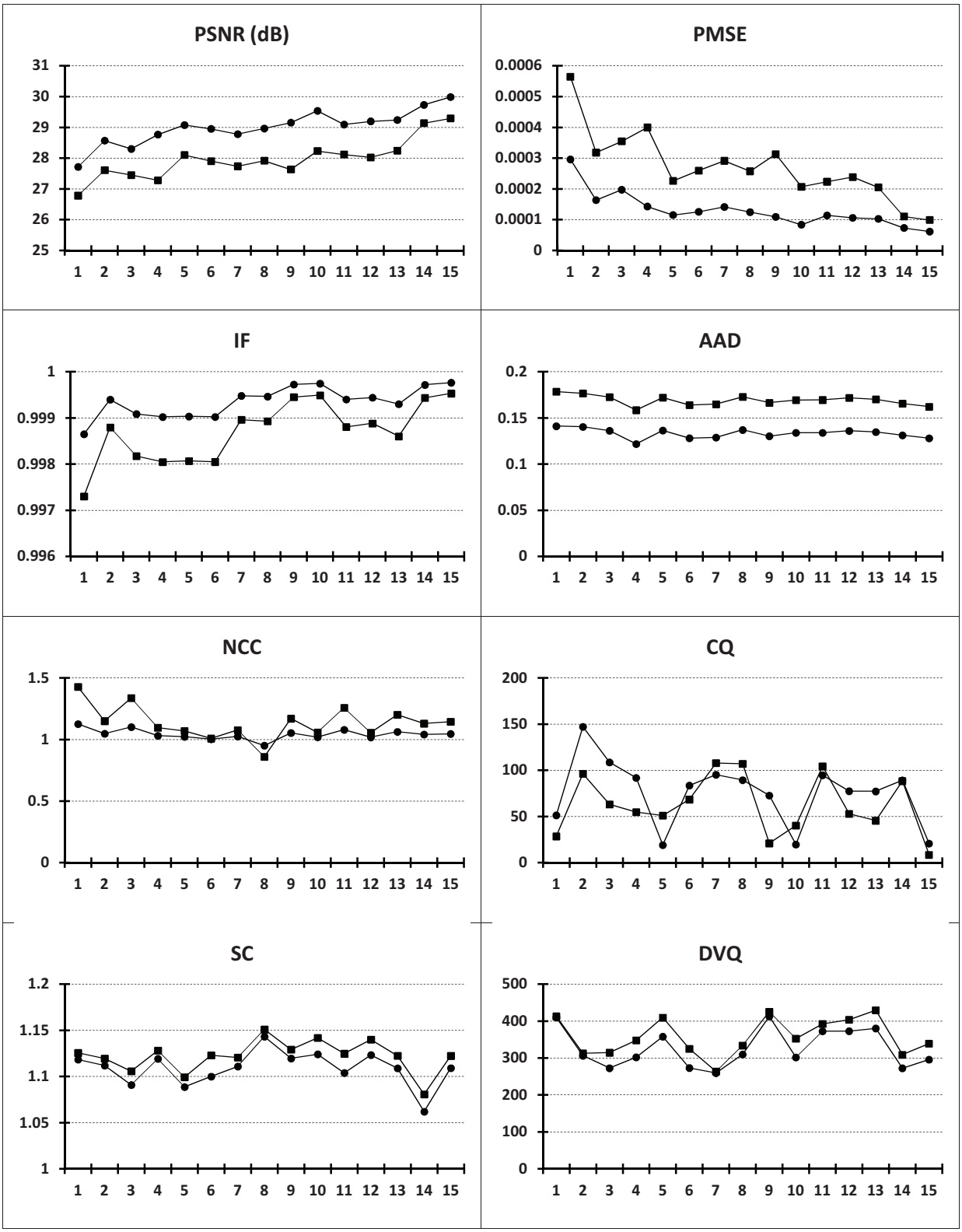


Figure.App.D-4 Effects of {-1, 1} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 100 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

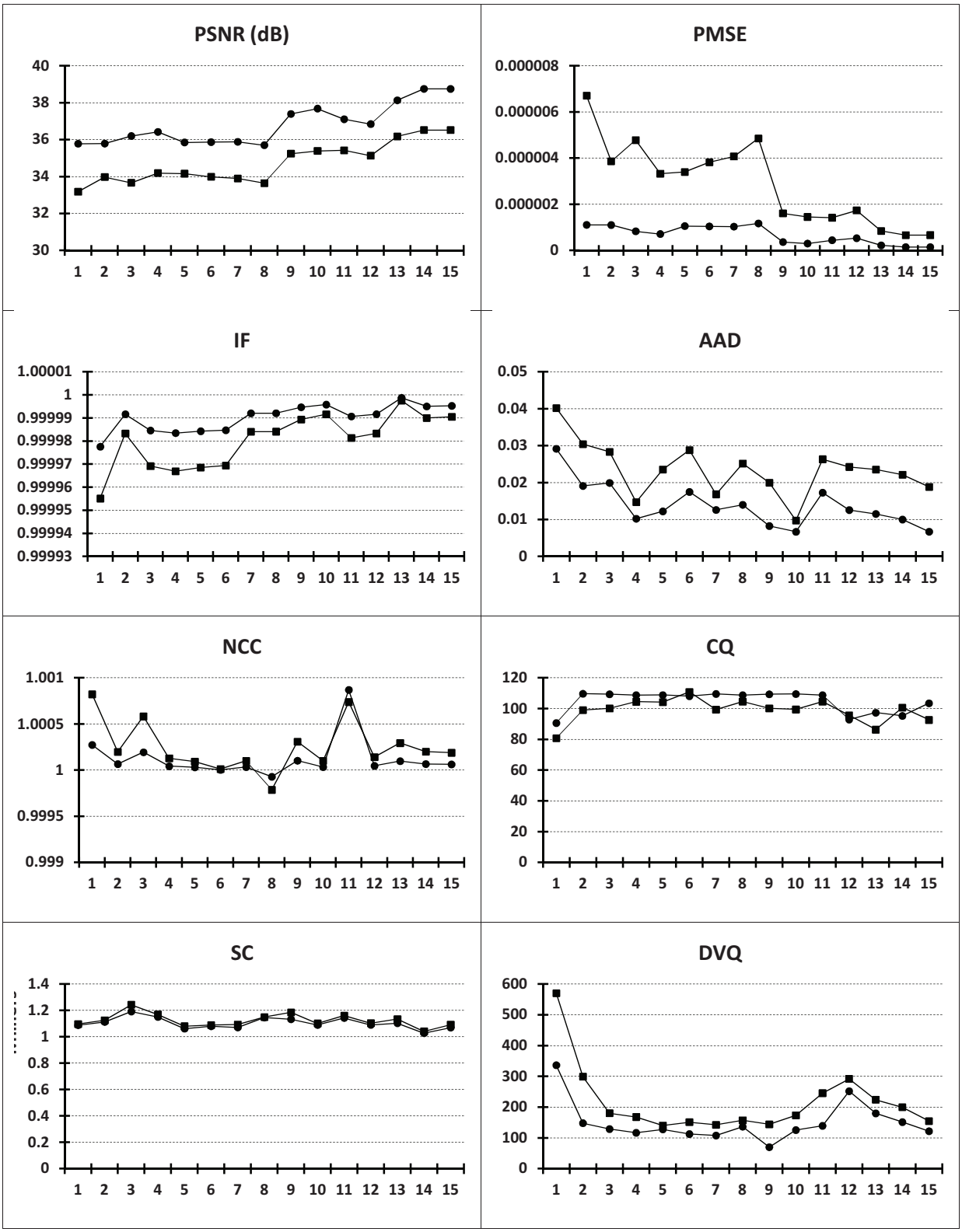


Figure.App.D-5 Effects of five symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 5 macroblocks per frame subjected to errors

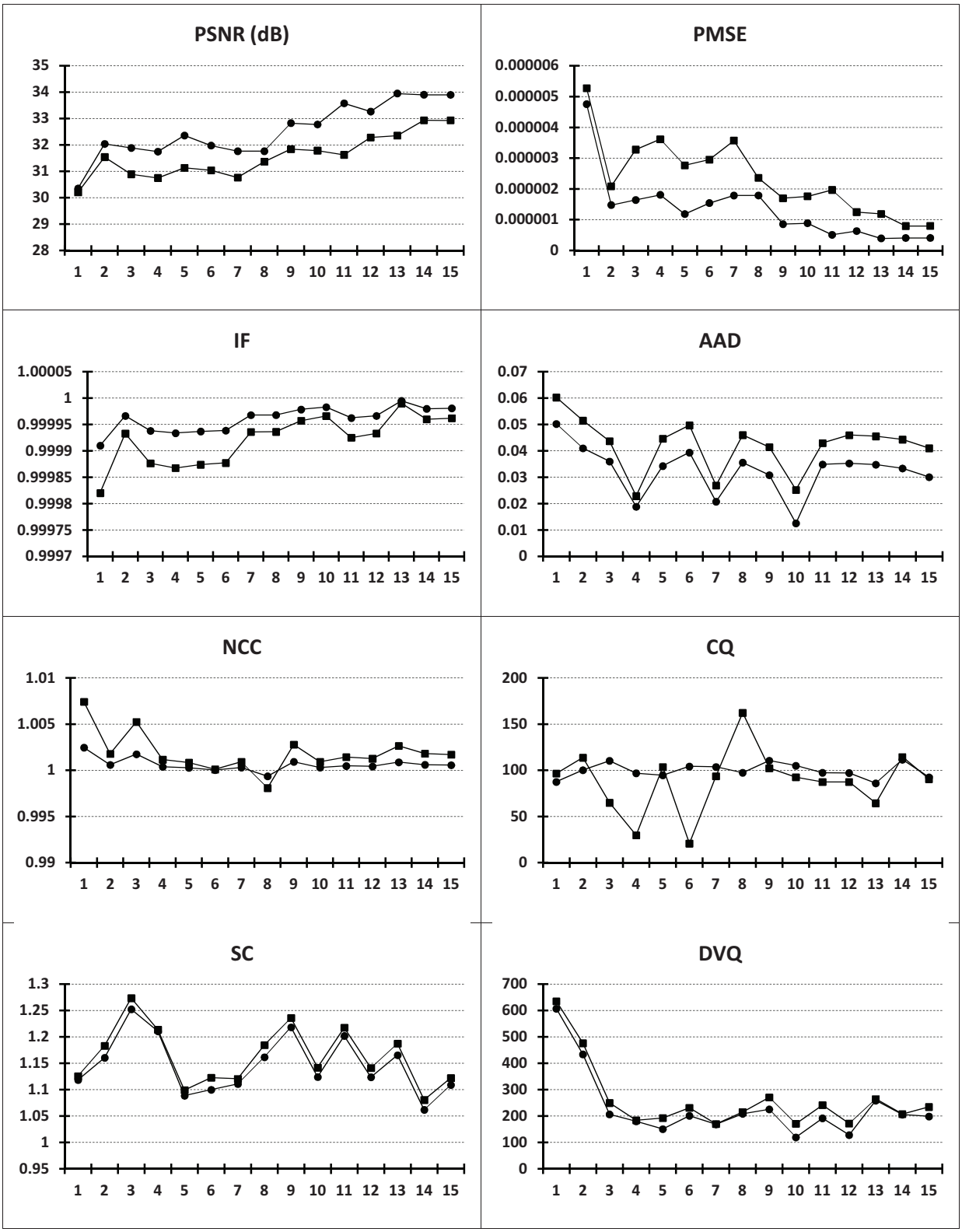


Figure.App.D-6 Effects of five symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 10 macroblocks per frame subjected to errors

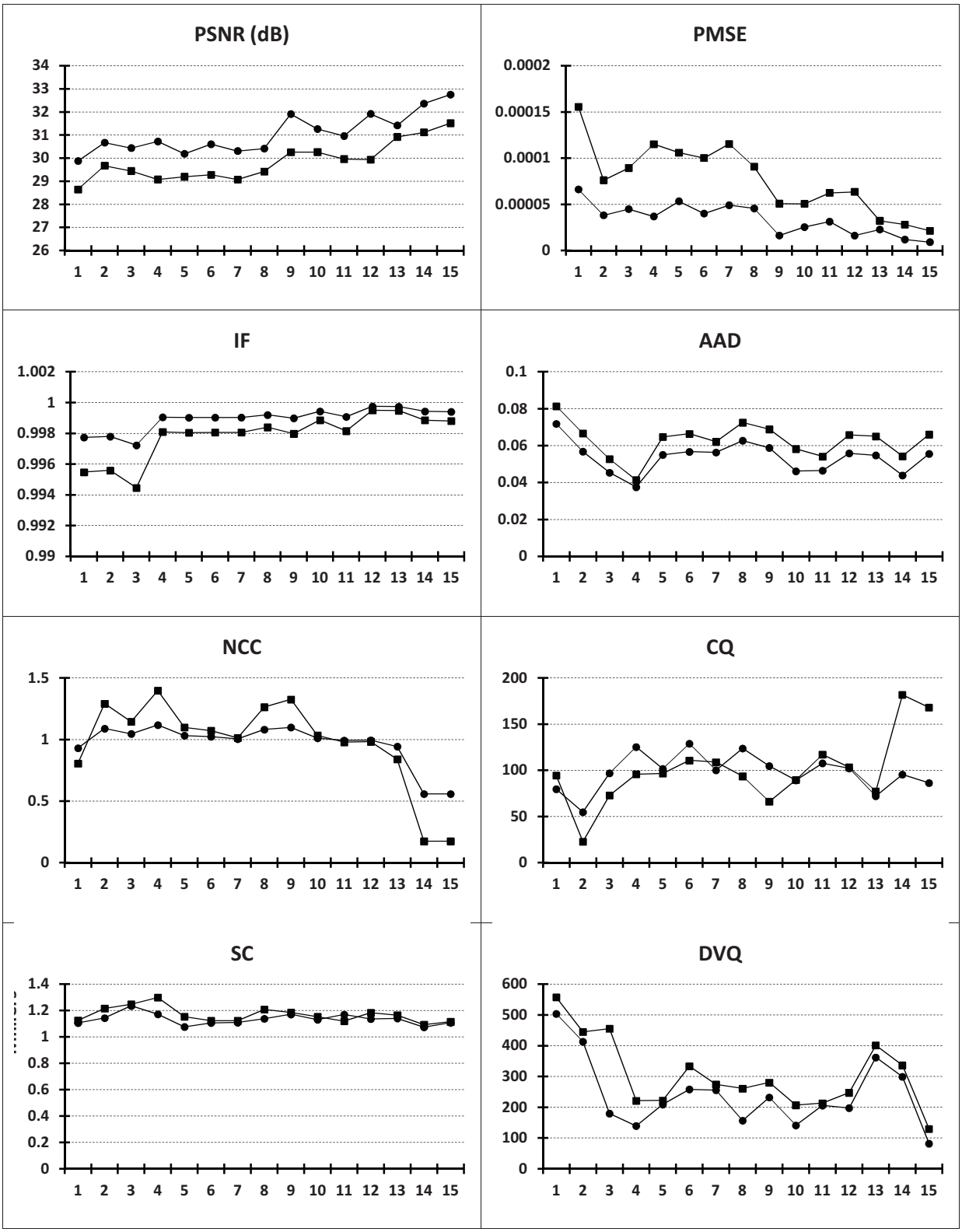


Figure.App.D-7 Effects of five symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 50 macroblocks per frame subjected to errors

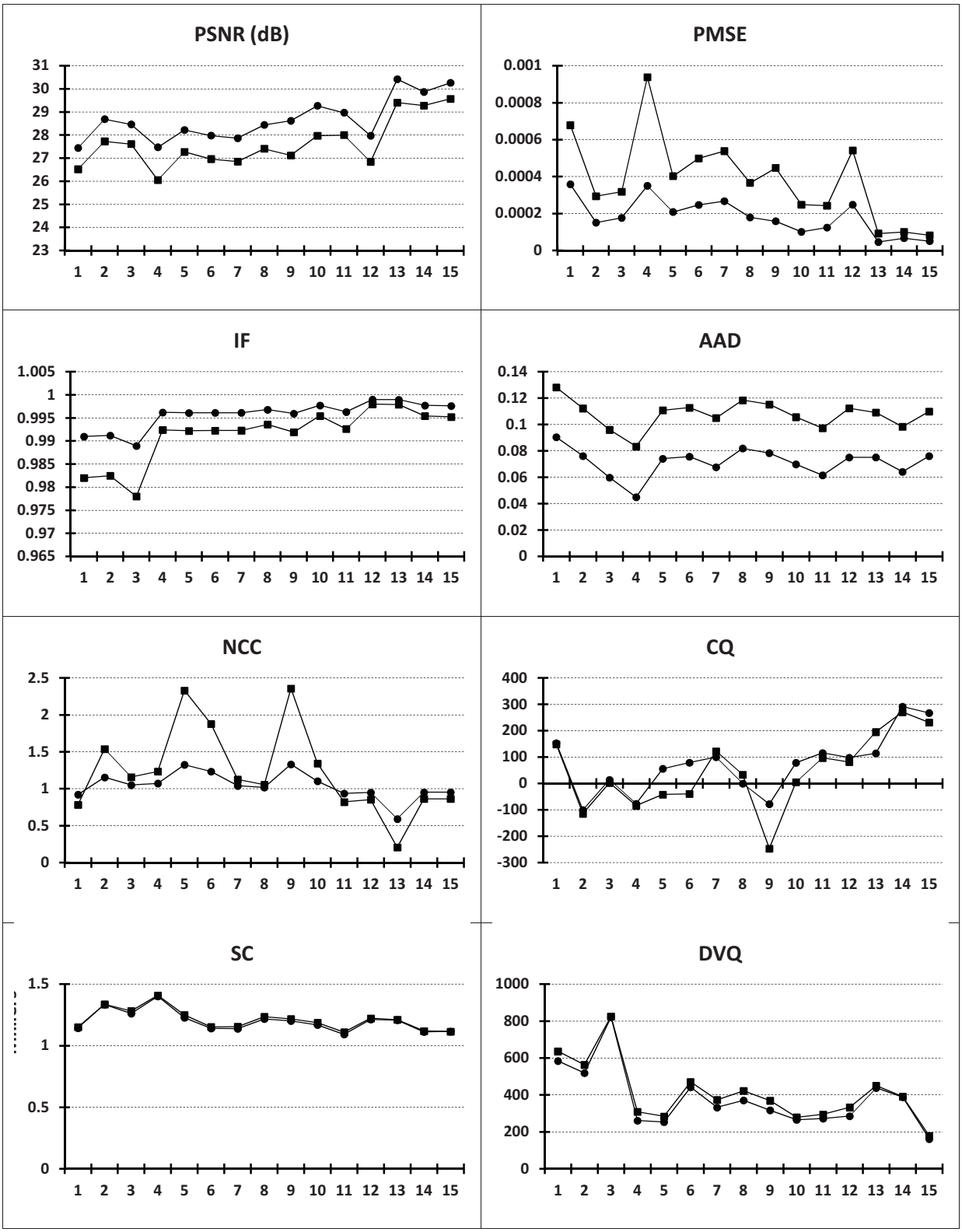


Figure.App.D-8 Effects of five symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 100 macroblocks per frame subjected to errors

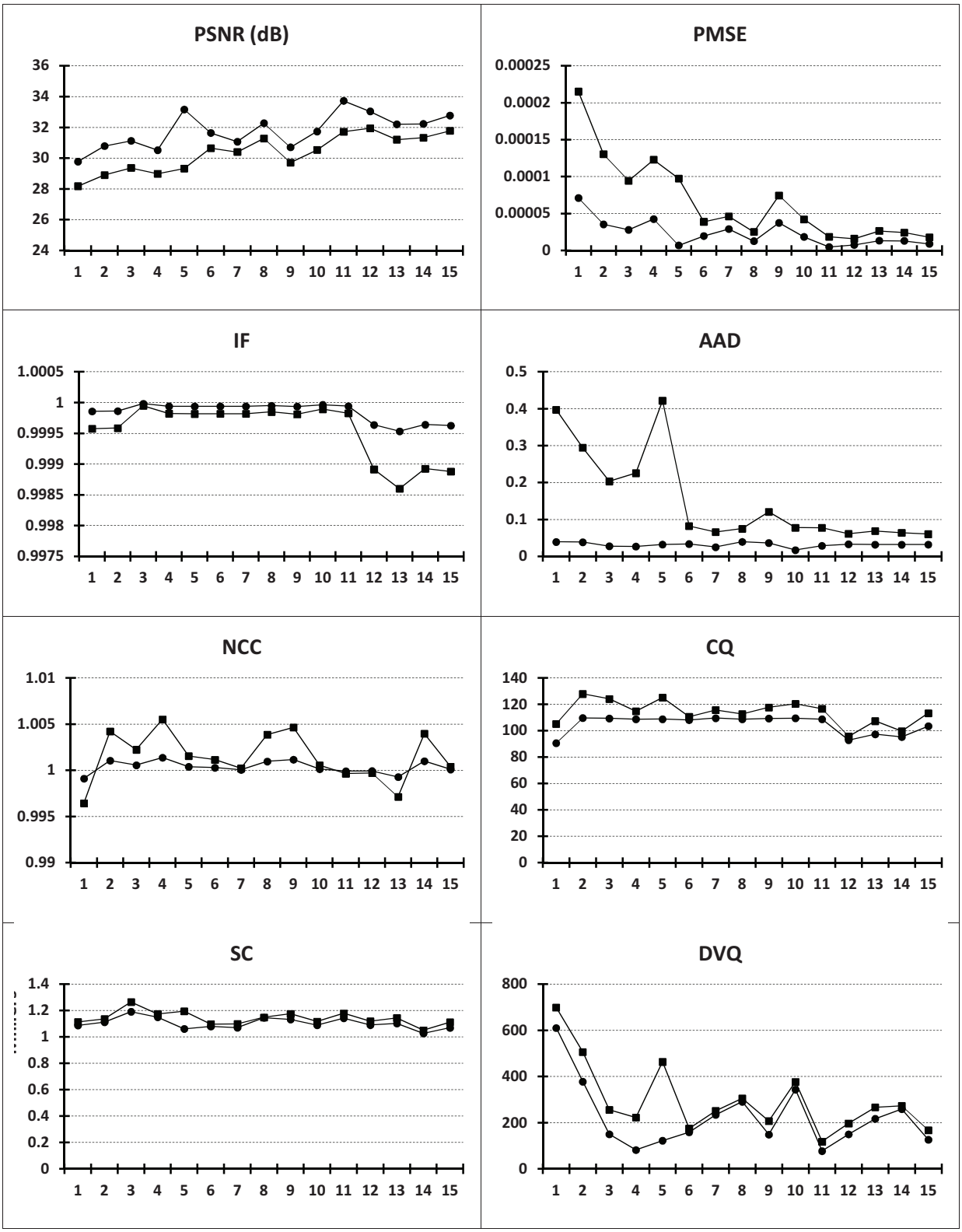


Figure.App.D-9 Effects of {-1, 1} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

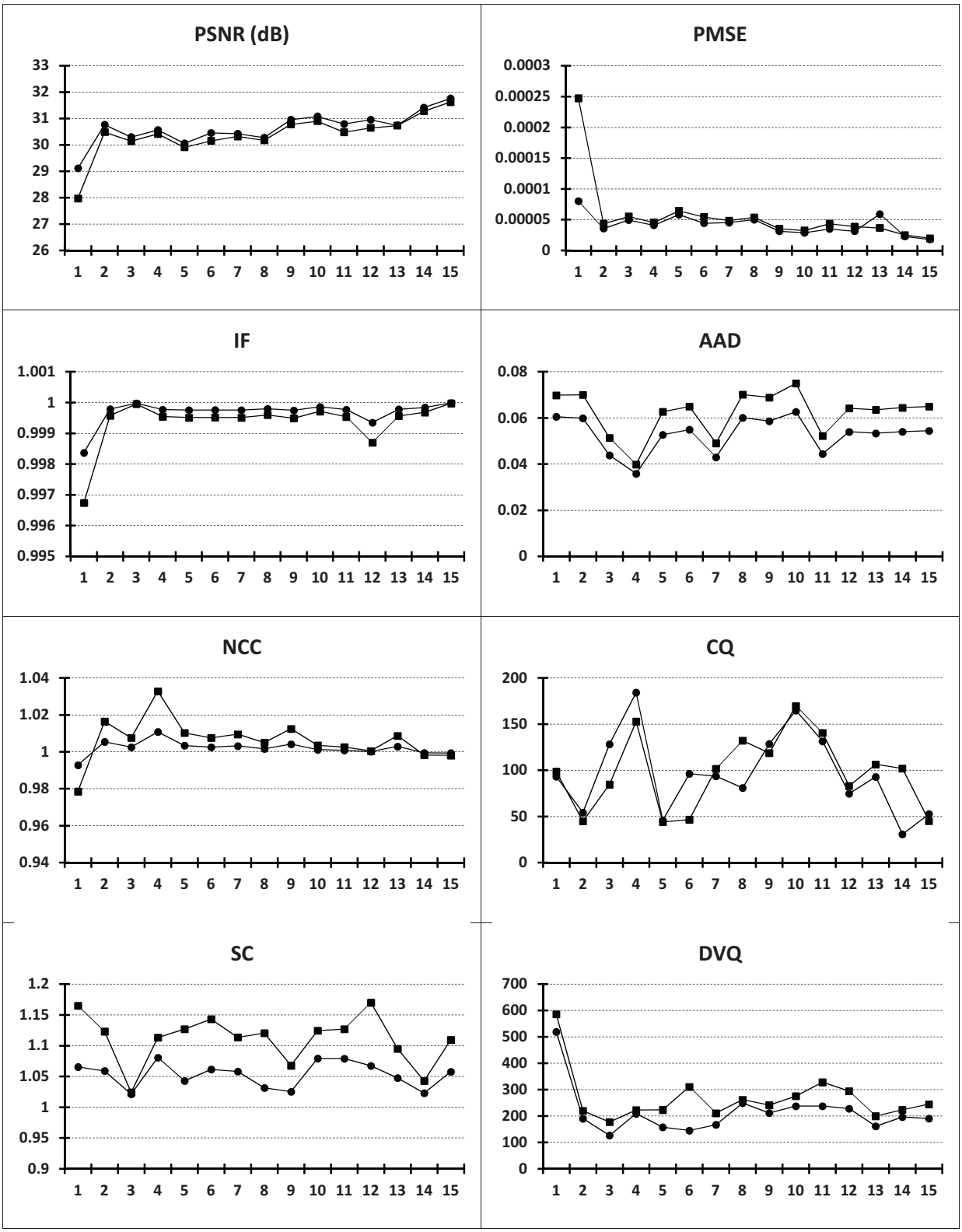


Figure.App.D-10 Effects of {-1, 1} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 10 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

