# Recognizing Song–Based Blink Patterns: Applications for Restricted and Universal Access

Tracy Westeyn & Thad Starner
GVU, College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
{turtle, thad}@cc.gatech.edu

## Abstract

*We introduce a novel system for recognizing patterns of eye blinks for use in assistive technology interfaces and security systems. First, we present a blink-based interface for controlling devices. Well known songs are used as the cadence for the blinked patterns. Our system distinguishes between ten similar patterns with 99.0% accuracy. Second, we present a method for identifying individual people based on the characteristics of how they perform a specific pattern (their "blinkprint"). This technique could be used in conjunction with face recognition for security systems. We are able to distinguish between nine individuals with 82.02% accuracy based solely on how they blink the same pattern.*

## 1. Introduction

A small population of people possess myopathies (muscular diseases or injuries) that inhibit their voluntary muscle movement, or in less extreme cases, limit their motor skills. For example, 30,000 people in the United States are afflicted with Amyotrophic Lateral Sclerosis (ALS, or Lou Gehrig's disease) [3]. In the United States two percent of brainstem stroke victims (2,500 people per year) survive paralyzed in the locked-in state [4]. Patients in this condition, known as Locked-in Syndrome, suffer from complete paralysis of all voluntary muscles except for those that control eye movement. People suffering from such crippling diseases still possess the cognitive abilities for communication and other functions, but they lack a method for interacting with the people and environment surrounding them.

Eyegaze tracking devices such as EagleEyes [10] can provide a practical method of communication for some individuals; however, eye tracking is not suitable for all members of this population. People with stroke related myopathies often have eye control problems. Nystagmus (jerky, involuntary eye movement) and disconjugate ocular movement (the inability to move the eyes as a pair) can be problematic for eyegaze systems and can result in tracking failure [12]. However, such conditions do not hinder a person's ability to blink. Research suggests that blinking patterns, such as variations of Morse code, can be an effective form of communication for a certain population of the disabled community [7]. However, there are drawbacks to Morse code based systems. On average, learning Morse code requires 20 to 30 hours of training. In addition to the training time, communication with Morse code is cognitively demanding for a novice [2]. While rates of thirty words per minute [7] can be achieved through mechanical devices for experienced telegraphers, it is not clear what communication rates can be achieved through blinking for the disabled population.

In this paper we present two applications: an interface for universal access, "BlinkI", and one for restricted access, "Prescott." BlinkI can be used as the basis of an assistive system for people with severe disabilities, allowing blink patterns to be used in communication or to control devices in the environment. Prescott, on the other hand, uses the pattern of blinks to identify the individual performing the pattern, rather than the pattern itself. Both applications are based on the idea of using songs as a cadence for the blinking. Songs based on the functionality of the device may be easier to associate as an interface command. For example, the opening rhythm of "Frosty the Snowman" could be used to lower the setting of a thermostat.

Using songs as the basis of blinked commands allows for more information to be encoded: not only does the blink itself provide information (i.e., number of blinks per sequence), but its temporal relation with other blinks in the pattern (i.e., the rhythm) also encodes information. The interval between the blinks can be used to distinguish between sequences with the same number of blinks (see Figure 1).

This method of encoding can only be practical if it is expressive enough to span a large number of sequences. We believe that, due to effects of rythm and syncopation, bit rates of up to 8 bps, or 480 bits per minute, can be acheived through blinking songs. The ultimate goal of our system

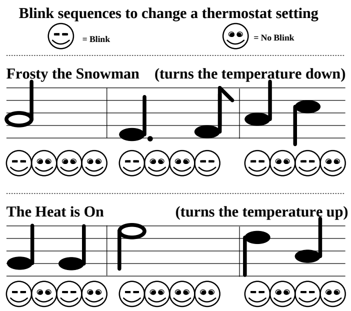


**Figure 1.** *Two song based blink sequences, each with 5 blinks, for controlling a thermostat. While each sequence has the same number of blinks, the temporal relationship allows the two patterns to be distinguishable.*

is to provide a signifiant fraction of normal converstaional rates, at 126–172 words per minute, or about 3,780 bits per minute (assuming 5 bits/character). Meanwhile, it suffices that we have the capability to provide more bandwidth than single switch communication devices. Such devices provide the minimum rate tolerable for interactive conversation at three to nine words per minute [8], or about 270 bits per minute.

