Classifying postures of freely moving rodents with the help of Fourier descriptors and a neural network

D. J. HEEREN and A. R. COOLS
Nijmegen Institute of Neuroscience, Nijmegen, The Netherlands

A computerized method for classifying the postures of freely moving rodents is presented. The behavior of the rats was recorded on videotape by means of a camera hanging perpendicular to an open field. An automatic tracking system (10 images/sec) was used to transform the video images of postures into a binary image, thereby providing silhouettes in a computer format. The contours of these silhouettes were used for determining their characteristic features with the help of a Fourier transformation. The resulting features were classified with the help of a Kohonen network composed of 32 neurons. The four best *winning neurons*, rather than the usual one, were used for the classification. The resolution (11,090 distinct classes of postures), reliability (96.9%), and validity of this method were determined. With the use of the same approach, the effectiveness of this method for classifying behaviors was illustrated by analyzing grooming (247 grooming images vs. 4,950 nongrooming images). We found 15.4% false positives and 2.5% false negatives.

Today, several techniques are available for digitizing video images of an animal in space. These techniques are primarily used for analyzing spatiotemporal displacements of animals (Bonatz, Steiner, & Huston, 1995; Cools, Brachten, Heeren, Willemen, & Ellenbroek, 1990; Cools, Ellenbroek, Gingras, Engbersen, & Heeren, 1997; Dai & Carey, 1994; Gallagher, Burwell, & Burchinal, 1993; Kernan, Mullenix, & Hopper, 1988; Samsdodd, 1995; Spruijt & Gispen, 1983; Spruijt, Hol, & Rousseau, 1992). Still, these techniques have not yet resulted in an automatic and computerized method that allows a reproducible, reliable, valid, and objective analysis of *behaviors* and/or *postures*. The goal of the present study was to develop a computerized method that provides an objective description and classification of rodent postures that can ultimately lay the foundation for classifying behaviors.

An essential prerequisite for the development of such a method is pattern recognition. Pattern recognition is a generic topic in the field of artificial intelligence (AI): Using the large speed and enormous storage capacity of desktop computers, very sophisticated software programs have been developed for pattern recognition. For that reason, we used pattern recognition as it is assessed in AI. The cornerstones of pattern recognition that we applied are as follows.

1. Data were acquired and transformed into the input domain of the computer. Thus, video images of rodent postures were made and, subsequently, transformed into an adequate computer format. In practice, the actual postures were transformed into silhouettes, of which the contours were used for further processing.

2. The transformed data were reduced in order to increase the computational speed of the required processing. Relevant features were extracted, thereby removing redundant and irrelevant data. For this purpose, we used Fourier transformation. In comparison with other techniques that allow feature extraction, Fourier transformation has a number of unique advantages: It provides data reduction without losing information and results in a number of harmonics, of which each higher harmonic provides more details about the original contour. The contours were described as two periodic changing signals in time, because Fourier transformation requires such signals.

3. The extracted features were classified in such a way that the output that was requested resulted. Thus, the extraction should result in postures that are biologically relevant according to the expert. Neural networks are very apt for the construction of classification systems. In contrast to other statistical classifiers, a neural network is model free: It does not require knowledge about the underlying model (i.e., linear, quadratic, etc.). For that reason, we preferred the assessment of neural networks. In principle, there are two distinct types of networks: a supervised and an unsupervised network. The so-called supervised network requires a priori information about the desired output: Given the goal of the present study, the desired output would be classes of postures that had been previously classified with the help of ethological techniques (cf. Rousseau, Van Lochum, Gispen, & Spruijt, 2000). In contrast to the supervised network, the so-called unsupervised network produces an output according to its own rules; this kind of network has the advantage that

Correspondence concerning this article should be addressed to A. R. Cools, Department of Psychoneuropharmacology, Nijmegen Institute of Neuroscience, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands (e-mail: a.cools@pnf.kun.nl).
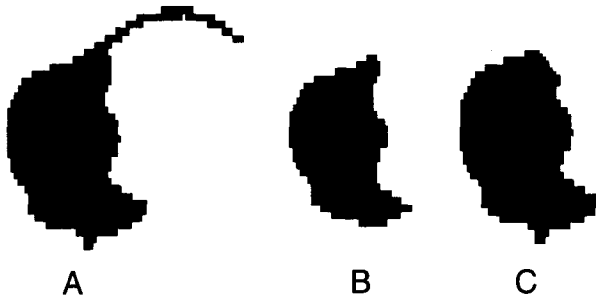
Figure 1. The original silhouette (A) was eroded until the silhouette lost its tail (B) and was dilated in order to regain the original size of the silhouette (C).