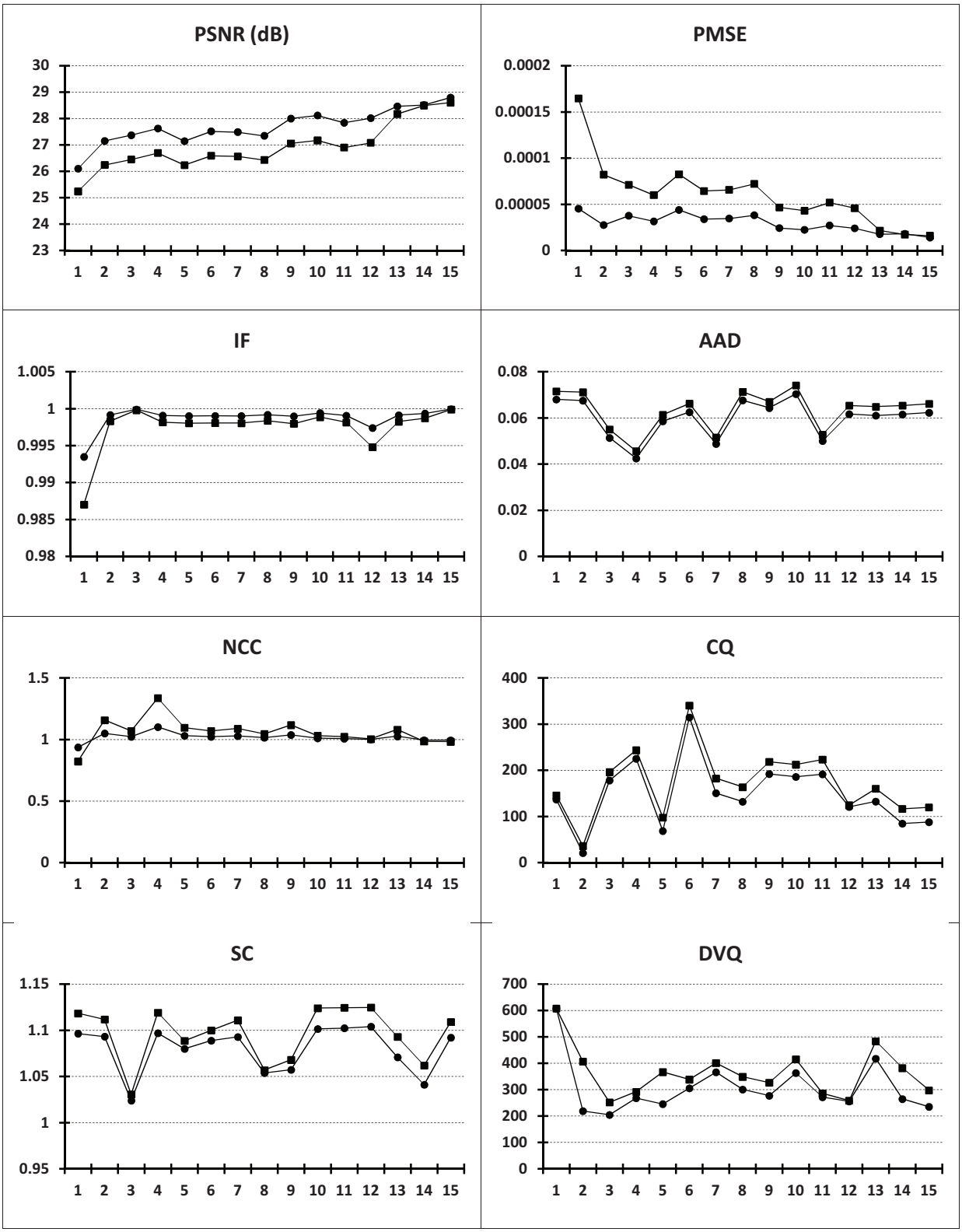


Figure.App.D-11 Effects of {-1, 1} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

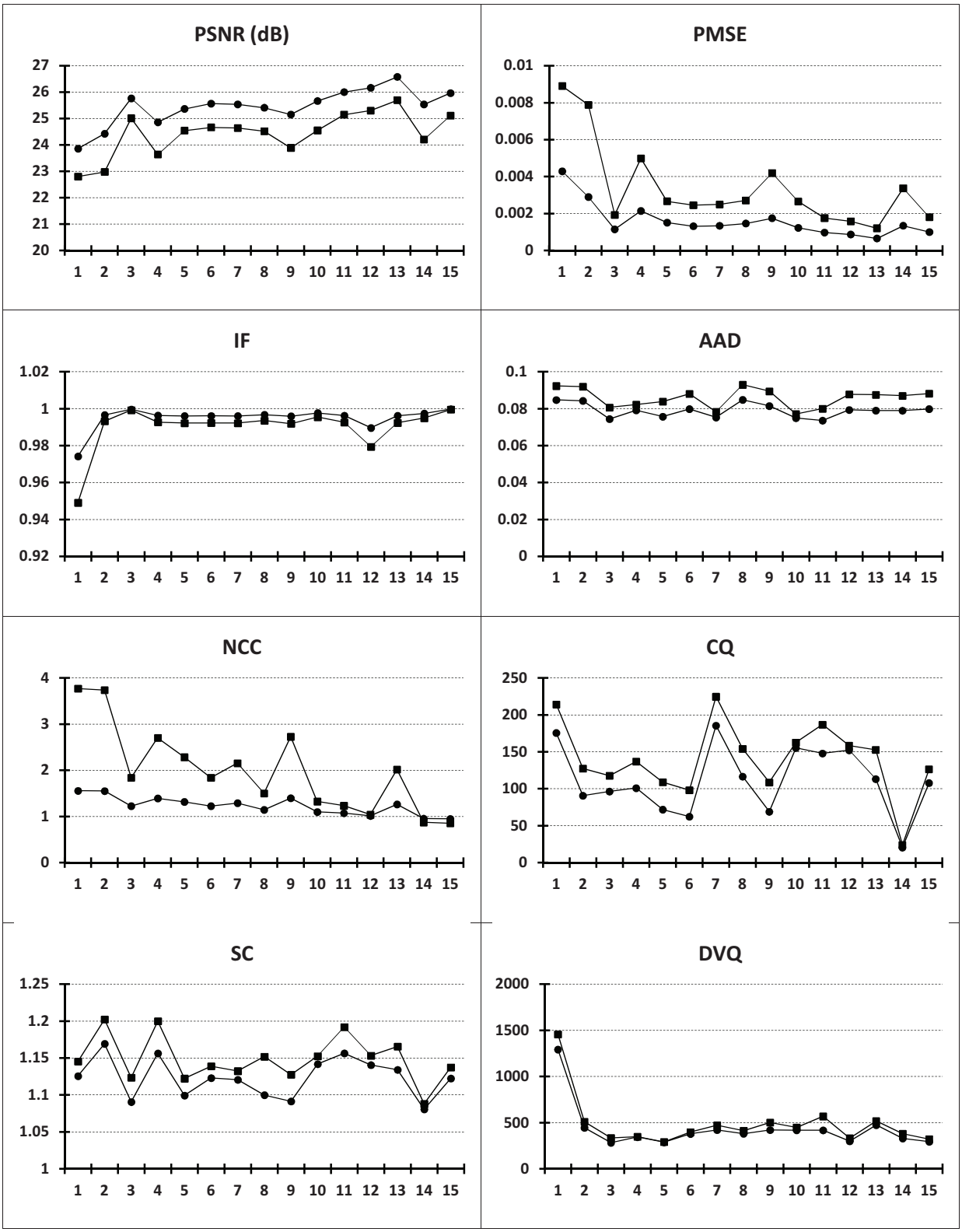


Figure.App.D-12 Effects of {-1, 1} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 100 macroblocks per frame subjected to errors

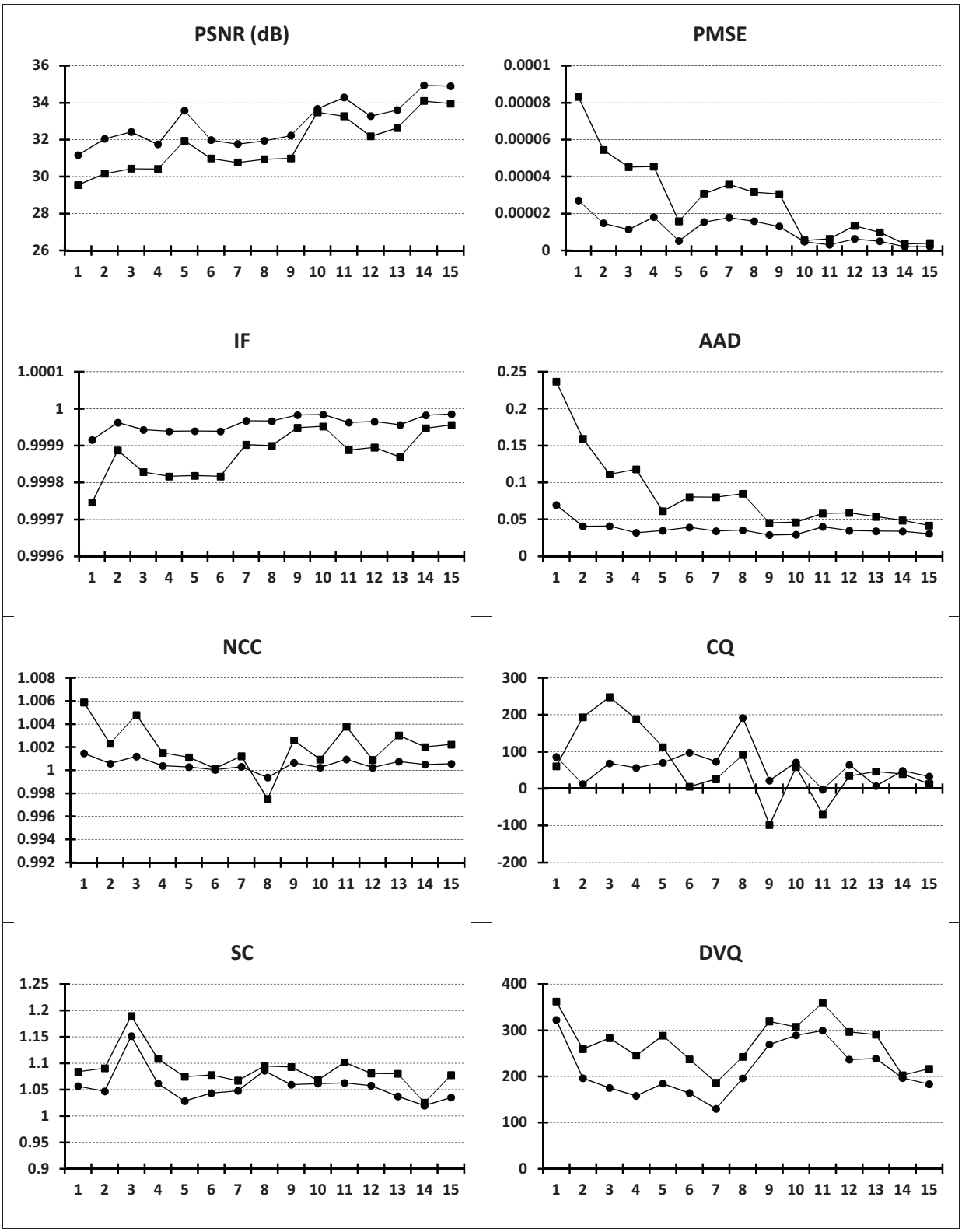


Figure.App.D-13 Effects of {-1, 1} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

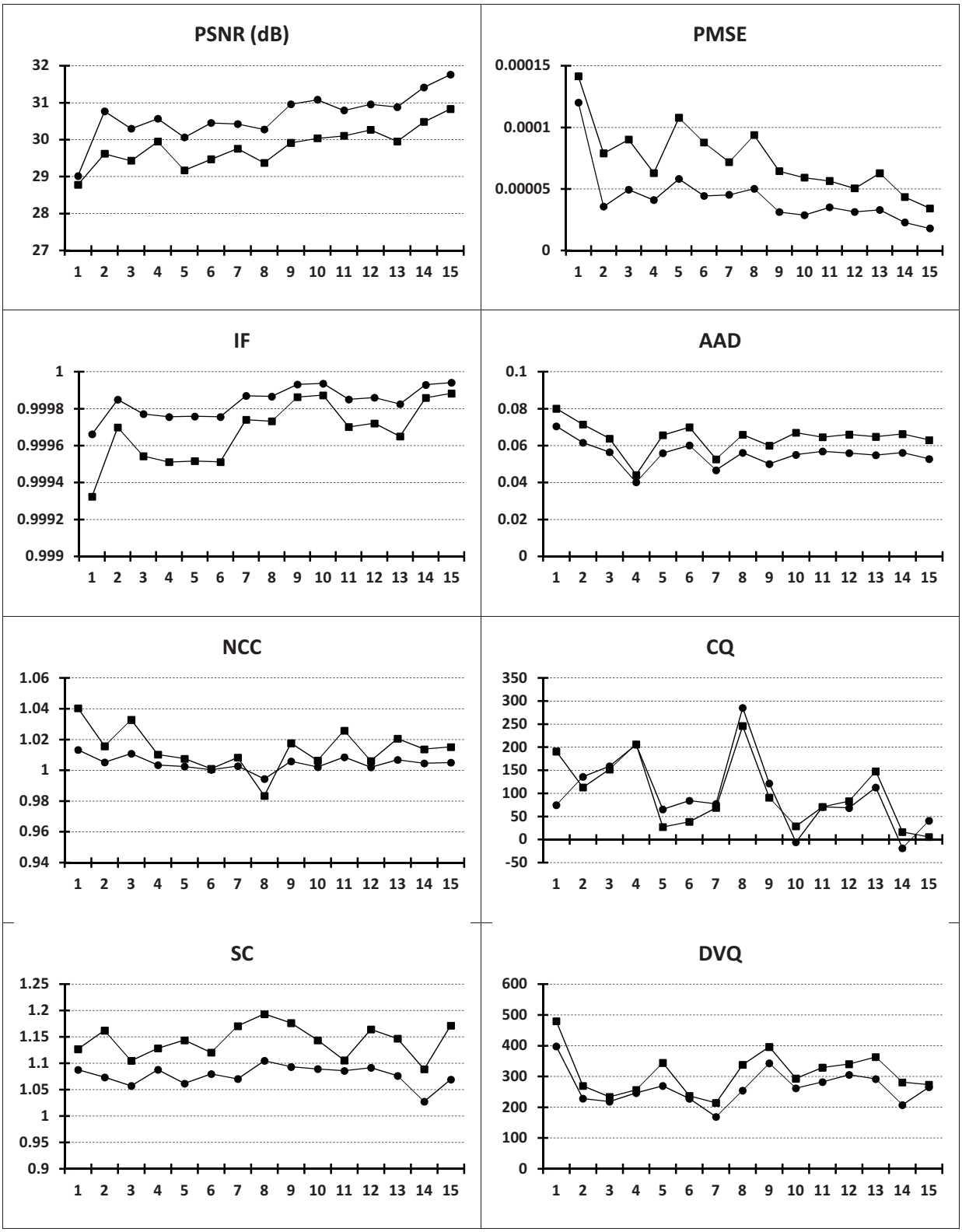


Figure.App.D-14 Effects of five symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 10 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

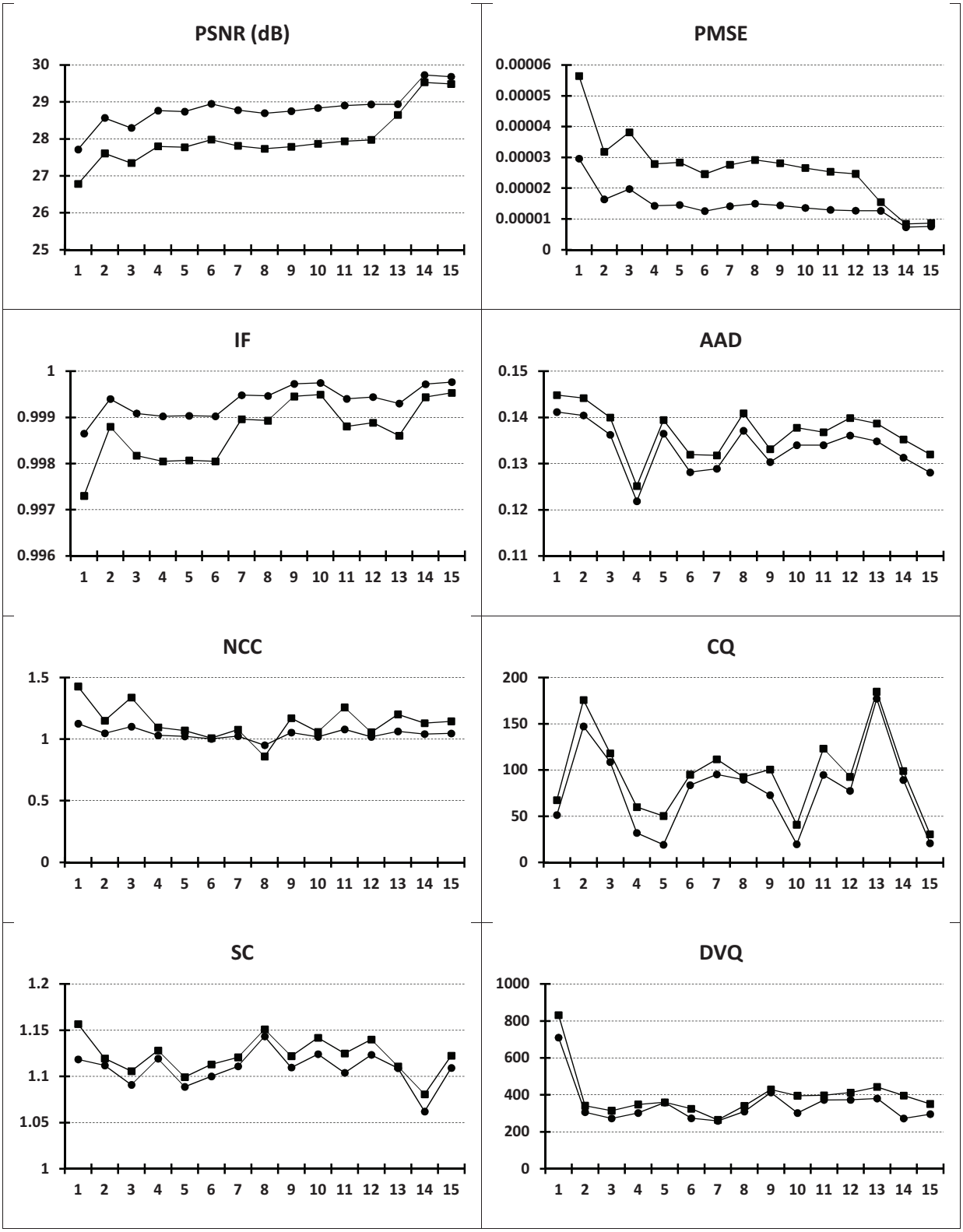


Figure.App.D-15 Effects of {-1, 1} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

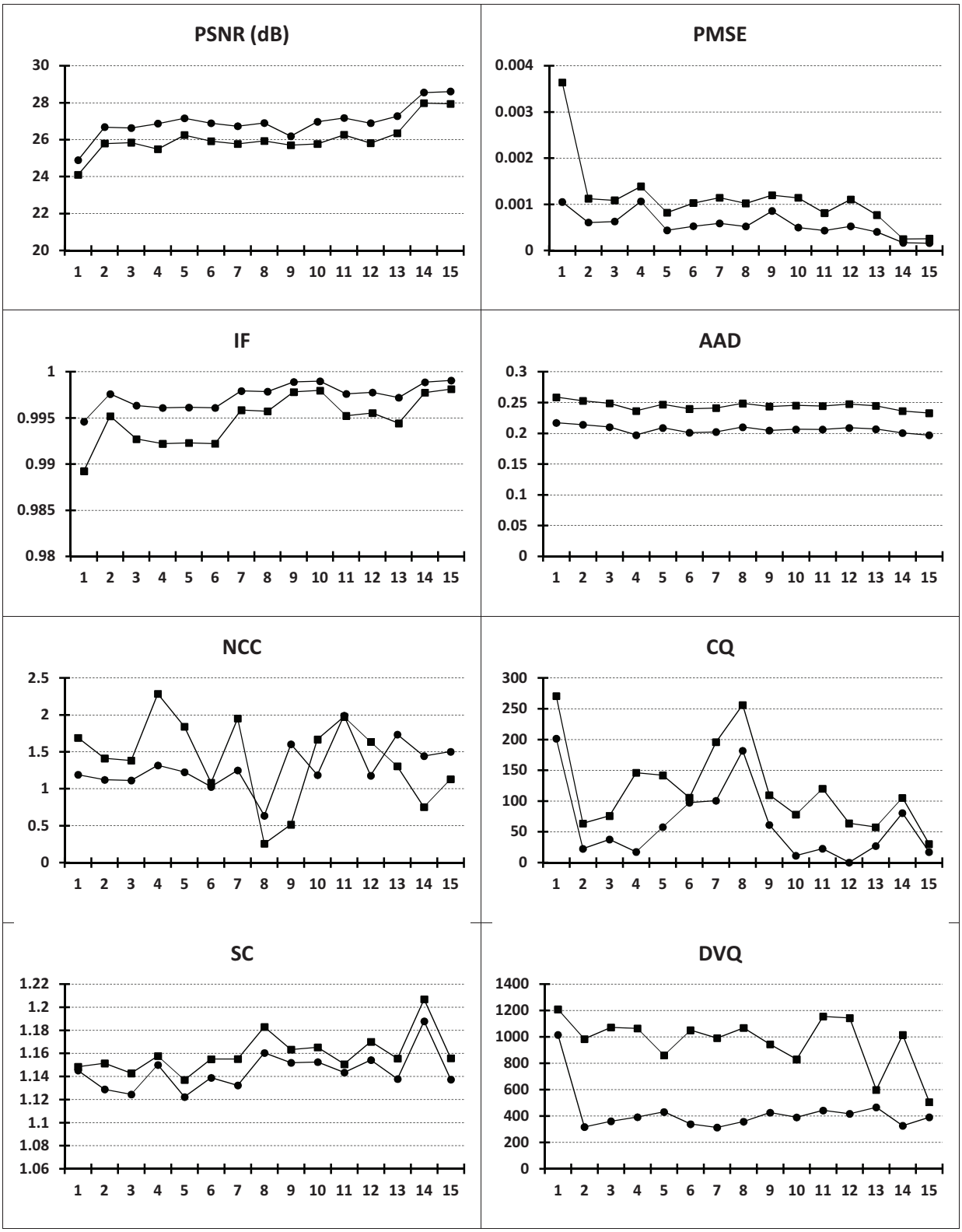


Figure.App.D-16 Effects of {-1, 1} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 100 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

