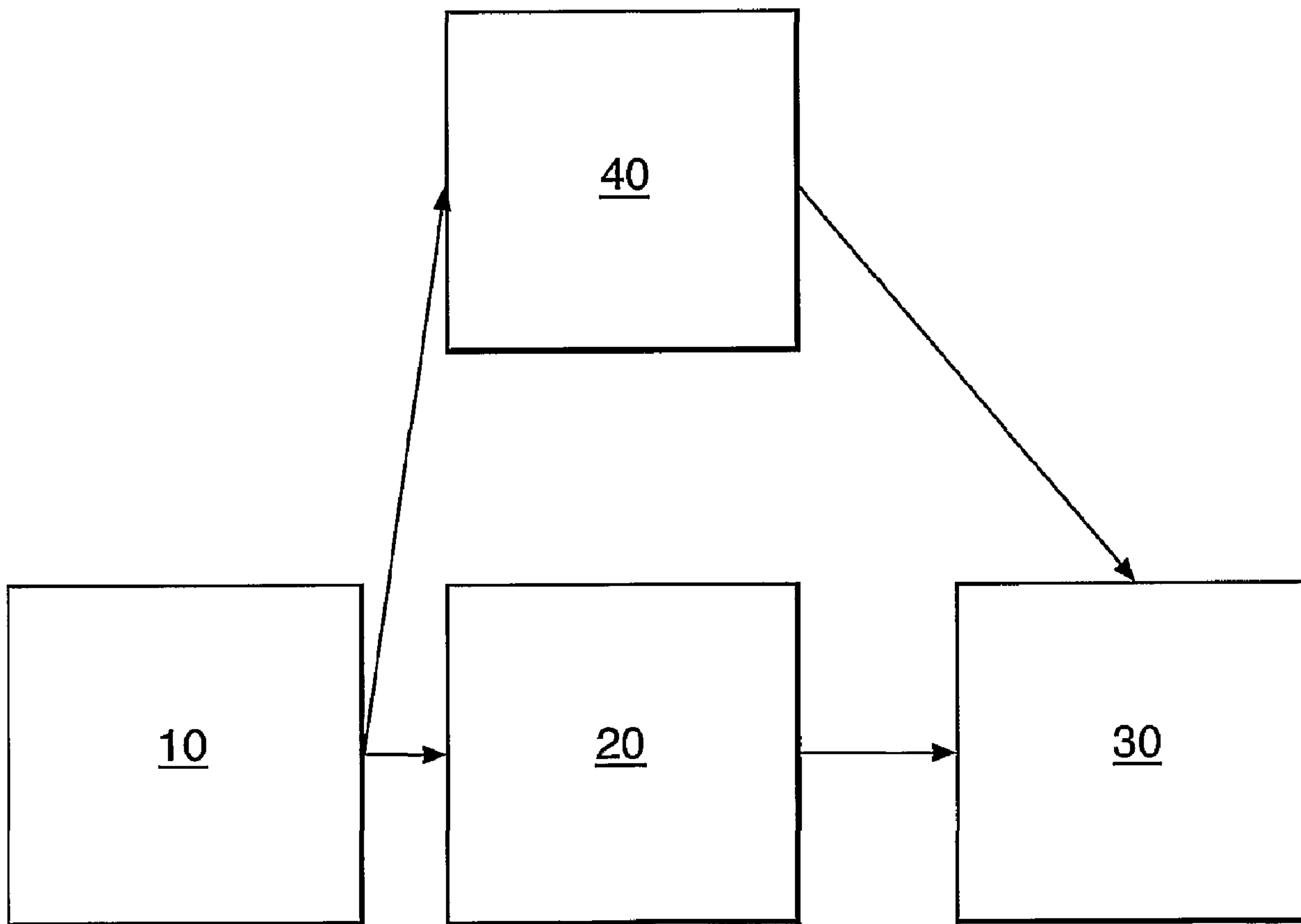




(86) Date de dépôt PCT/PCT Filing Date: 2001/05/29
 (87) Date publication PCT/PCT Publication Date: 2001/11/29
 (45) Date de délivrance/Issue Date: 2010/11/02
 (85) Entrée phase nationale/National Entry: 2002/11/07
 (86) N° demande PCT/PCT Application No.: GB 2001/002374
 (87) N° publication PCT/PCT Publication No.: 2001/091473
 (30) Priorité/Priority: 2000/05/26 (GB0012992.4)

(51) Cl.Int./Int.Cl. *H04N 17/00* (2006.01)
 (72) Inventeur/Inventor:
 BOURRET, ALEXANDRE JEAN, GB
 (73) Propriétaire/Owner:
 BRITISH TELECOMMUNICATIONS PUBLIC LIMITED
 COMPANY, GB
 (74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : PROCÉDE POUR TESTER LES SEQUENCES VIDEO
 (54) Title: METHOD FOR TESTING VIDEO SEQUENCES



(57) Abrégé/Abstract:

A non-intrusive test method for video sequences is described that correlates a number of quality indicators derived from a sequence with a matrix of constants to predict a quality score for the sequence.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number
WO 01/91473 A1

- (51) International Patent Classification⁷: **H04N 17/00**
- (21) International Application Number: PCT/GB01/02374
- (22) International Filing Date: 29 May 2001 (29.05.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
0012992.4 26 May 2000 (26.05.2000) GB
- (71) Applicant (*for all designated States except US*): **BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY** [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).
- (72) Inventor; and
- (75) Inventor/Applicant (*for US only*): **BOURRET, Alexandre, Jean** [FR/GB]; 26 Withipoll Street, Ipswich, Suffolk IP4 2BZ (GB).
- (74) Agent: **SEMOS, Robert, Ernest, Vickers**; BT Group Legal Services, Intellectual Property Department, Holborn Centre, 8th floor, 120 Holborn, London EC1N 2TE (GB).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR TESTING VIDEO SEQUENCES

(57) Abstract: A non-intrusive test method for video sequences is described that correlates a number of quality indicators derived from a sequence with a matrix of constants to predict a quality score for the sequence.

WO 01/91473 A1

METHOD FOR TESTING VIDEO SEQUENCES

This invention relates to a test method that is can assess the quality of video sequences, in particular digitally encoded video signals.

5 In order to reduce the bandwidth necessary to carry video sequences, compression algorithms such as H263 or MPEG rely on a set of well known methods. The first of these is to suppress the information present in each frame that is not perceptually relevant. The result of this operation may not be perceptible to the end user, perceptible but not perceptually relevant, or disruptive
10 to the information content of the frame. In most cases, the effect of the information suppression will be to modify the appearance of textured objects, giving a greater repetitiveness or simplicity.