it can even classify "unknown" postures. Because manipulations such as gene therapy and the administration of drugs often produce fully abnormal and, accordingly, new behaviors, we preferred the assessment of the unsupervised network of Kohonen (Hecht-Nielsen, 1989).

The present study provides an expert database that automatically and objectively classifies postures of freely moving rodents. The kind of question put forward by the scientist determines whether the resulting posture or set of postures is biologically relevant. The study also illustrates that the newly developed method is suitable for constructing an automatic and objective classification of behaviors.

MATERIAL AND METHOD

Raw Data

The material that was used for testing the newly developed method was collected in two distinct sets of experiments. The ultimate aim of these experiments was to investigate the behavioral changes that were elicited by bilateral injections into the neostriatum. In Experiment 1, dexamphetamine was injected; in Experiment 2, the solvent of dexamphetamine sulpiride—namely, distilled water—was injected. Each experiment was done with 6 male Wistar rats, weighing 230–260 g at the time of the experiment, and lasted 50 min.

Data Acquisition

In general, the behavior of each rat was recorded on videotape by means of a camera (Sony DXC-107P) hanging perpendicular (height, 160 cm) to an open field (100 × 100 cm). Thus, each experiment provided six tapes, each of these lasting 50 min. A computerized automated tracking system (CCD camera; video processor, Fast FSP60; computer, Pentium 166, Windows 95; Cools et al., 1990) was used to capture 10 images per second. Each image consisted of 512 × 512 pixels, and each pixel had 256 intensity levels. This system was used to remove noise and to transform the video image of the rodent's posture into a binary image (rat, 1; background, 0). Thus, postures were transformed into binary images, forming silhouettes. Run length encoding was used to compress the data to approximately 6 Mb. The resulting data set was saved on hard disk.

Data Transformation

As was mentioned in the introduction, data reduction is essential for the computational speed that is required for the processing. Because there is a lot of redundancy in the data that slows down the processing and, in addition, hampers the classification, the data were reduced as follows.

First, each silhouette encompassing the whole rat was reduced to a silhouette of the body without a tail. For that purpose, the original silhouette (Figure 1A) was eroded (peeled off) until the silhouette had lost its tail (Figure 1B). Next, the silhouette was expanded (dilated) in order to regain the original size of the silhouette (Figure 1C). Subtraction of the latter silhouette from the original revealed the base of the tail (Figure 2B). Eroding and dilating are standard image-processing methods. The base was used as an anchor for the x-y position of the silhouette in the open field. Next, a contour of the silhouette was made (Figure 2C) and transformed into two periodically changing signals in time (Figure 3).

Second, the Fourier transformation was assessed to determine the characteristic features of the various contours. The method of Kuhl and Giardina (1982) was used for the calculation of the Fourier descriptors of the contours. This method produces harmonics, being characteristic features of the contours that are invariant for (1) the x-y position of the base of the tail in space, (2) the orientation of the rodent, and (3) the size of the rodent. At this step of the transformation, the contour is expressed in two sets of Fourier descriptors (x and y), where the first pair of harmonics (an ellipse) provides a raw approximation of the contour (Figure 2D). Each additional pair of harmonics provides more details and, accordingly, improves the approximation of the actual contour (Figure 2). Adding too many pairs of harmonics leads to details, such as ears, that are irrelevant for classifying the resulting contours in terms of biologically relevant postures (Figures 2G–2H). Using the unique advantage of the Fourier transformation that a contour can be reconstructed with the help of the harmonics, we assessed this so-called inverse Fourier transformation, to match the resulting contour with the original one: In the case of a mismatch, higher harmonics were added until the required match was found. With the use of this method, it was found that the first five pairs of harmonics were sufficient in this respect and were, accordingly, used for further analyses (Figure 2F).