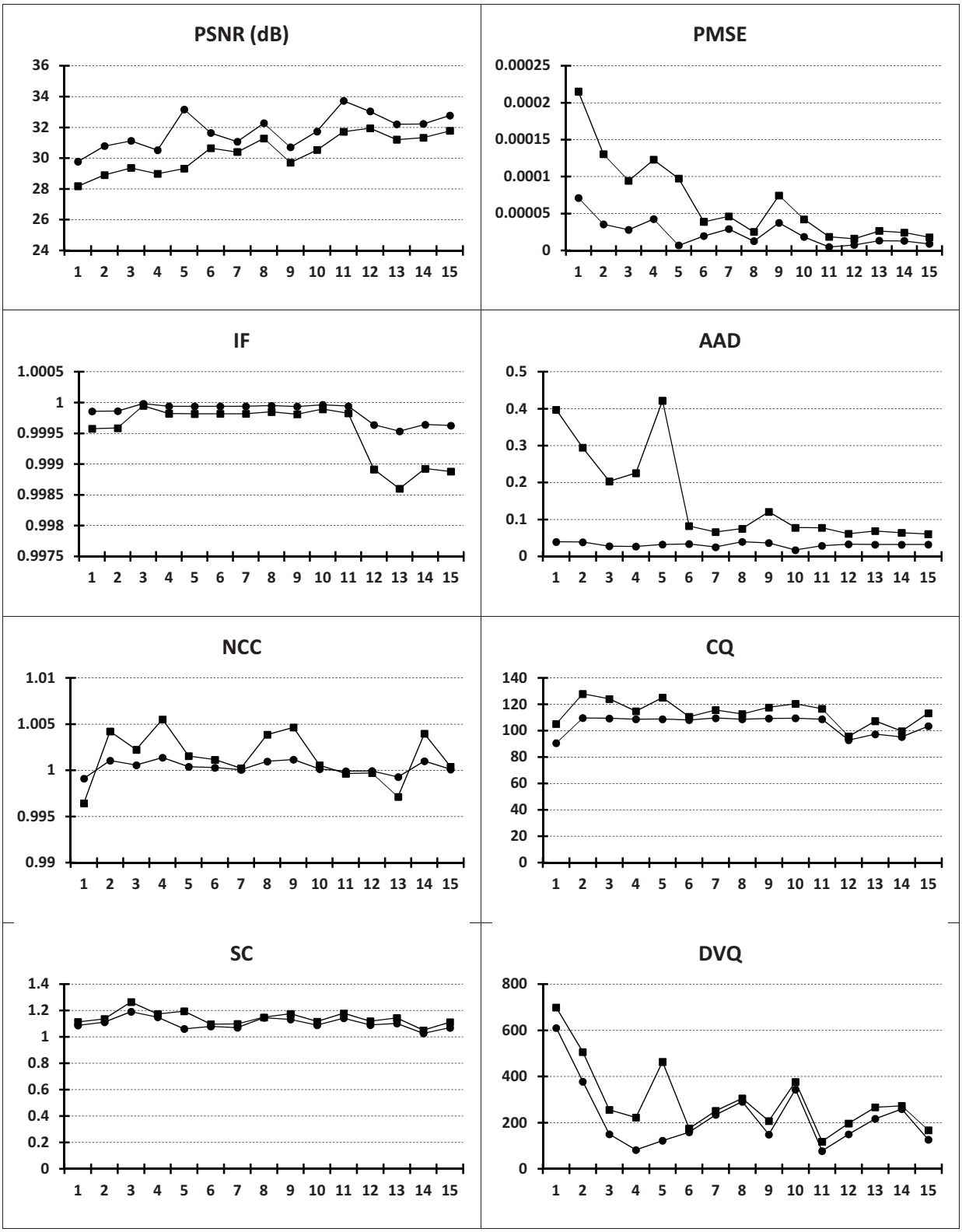


Figure.App.D-17 Effects of five symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

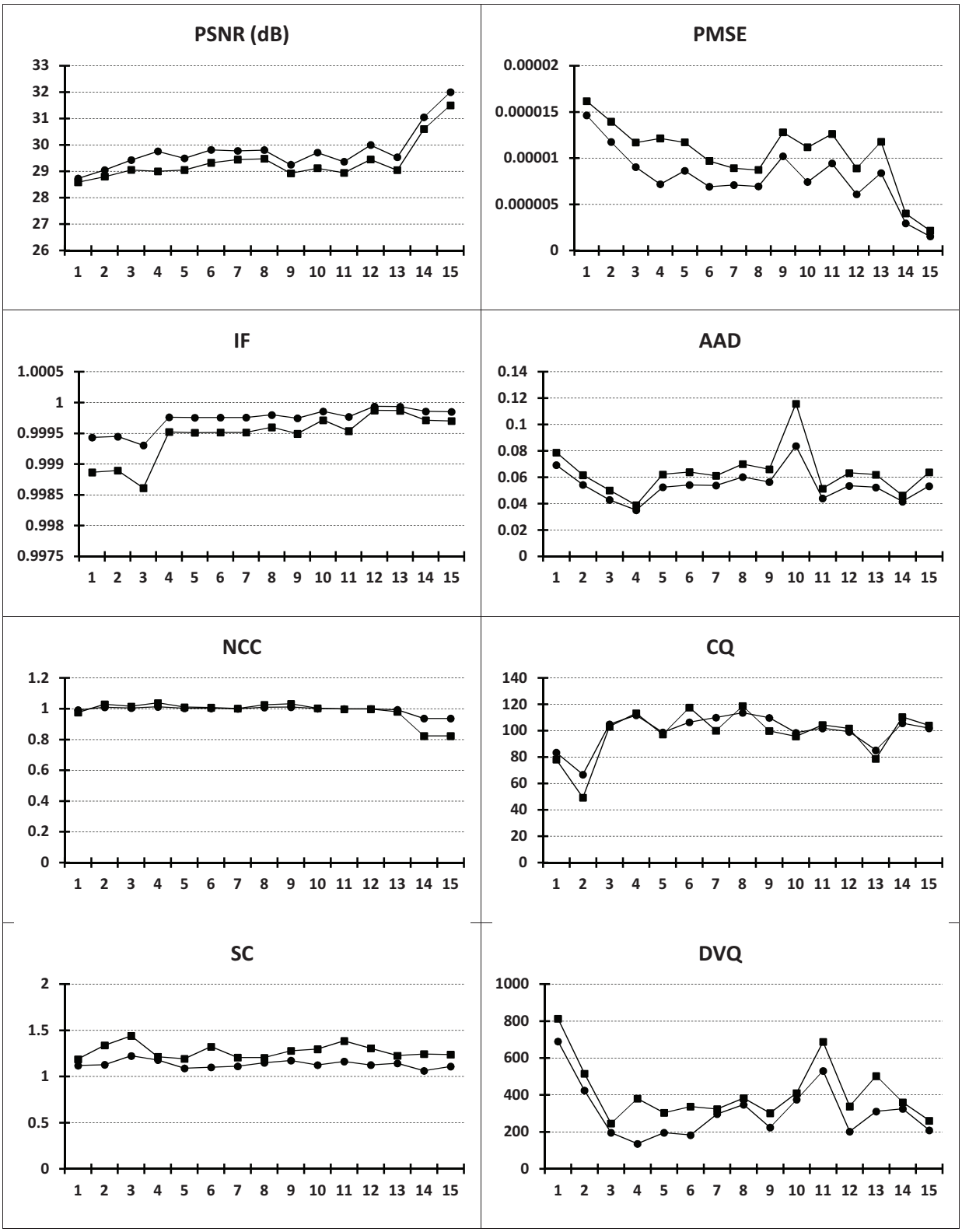


Figure.App.D-18 Effects of five symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 10 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

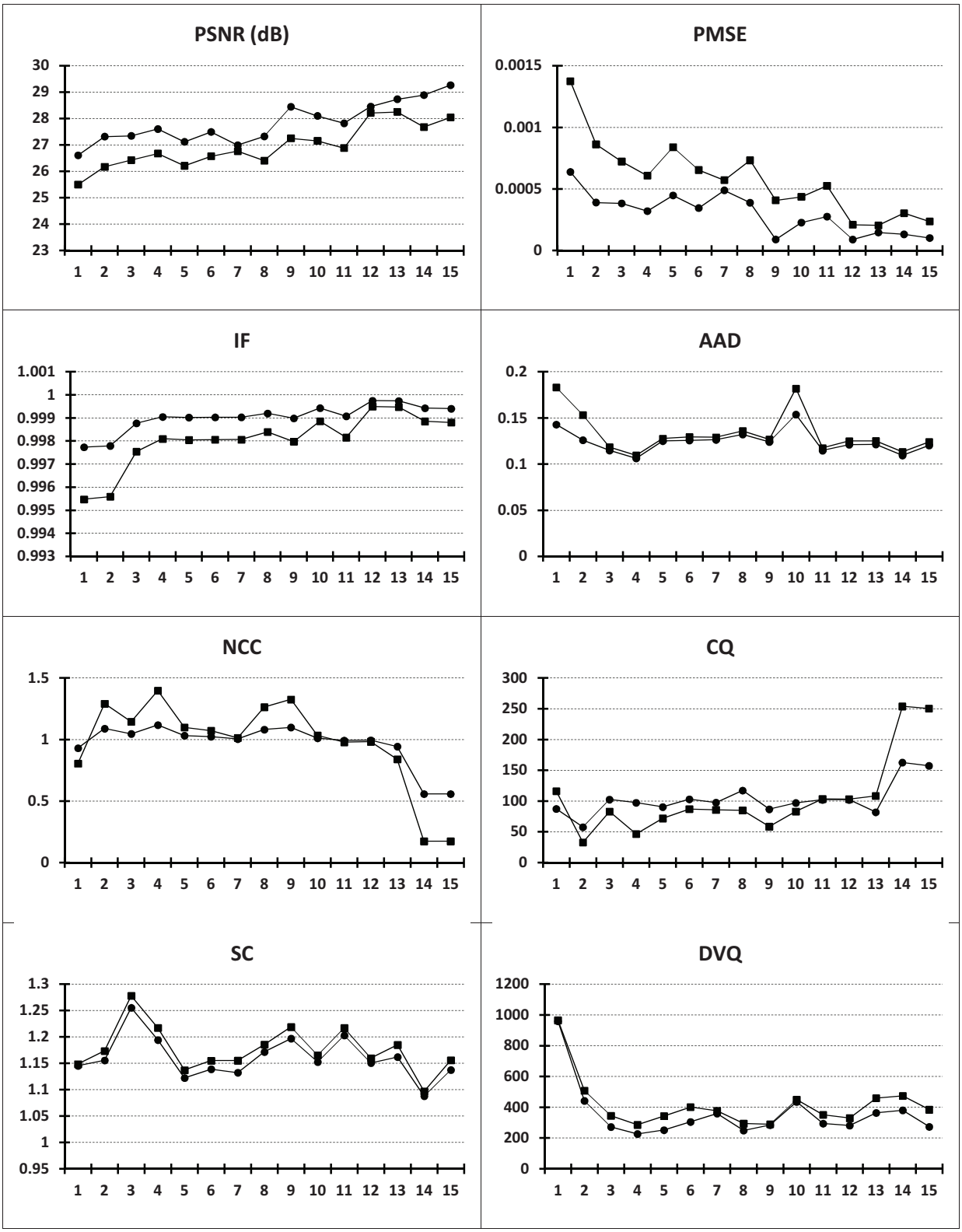


Figure.App.D-19 Effects of five symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 64 kbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

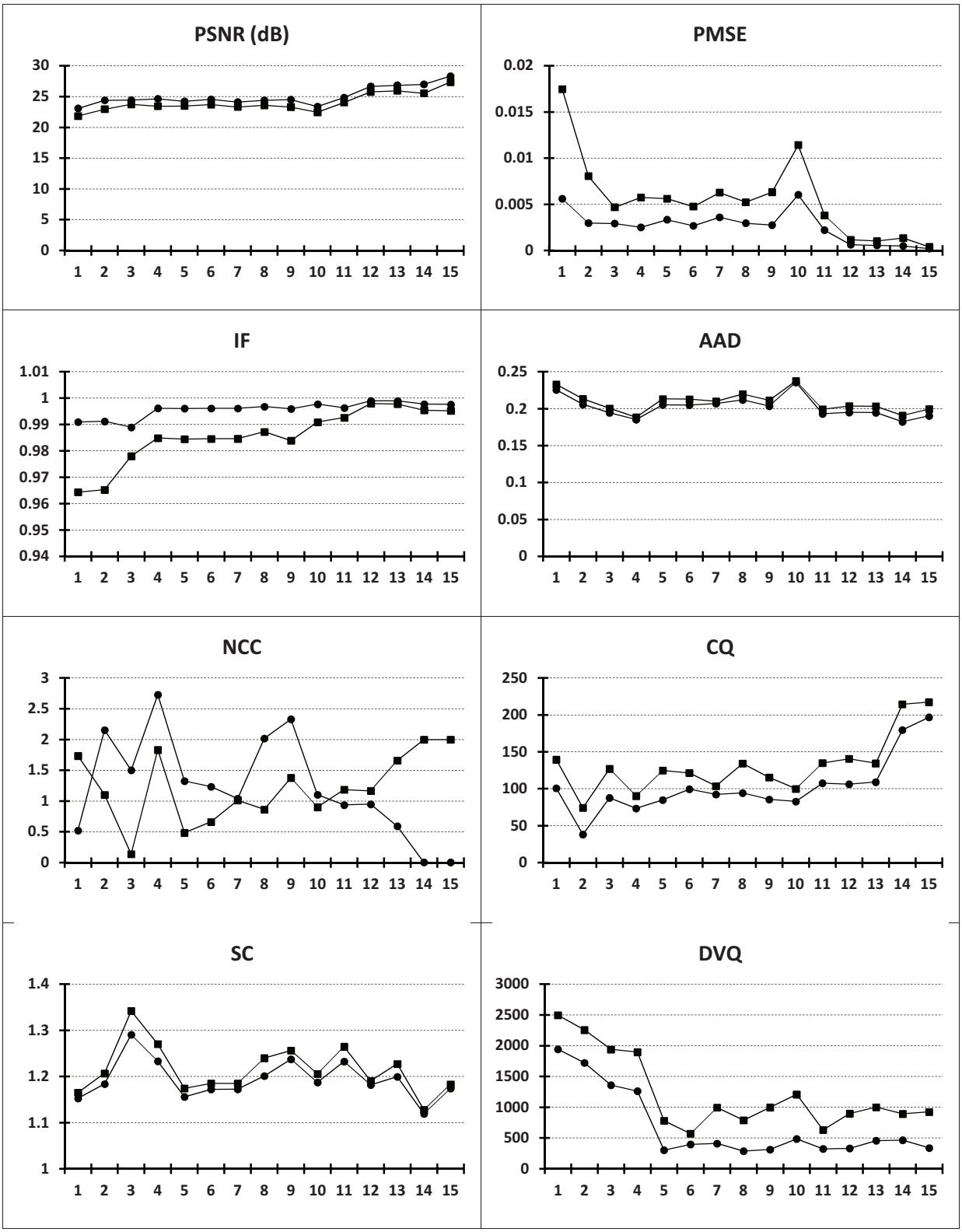


Figure.App.D-20 Effects of five symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ▲). Quality measures evaluated for video SD encoded at 64 kbps, and with 100 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

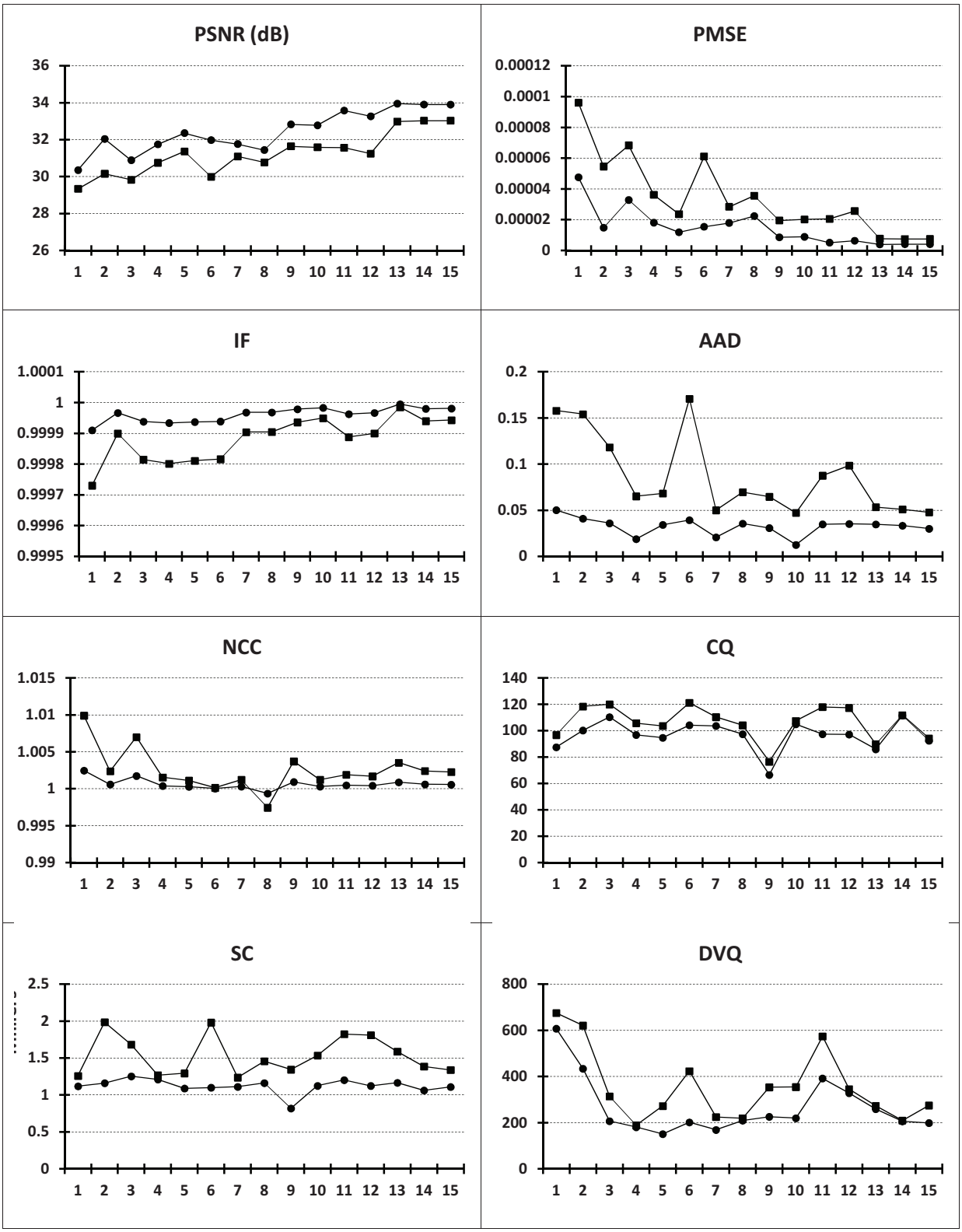


Figure.App.D-21 Effects of five symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

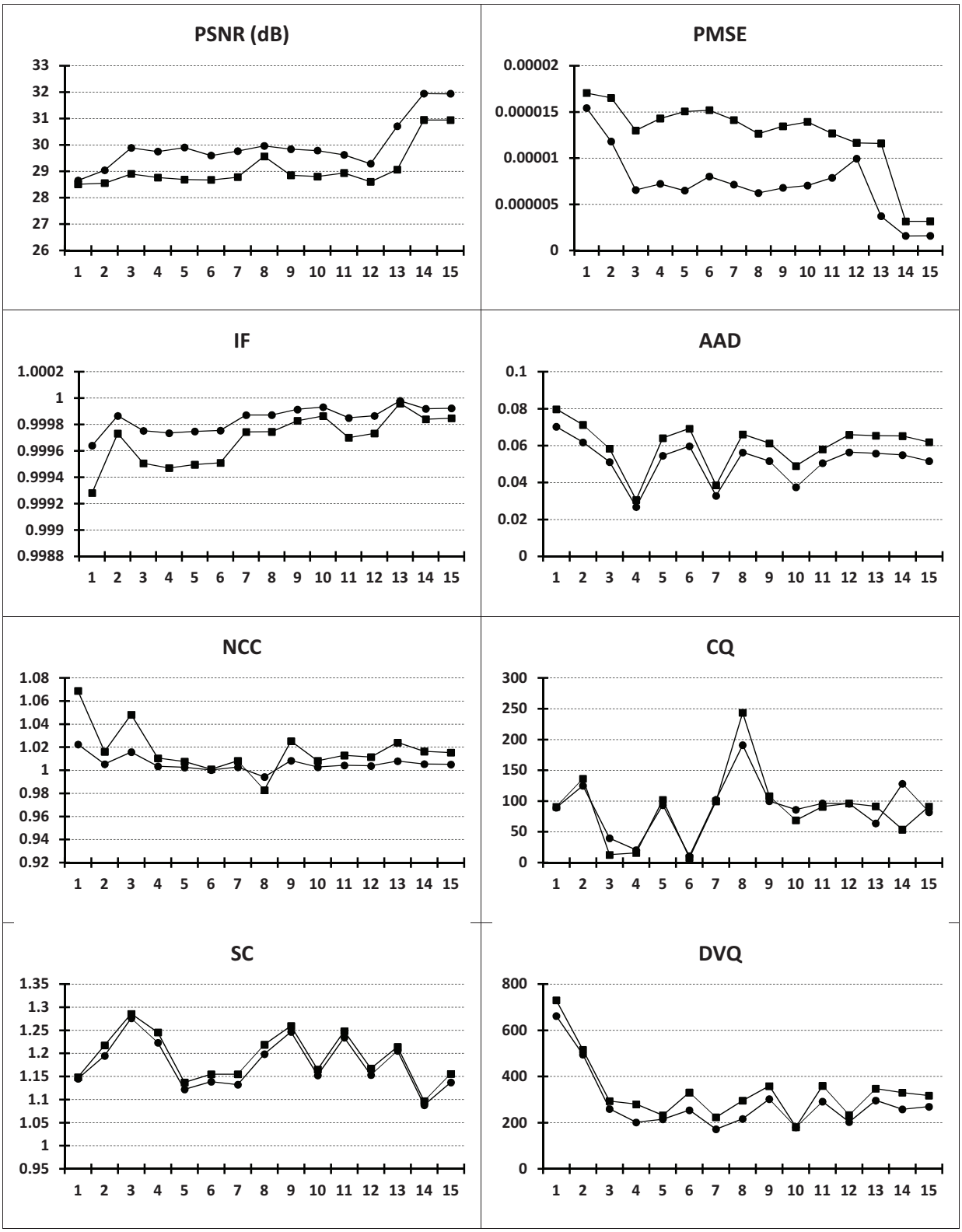


Figure.App.D-22 Effects of five symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 10 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

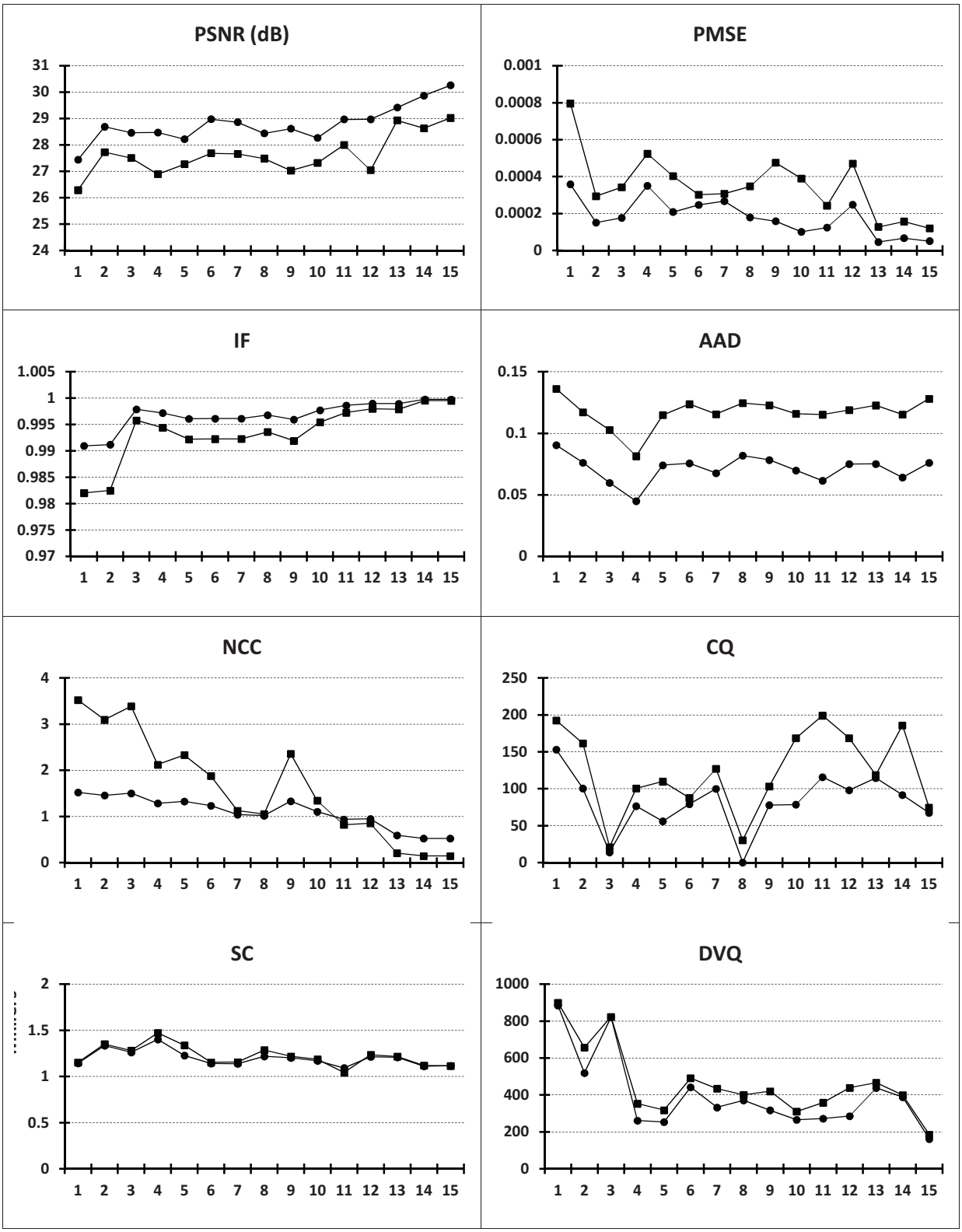


Figure.App.D-23 Effects of five symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