In real world applications, where the video quality has to be assessed continuously, it can be very difficult or costly to provide a system with a test
15 signal, and the test signal will be transmitted through the system under test. For these reasons, non-intrusive systems have been investigated, where only a reconstructed signal is needed. How can we know the quality of a video sequence be assessed without being able to determine what has been added to or removed from it? However, as a subjective viewer knows what quality of video sequence
20 can be expected from a television set, most people are able to make a judgement of quality without referring to the original image. More advanced or experiences users of the system will also know the type of degradation introduced by the medium they use, and will be able to spot them to make a more educated estimation of the quality. Similarly, it is known that each family of compression
25 algorithm tends to introduce its own type of distortion and what can be expected from natural images. This information can be used to determine what amount of distortion has been added to a sequence without referring to its original.

According to a first aspect of the invention there is provided a method of comparing two frames from a video sequence, the method comprising the steps
30 of;

- (i) taking a first frame from said video sequence;
- (ii) taking the subsequent frame from said sequence;

(iii) comparing one pixel from the first frame with the corresponding pixel from the
5 subsequent frame; and

(iv) repeating step (iii) for all pixels in a test area of the first frame.

If the results of the comparison stage indicate a large number of substantially equal pixels then a dropped frame may be inferred. If the results of the comparison stage indicate a small number of substantially equal pixels then a scene cut may be inferred.

10 According to a second aspect of the invention there is provided a method of predicting a quality score for a video sequence using quality parameters derived from a video sequence as described above.

The invention will now be described, by way of example only, with reference to the
15 following figures in which;

Figure 1 shows a reference video frame;

Figure 2 shows the video frame of Figure 1 after having been compressed;

Figure 3 shows a schematic depiction of apparatus for performing a method according to the present invention;

20 Figures 4a – c show the variation in unchanged pixel proportion with time for a number of different compression schemes;

Figure 5 shows the correlation of predicted and subjective quality scores for a number of video sequences; and

25 Figures 6a – b show the correlation of predicted and subjective quality scores for a number of subsets of video sequences.

Most video sequences used come from natural sources, such as a studio or external scene captured by a video camera. This type of transceiver produces white noise, which implies that the value for a given pixel is likely to change over time even if the scenery being
30 analysed is still and in a perfectly controlled environment. Because of its low amplitude and high spatial frequency, this noise is the first feature that will be modified or removed by compression algorithms. For this reason, absence of noise is an indicator of compression level. This effect is shown in the reference and degraded frames shown in Figures 1 and 2 respectively. The details present in the tarmac textures, such as the light presence

of noise and the gradient have been simplified and replaced by a "flat" area of the average luminance and colour.

In most compression algorithms, this simplification operation is done within blocks of pixels, most often of 8x8 pixels. As a consequence, disparities
5 tend to appear between adjacent regions, making the block structure of the image analysis visible. This effect, referred as blockiness, can easily be measured (see *"Image quality measure based on local visual properties"*, P Franti, Picture Coding Symposium 97, 10-12 Sept. 1997, no.143, pp217-20, or M Ghanbari, *"Blockiness Detection for MPEG2-coded video"*, IEEE Signal Processing Letters,
10 7:8 (August 2000), pp 213-215) even without having access to the reference sequence.

A second such method used to reduce bandwidth requirements for video transmissions is movement estimation. This technique involves analysing the motion present between one frame and the next frame in the sequence, so that
15 movement vectors can be transmitted to the decoder in order to reconstruct the sequence from a starting frame. In order to avoid the accumulation of errors over time, motion estimation is not used between every frame. Instead, a temporal pattern can be used, as in the MPEG-2 compression scheme, where original, interpolated and predicted frames are interleaved in a predefined sequence.

20 The third method used is to reduce the frame rate of the sequence. This is used for lower bitrate applications such as videoconferencing or video over IP, and is simply done by dropping one or more frames from a sequence.

It appears from these three main effects that motion present in a sequence is an important indicator of the image quality. Furthermore, the motion
25 present in a sequence can be determined without the help of the original sequence, enabling the creation of non intrusive assessment systems that can estimate the quality of video systems.

Figure 3 shows video capture unit 10, which is suitable for capturing a series of frames from a video sequence. Each frame from the series is sent to
30 blockiness detector 40 and also each frame is transmitted to the comparator 20, along with the subsequent frame from the video sequence. Both the blockiness detector 40 and the comparator 20 are connected to a data processing unit 30, so

that outputs from the blockiness detector and the comparator can be passed to the data processing unit.

The blockiness detector 40 analyses the pixels of each video frame received from the video capture unit. As has been described above the video
5 compression technique divides the video image into a number of blocks and a side effect of simplifying the content of each block is to increase the disparity between the blocks, making them clearly visible in the reconstructed image. Different approaches have been tried to detect blockiness (see Franti or Ghanbari *op cit*) but the preferred technique for detecting blockiness for use in the present invention is
10 to compare the variation of a pixel property (for example the luminance) every 8 pixels with the sum of variations within these 8 pixels. A ratio indicating a higher activity every 8 pixels than within 8 pixel blocks triggers an alarm as it tends to indicate an abrupt change in the pixel property in the transition from a block of very similar values to the next block. Alarms are counted across the whole frame,
15 and the sum is used as an indicator of blockiness for the image. These blockiness indicator values are then transmitted to the data processing unit, along with a reference to the frame to which they correspond.

The two video frames are then sent to comparator 20 which compares the two frames pixel by pixel using one of the properties of a pixel as a basis for the
20 comparison. The results of the comparator are fed to data processing unit 30. The comparator counts all those pixels where there has been no significant change in the pixel property (i.e. the luminance of the pixel in the subsequent frame is within a pre-determined threshold from the luminance of the pixel in the first frame) and determines the proportion of pixels for which the pixel property has not changed.
25 The proportion of pixels in which there has been no change of pixel property and/or the variation in the proportion of pixels in which there has been no change of pixel property can be used to assess the quality of the video compression. Data processing unit 30 stores the pixel proportion values along with a reference to the frame to which they correspond.

30 Most video capture systems introduce some noise into the image. For instance, the luminance of a given pixel produced by a video camera which is capturing a stationary image, for example background scenery, will vary in time. As the video compression rate is increased, i.e. the bandwidth of the transmitted

video stream is decreased, the variations of a given pixel tend to be suppressed, and the pixel properties (for example chrominance, luminance, etc.) tend towards a constant value. A result of this is that stationary objects present in a compressed video sequence tend to look unnaturally still. This is shown in Figure 4, where

5 Figure 4a shows the variation of the proportion of unchanged pixels for the reference signal, Figure 4b shows the variation of the proportion of unchanged pixels when that signal has been compressed at 4Mb/s and Figure 4c shows the variation of the proportion of unchanged pixels when the reference signal has been compressed at 1Mb/s. Thus this variation can also be used to indicate the level of

10 compression that has been applied to a reference signal. Accordingly the data processing unit has the facility to calculate a value which gives an indication of the variation of pixel proportion values# e.g. by performing a fast Fourier transform (FFT) of the pixel proportion values, calculating the derivative of the pixel proportion value, etc. The preferred indicator is to measure the average and

15 standard deviation of the proportion over a 500m window.