Each harmonic consists of a real number and an imaginary number. Because five harmonics were used for describing a contour, each contour was characterized by 5 × 2 values for the x projection and 5 × 2 values for the y projection. Only 17 of the 20 numbers were used for the approximation of the contour, because 3 of these numbers were 0 or 1, as the result of the fact that the approximation was made invariant for the orientation and size of the rodent (see above).
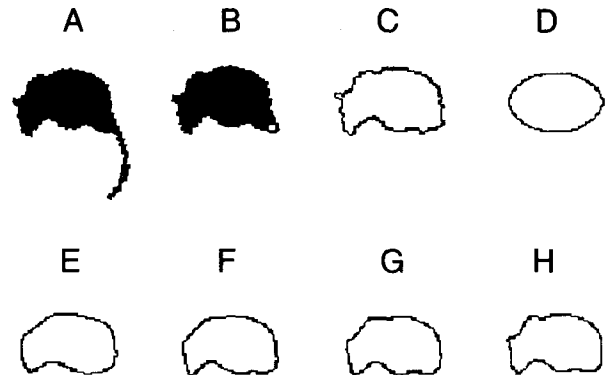

Figure 2. The tail of the silhouette (A) was removed by erosion and dilation (B). Next, the contour (C) was used for the calculation of the Fourier descriptors. Increasing the number of harmonics of the Fourier descriptors from one to nine (D = 1, E = 3, F = 5, G = 7, and H = 9) shows that the first five harmonics were sufficient to describe the contour of a silhouette. Using more than five harmonics (G, H) will add details that are not relevant (ears) for classifying postures.
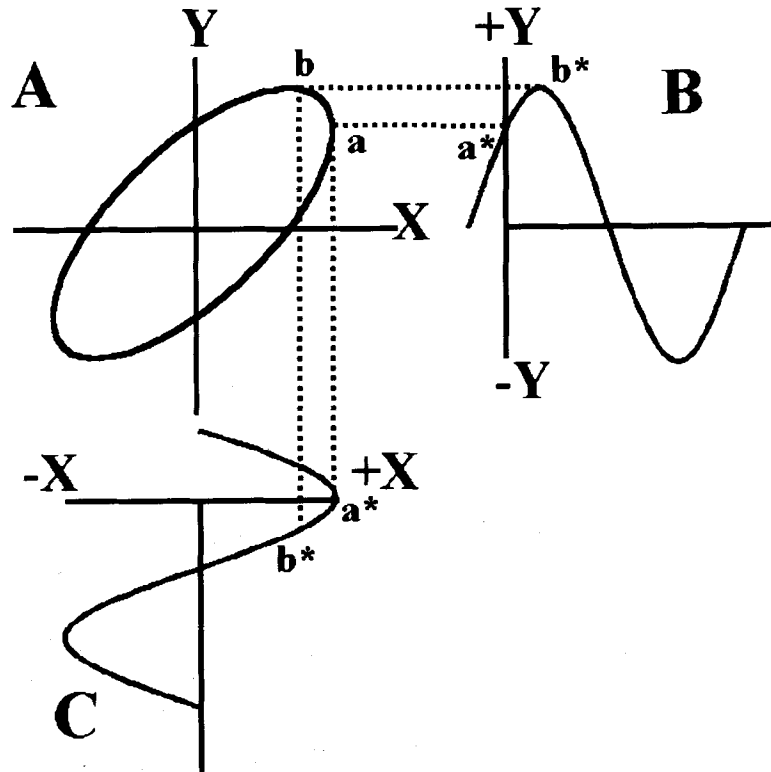
Figure 3. Projection of each point of Contour A on the x-axis and y-axis provides two changing signals in time (B, C). For the sake of simplicity, we have taken an ellipse as the contour.

Kohonen Network: General

The resulting contours were classified with the help of a Kohonen neural network. This network classifies contours according to its own rules, implying that a human expert has to determine the biological relevance of each resulting contour by using one of the following methods: (1) The expert uses the library of ethologically accepted postures in order to match the resulting contours, or (2) the expert designs an experiment that allows him to establish whether the resulting contour provides biologically relevant information. The originally stored image or the approximation of the contour can be used for the labeling of the postures, whereby the definition used depends on the objective: building a relational database that provides yes/no answers or building a fuzzy database that provides answers with a calculated confidence factor. A relational database is apt for classifying contours in terms of biologically relevant postures, whereas a fuzzy database is required for classifying behaviors (see below).