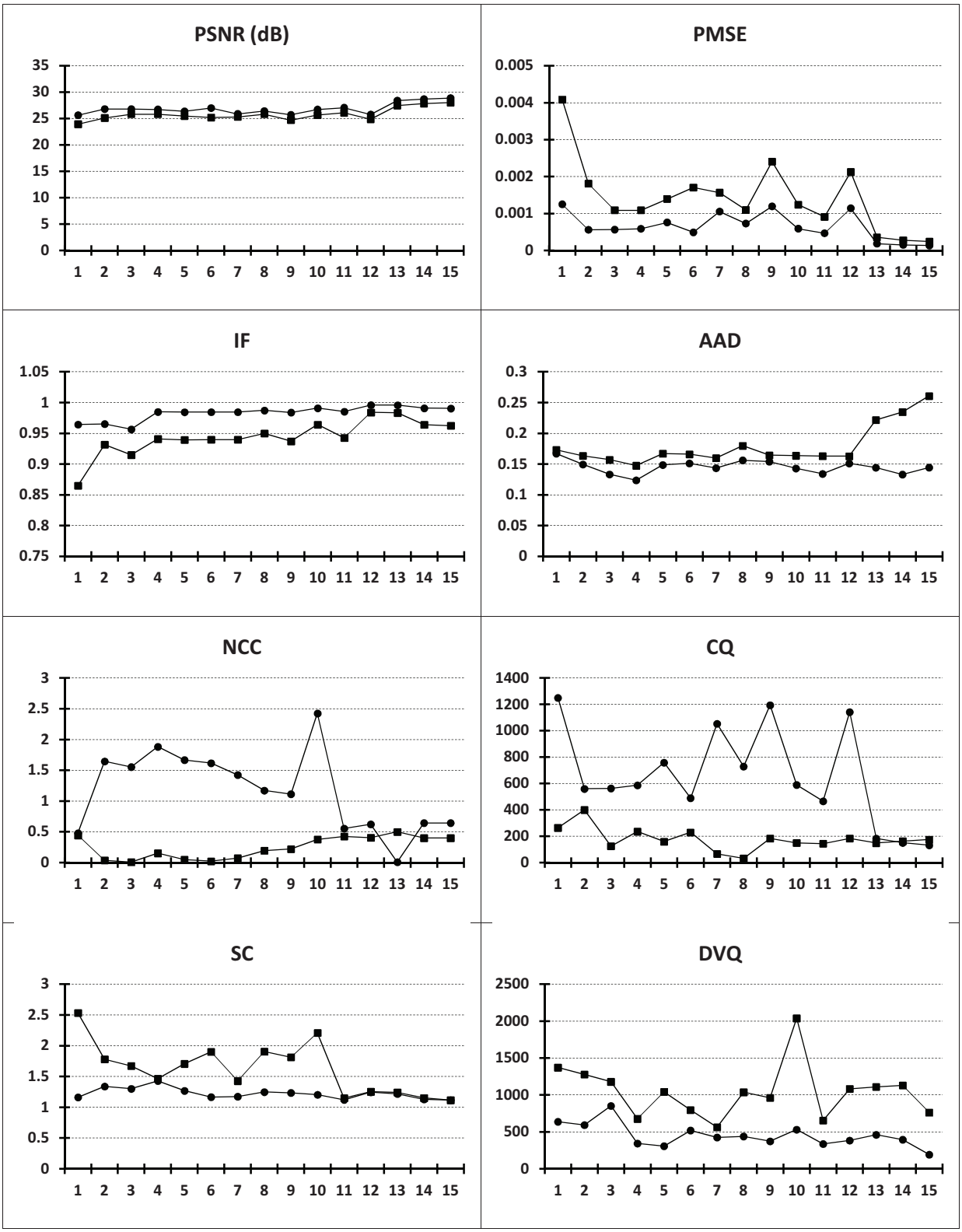


Figure.App.D-24 Effects of five symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video SD encoded at 256 kbps, and with 100 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

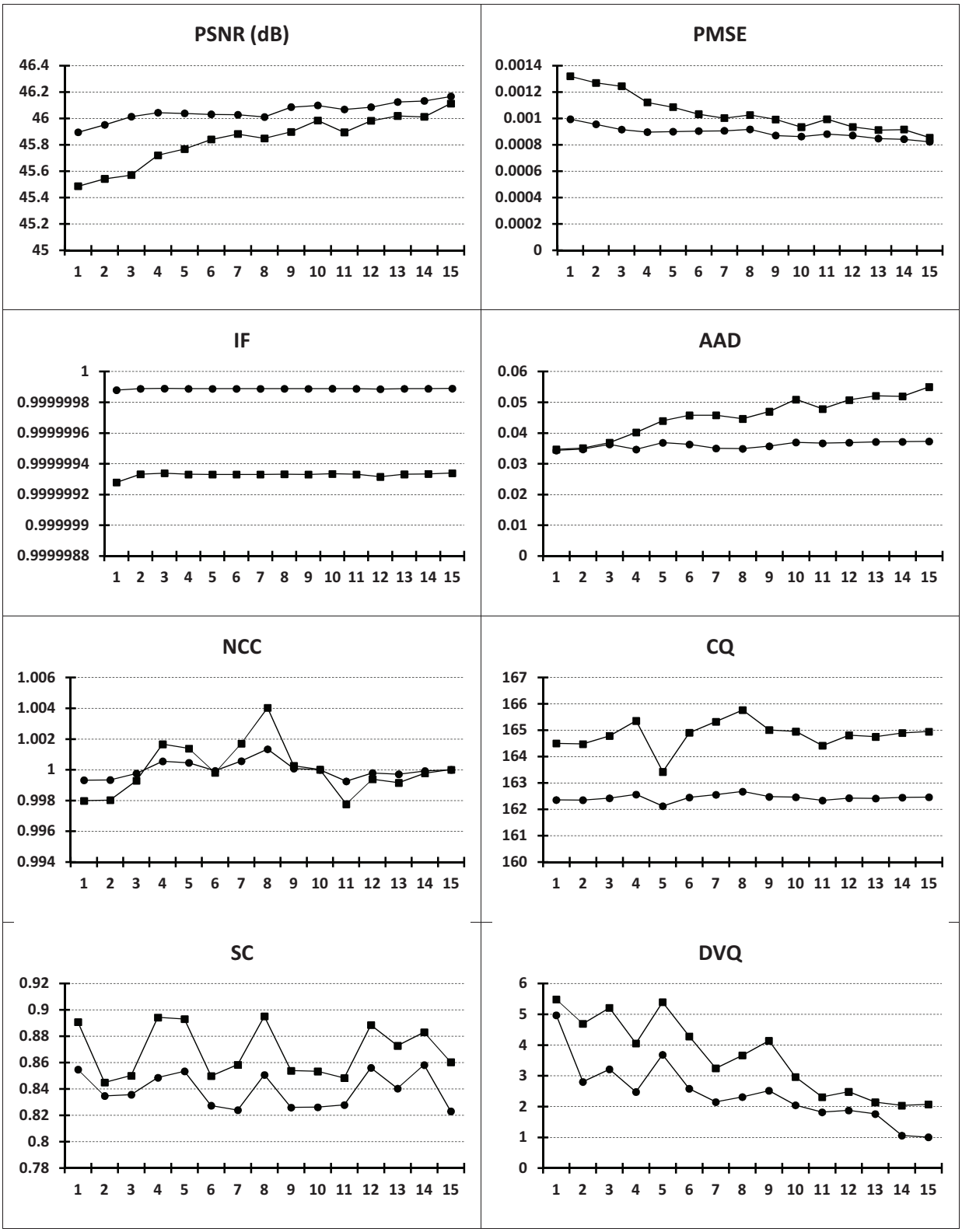


Figure.App.D-25 Effects of {-2,2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

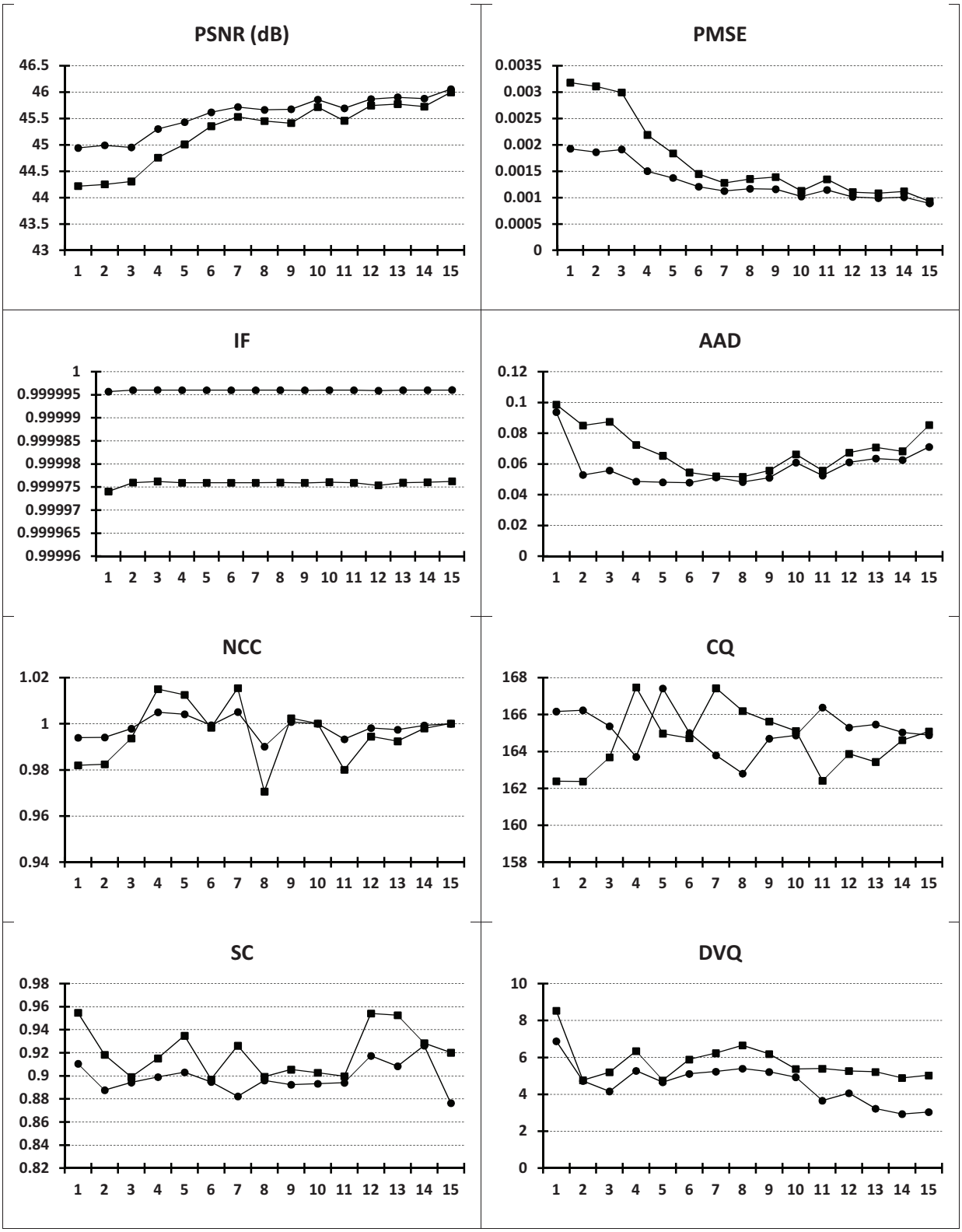


Figure.App.D-26 Effects of {-2,2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

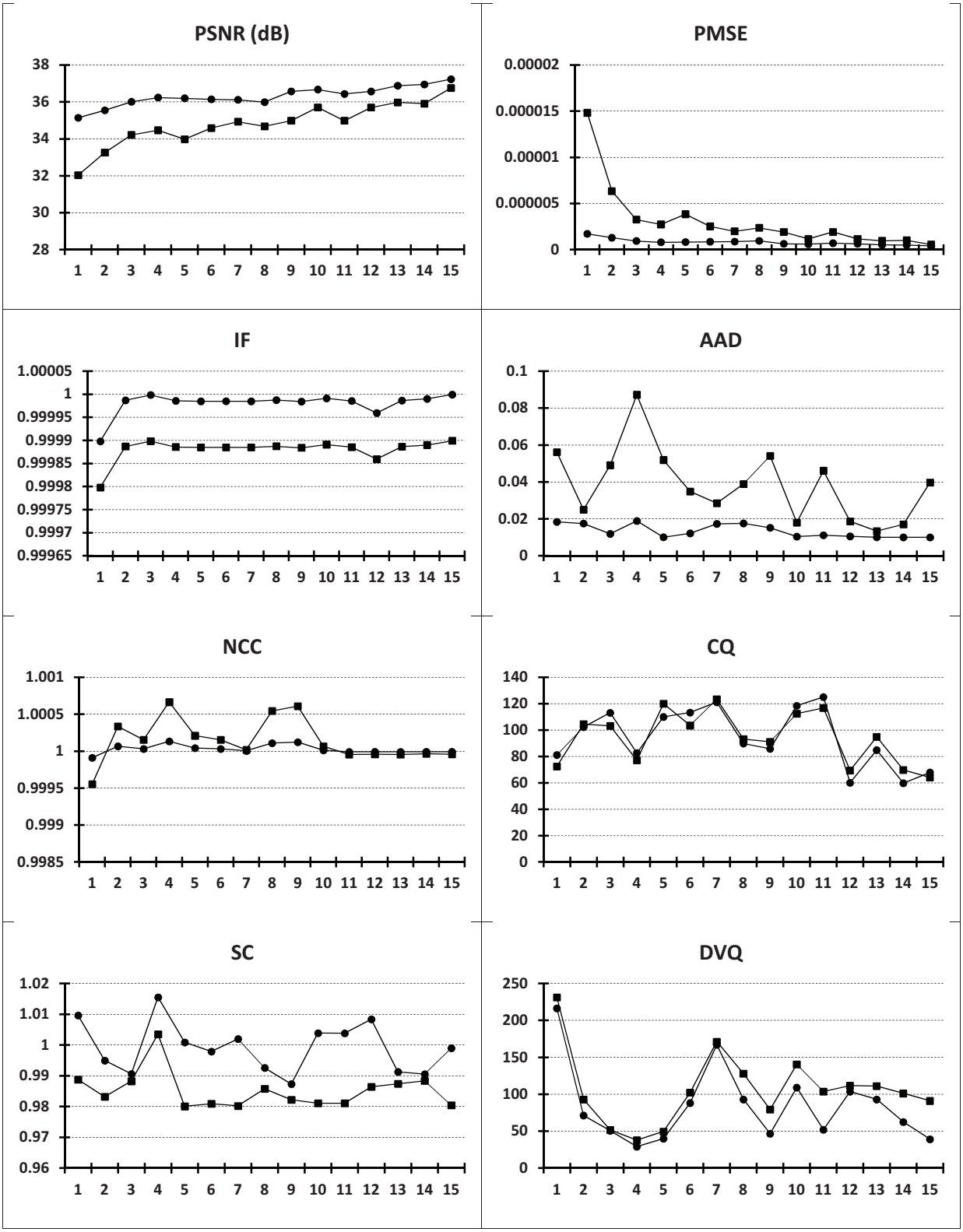


Figure.App.D-27 Effects of {-2,2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 500 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

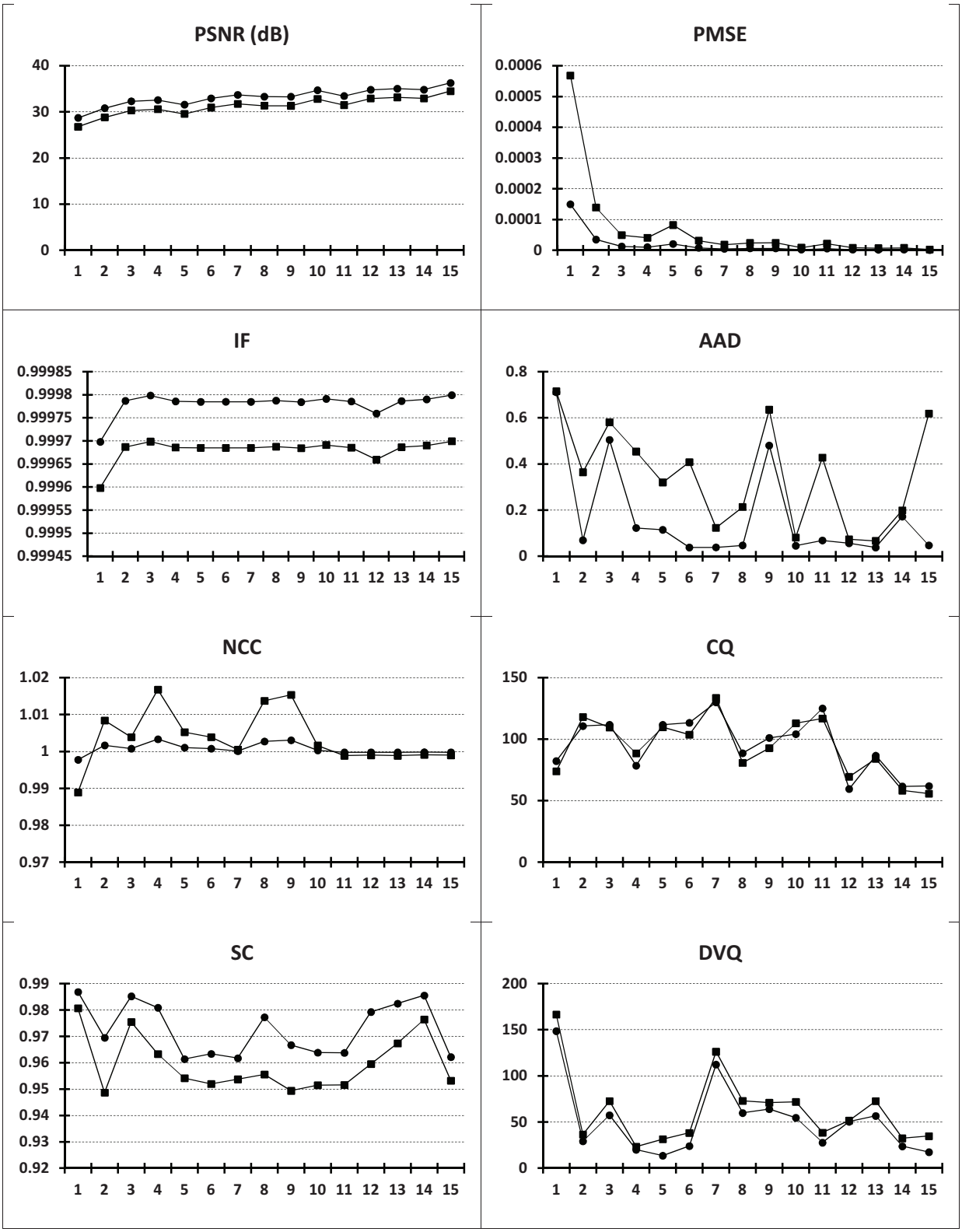


Figure.App.D-28 Effects of {-2,2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 1000 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

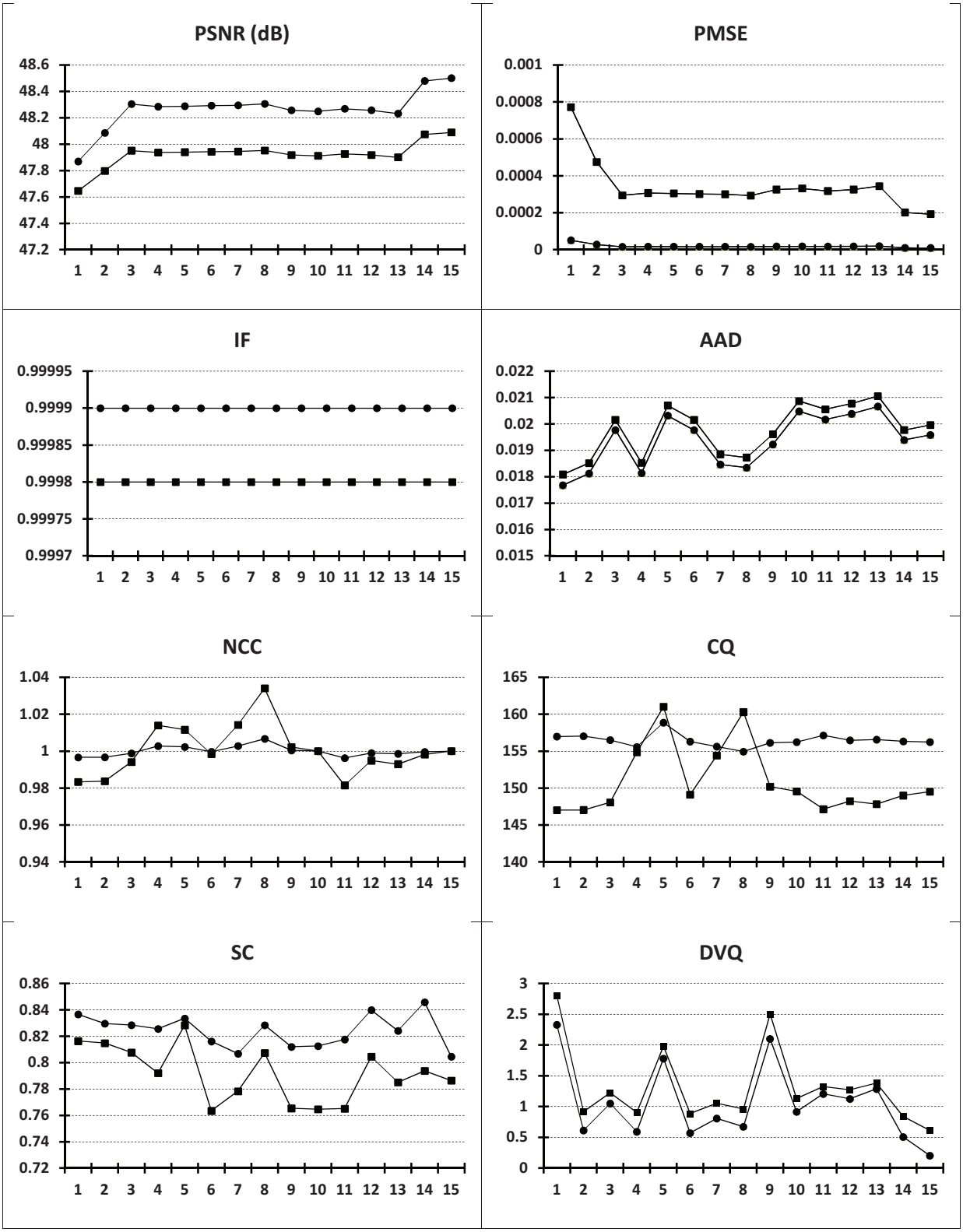


Figure.App.D-29 Effects of {-2,2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

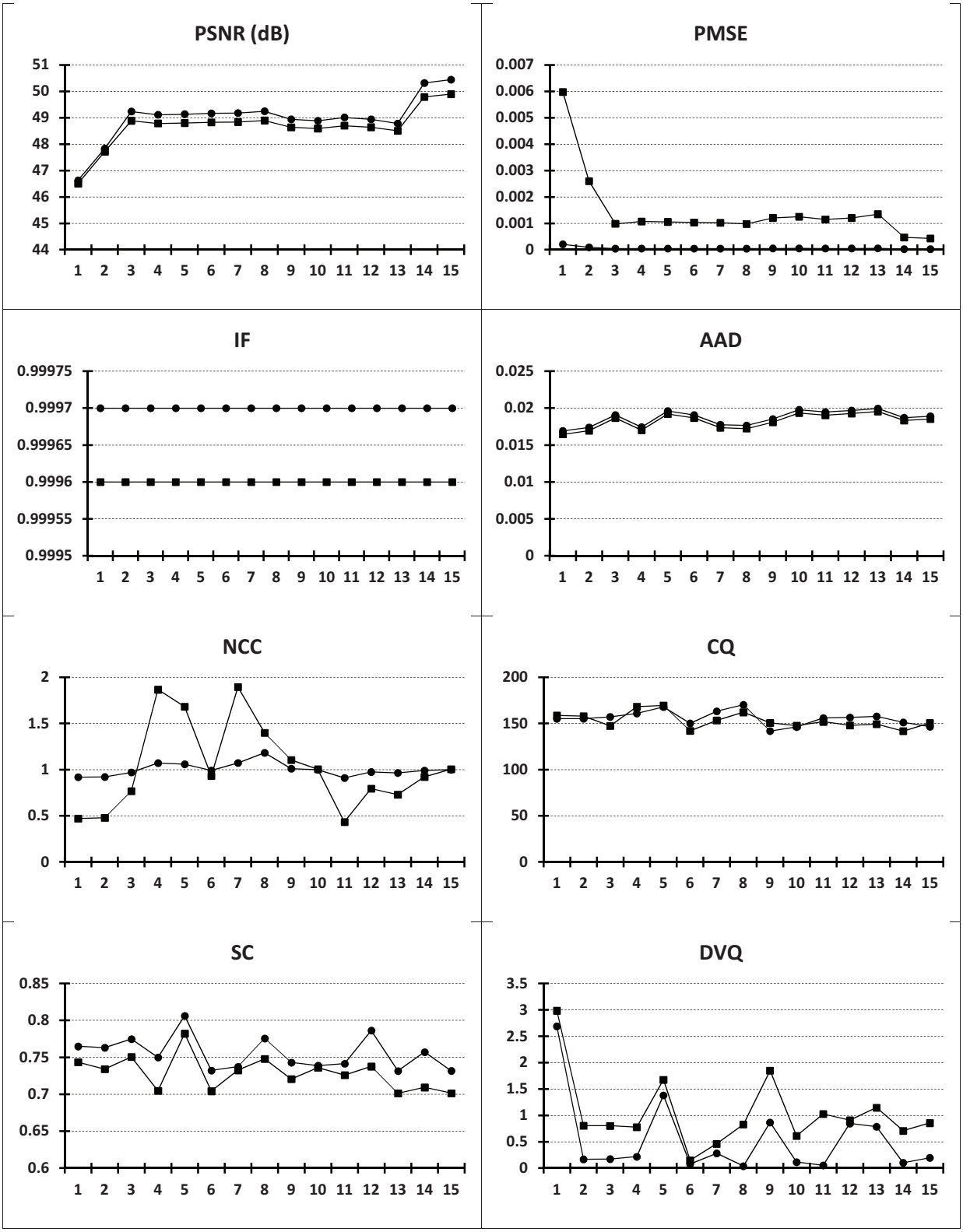


Figure.App.D-30 Effects of {-2,2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