If only a small number of pixels, for example less than 4%, have remained the same then it can be assumed that there has been a cut in the video sequence (i.e. the same scene is being viewed from a different camera angle, or a new scene is now being viewed). As it is not possible to proceed with the analysis of

20 the change in pixel values for these frames because of the scene cut, the analysis may be discontinued and the next frame in the sequence may be selected for comparison.

If a very large number of pixels, for example greater than 98%, have remained the same then there is a high probability that a frame has been dropped,

25 i.e. the same frame has been repeated due to the failure of the subsequent frame in the sequence to be received. The ratio of dropped frames is recorded by the data processing unit, and their position in the sequence.

If a large number of dropped frames are detected within a video sequence then this may be an indicator that the video sequence is being transmitted across

30 an unreliable transport mechanism (in this case, if a video sequence is being transmitted 'live' across a communications network, it may be possible to send a signal to the transmitting codec to decrease the frame rate in use (if the frame rate

received is less than the frame rate in use) or to switch to a transport mechanism having a greater quality of service.

These three indicators, the amount of blockiness, the proportion of unchanging pixels and the ratio of dropped frames can be used to generate a
5 model that can be used to assess the quality of a received video sequence in a non-intrusive manner, that is without needing to insert any test signals, or without having access to the initial video sequence. A matrix was formed from each of the three indicators and a time-dependent variation of each indicator. The non-intrusive model was trained and calibrated using linear-regression techniques to
10 assess the qualities of a number of known video sequences which had previously been subjectively judged (that is, the quality of the sequences has been assessed by a number of human viewers).

The non-intrusive model, by design, is looking for known artefacts introduced by the codec in use. For this reason, the characteristics of the codec in
15 use must be known and the model trained for the particular configuration being assessed. This requirement is incompatible with applications such as assessment of quality improvement of compression algorithms or tests on new coding schemes for instance. The model could easily be tricked by hiding such artefacts, without improving the overall quality (the most obvious example would be the use
20 of a blockiness smother algorithm, as described by Ghanbari *op cit*). On the other hand, this approach can be particularly useful to telecom operators who are not changing coding algorithms on daily basis, but who want to monitor the performance of their network.

The first database used to calibrate and train the non-intrusive model is
25 the series of sequences used for the VQEG contests. It consists of a number of sequences which are repeated but at different levels of compression. The resolution used in the sequences and the levels of degradation are consistent with broadcast applications (for the purpose of this calibration, some degradations are discarded, because they are based on analogue techniques. The second database
30 relates to videoconference applications. The format used is either CIF or QCIF, the codec used to generate the degradations are H263 or MPEG4, and IP or UMTS error pattern have been inserted into the data stream before the decoding stage. The main difference between the two databases lies in the way the sequences

have been previously assessed. The opinion scores of the VQEG sequences, which are about 10 seconds long in average, have been taken at the end of each presentation. The second database, which contains longer sequences and shows rapid variations in quality, has been assessed continuously. As test material is
5 complicated to produce, the databases were split into a calibration set and a test set.

The test set of each database were processed through the model, which produces a set of parameters for each frame it contains (blockiness, pixel variation and dropped frame ratio). An extra set of parameters is derived from these initial
10 data, by computing their mean and mean and standard variation on blocks of one second. These account for the temporal variations, which are linked to the compression rate. A linear regression is used to obtain the best combination between the extracted parameters and the mean opinion score as previously measured for the given sequences. The combination is then tried with on the
15 parameters extracted for the rest of the database. The produced is compared with the measured mean opinion score, giving us an indicator of the model performance. Once the model linear regression technique has been applied a number of times a composite node is arrived at that can be used to predict the subjective testing of the sequences in the test set.

20 The model was also calibrated using a subset of the Eurescom Aquavit database. This bank of subjectively assessed sequences represents a variety of codec and compression rate at levels compatible with multimedia applications and Figure 5 shows the correlation between the predicted score generated by the non-intrusive model and the mean subjective score for the test set of sequences.
25 Figure 6a shows the variation of correlation between predicted score and subjective score with the frame rate used in the sequence and Figure 6b shows the variation of correlation between predicted score and subjective score with the bandwidth used to code the sequence.

It will be realised that another pixel property other than luminance, for
30 example chrominance, hue or one of the RGB signals, could be used to assess the inter-frame variation of pixels. Additionally, all of the pixels in the frame could be compared with their respective pixels in a second frame or a specific sub-set of pixels could be compared with their respective pixels in the second frame.

This organisation of the coded sequence has an effect on the distribution of the motion. In a non coded sequence, the amount of inter-frame change will change slowly in relation to the variation of motion present in the image. However, in the coded sequence, the inter-frame change varies quickly around its average value, with a period related to the coding sequence used by the encoder.

Profiles can be incorporated into the model, according to the type of video sequences processed. Requirements and expectations can be very different according to the type of service or images transmitted. At the equivalent level of quality, a video sequence can be acceptable in video conference but not in broadcast. The bandwidth requirement are much higher for sport than head and shoulder sequences. Profiles take these aspects into account before to make a judgement about the quality.

The measurement tool can be used in different ways. By providing a means to measure the quality of refreshment in video sequences, it allows comparison between a reference signal and the same signal once gone through by the coding and decoding process. But it can also provides useful information about a sequence without the help of any reference.

If processing power is limited (for example if a real-time assessment is required or the model is being incorporated into a device with limited capabilities then a less-sophisticated model can be derived by using fewer parameters; the nature of the system to be measured will influence which quality indicator, or its time dependence, can be excluded from the model and the optimal combination can easily be determined by experimentation and further application of linear regression techniques.

It will be readily understood that the present invention could be implemented in software alone, i.e. running on a general purpose computer such as personal computer (which may be supplied on a data carrier such as, for example a floppy disc or a CD-ROM), in hardware alone or in a combination of hardware and software (for example programmable DSPs) and that the current invention should not be limited by the form of the implementation.