Kohonen Network: Classification

Generally, the learning and recall algorithms for a Kohonen network are based on the geometrical layout of the processing elements, the so-called *neurons*. Briefly, an input vector (data) is presented to the network. Next, the euclidian distance between the input vector and the weight vector of each of the neurons is calculated, and the neuron that is closest to the input vector is the so-called *winner*. As a result of training the network with a particular training set of data, the input vectors are attributed to a particular neuron that represents a cluster of typical features (Figure 4: Neuron 1 is the winner of Input Vectors A–C). This trained network is subsequently used for classifying new input vectors, with the help

of these neurons. The result is that each new input vector is attributed to a particular neuron—that is, the winning neuron—and all input vectors that are attributed to the same neuron are grouped together into one class. For a more detailed description, the reader is referred to Kasabov (1996). This approach has two disadvantages: (1) a network with N neurons can maximally distinguish N distinct classes, and (2) the input vectors can only be grouped into classes that result from training the network, implying that no new classes are created. These problems have to be solved, because (1) the actual number of distinct postures is unknown, and (2) several manipulations with rodents can produce types of distinct postures/silhouettes/contours that do not occur in naive animals.

Therefore, we increased the resolution of the classification system by creating the possibility of classifying a single contour with the help of four neurons rather than just one neuron—namely, the winning neuron, as well as the second, third, and fourth best neurons. Accordingly, each contour got a four-digit code rather than a single-digit code. Thus, this approach results in a network with N neurons that can theoretically distinguish $N \times (N - 1) \times (N - 2) \times (N - 3)$ distinct classes. Furthermore, this network can create new classes. Let us assume that Input Vector A is attributed to Neuron 1 after training, the new Input Vector B will also be attributed to this neuron, given that a single-digit code is used for the classification (Figure 5); in that case, Input Vectors A and B are grouped into a single class. In contrast, using the four-digit code results in the creation of a new class for Input Vector B (Figure 5B: 1–2–4–3), apart from the already existing class for Input Vector A (Figure 5A: 1–4–3–2).

Thirty-two neurons were incorporated into the network, after it was found that networks with either 16 or 64 neurons did not comply with our aim of classifying the resulting contours in terms of bi-
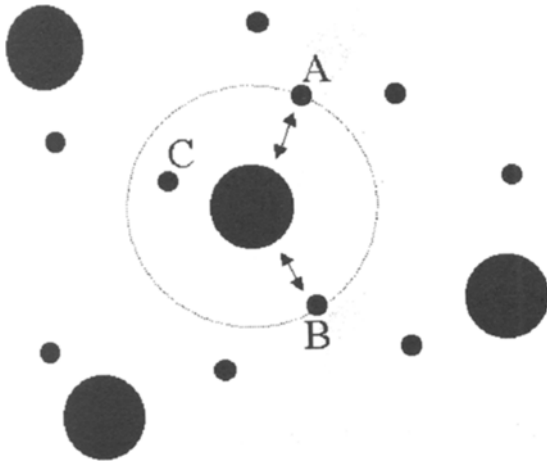
Figure 4. The network cannot distinguish Input Vectors (Contours) A and B, because both input vectors have exactly the same distance to the winning neuron.

ologically relevant postures. The network was trained by repeatedly presenting a particular set of data until an equilibrium was reached: In this respect, 500 iterations turned out to be sufficient. The sequence of the samples in the training set was randomized, to prevent the initial set of data from creating a bias in the system. The learning rate ($\alpha = e^{-\text{iter}/100}$) was decreased after each iteration. Next, a new or the same set of data was analyzed with the trained network.