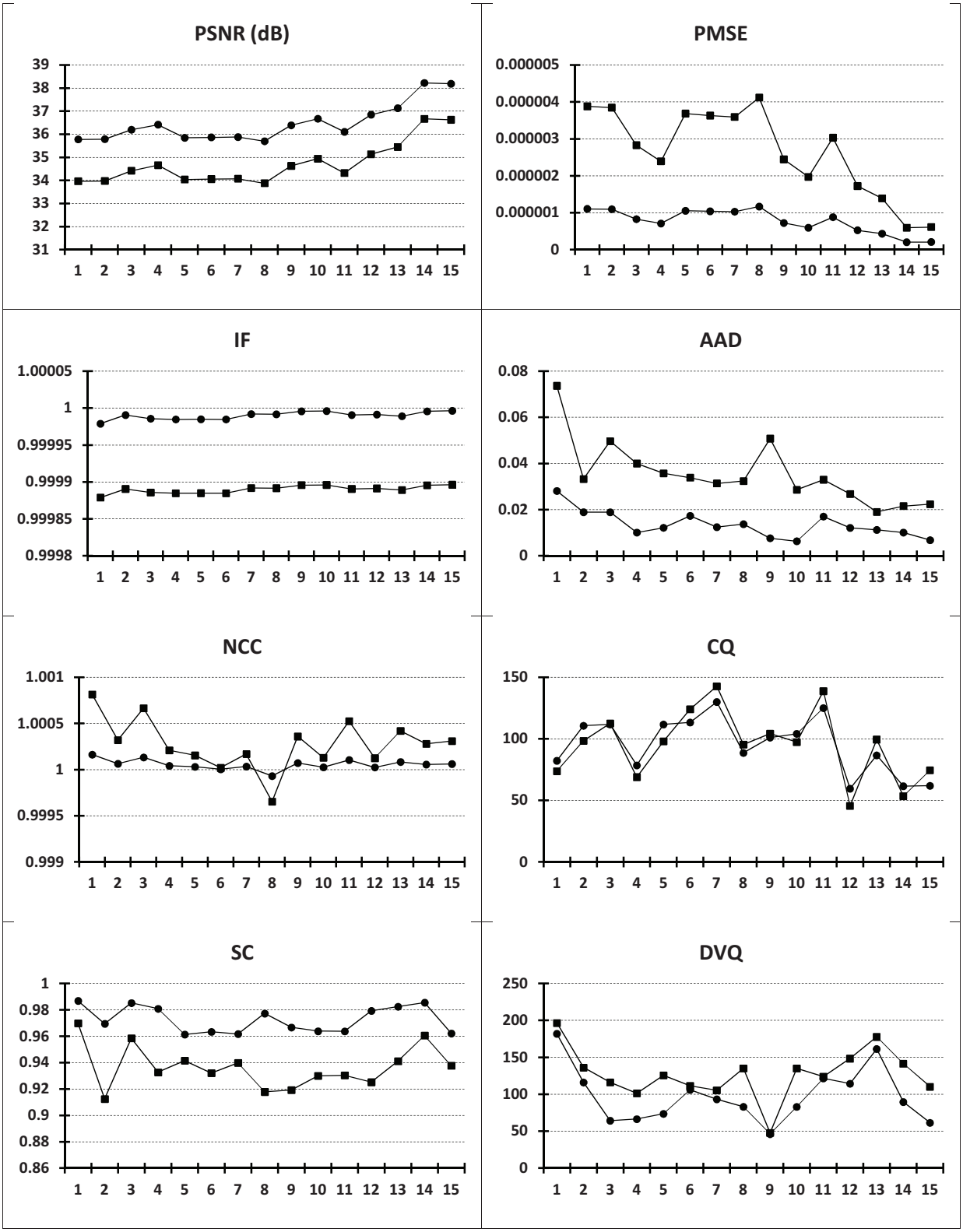


Figure.App.D-31 Effects of {-2,2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 500 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

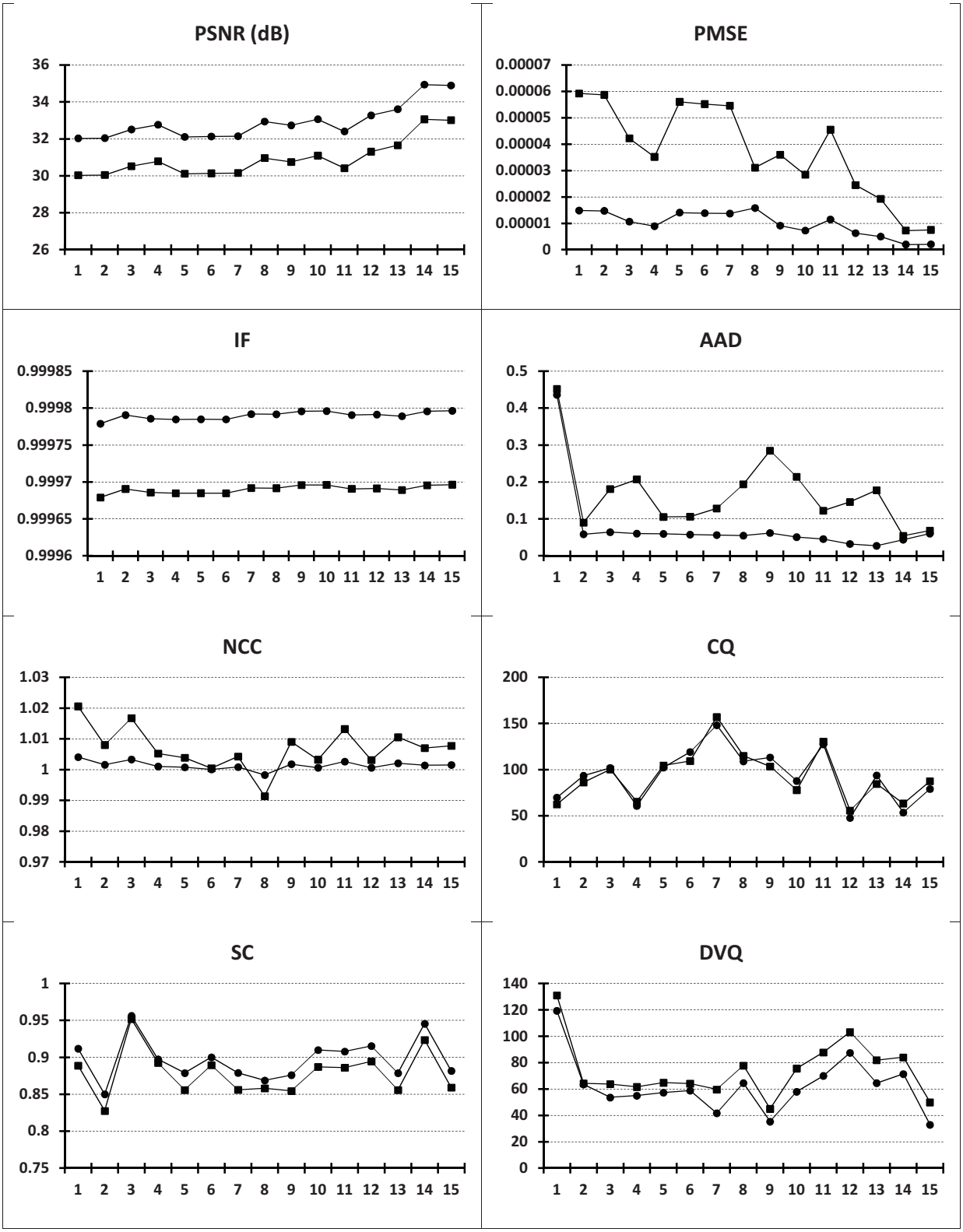


Figure.App.D-32 Effects of {-2,2} additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 1000 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

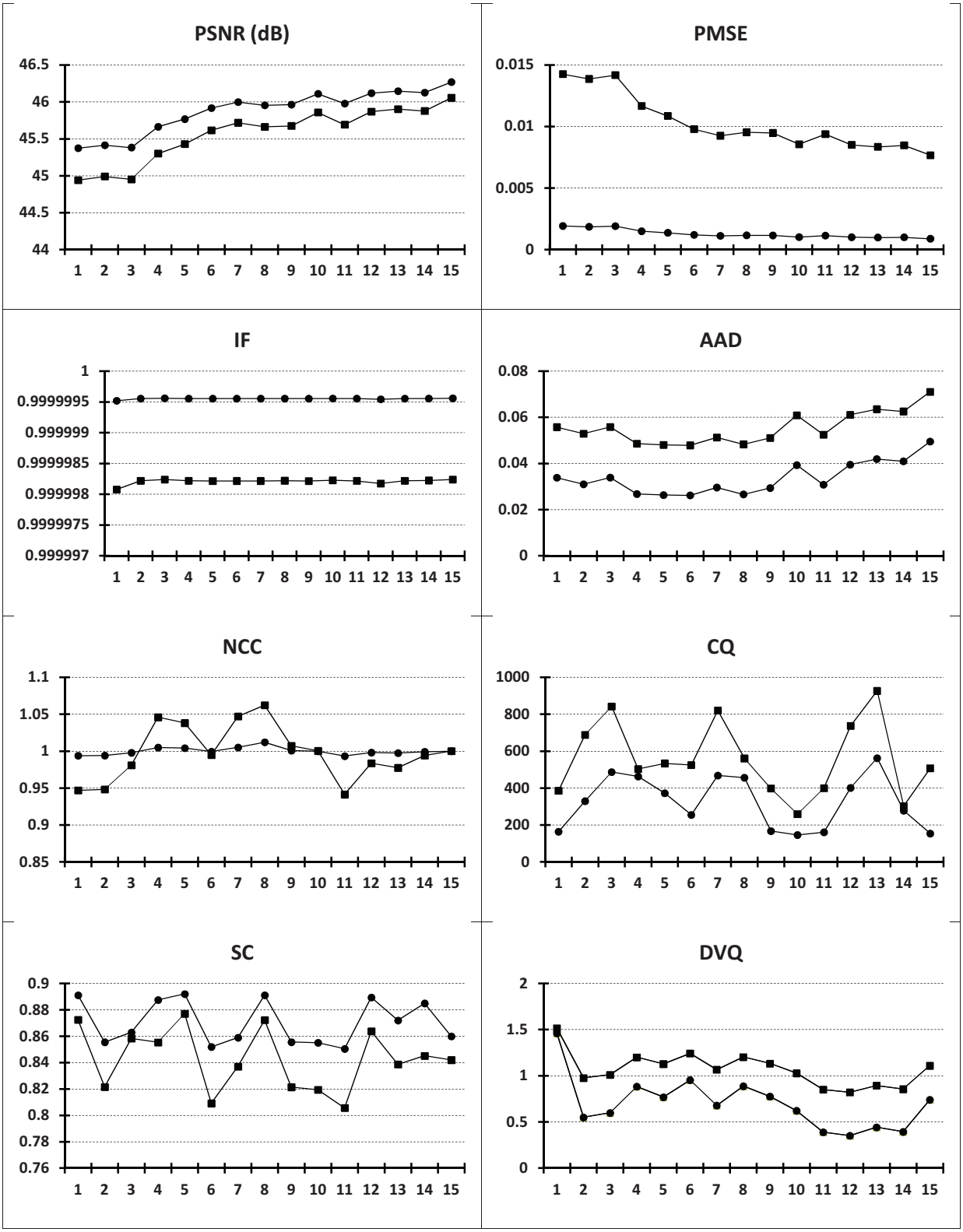


Figure.App.D-33 Effects of eight symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

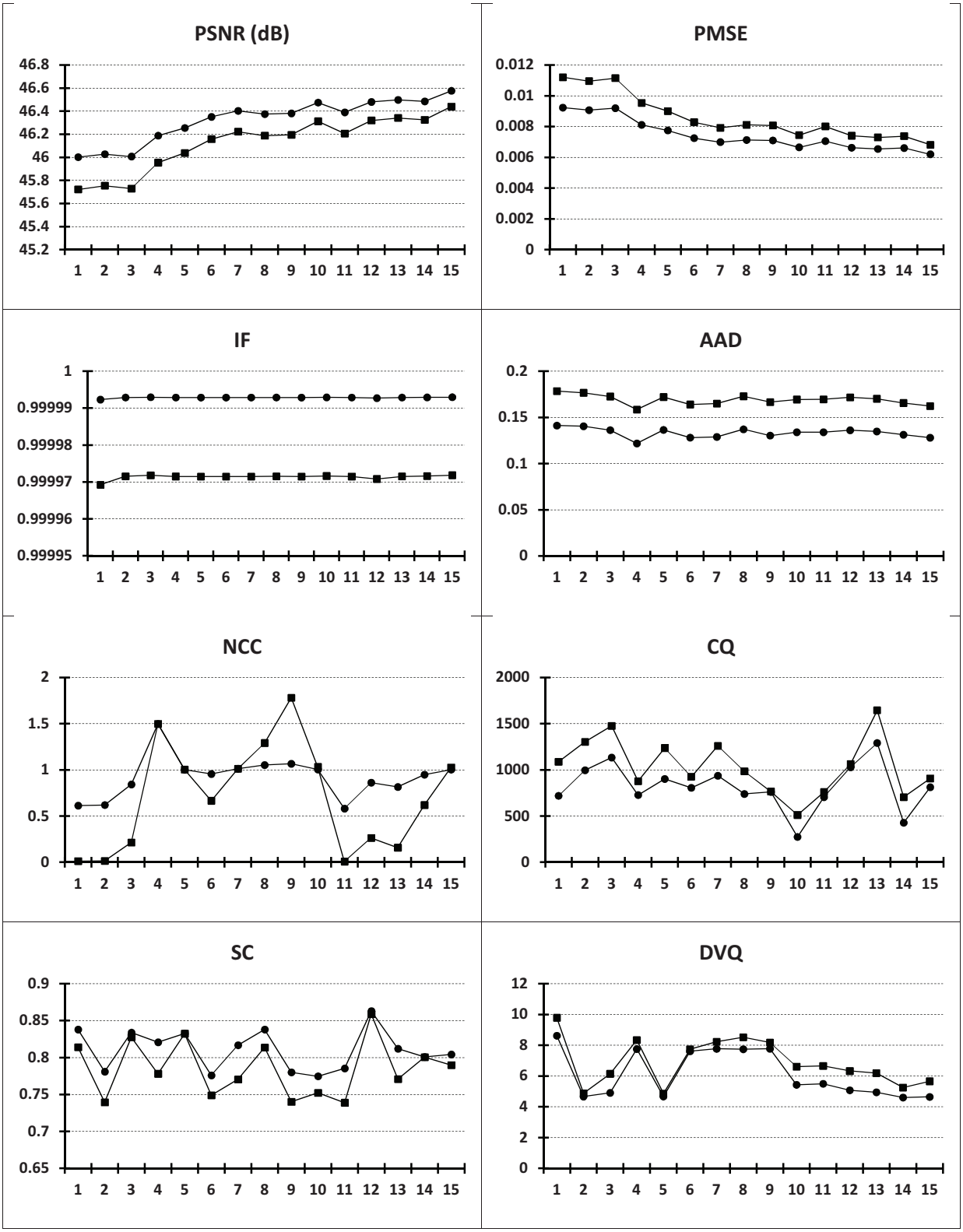


Figure.App.D-34 Effects of eight symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

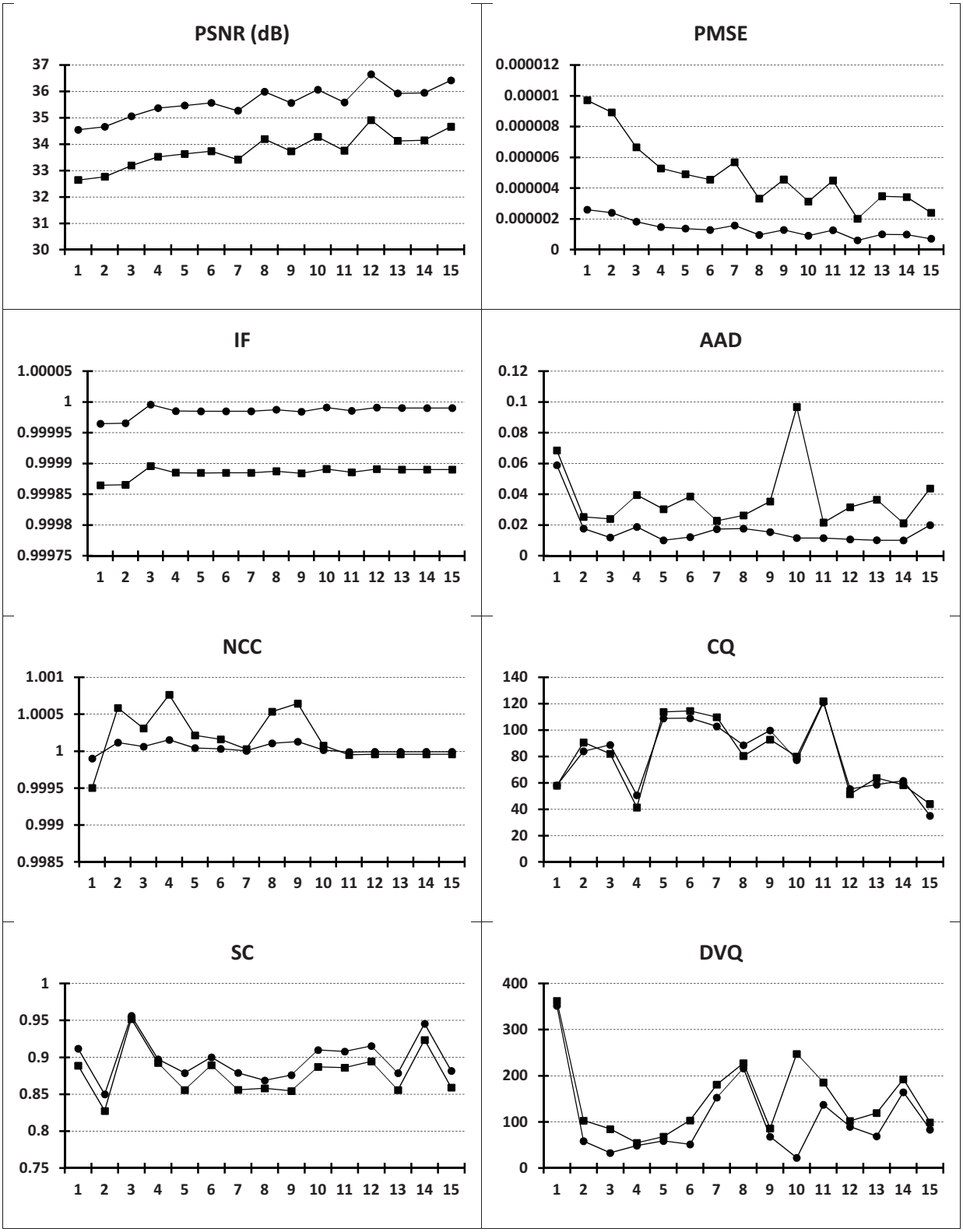


Figure.App.D-35 Effects of eight symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 500 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

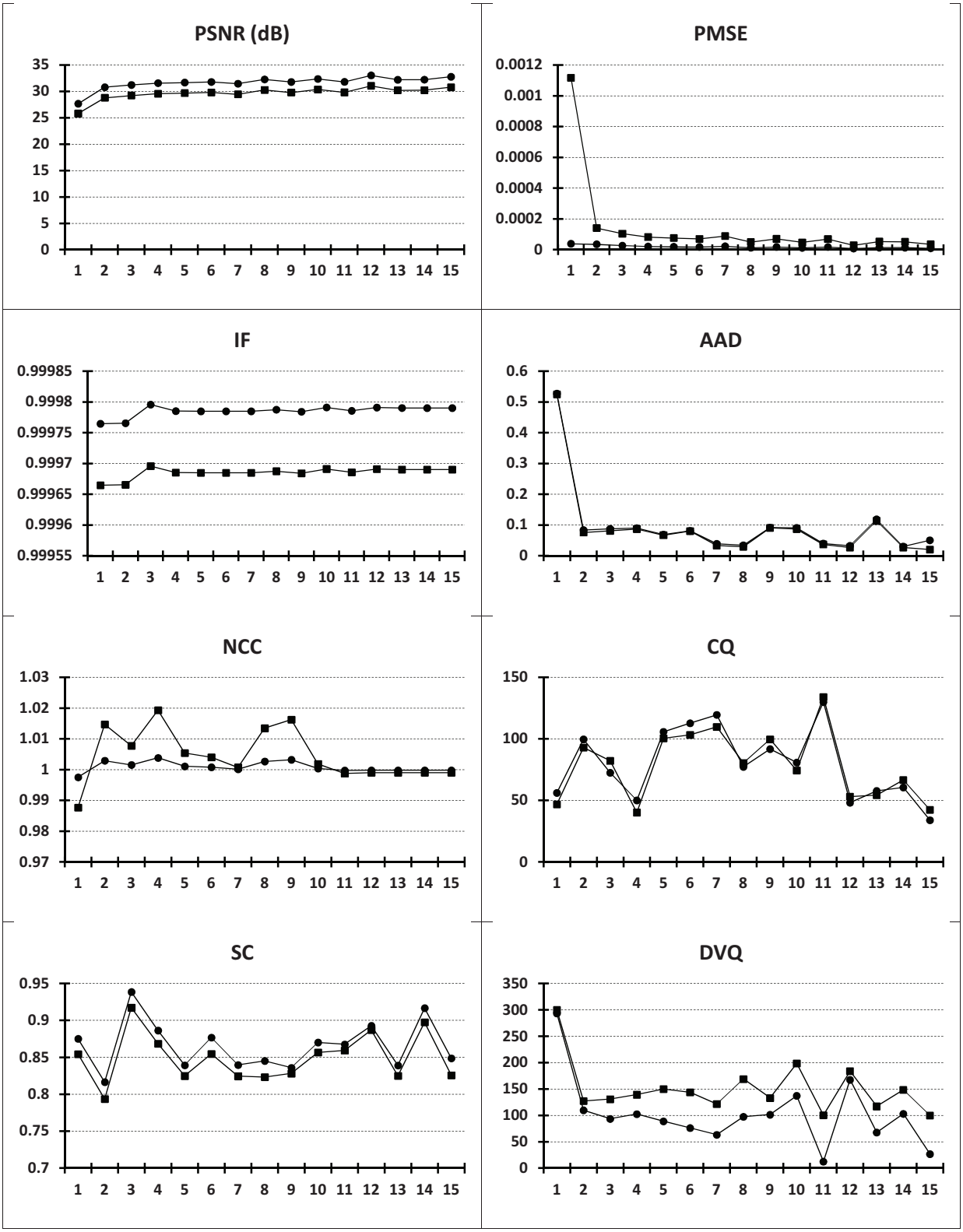


Figure.App.D-36 Effects of eight symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 1000 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