The remainder of the paper is arranged as follows: first, we will address the hardware requirements, algorithms used and metrics for evaluation; second, we will discuss in detail experiments and results involving blink pattern recognition using BlinkI; third, we will present work involving the issues and techniques concerning Prescott as well as preliminary results; last we present an overall discussion, related work, and our conclusions.

## 2. Blink Detection

In order to explore the feasibility of recognizing distinct song–based blinked patterns, we constructed a prototype video capture system. While more precise methods of blink detection exist, such as using electromyography to detect electrical signals sent from the muscles that control eye blinks, we decided to build a cost–effective system that did not require augmentation of the user. The minimal hardware required by our system is a low resolution camera, a standard computer (PentiumIII 600MHz with 256MB RAM), and a method to ensure approximate head alignment of the user.

For our initial experiments, we collected blinking data from ten different users. The participants were required to situate themselves in front of a camera and a video monitor. They then had to align themselves in such a way that the video image of their face matched a template affixed to the video monitor. Once aligned, the participant blinked several repetitions of the desired pattern. The video was recorded and later post–processed to ensure that no frames of data were lost. Data was collected from all individuals in three separate locations over the course of two months.

When blinking song–based sequences, patterns of rhythm are indicated only by the duration of time that the eye remains open between blinks, rather than how long the eye is kept closed. This allows users to blink rhythms using a natural blink. Therefore, during a blink, the eyelid is always in motion. We use the well–known technique of optical flow [13] to detect frames of video where this motion occurs. Using optical flow helps to provide robustness by producing features that remain somewhat consistent across varying environments.

Optical flow represents the movement of a pixel as a velocity vector expressing the magnitude ($\rho$) and direction of change ($\theta$) from a previous frame. Although some head movement is inevitable, the motion of blinking will be much more rapid in comparison. This fact allows us to filter out pixels with low velocities, and keep only those associated with blinking. The mean velocity of the pixels for each frame is calculated and used to determine whether that frame was part of a blink or not. Once data for an entire pattern is collected, the number of blinks, spaces, and the duration of each is calculated.

By having our users align their faces to a template placed on the monitor, we can calculate optical flow for only a small portion of the video (a 64x64 pixel region over the left eye). Using the template allows optical flow calculation to remain practical; however this may not be required: Betke et. al. showed that the motion of the eyelids is sufficient for detection of the eyes' location in video [5].

## 3. Blinked Pattern Classification

A sequence of blinks is identified by the temporally ordered collection of interval lengths between blinks. This encodes the number of blinks and the duration (in frames of video) of the length between the blinks. For example, the 7–blink sequence for "The Star Spangled Banner" would be represented by the string "6,4,12,10,12,32". Because there are six elements in the string we can infer that there are seven blinks in this pattern (see Table 1). Inspection of the relative sizes of the intervals demonstrates that the first two blinks in the sequence occured rapidly compared to the rest of the sequence which was more evenly spaced.

After a sequence is detected, it is encoded according to the process described above. An observed sequence of blinks is classified by using a nearest neighbor classification scheme. Because of temporal variations, the sequences cannot be compared directly. For example, just because a string has seven elements it does not mean the string represents "The Star Spangled Banner". Not only must the number of elements match, but the elements of the string must also encode the same relative duration of two quick blinks followed by five longer blinks. A direct comparison would also have difficulty dealing with missing or extraneous blinks that might occur in the sequence as a result of either system error, user error, or both. If performed quickly enough, two blinks can often appear as one blink to the segmentation algorithm.

Dynamic time warping (DTW) [9] allows two temporal sequences to be compared and can be used as a metric for nearest neighbor classification. DTW provides a measure of similarity between two sequences that factors in both the length of the sequence and the temporal relation of the elements in the sequence. In order for a match to score well, it must have the same number of blinks and the same relative duration of space between each blink in the sequence. DTW is also flexible enough to cope with missing or extraneous blinks included into the sequence. Assuming that the relative timing is intact until the extraneous blink occurs, the score will only be slightly effected. The correct sequence plus or minus a blink should still match better than other sequences in the database.