RESULTS

Resolution of the Classification System

The resolution of the classification system that is applied in this study is determined by the number of neurons used for coding the postures. As was mentioned above, the network contained 32 neurons, of which a combination of 4 neurons was used to classify postures. Thus, each posture could be labeled by a four-digit code, theoretically producing thereby 863,040 [$N(N-1)(N-2)(N-3)$] distinct postures; this number greatly exceeds the 32 distinct postures that are found after using a single-digit code. In order to evaluate the resolution of this classification system, we trained the network by presenting a set of data from a so-called training rat (Experiment 1, Rat 1: 0–10 min: 6,000 samples). Next, a new set of data from this rat (Experiment 1, Rat 1: 0–50 min: 30,000 samples) was analyzed with the trained network. The trained network distinguished 11,090 distinct postures. As was mentioned, a human expert has to determine the biological relevance of each of these postures. In other words, the original problem of "splitting and lumping" in ethology is now a matter of labeling a particular posture with a one-digit, two-digit, three-digit, or four-digit code, depending on the number of digits necessary for distinguishing this posture from other biologically relevant postures (Figure 6).

Reliability of the Classification System

We consider a system among other things reliable when the outcome is more or less independent of its training set. In order to investigate this aspect of reliability, two distinct training sets were used to classify a single set of data. First, the network that was trained with the data from a training rat (Experiment 1, Rat 1, 0–10 min: 6,000 samples) was used to analyze the data of the so-called experimental rat (Experiment 1, Rat 2, 0–50 min: 30,000 samples): The network distinguished 9,562 postures. Second, the network was retrained with the data from the experimental rat (Experiment 1, Rat 2, 0–10 min: 6,000 samples) and, subsequently, was fed with the above-mentioned experimental set (Experiment 1, Rat 2, 0–50 min: 30,000 samples): The network now distinguished 9,261 distinct contours. Thus, the difference between both analyses was less than 3.1%, underlining the large reliability of the chosen approach. This result shows that the outcome was nearly independent of the training set used.

Effectiveness of the Classification System

The effectiveness of the method for detecting new postures was assessed as follows. First, the design of the resolution test was used. Thus, the network was initially trained with data from a training rat (Experiment 1, Rat 1, 0–10 min: 6,000 samples). Next, a new set of data from the same rat (Experiment 1, Rat 1, 0–50 min: 30,000 samples) was analyzed, and 11,090 silhouettes were distinguished. Finally, the data from another rat (Experiment 1, Rat 3, 0–50 min: 30,000 samples) were presented: The network detected 5,752 new postures among the 9,562 postures that were classified. In this case, too, it is
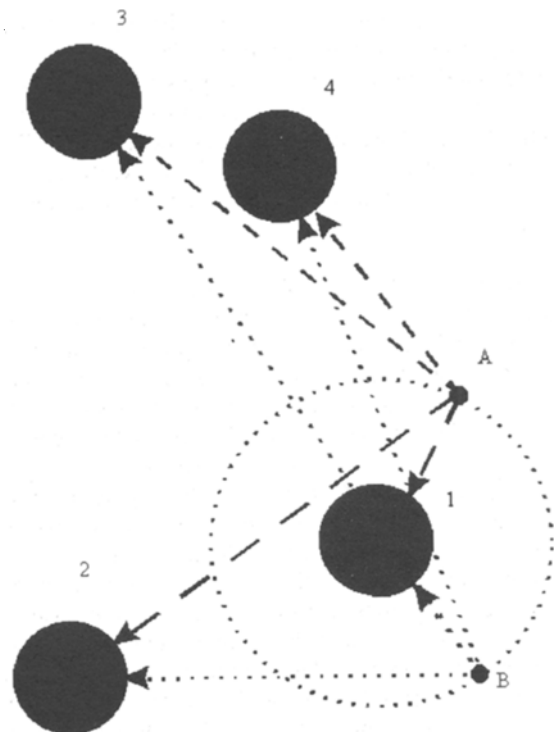