CLAIMS:

1. A test method of comparing two frames from a received video sequence to non-intrusively assess the quality of a received video sequence, the method comprising:
- 5 (i) taking a first frame from said received video sequence;
- (ii) taking the subsequent frame from said sequence;
- (iii) comparing one pixel from the first frame with the corresponding pixel from the subsequent frame to detect variation of a pixel property with respect to nearby pixels within a first neighborhood as compared to nearby pixels in a second
- 10 neighborhood beyond said first neighborhood to drive a blockiness indicator value for a compared frame;
- (iv) repeating step (iii) for all pixels in a test area of the first frame;
- (v) inferring a scene cut if the results of the comparison stage indicate a small number of substantially equal pixels;
- 15 (vi) ignoring comparisons indicating a scene cut in deriving a frame quality assessment using said blockiness indicator values; and
- (vii) providing said frame quality assessment as data output from said method.
- 20 2. A method as in claim 1 wherein said compared pixel property is a luminance of the corresponding pixels is compared.
3. A method as in claim 1 further comprising inferring a dropped frame if the results of the comparison stage indicate a large number of substantially equal pixels
- 25 and ignoring comparisons indicating a dropped frame.
4. A method of non-intrusively assessing the quality of a received video sequence, the method comprising:
- (i) comparing each frame in the received video sequence with a subsequent
- 30 frame by
- (a) taking a first frame from said video sequence;
- (b) taking the subsequent frame from said sequence;
- (c) comparing a pixel property of one pixel from the first frame with the same pixel property of a corresponding pixel from the subsequent
- 35 frame to detect variation of that pixel property; and
- (d) repeating step (c) for all pixels in a test area of the first frame to determine the magnitude of pixel variation for that test area;

- (ii) ignoring those comparisons indicating a dropped frame or a scene cut;
- (iii) calculating a quality figure for the video sequence from the remaining comparison values; and
- (iv) outputting data representing the quality of said sequence.

5

5. A method of assessing the quality of a video sequence, the method comprising:

- (i) comparing all of the frames in the video sequence with the subsequent frame in accordance with claim 3;

10

- (ii) ignoring those comparisons indicating a dropped frame or a scene cut; and

- (iii) calculating a quality figure from the remaining comparison values including an estimated value of compression applied to the received video sequence during transmission; and

15

- (iv) outputting data representing the quality of said sequence.

6. A method as in claim 5 further comprising:

- calculating a further quality figure associated with temporal variations in said quality figure; and

20

- using said quality figure and said further quality figure to assess the quality of said video sequence.

7. A method for assessing the quality of a received video sequence in a non-intrusive manner without access to the as transmitted video sequence, said method comprising:

25

- processing said video sequence to derive a value representing measured variation of a pixel property value from one frame to the next in said received video sequence;

30

- if measured pixel variation is above a predetermined first threshold, inferring that there has been a scene cut in the video sequence and omitting analysis of the video sequence for frames adjacent the detected scene cut;

- if measured pixel variation is below a predetermined second threshold, inferring that a frame has been dropped, and recording the occurrence of dropped frames as one quality measure;

35

- using said measurements to derive video sequence quality values for the amount of blockiness, the proportion of unchanging pixels and the ratio of dropped frames to generate a non-intrusive model representing quality of a received video

sequence in a non-intrusive manner without inserting test signals and without having access to an initial version of the received video sequence prior to its transmission.

8. A method as in claim 7 wherein generation of said non-intrusive model
5 comprises:

forming a matrix from each of indicators for blockiness, unchanging pixels and dropped frames and a time-dependent variation of each indicator.

9. A method as in claim 7 further comprising:
10 training and calibrating said non-intrusive model using linear-regression to assess qualities of plural known video sequences which were previously subjectively judged by a number of human viewers.

1/5

Fig. 1.

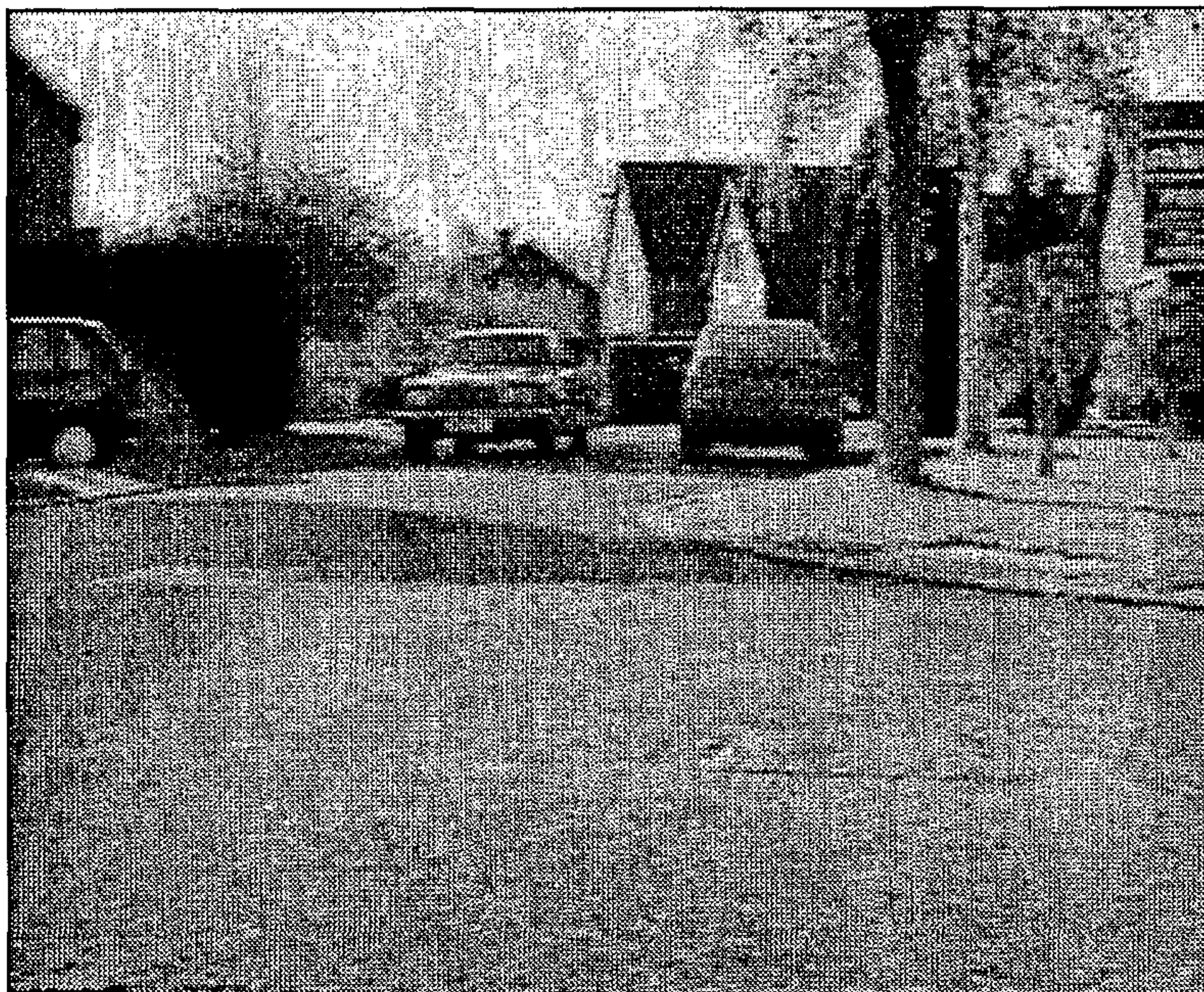


Fig. 2.