The performance measure for the system is based on the accuracy of recognition. Our metric for accuracy reflects that recognition is performed in isolation and only incorporates substitution errors (mistaking one class for another). If we let $S$ represent substitution errors and $N$ represent the total number of samples, then the measurement of accuracy is defined to be: $Accuracy = \frac{N-S}{N}$.

## 4. BlinkI Experiment: Blinked Sequence Recognition

We performed an experiment to determine if the BlinkI system could accurately distinguish between distinct blinked patterns. A single individual blinked a total of ten distinct patterns based on cadences of the opening rhythm for the songs listed in Table 1. These songs were chosen because of their simplistic rhythms, relative lengths, similarity to each other, and popularity. Most of the patterns are very similar in the number of blinks. The participant performed twenty to forty repetitions of each song. A total of 303 examples were collected. Using leave–out–one validation (all possible combinations of the 303 samples with one test sample and 302 training samples), the system correctly classifies 300 of the 303 examples. This is an overall accuracy of 99.0%. Table 1 shows the performance accuracy of the system.

We intentionally selected songs that were similar in order to test the robustness of the classification scheme. The songs were either similar in duration, rhythm, number of blinks, or some combination of the three. For example, the pattern based on "This Old Man" was confused for the pattern based on "God Save the Queen." These patterns are similar in both duration and rhythm (see Figure 2). In reality the patterns differ in length by three blinks; however, if the system accidently detects an extra blink or omits a blink, the sequences can become practically identical.

| Class | Song Title | Total Blinks | Accuracy |
|---|---|---|---|
| 1 | Star Spangled Banner | 7 | 97.6% |
| 2 | God Save the Queen | 11 | 100.0% |
| 3 | Happy Birthday | 13 | 100.0% |
| 4 | This Old Man | 14 | 97.4% |
| 5 | When Johnny Comes Marching Home | 14 | 97.1% |
| 6 | Three Blind Mice | 15 | 100.0% |
| 7 | Twinkle Twinkle Little Star | 15 | 100.0% |
| 8 | America the Beautiful | 15 | 100.0% |
| 9 | My Country Tis of Thee | 17 | 100.0% |
| 10 | When the Saints Go Marching In | 17 | 100.0% |
| Overall | | | **99.0%** |

**Table 1.** *Each blink corresponds to a note of the song. The accuracy column denotes the results of leave–one–out validation over 303 examples, where 300 examples were correctly classified.*

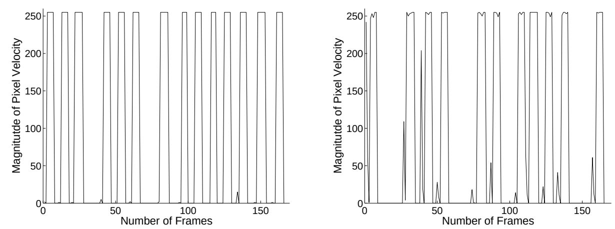


**Figure 2.** *Visualization of data for the patterns of "This Old Man" (left) and "God Save the Queen" (right). Both patterns have similar rhythm and duration.*

In practice, we can use user feedback to help choose distinguishable patterns. When a locked–in individual enrolls a new data pattern into the system, the system can use DTW to determine the similarity of the new pattern to patterns existing in the current database. If the pattern matches too closely with existing patterns the system can recommend that the user enroll a different pattern. This process will help the system to remain more robust by avoiding potential errors in classification.

## 5. Prescott: Person Recognition using Blinkprints

We have demonstrated the capability to recognize distinct blinked codes. We will now motivate how a similar system can be used to augment security interfaces to help restrict unauthorized access.

Imagine a secure area located in a major airport. To ensure that only authorized personnel have access to the area, a numerical keypad controls the locking mechanism on the door. To unlock the door the correct code must be entered on the keypad. Because access is based solely on entering the correct number, an unauthorized person can foil the system by observing the correct code and then entering it. Biometrics may be added to the system to improve security; for example, a camera can be placed in front of the door and access can be controlled based on face recognition and entering the correct personal identification number (PIN). However, this system is flawed: the face recognition system can be fooled by placing a photograph of an authorized person in front of the camera.