Figure 5. Using the arrangement of the four best winning neurons, it is possible to distinguish Input Vector A (1–4–3–2) and Input Vector B (1–2–4–3).
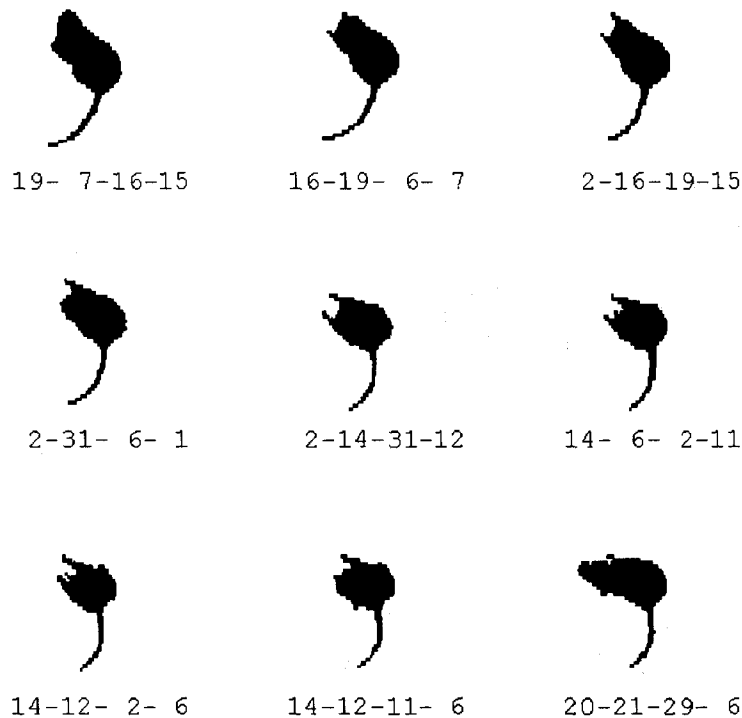
**Figure 6. Nine successive video images (0.1 sec/image) of the behavior *rearing*
show that the method even distinguishes nearly identical images by assigning dis-
tinct codes. Note that it is sometimes the third best winning neuron that reveals the
difference (see 14–12–2–6 and 14–12–11–6).**

the human observer that has to determine the biological
relevance of each of these postures. Theoretically, it is
possible that the postures in the samples from the 2nd rat
are all within-class variations of postures in the samples
from the 1st rat. However, this does not occur when all of
the postures that are considered to be biologically relevant
according to the human expert are described with a four-
digit code.

Fail-Safeness of the Classification System

We consider a classification system of animal postures
fail-safe when it produces neither false positives, implying
that it does not produce within-class variations of pos-
tures, nor false negatives, implying that it does not lump
postures that have to be considered as distinct postures in
accordance with their biological relevance.

As long as the system is used for classifying postures
in accordance with ethological principles, this fail-safe
problem does not exist, for it is the human expert who
determines whether a one-digit, two-digit, three-digit, or
four-digit code has to be attributed to a particular posture.
Using the four-digit codes, the human expert matches all
the available postures that have the same four-digit code,
excepting the fourth digit; he will lump these as long as
they have to be considered to be within-class variations
of ethologically identified postures, but he will split these
as long as they have to be considered to be distinct types
of ethologically identified postures. Of course, it is also

possible to use codes that have only three digits (or even
two digits) for the classification. In principle, the same
procedure has to be followed in that case.

If the classification system, however, is used for de-
tecting and classifying spontaneously occurring or exper-
imentally induced changes in behaviors, especially pos-
tures, it is necessary to prevent the occurrence of false
positives. In order to investigate this aspect of the classi-
fication system, the outcome of the classification system
was compared with the results of a frame-by-frame analy-
sis of the videotapes. A series of 100 identical four-digit
codes was selected and randomly mixed with a series of
four-digit codes, of which the fourth digit varied. The
video images that belonged to these codes were visually
analyzed by a human expert (A.R.C.). His task was to split
the video images into two classes—namely, a class with
identical video images (Class A) and a class with dissim-
ilar video images (Class B). Subsequent analysis revealed
that all identical four-digit codes represented video im-
ages that were clustered in Class A, implying that the
system did not produce false positives. Given these results
it appears that the system's ability to distinguish distinct
postures is equal to or even better than that of humans.