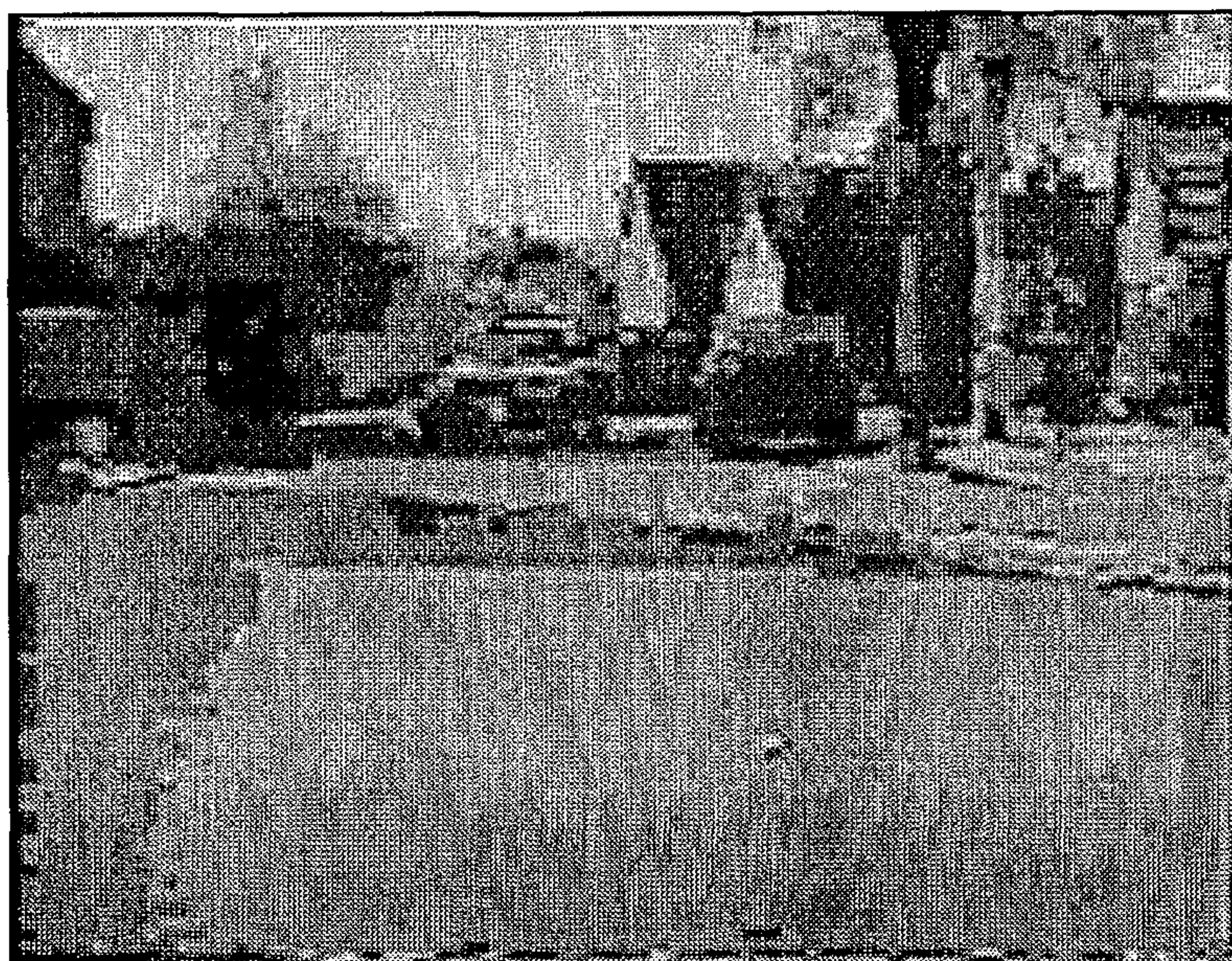
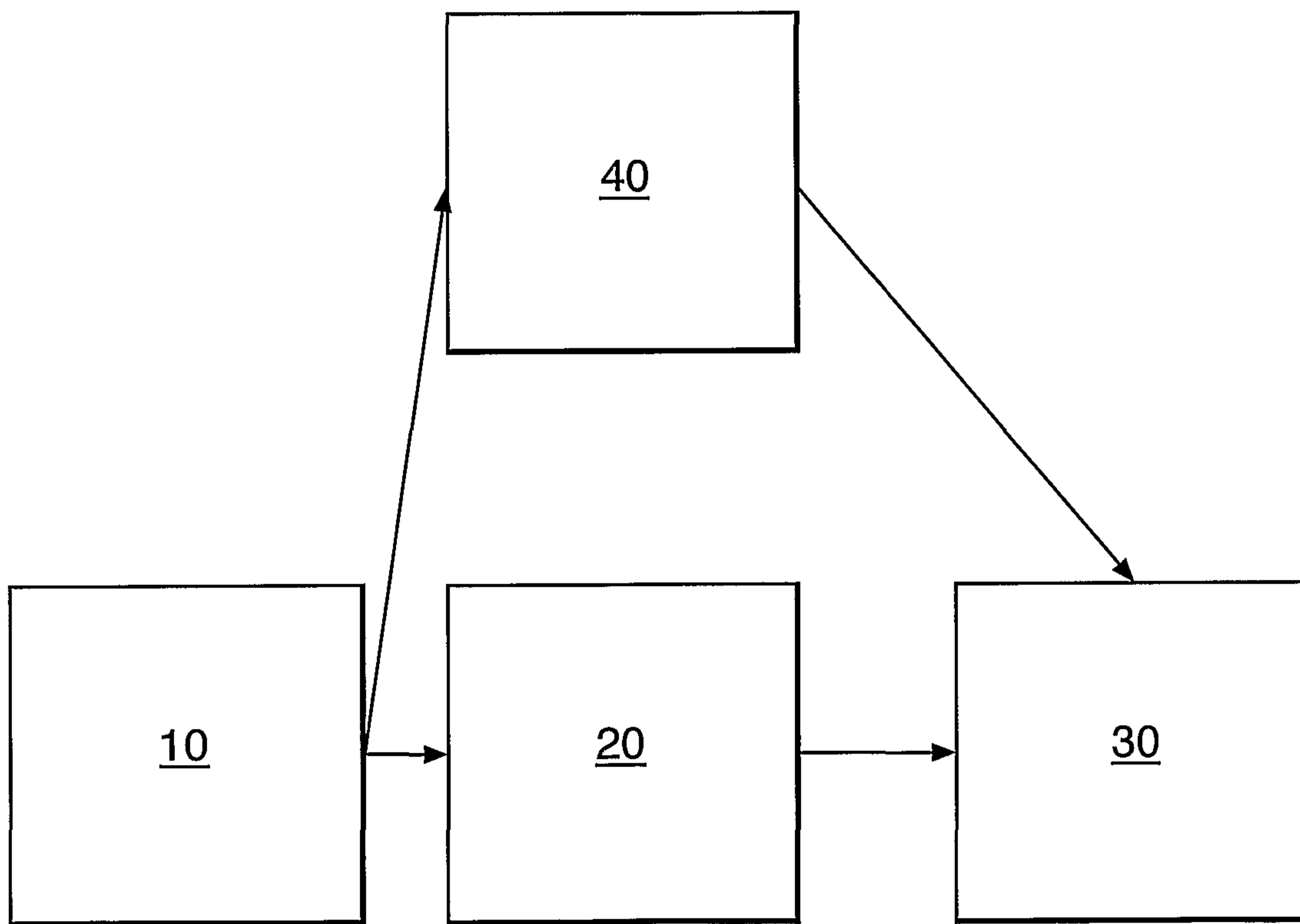