To address this issue, one might replace the numeric keypad with a system that requires the person to blink a specific pattern. This system could utilize the hardware already in place for face recognition. Several benefits could be introduced by such an augmentation. First, replacement of the keypad would allow hands–free entry through the door. Second, the rapid movement of the eyes during a blink can be used to localize the position of the head in the video [5], which can be beneficial to the face recognition portion of the system. The blinking can also reduce the probability of someone deceiving the face recognition by placing a photograph in front of the camera because the face is now required to have a dynamic component. Third, a personal blink pattern may be more difficult for a third party to observe because simply looking over the shoulder will not work; an observer must be watching the user from a position similar to the camera's. Since the user will be facing the camera to perform his blinks, he will be more likely to notice a person trying to observe his code. Fourth, a PIN based on a song may be easier for the user to remember.

However, this system may still be vunerable if a PIN is compromised. Someone might be able to foil the system by crafting a mask from a photograph of an authorized person. The person could remove the eyes from the photograph and blink the code while covering his face with the photograph. In this fashion both the face recognition and the blink pattern recognition can be foiled. However, it is possible that the way in which a subject blinks a pattern may provide information about the subject's true identity, providing a biometric check against the PIN.

Using the same prototype system from previous sections, we investigated if the intrinsic properties of how a person blinks a specific pattern, their "blinkprint"—a "blinking fingerprint"— can be used to perform identification. In this case, recognition would depend on more than just the pattern itself; it could also be dependent on the time beween blinks, how long the eye is held closed at each blink, or other physical characteristics the eye undergoes while blinking. Figure 3 shows the difference between two users blinking the same pattern. The next section will discuss an experiment exploring this possibility.

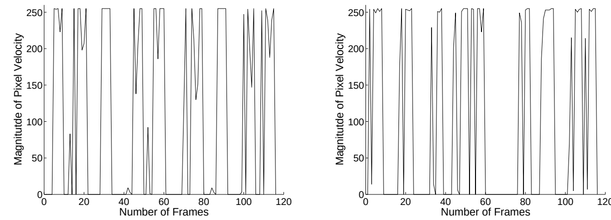


**Figure 3.** *A visual representation of the same 9–blink pattern, "— — — . . — — . ." lasting 4.2 seconds performed by two different people.*

## 6. Experiment: Person Recognition

For this experiment we wanted to determine if it is possible to recognize people based on the intrinsic properties of how they blink a fixed pattern. Nine participants, some with glasses, blinked the same pattern "— — — . . — — . ." where '—' and '.' represent long and short durations between blinks respectively. A total of 500 examples were collected between all participants. Each person provided a minimum of 30 examples.

The system could attempt to identify a user solely based on the length of time it takes a person to blink the sequence. However, while this method may work for very small data sets, it will not scale to larger populations. In fact, comparison of sequence times would be insufficient for identification in our nine participant sample set. For example, two of the participants, on average, perform the sequence in 4.2 seconds with standard deviations of 0.31 seconds and 0.41 seconds respectively (see Figure 3). Classification was first attempted using our DTW based prototype; however, using the sequence of intervals between the blinks proved insufficient to represent the intrinsic properties of an individual's blink. We hypothesize that the forgiving nature of DTW generalizes the characteristics of the data required to distinguish between individuals.

For this more complex recognition task we explored the use of hidden Markov models (HMMs). HMMs provide a more robust method of modeling and recognizing temporal sequences [14]. We postulate that individual characteristics of blinking will be encoded in the output probabilities of the HMMs. Instead of segmenting the sequence into clusters of blinks and spaces, the overall pixel velocity features per frame are represented as a two element vector $(\rho, \theta)$ and input into the HMMs.

Our first HMM was designed with one emitting state per blink, for a total of nine emitting states. This model provided poor classification. Experiments were run systematically reducing the number of states, using the Georgia Tech Gesture Toolkit [15]. A three-state HMM (with only one emitting state) provided 68.88% accuracy for discrimination between individuals. The single emitting state of this model represents the observation data as a Gaussian and encodes the length of the sequence in the transition probability involving the non–emitting initial and final states. This topology suggests that a Gaussian model of the pixel velocity features may be sufficient for representing the data.

We represent each frame of a blinked sequence by the average change in pixel velocity and direction over the eye for that frame: $(\rho, \theta)$. For each example $i$, the sequence of all frames is represented as a three–dimensional Gaussian $G_i : (\bar{\rho}, \bar{\theta}, S)$, where $\bar{\rho}$ and $\bar{\theta}$ represent the average angle and velocity values for the frames of a sequence, and $S$ is the total number of frames in the sequence.