Validity of the Classification System

We consider a classification system to be valid when
the classifier has the ability to generalize. This implies
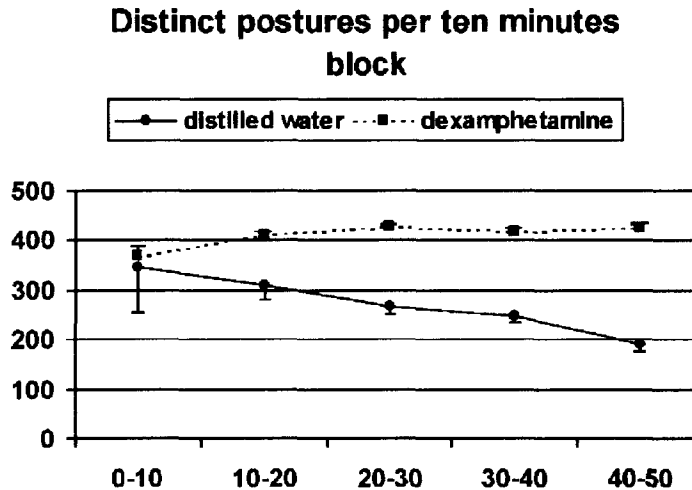that the training set has to produce classes that not only

Distinct postures per ten minutes block

Figure 7. Means and *SEMs* are given for a control group (*n* = 6) and a treated group (*n* = 6). The two groups differ significantly in time (see the text).

classify the data of the training set, but also classify independent data. In order to illustrate this aspect of our classification system, we used the following data for our training set: The initial 10 min of Rats 1 and 2 per experiment (*n* = 3) were lumped and formed the training set for the network (Experiment 1, 2 rats, 0–10 min + Experiment 2, 2 rats, 0–10 min + Experiment 3, 2 rats, 0–10 min: 360,000 samples). We used the initial 10 min of Rats 3–6 of Experiment 1 as independent data. Only the winner and the second best neurons were used for coding the postures in this case. After training the network, we determined the number of unique classes of postures for Rats 1 and 2 (Experiment 1) in the data of the training set and compared these unique classes with those for Rats 3, 4, 5, and 6 of Experiment 1 (independent data set). This analysis revealed that 94% of the total number of distinct classes for Rat 1 of the training group (273) was also found in the independent set of data (*n* = 4), showing that the classifier had not specialized itself to the peculiarities of the training set. A similar analysis with Rat 2 of Experiment 1 revealed that even 98% of the total number of distinct classes for Rat 2 of the training group (272) was also found in the independent set of data (*n* = 4). In sum, both sets of results clearly show that the present classification system has the ability to generalize.

Classification of Postures: An Illustration

In order to illustrate the usefulness of the present classification system, the behavioral effects of two manipulations were analyzed, with the help of this system. For that purpose, the data sets of Experiment 1 (striatal injections of dexamphetamine; *n* = 6) and Experiment 2 (striatal injections of the solvent of dexamphetamine; *n* = 6) were analyzed, after the network was trained with the set that was also used for testing the validity of the system (see above). Figure 7 shows that dexamphetamine significantly enhances the total number of distinct classes of

postures in time [two-way analysis of variance, $F(4,50)$ = 9.5, $p < .001$]. This figure also illustrates that a training set encompassing the data of 6 rats is representative for the whole population, for the *SEM* is extremely small in both groups of rats.

Classification of Behaviors

To illustrate that the present classification system of *postures* lays the foundation for a classification system of *behaviors*, the network was used to trace the display of grooming. First, a frame-by-frame analysis (6,000 samples) was used to identify the display of grooming in time; this was done with the data of Rat A and Rat B. Next, the network was trained with the data of Rat A (6,000 samples). Then, the trained network was used to trace the codes of the grooming postures in Rat A; grooming postures that lasted no longer than 0.2 sec were removed, because grooming is by definition, a behavioral item that lasts 0.2 or more seconds. Even when just one neuron was used for coding these postures, 0.6% false positives (1 out of 155) and 0% false negatives (0 out of 2,845) were found. To validate this approach, the trained network was used to trace the codes of grooming postures in Rat B; again, grooming postures that lasted no longer than 0.2 sec were removed. The network found 15.4% false positives (38 out of 247) and 2.5% false negatives (123 out of 4,950) in case just one neuron for the coding was used. The presence of false positives was due to the fact that a particular posture that was found to be specific for the grooming sequence of Rat A also occurred in the sequence that coded pivoting in Rat B. Thus, the relational database that was used provided a yes/no answer, implying that 100% of the selected postures represented grooming despite the 15.4% false positives. Using a fuzzy database, however, would have made it possible to state that it was grooming that occurred with a calculated confidence factor of <.5 in these cases, illustrating the advantage of a fuzzy

database over a relational database. The presence of false negatives was due to the fact that the training set of Rat A did not incorporate all of the distinct types of grooming. Because it was not the primary aim of the present study to provide a classification system for *behaviors*, we did not further improve the present classification system. Nevertheless, it will be evident that the classification of behavioral items can be simply improved by (1) incorporating the sequence of postures in time and (2) adding data that contain all the distinct types of grooming.