Fig.3.



3/5

Fig.4a.

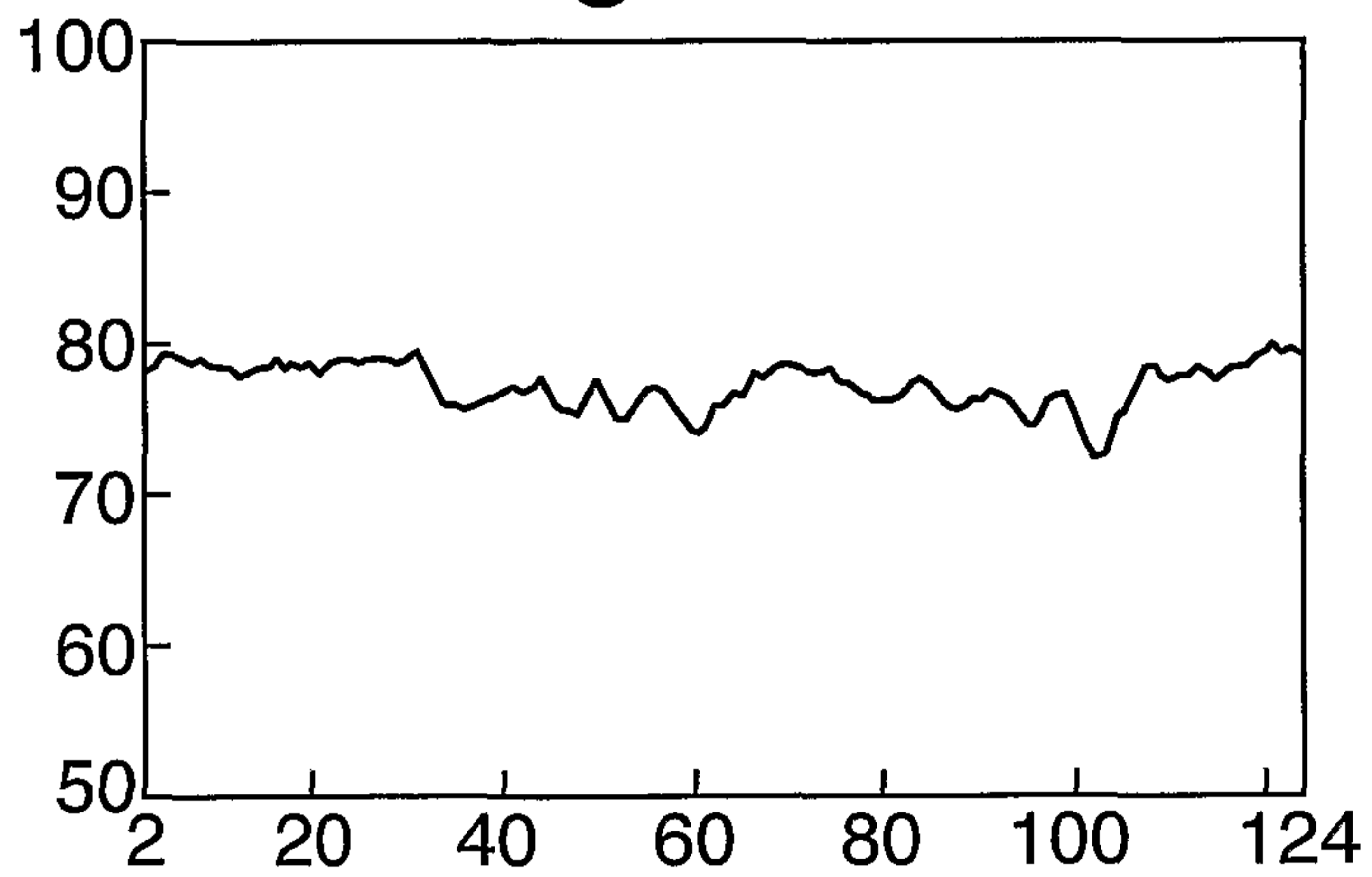


Fig.4b.