The discrimination power of this representation was investigated using six different experiments: $k$–nearest neighbor using Euclidean distance, Mahalanobis distance, and probabalistic measure; nearest class centroid using Euclidean distance and Mahalanobis distance; and most likely centroid. The results from all experiments were averaged over a total of 100 iterations of cross–validation. For each iteration, two thirds of the examples from each class were used as training points.

For the $k$–nearest neighbor experiments, each test sequence $j$ is modeled as a three–dimensional Gaussian $G_j$, as described above. Its classification is then determined by the most frequent class, or person, of the $k$ "closest" training examples $G_i$. For these experiments $k$ was chosen to be 10. The closeness of a point is determined by either a distance metric or a measure of how well a test point is represented by a Gaussian model of each training example. We used the Euclidean and Mahalanobis distances between the means for the former, and the maximum likelihood for the latter.

For the nearest class centroid experiments, we modeled all of the training examples $G_i$ for a class $\omega$ as a single Gaussian $G_\omega$. The classification of the test sequence $j$ is determined either by finding the $G_\omega$ whose mean is closest to the mean of $G_j$ as measured by the Euclidean or Mahalanobis distance, or by evaluating the probability of each $G_\omega$ given $G_j$ and picking the most likely class.

Experiments using the nearest class centroid with a maximum likelihood metric (the likelihood of a test sequence being represented by the class model) provided the most discrimination with an accuracy of 82.02%. While this level of accuracy is not currently high enough to utilize blinkprints as a stand–alone biometric, these results encourage its use as a component in multimodal biometric systems. For example, blinkprints could be used to boost the accuracy of face recognition by helping to verify the results. As shown in Table 2, when the correct result is assumed to be one of the top three most likely candidates, accuracy increases to 95.57%.

| Experiment | N=1 | N=2 | N=3 |
|---|---|---|---|
| Nearest Centroid, Euclidean | 71.60 | 72.90 | 76.16 |
| Nearest Centroid, Mahalanobis | 76.74 | 77.84 | 79.05 |
| **Most Likely Centroid** | **82.02** | 83.14 | 84.24 |
| **$k$–Nearest Neighbor, Euclidean** | 77.10 | 91.22 | **95.57** |
| $k$–Nearest Neighbor, Mahalanobis | 75.53 | 89.09 | 94.40 |
| $k$–Most Likely Neighbors | 78.75 | 89.85 | 93.61 |
| HMM 3-state | 68.88 | 82.00 | 92.80 |

**Table 2.** *Average results from 100 iterations of 7 different experiments. Columns indicate results when the correct classification is within the top N–best classifications.*

## 7. Related Work

Several systems have been developed for communication based on tracking eye gaze [1]. The Eyegaze System was studied to determine the feasibility of eye tracking as a means of communication for people afflicted with Locked-in syndrome [6]. Seventy–five percent of the participants were able to perform a subset of the activities supported by the system including typing, controlling environmental devices, voice synthesizing, telephone usage, and playing games.

Grauman et al [11] constructed an assistive technology system based on correlation that accurately tracks the eyes and measures the duration of eye blinks continuously in real time (27 to 29 fps). The correlation between the image and templates of open and closed eyes distinguishes between natural "short" eye blinks and voluntary "long" eye blinks allowing the use of deliberate blinks to trigger a mouse click. Their system required no special training and can continually detect blinks in a desk environment setting with an accuracy of 96.5%. The system can correctly classify detected blinks as natural or voluntary blinks 93% of the time.

Currently our system only performs isolated recognition. We would like to incorporate both the eye gaze tracking and the template matching method described above into our system to help aid with continuous recognition.

## 8. Discussion and Future Work

We have shown that both systems discussed above can perform reasonably well in isolation; however, if these systems are to be used in a non–laboratory environment the issue of continuous recognition must be addressed. During experimentation participants pressed a key to notify the system when to start and stop recognition. While this type of system can be implemented for security applications, it is impractical for applications involving assistive technology interfaces because individuals interacting with the system will most likely be limited to blinking. Use of a specific deliberate blink or eye gaze gesture could be used as a mechanism for notifying the system when to begin the recognition process.