Miscellaneous

D.J.H., at the Department of Psychopharmacology of the Nijmegen Institute of Neuroscience, developed all the software. The system consists of three separate programs: (1) a program for data acquisition that collects 10 images per second (real time); (2) a program for determining the contours and calculating the Fourier descriptors that requires about 2 h for off-line analysis of 30,000 samples; and (3) a program for classification that takes 1 min for the analysis of 30,000 samples. Once the three programs are combined, on-line real-time processing of 5 images per second is feasible. By using a faster computer and faster algorithms, 10 images per second will become feasible in the near future.

CONCLUSION

The method described provides a fully objective method for automatically analyzing postures and behavioral items in a reproducible, reliable, and valid manner. Because no a priori information about the postures has to be given, this method is also suitable for delineating new postures that are not present in the training data.

REFERENCES

BONATZ, A. E., STEINER, H., & HUSTON, J. P. (1995). Video image analysis of behavior by microcomputer: Categorization of turning and locomotion after 6-OHDAA injection into the substantia nigra. *Journal of Neuroscience Methods*, **22**, 13-26.

COOLS, A. R., BRACHTEN, R., HEEREN, D., WILLEMEN, A., & ELLENBROEK, B. (1990). Search after neurobiological profile of individual-specific features of Wistar rats. *Brain Research Bulletin*, **24**, 49-69.

COOLS, A. R., ELLENBROEK, B. A., GINGRAS, M. A., ENGBERSEN, A., & HEEREN, D. (1997). Differences in vulnerability and susceptibility to dexamphetamine in Nijmegen high and low responders to novelty: A dose-effect analysis of spatiotemporal programming of behaviour. *Psychopharmacology*, **132**, 181-187.

DAI, H., & CAREY, R. J. (1994). A new method to quantify behavioral attention to a stimulus object in a modified open-field. *Journal of Neuroscience Methods*, **53**, 29-34.

GALLAGHER, M., BURWELL, R., & BURCHINAL, M. R. (1993). Severity of spatial learning impairment in aging: Development of a learning index for performance in the Morris water maze. *Behavioral Neuroscience*, **107**, 618-626.

HECHT-NIELSEN, R. (1989). Learning laws: Self-adaptation equations. In R. Hecht-Nielsen (Ed.), *Neurocomputing* (pp. 45-78). Menlo Park, CA: Addison-Wesley.

KASABOV, N. K. (1996). Theoretical and computational models o neural networks. In N. K. Kasabov (Ed.), *Foundations of neural net works, fuzzy systems, and knowledge engineering* (pp. 252-329). Cambridge, MA: MIT Press.

KERNAN, W. J., JR., MULLENIX, P. J., & HOPPER, D. L. (1988). Pattern recognition of rat behavior. *Pharmacology, Biochemistry & Behavior*, **27**, 559-564.

KUHL, F. P., & GIARDINA, C. R. (1982). Elliptic Fourier features of closed contour. *Computer Graphics & Image Processing*, **18**, 236-258.

ROUSSEAU, J. B. I., VAN LOCHUM, P. B. A., GISPEN, W. H., & SPRUIJT, B. M. (2000). Classification of rat behavior with an image processing method and a neural network. *Behavior Research Methods, Instruments, & Computers*, **32**, 63-71.

SAMSDODD, F. (1995). Automation of the social interaction test by video-tracking system: Behavioural effects of repeated phencyclidin treatment. *Journal of Neuroscience Methods*, **59**, 157-167.

SPRUIJT, B. M., & GISPEN, W. H. (1983). Prolonged animal observatio by use of digitized videodisplays. *Pharmacology, Biochemistry & Behavior*, **19**, 765-769.

SPRUIJT, B. M., HOL, T., & ROUSSEAU, J. (1992). Approach, avoidance and contact behavior of individually recognized animals automati cally quantified with an imaging technique. *Physiology & Behavior* **51**, 747-752.