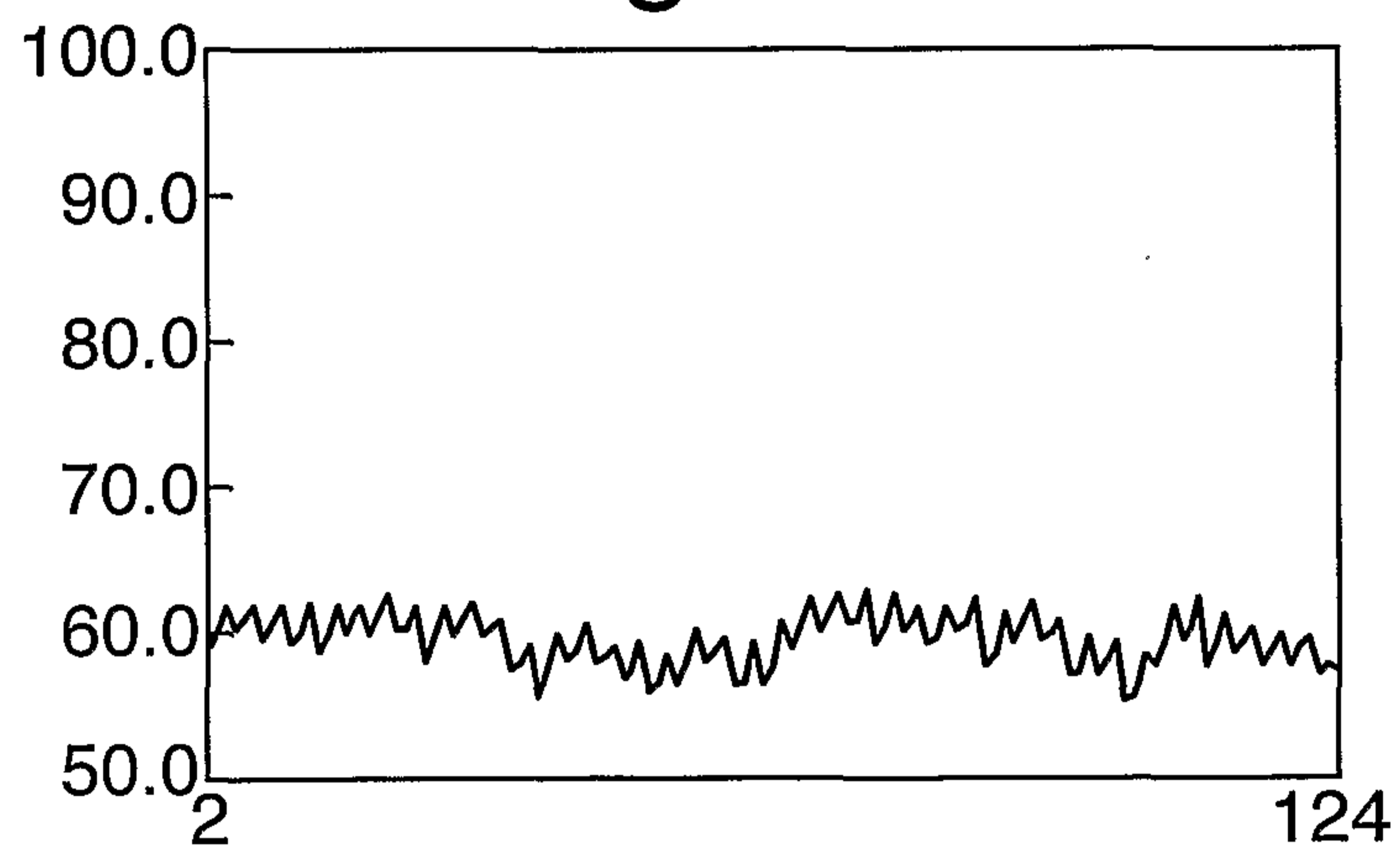
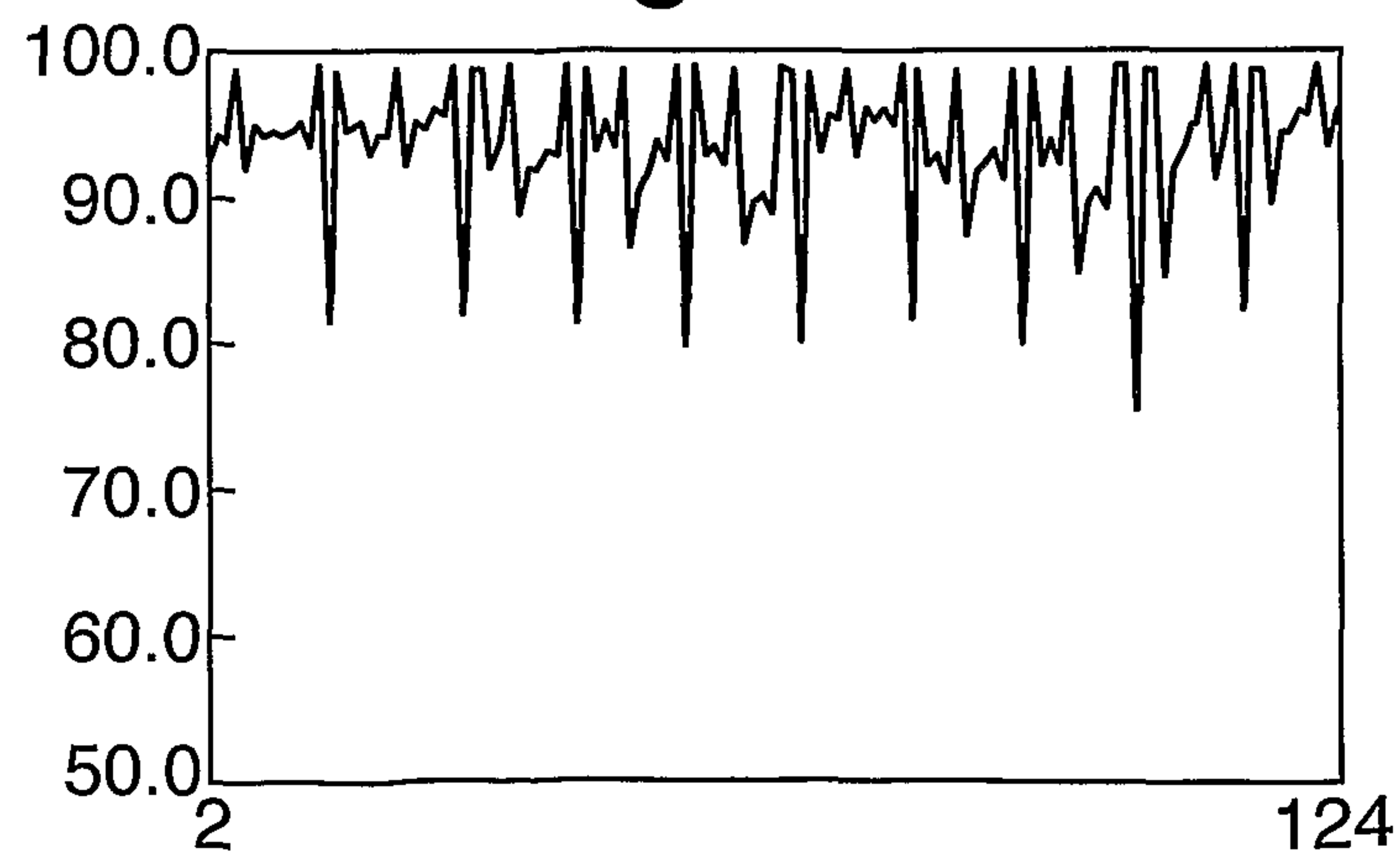


Fig.4c.