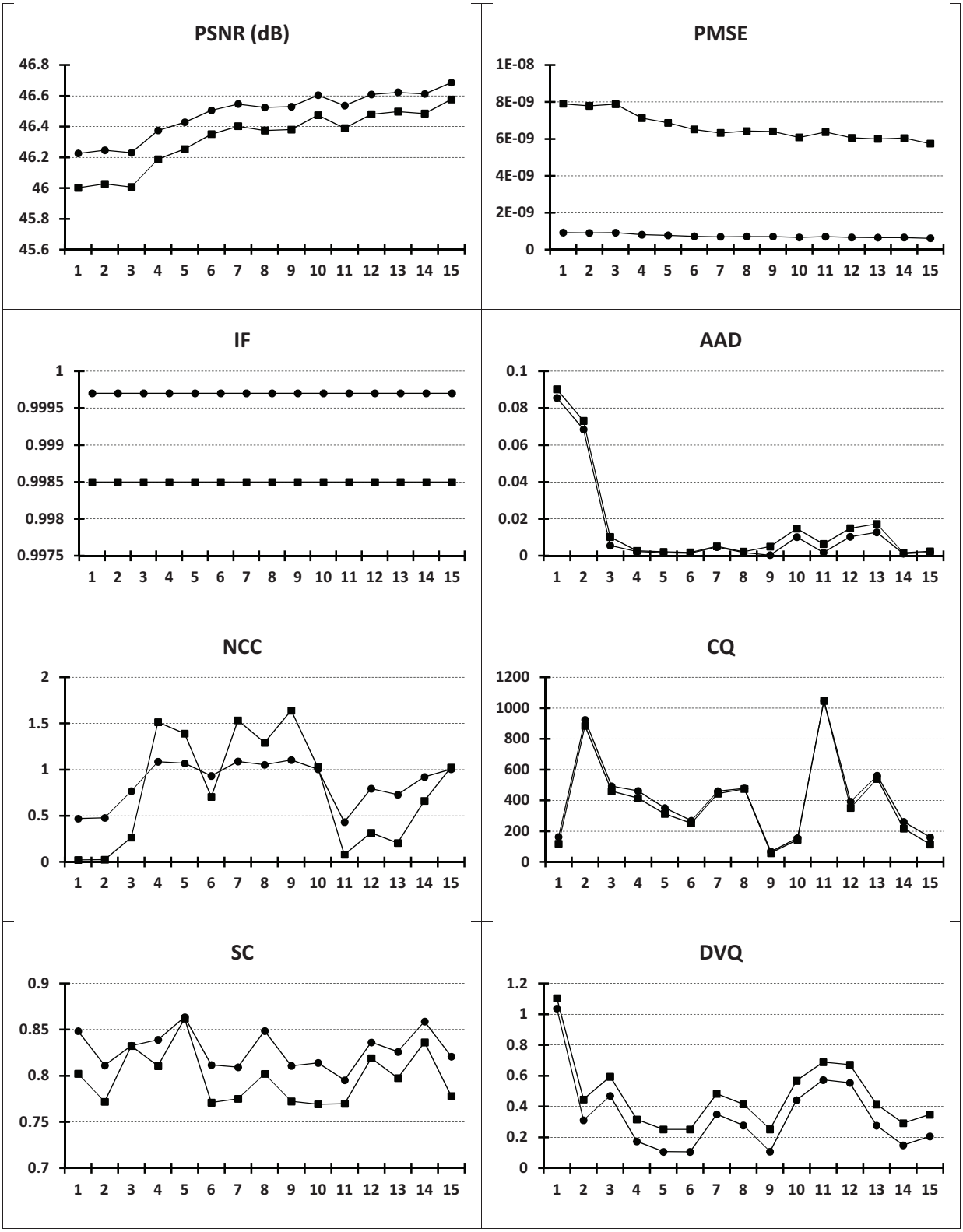


Figure.App.D-37 Effects of eight symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

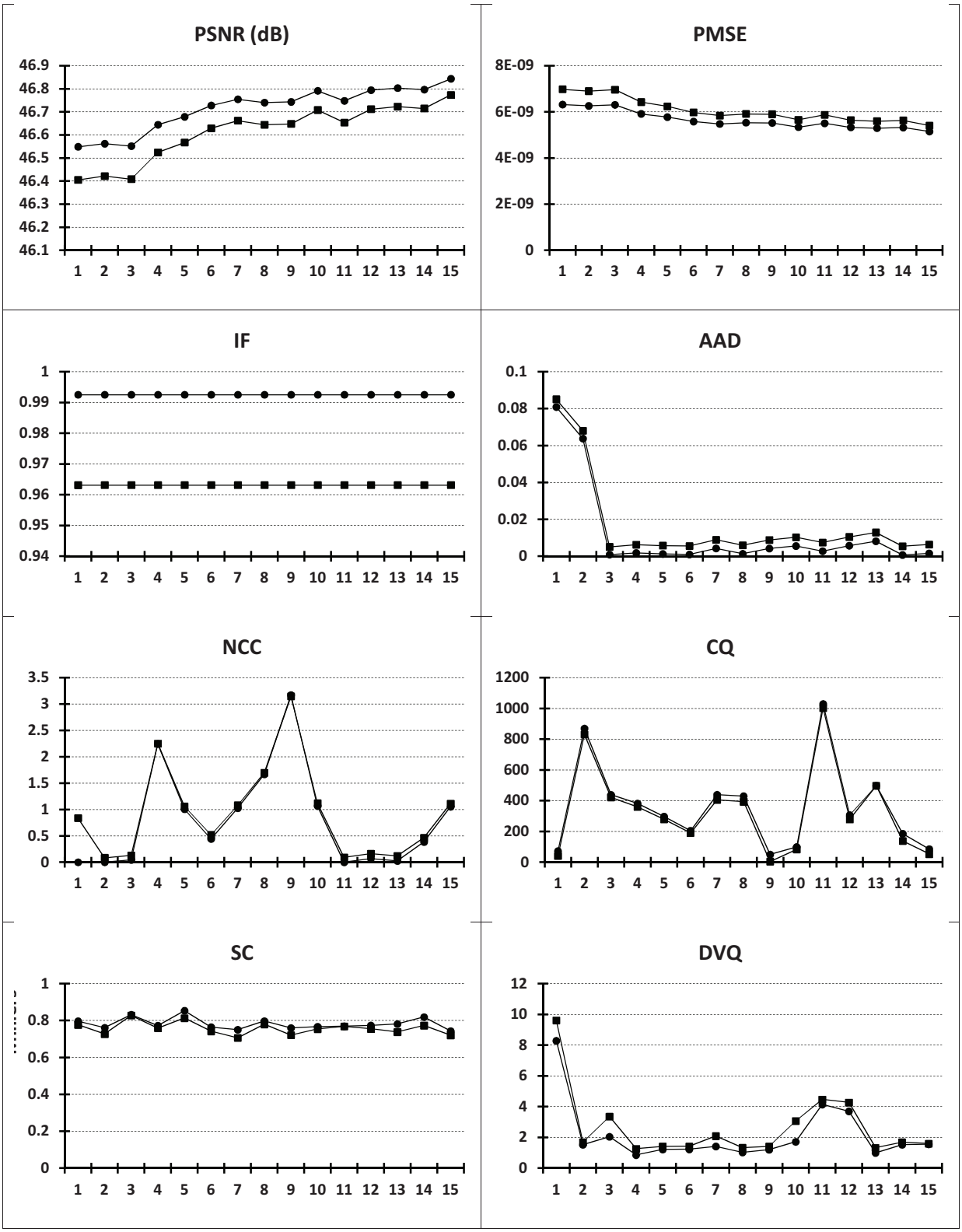


Figure.App.D-38 Effects of eight symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

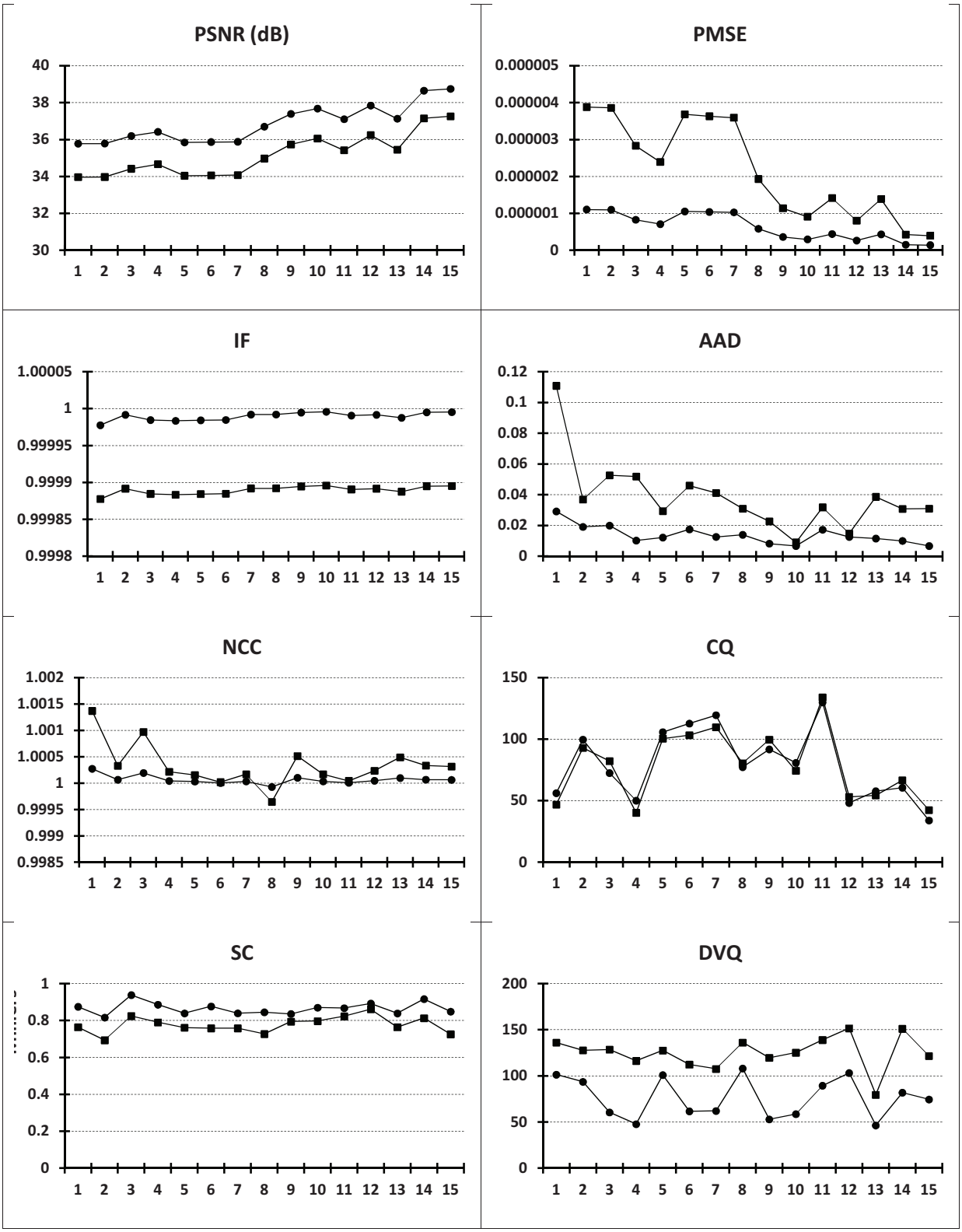


Figure.App.D-39 Effects of eight symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 500 Mbps, and with 100 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

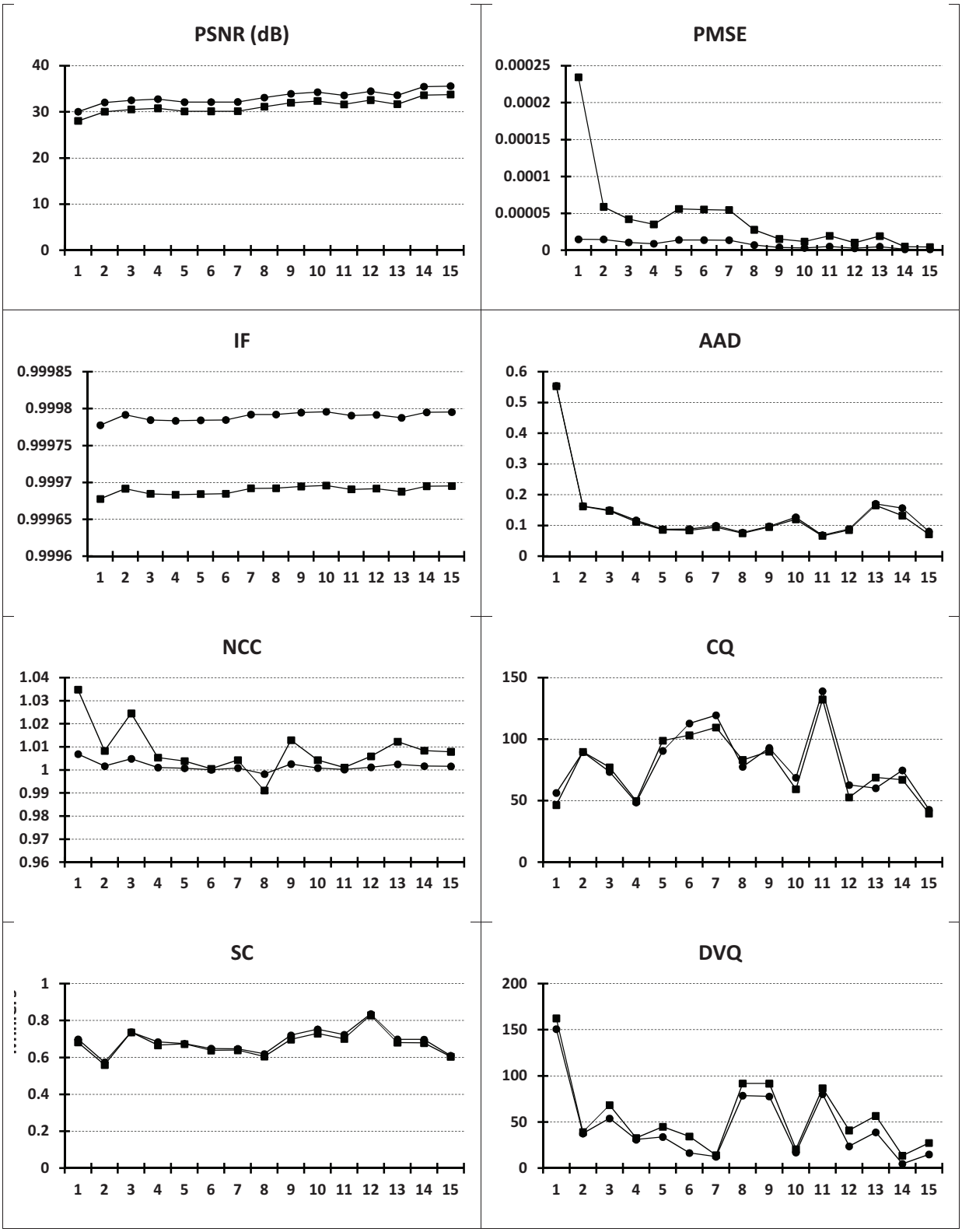


Figure.App.D-40 Effects of eight symbols additive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 1000 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

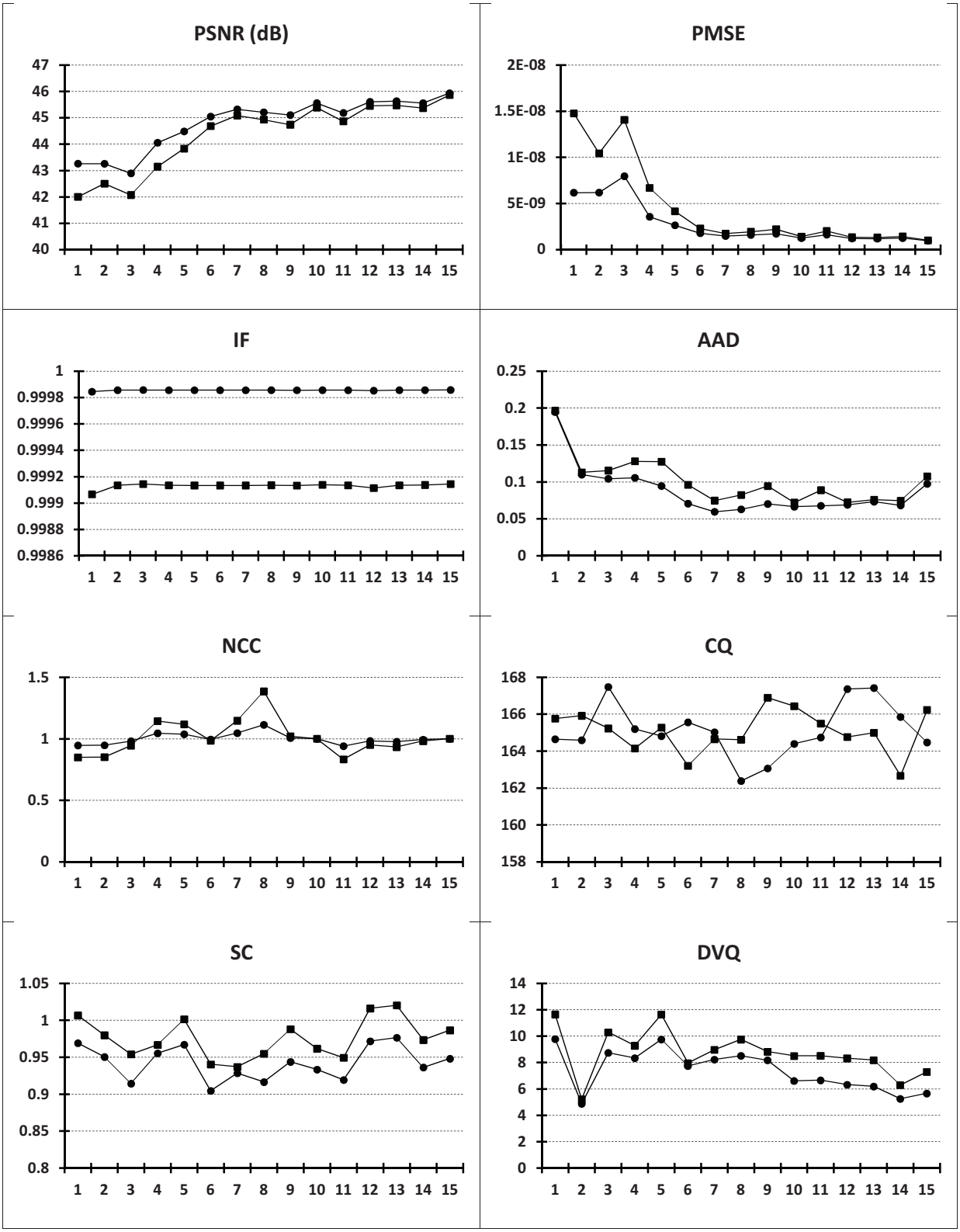


Figure.App.D-41 Effects of {-2,2} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

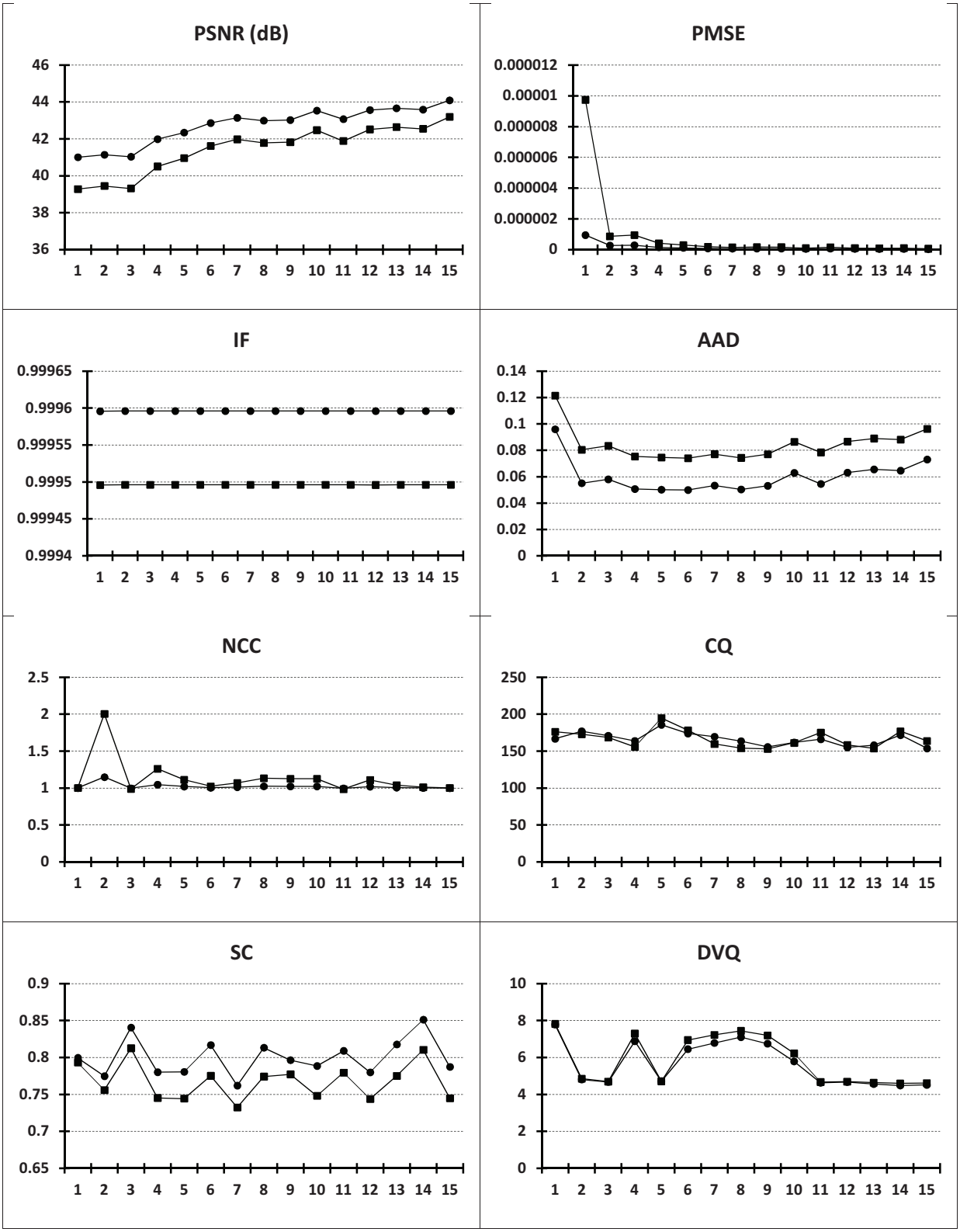


Figure.App.D-42 Effects of {-2,2} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

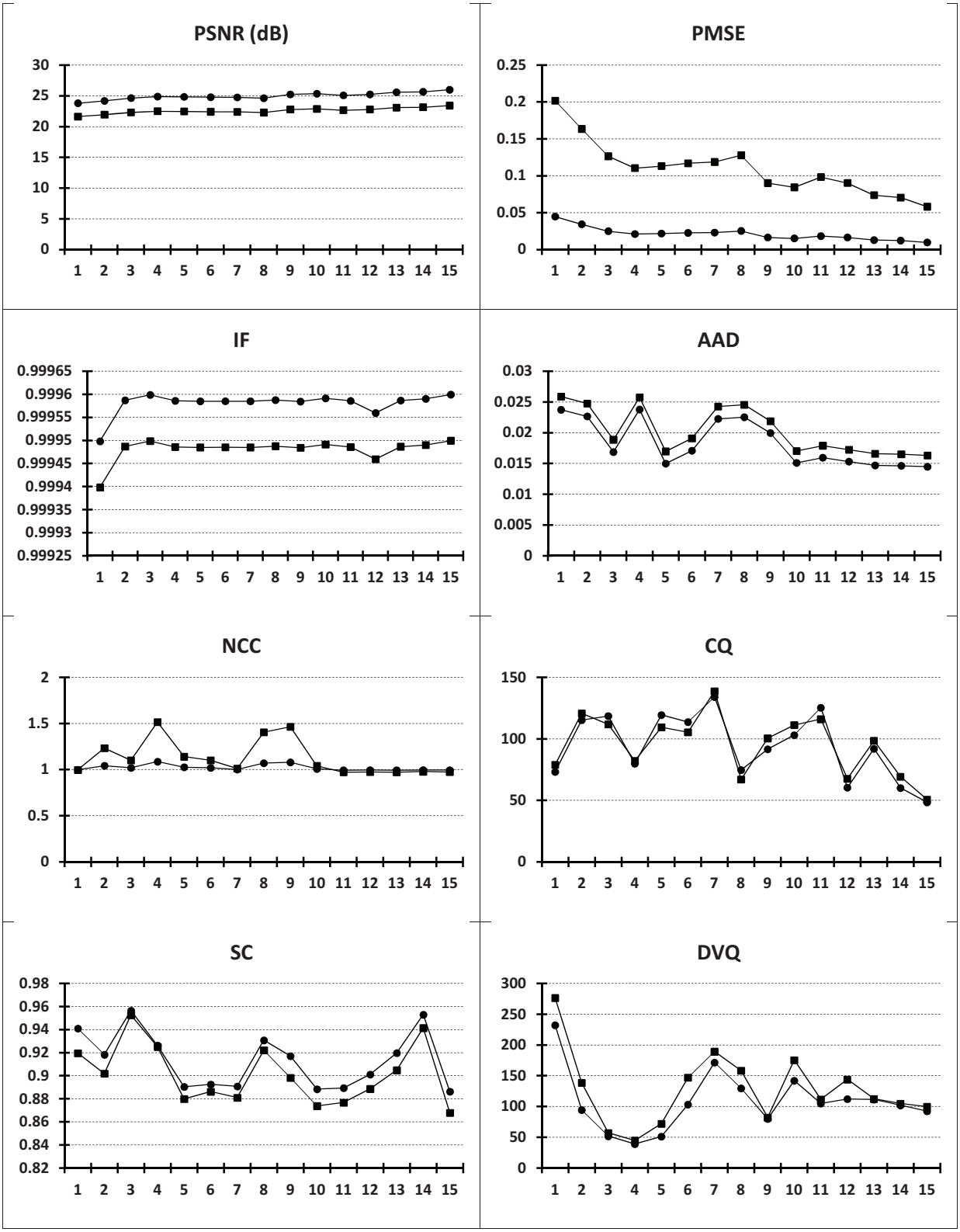


Figure.App.D-43 Effects of {-2,2} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 500 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