The practicality and usability of blink sequences will be influenced by the blinking capabilities of the individual. We would like to conduct studies to determine the limits of blinking as a form of communication. Research is needed to determine how rapidly a person is capable of blinking, the consistency with which a person can blink, and how long a person can blink before becoming fatigued.

Another issue to be addressed is the scalability of the system. Our experiments show encouraging results for a limited number of subjects and samples. More experimentation is required to determine how well these methods will perform as the number of users or patterns to identify increases.

We have demonstrated a system where a classification scheme based on DTW can be used to identify a specific blinked pattern. Once the specific pattern is identified, Gaussian models over the data can be used to determine which individual is performing the pattern. Currently these models ignore important factors such as the shape of the eye. We are interested in exploring more complex models that would take this information into account.

Currently the system shows improvement when the top 3 choices are examined. This may be because there is confusion between two classes that could be resolved with a less general representation. Instead of Gaussian models, blinkprints may be better represented by non-parametric models that could allow for greater discrimination.

## 9. Conclusion

We have demonstrated two applications that utilize patterned blink recognition. The first application is a blink–based interface for controlling devices that uses songs as the cadence for the blinked patterns. It is designed to facilitate universal access for people with disabilities so severe that they only retain control of their eyes. Our system distinguishes between ten similar patterns with 99.0% accuracy using classification techniques based on dynamic time warping. Our second application presents a method for identifying individual people based on the intrinsic characteristics of how they blink a specific pattern (their blinkprint). This system can be used to help aid the accuracy of facial recognition systems and other security based interfaces. We were able to distinguish between nine individuals blinking the exact same pattern with 82.02% accuracy. Both prototype systems have provided encouraging preliminary results and will be further developed in the future.

## 10. Acknowledgments

Funding was provided in part from the Graphics, Visualization and Usability Center Seed Grant program and by NSF career grant # 0093291.

## References

[1] *http://www.cs.uta.fi /hci/gaze/*.

[2] D. Anson. *Alternate Computer Access, a guide to selection*. Davis and Co, 1997.

[3] E. Barker and R. Hobdell. *Neuroscience Nursing*. Mosby-Year Book Inc., St. Louis, MO, 1994.

[4] H. Barnett, J. Mohr, B. Stein, and F. Yatsu. *Stroke: Pathophysiology, Diagnosis, and Management*. Churchill, Livingstone, 2nd Ed., 1992.

[5] M. Betke, W. Mullally, and J. Magee. Active detection of eye scleras in real time. In *IEEE CVPR Workshop on Human Modeling, Analysis and Synthesis*, June 2000.

[6] J. E. Chapman. Use of an eye-operated system in locked-in syndrome. In *CSUN's Sixth Annual International Conference: Technology and Persons with Disabilites*, 1991.

[7] A. Cook and S. Hussey. *Assistive Technologies Principles and Practice, Second Edition*. Mosby, 1984.

[8] J. Darragh and I. Witten. *The Reactive Keyboard*. Cambridge Series on Human-Computer Interaction. Cambridge University Press, Cambridge, 1992.

[9] T. Darrell and A. Pentland. Space-time gestures. *Proc. Comp. Vis. and Pattern Rec.*, pages 335–340, 1993.

[10] J. Gips, P. DiMattia, F. Curran, and P. Olivieri. *Using EagleEyes*. Oldenbourg, Vienna, 1996.

[11] K. Grauman, M. Betke, J. Gips, and G. R. Bradski. Communication via eye blinks detection and duration analysis in real time. In *IEEE CVPR*, Dec 2001.

[12] P. Kennedy, R. Bakay, M. Moore, K. Adams, and J. Goldthwaite. Direct control of a computer from the human central nervous system. *IEEE Transactions on Rehabilitation Engineering*, 8(2), June 2000.

[13] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI81*, pages 674–679, 1981.

[14] T. Starner, J. Weaver, and A. Pentland. Real-time American Sign Language recognition using desk and wearable computer-based video. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 20(12), December 1998.

[15] T. Westeyn, H. Brashear, A. Atrash, and T. Starner. Georgia Tech Gesture Toolkit: Supporting experiments in gesture recognition. In *Fifth International Conference on Multimodal Interfaces*, pages 85–92, 2003.