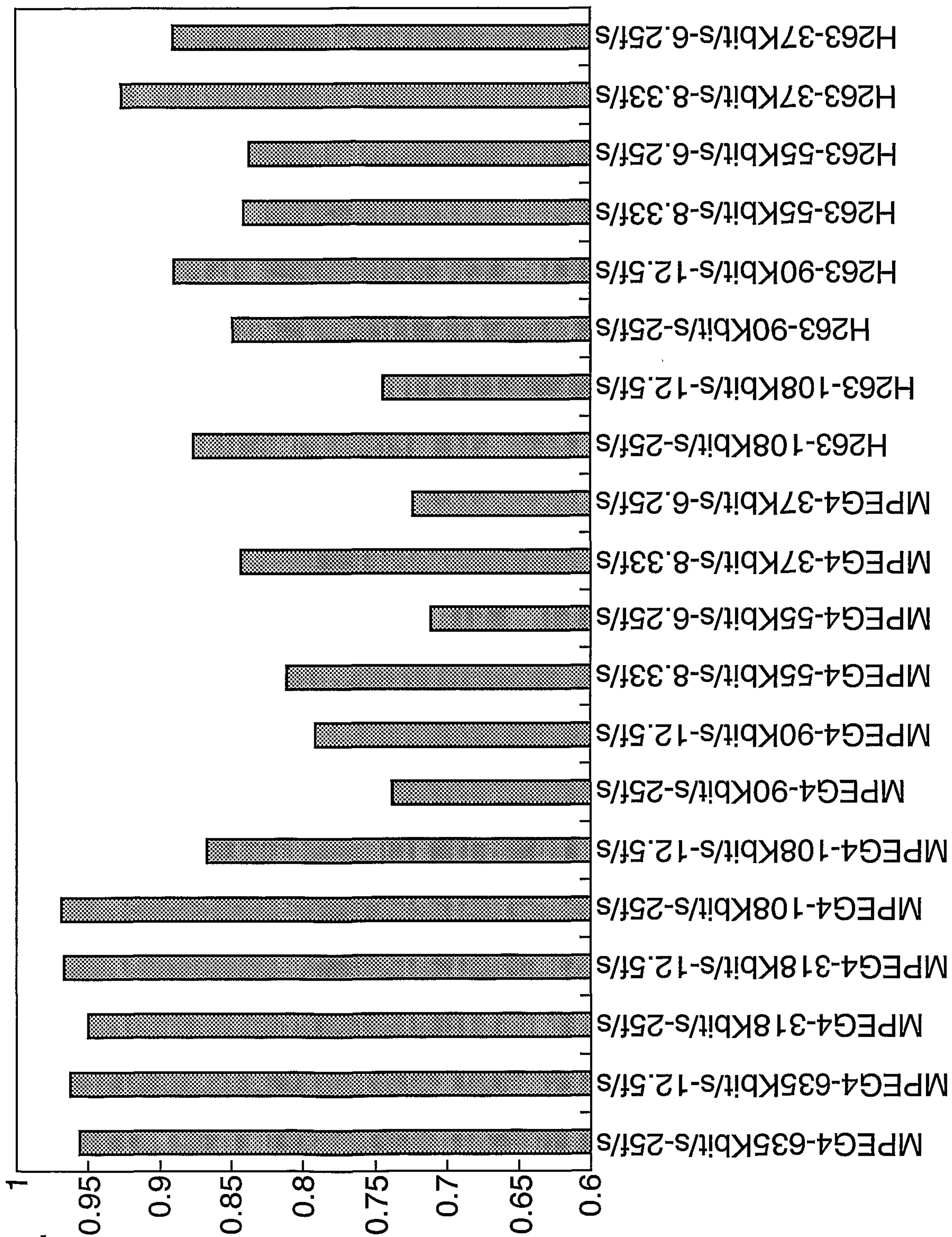


Fig.5.

5/5

Fig.6a.

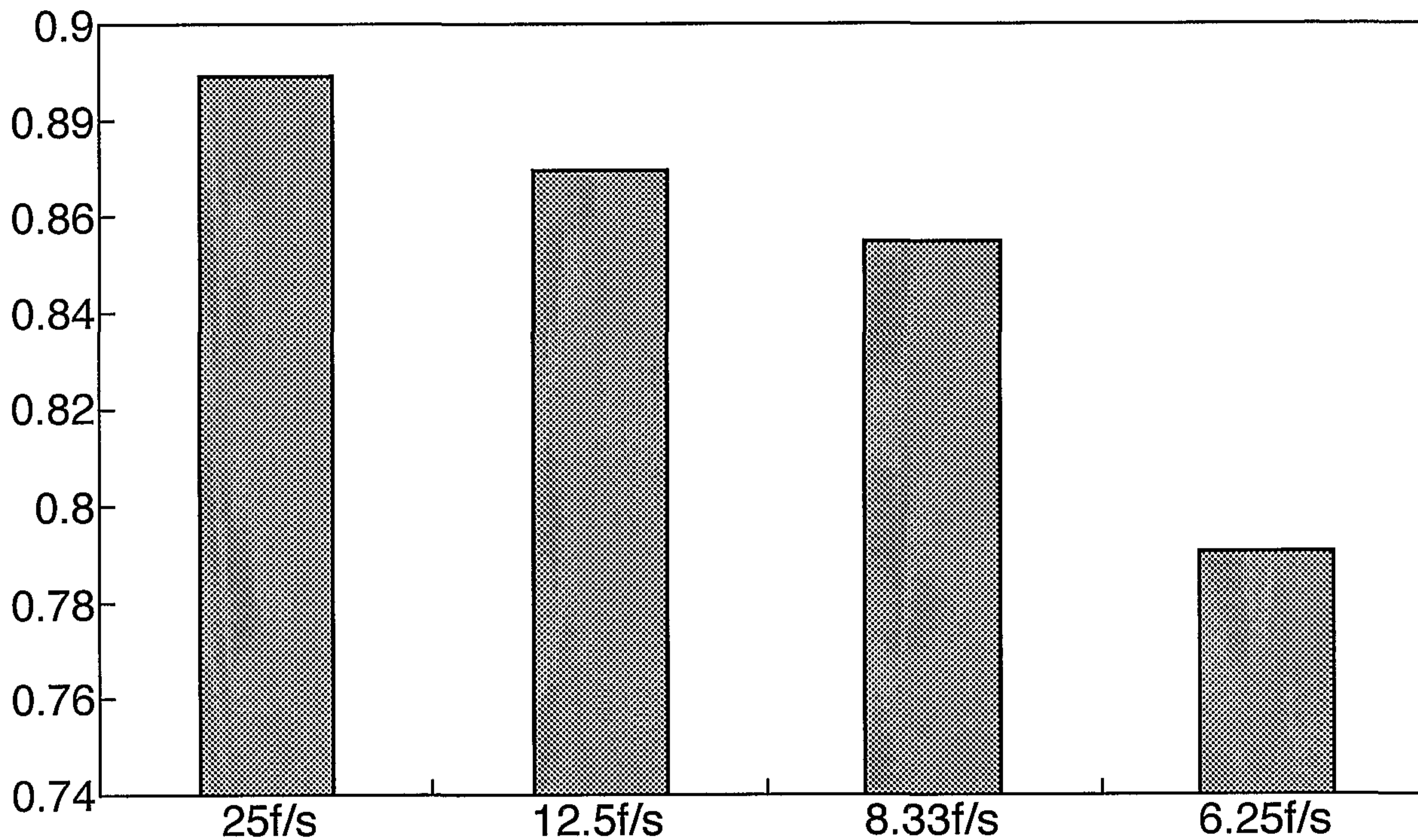


Fig.6b.