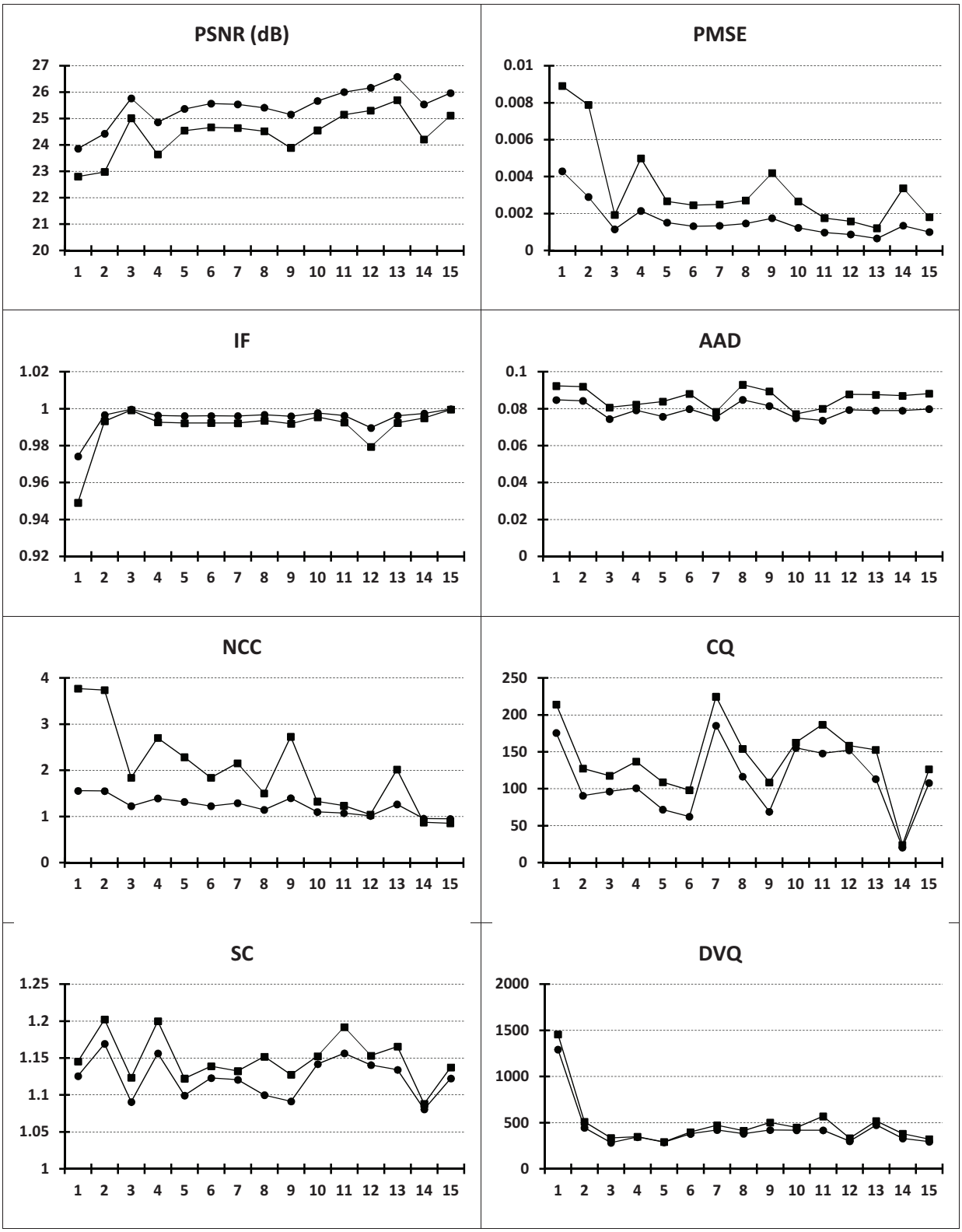


Figure.App.D-44 Effects of {-2,2} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 1000 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

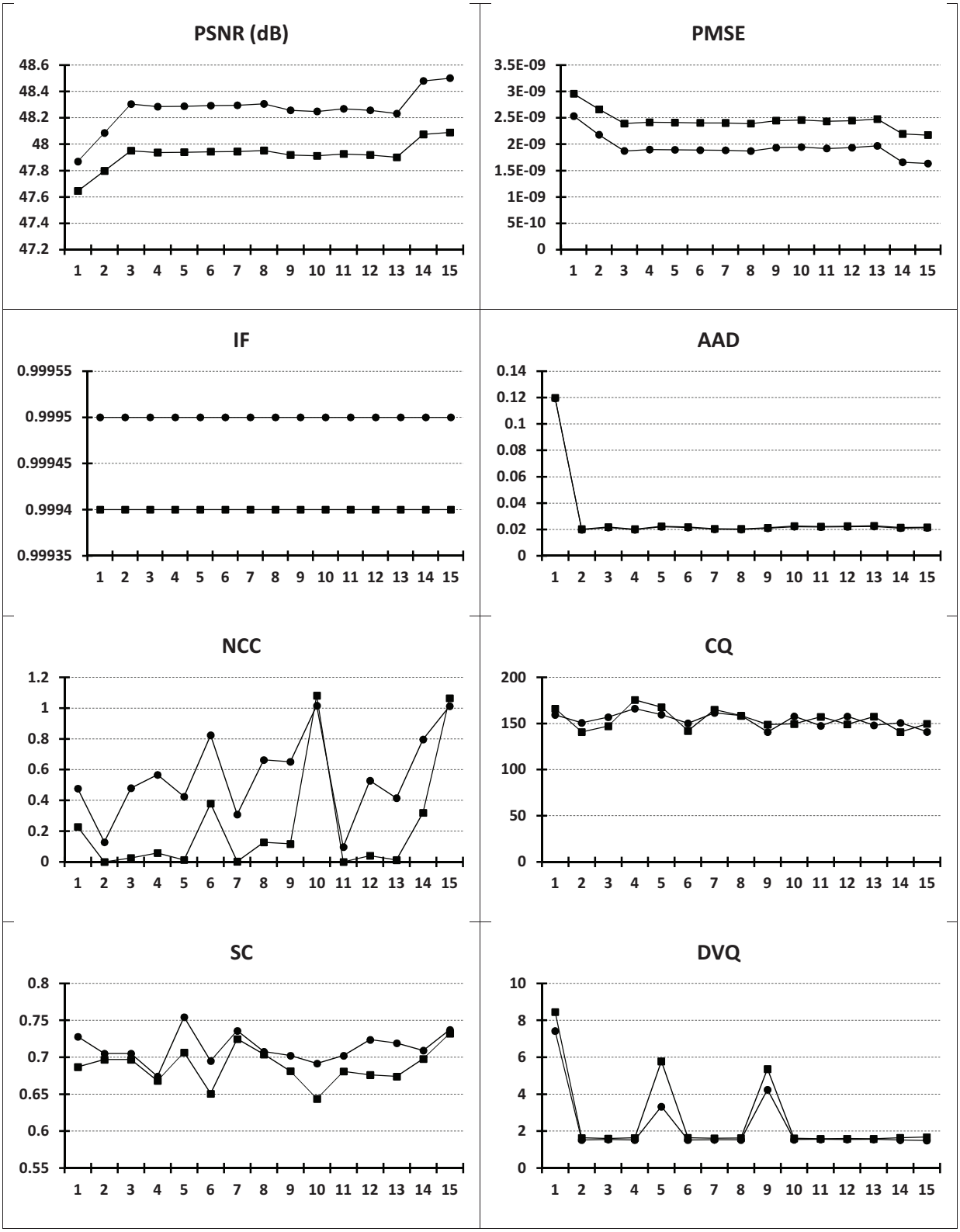


Figure.App.D-45 Effects of {-2,2} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

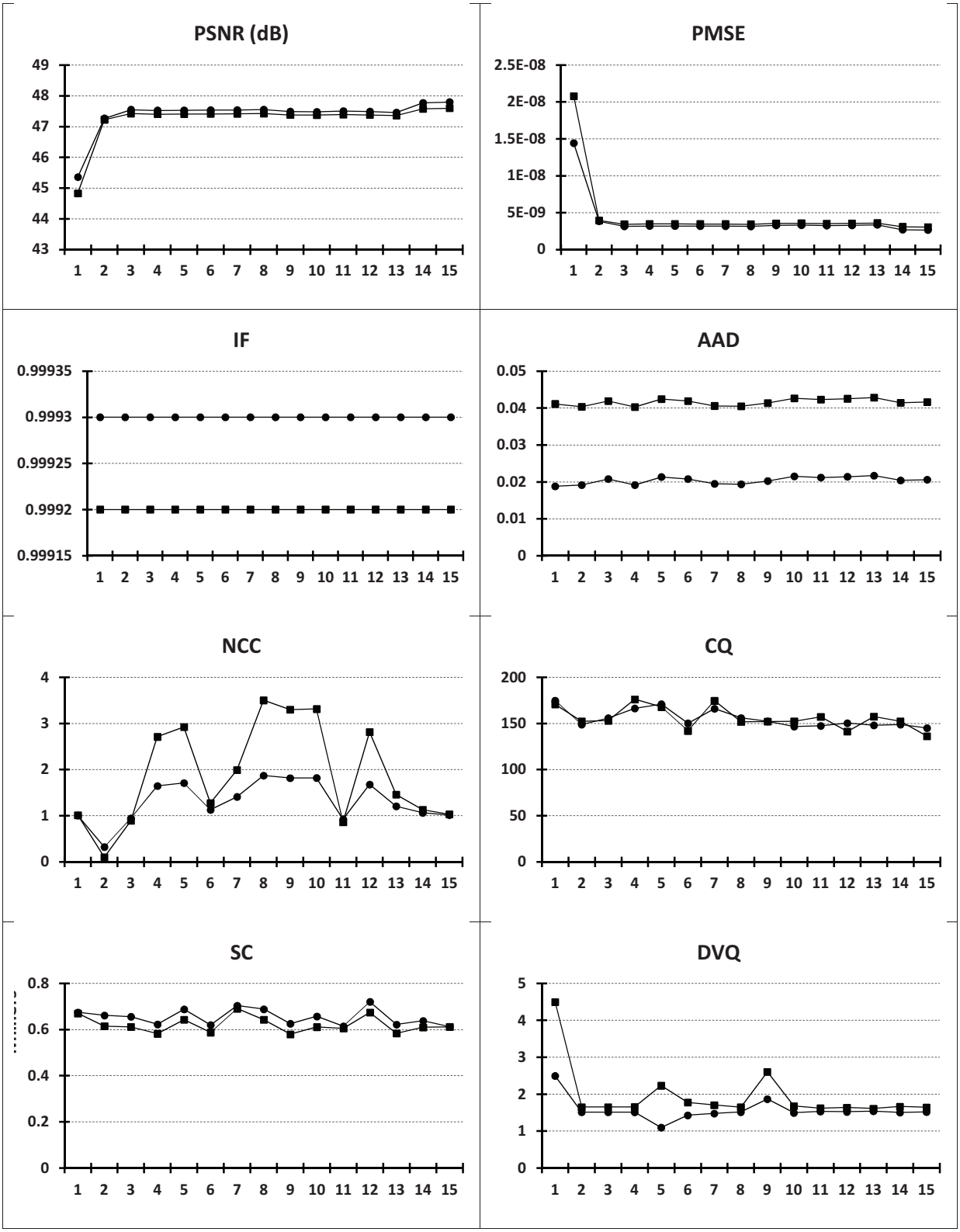


Figure.App.D-46 Effects of {-2,2} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

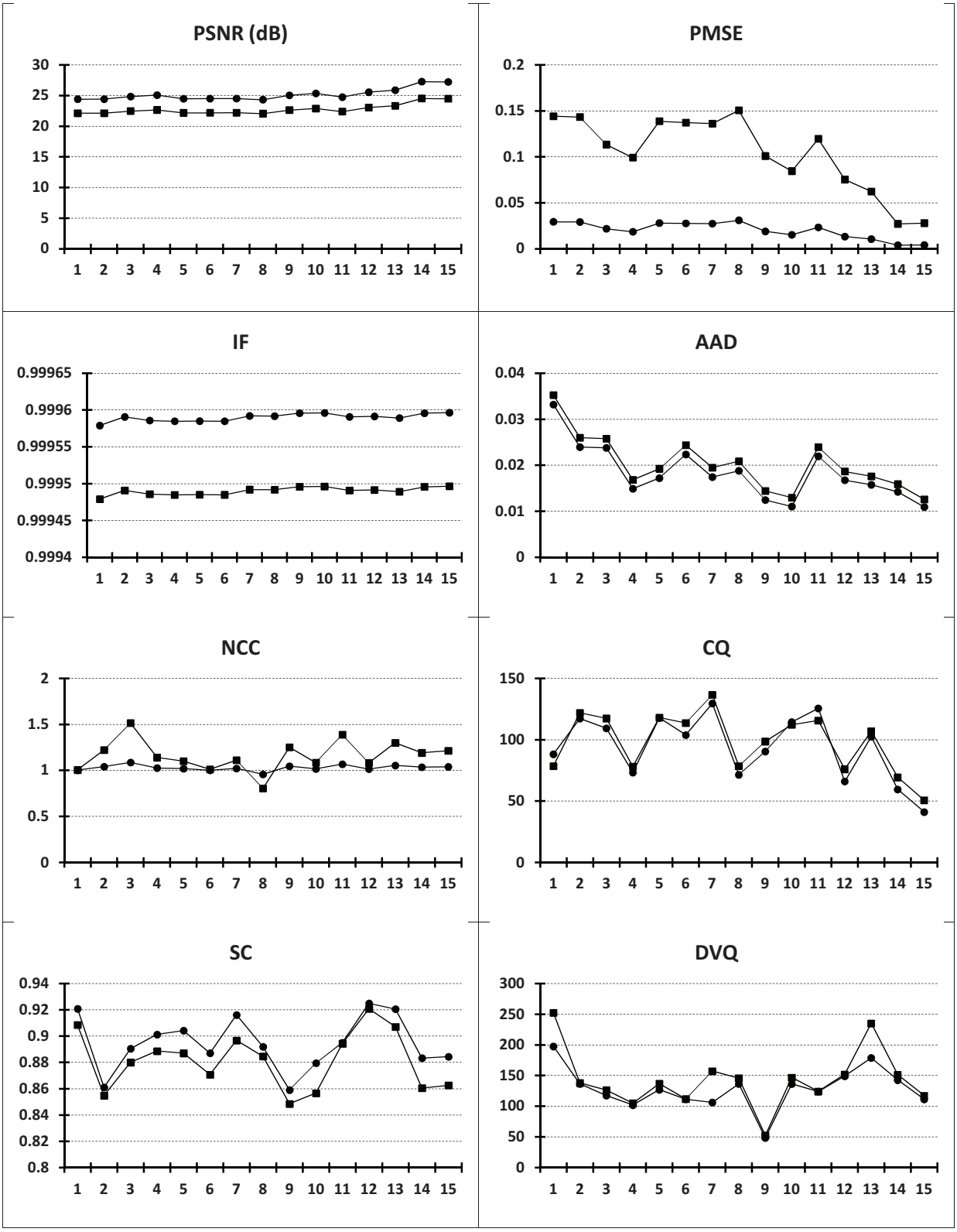


Figure.App.D-47 Effects of {-2,2} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 500 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

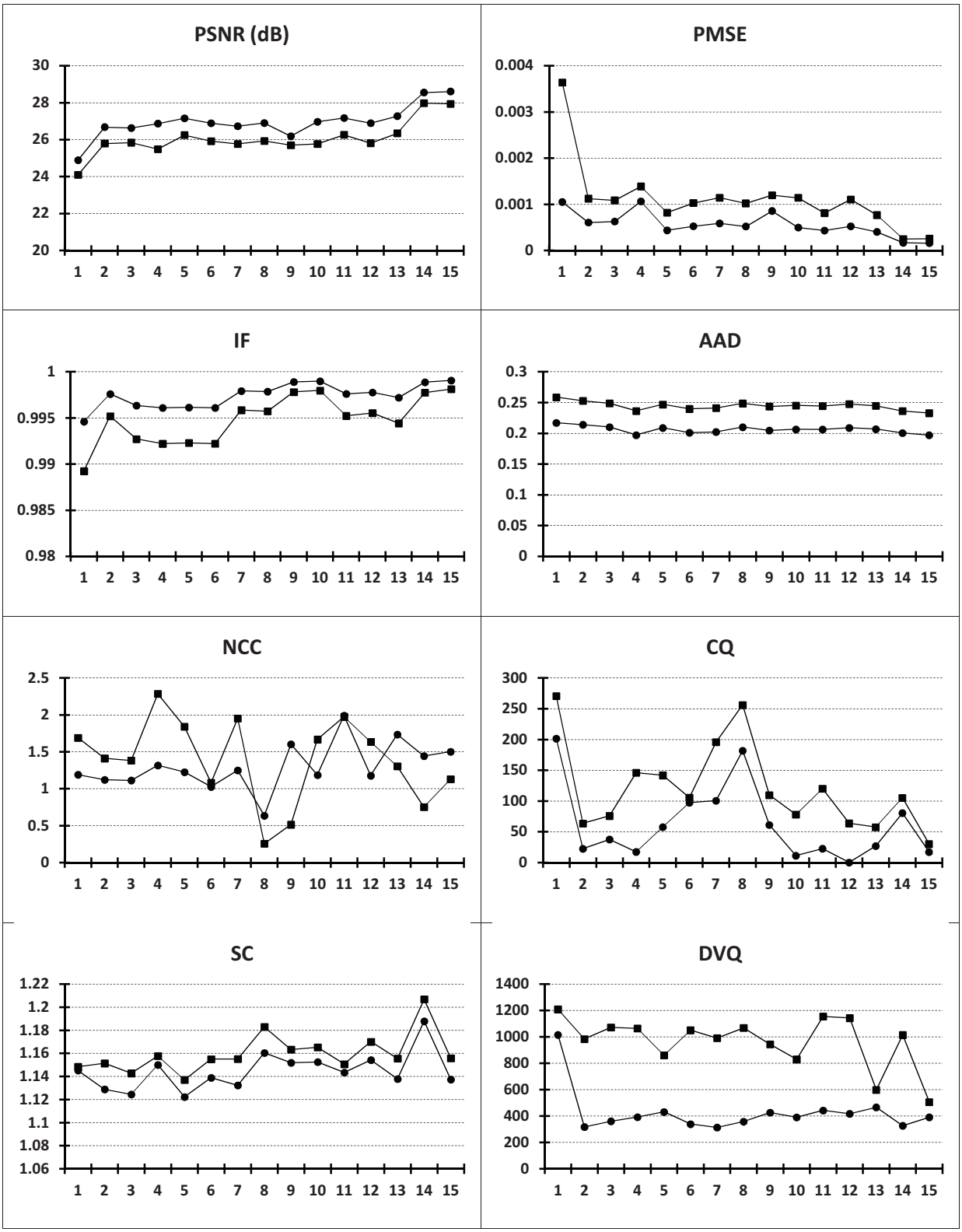


Figure.App.D-48 Effects of {-2,2} substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 1000 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

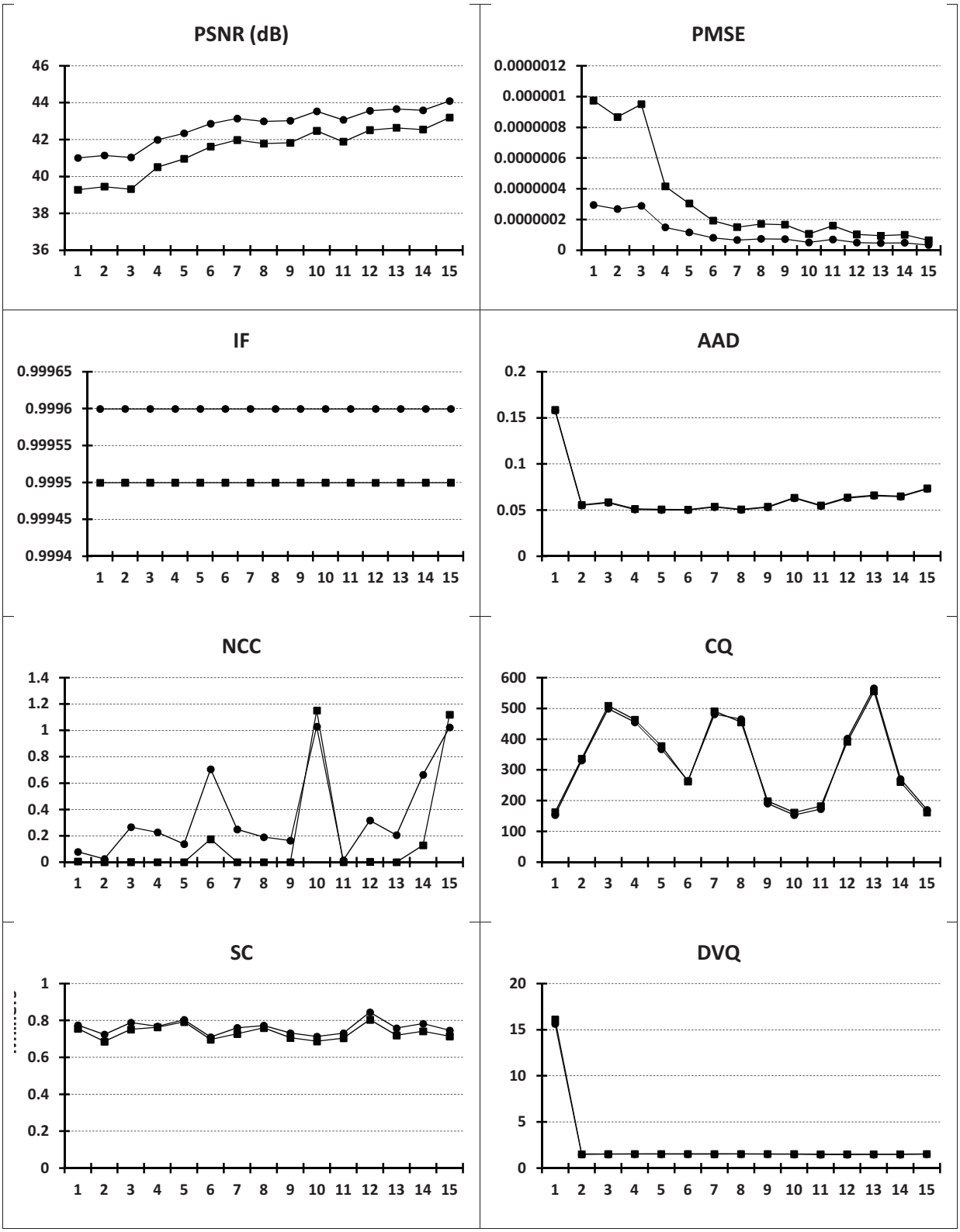


Figure.App.D-49 Effects of eight symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

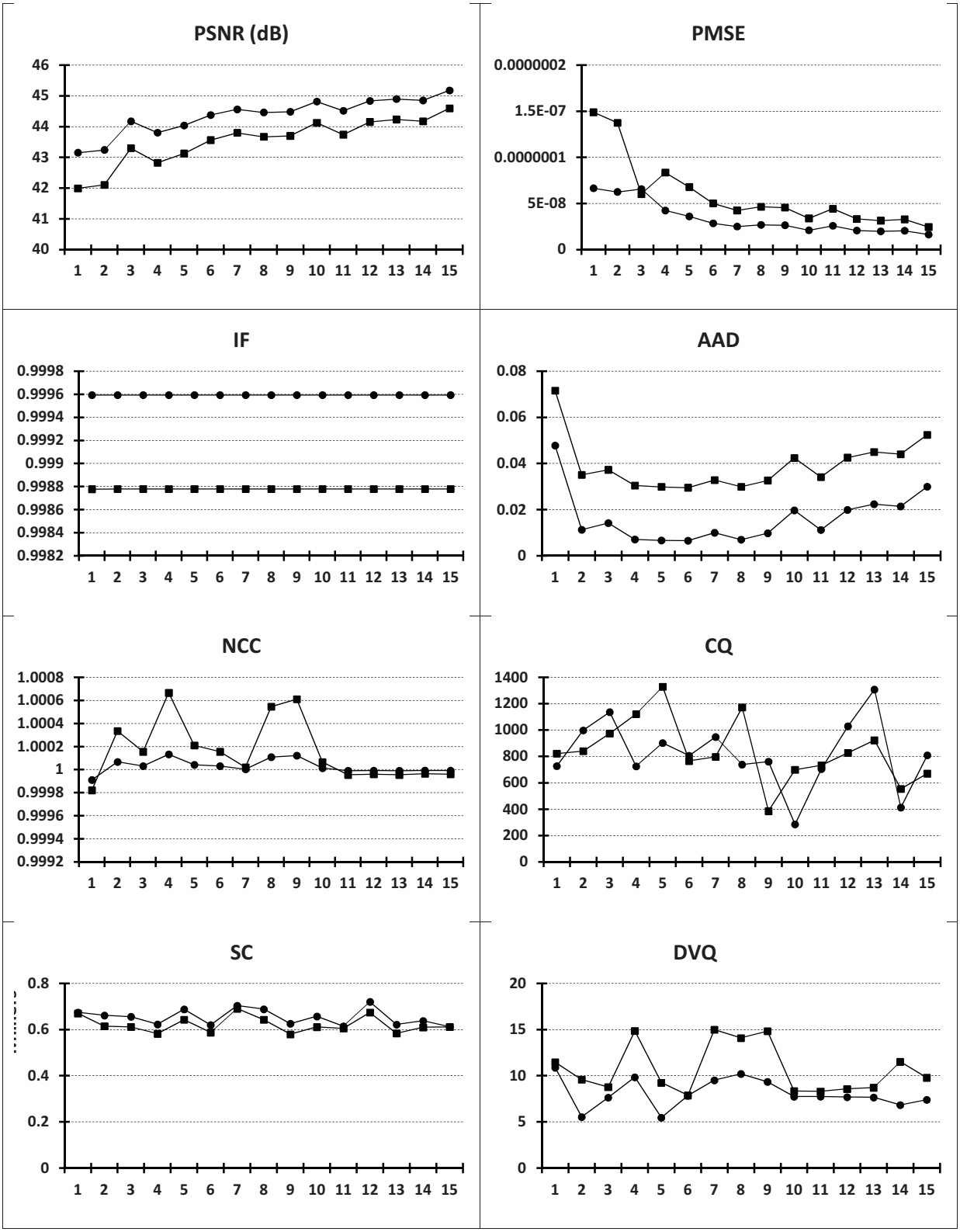


Figure.App.D-50 Effects of eight symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

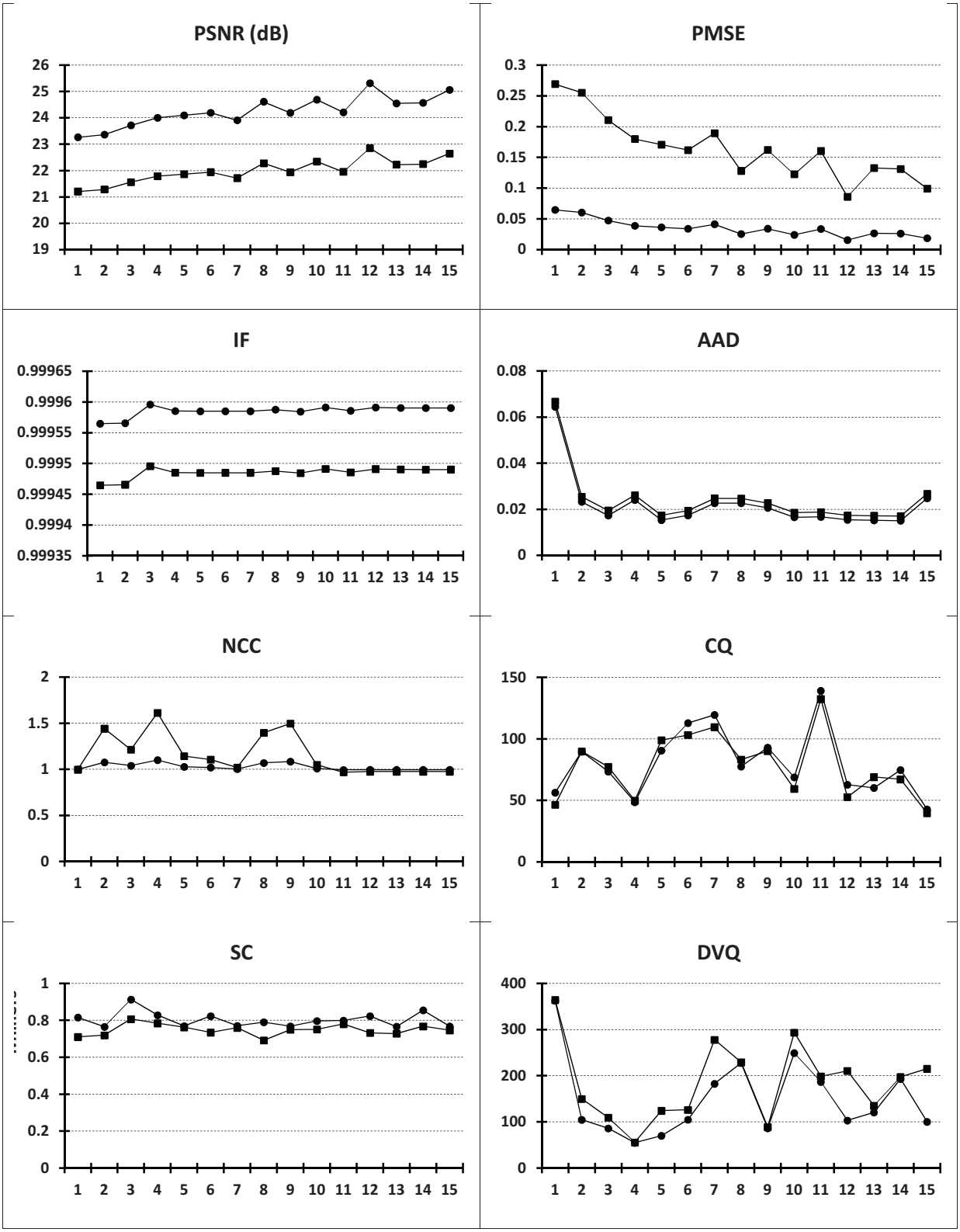


Figure.App.D-51 Effects of eight symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 500 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

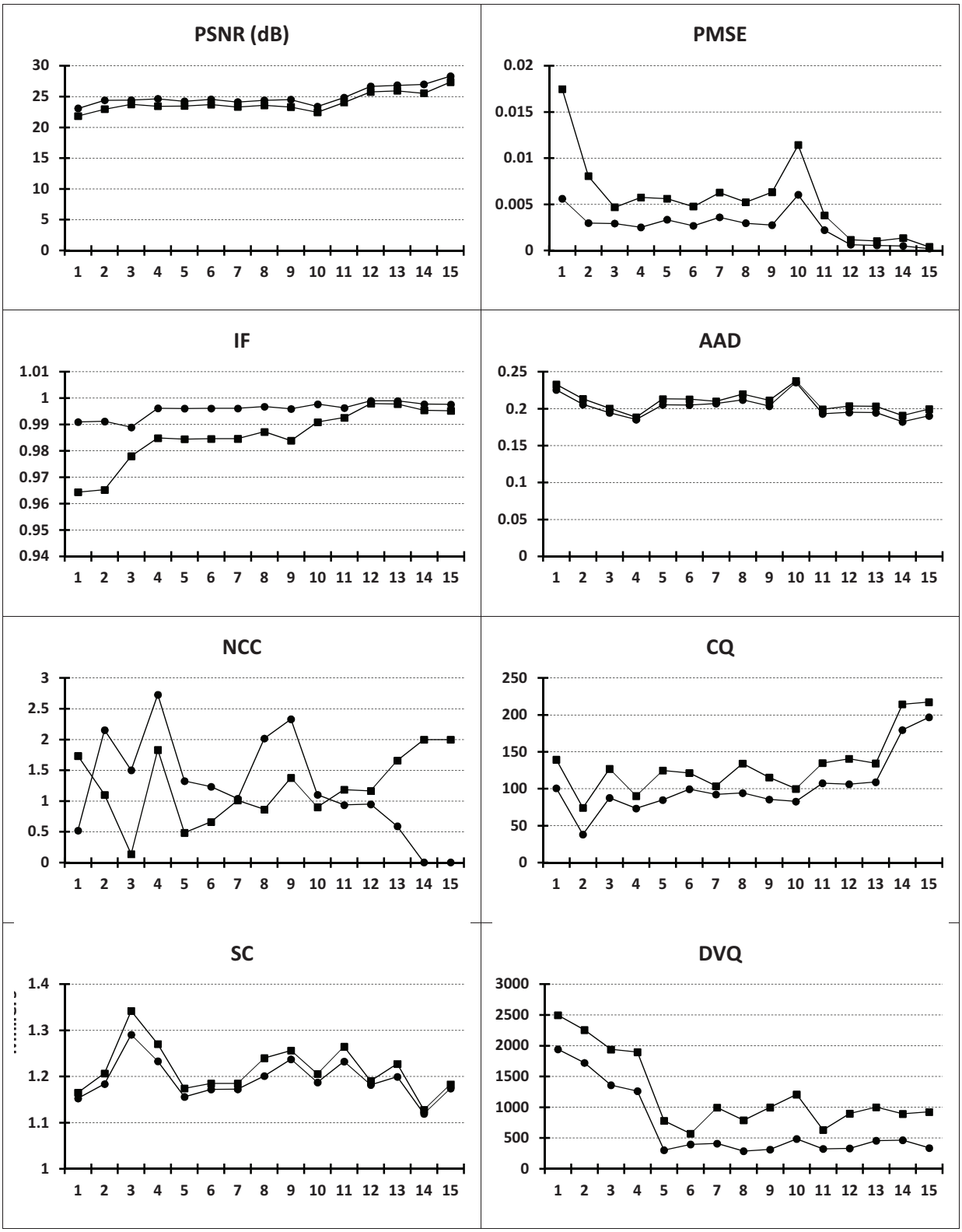


Figure.App.D-52 Effects of eight symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 5 Mbps, and with 1000 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

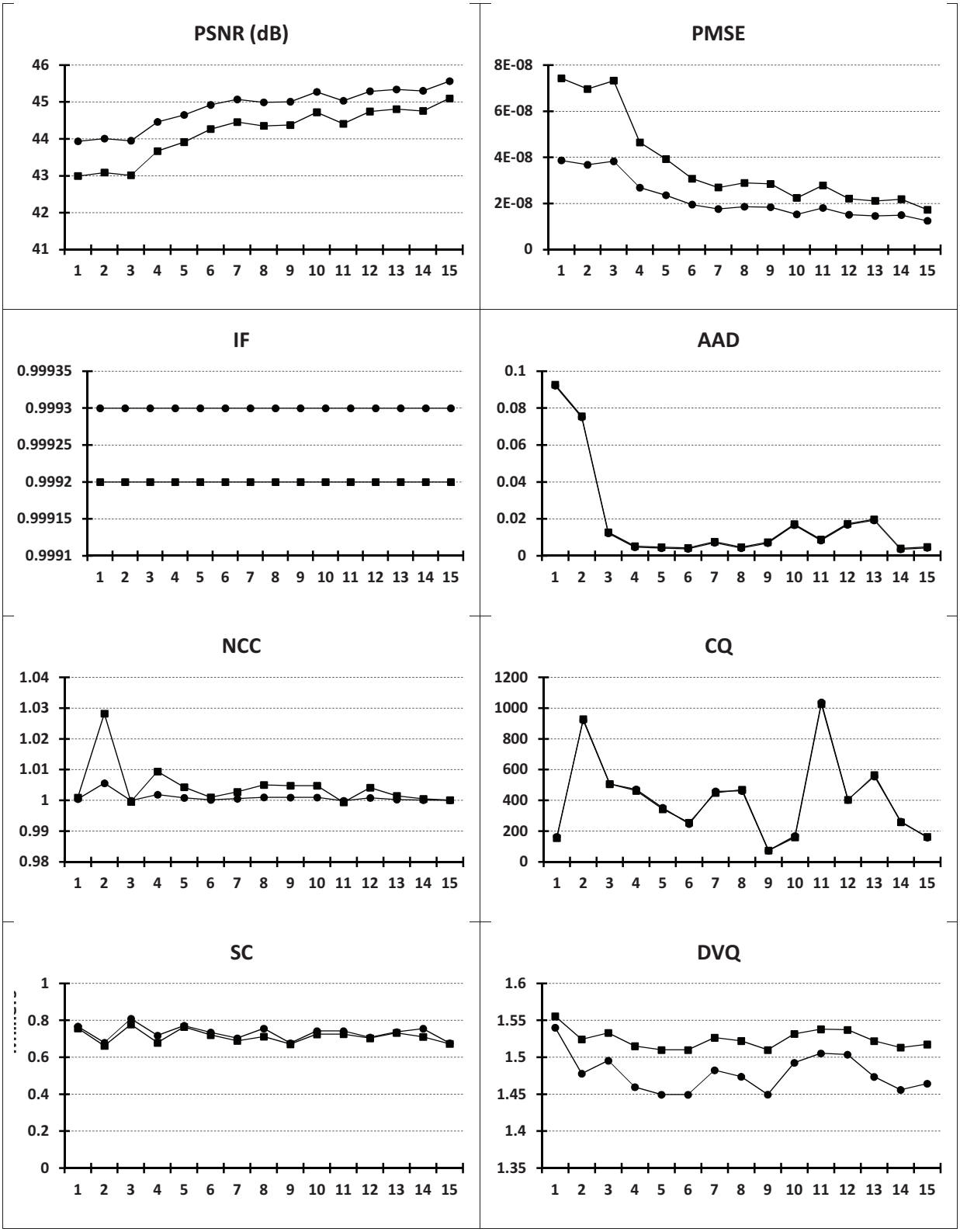


Figure.App.D-53 Effects of eight symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 5 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

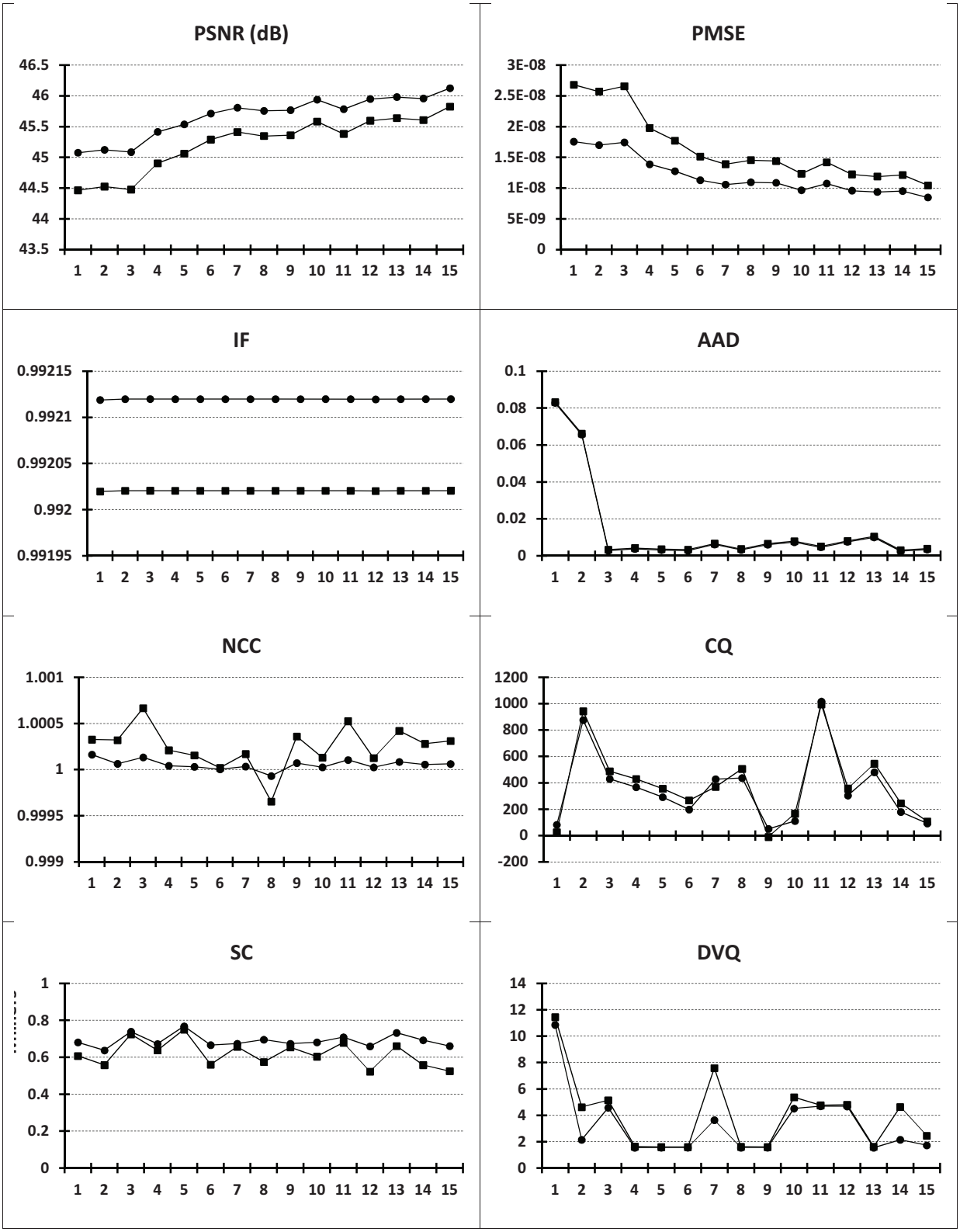


Figure.App.D-54 Effects of eight symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 50 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

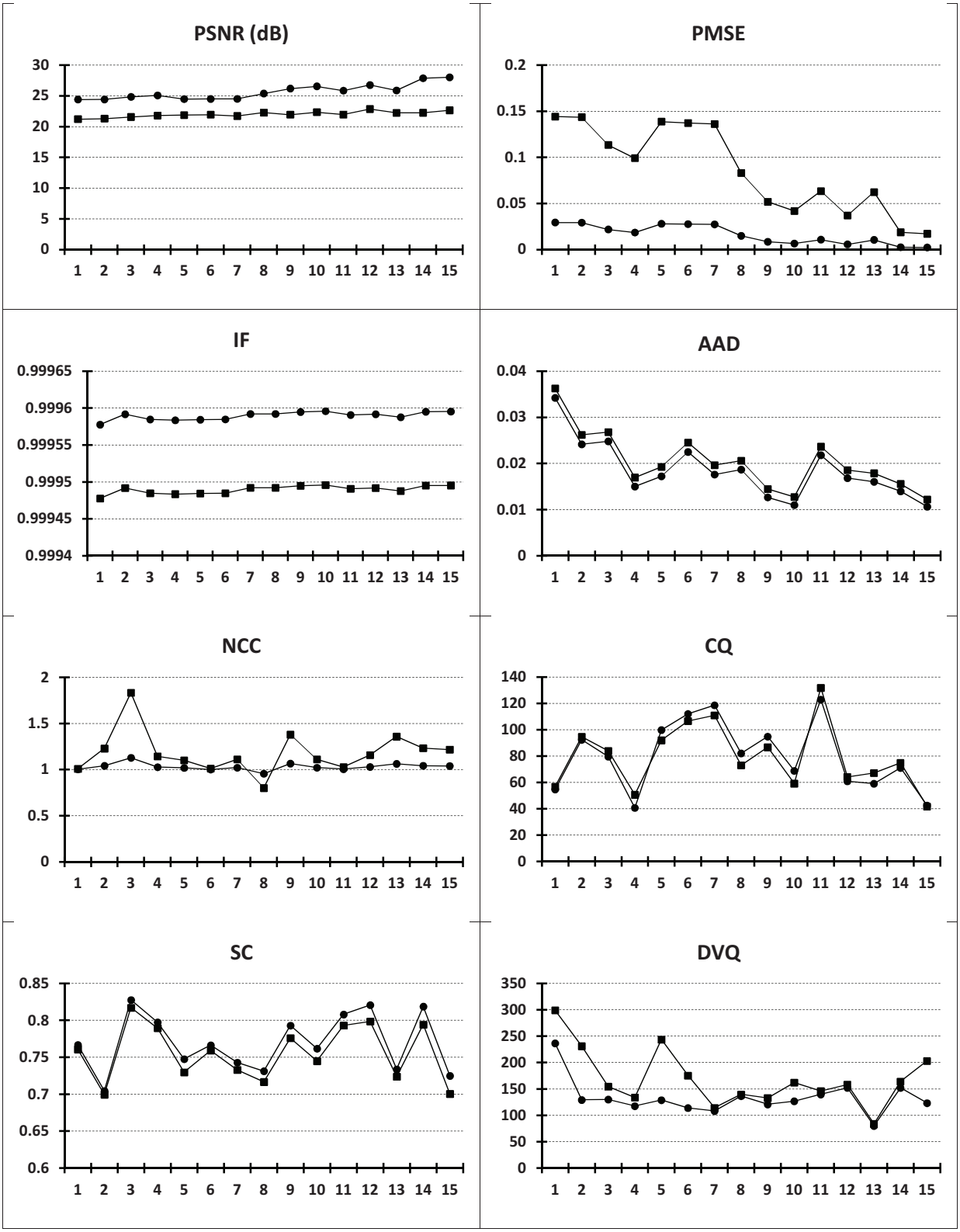


Figure.App.D-55 Effects of eight symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 500 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.

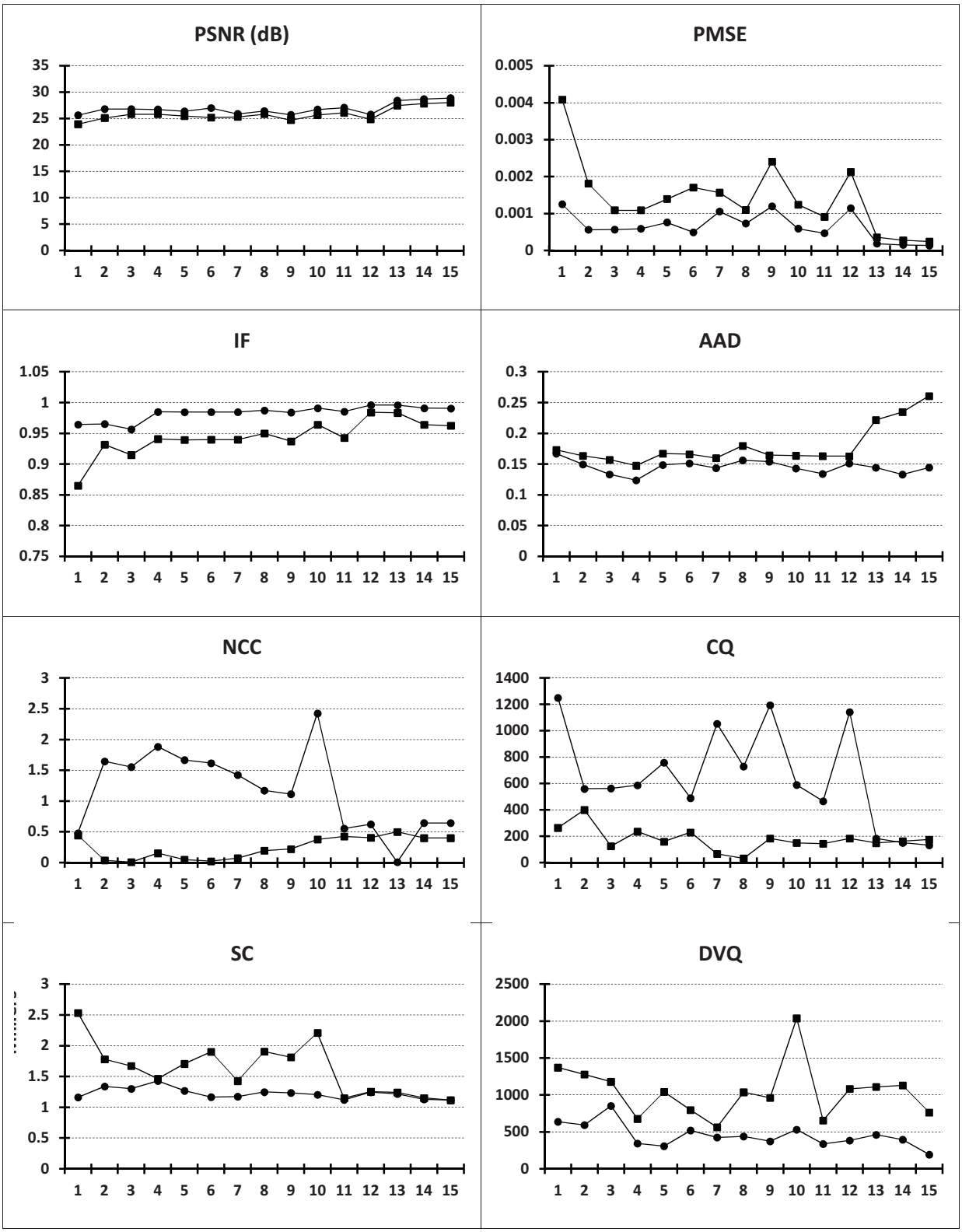


Figure.App.D 56 Effects of eight symbols substitutive noise, altering one sub-macroblock (the plot in ●) and 16 sub-macroblocks (the plot in ■). Quality measures evaluated for video HD encoded at 10 Mbps, and with 1000 macroblocks per frame subjected to errors. On the abscissa: the investigated coefficient.