Automated Identification of Optically Sensed Aphid (Homoptera: Aphidae) Wingbeat Waveforms

A. MOORE[1] AND R. H. MILLER[2]

Ann. Entomol. Soc. Am. 95(1): 1–8 (2002)

ABSTRACT An optical sensor was used to make digital recordings of wingbeat waveforms for the five most common aphids found on Guam: *Aphis craccivora* Koch, *A. gossypii* Glover, *A. nerii* Fonscolombe, *Pentalonia nigronervosa* Coquerel, and *Toxoptera citricida* (Kirkaldy). Wingbeat frequencies for each species overlapped all other species. However, mean wingbeat frequencies were significantly different for all species. Wingbeat frequencies and harmonic patterns were extracted from the recordings and submitted to cluster analysis, which failed to separate species completely. Several nearest neighbor and probabilistic neural network classifiers were built using time series, frequency spectra, wingbeat frequencies, and harmonic patterns as input variables. These classifiers were evaluated by having them identify wingbeat waveforms from aphids collected and recorded after their construction. The best performing classifier model was a probabilistic artificial neural network trained using 256-bin frequency spectra as input. Sixty-nine percent of the waveforms presented to this network were identified correctly. This study demonstrates the feasibility of developing an insect flight monitor that automatically counts and identifies individual flying insects. Essential components of the monitoring system are a photosensor, a multimedia personal computer, and software that identifies wingbeat frequency spectra using an artificial neural network.

KEY WORDS insect wingbeat frequency, insect flight monitor, artificial neural network, nearest neighbor classifier, automated classification, aphid

ACQUISITION OF BIOLOGICAL field data are usually expensive, in terms of time and labor, due to the lack of automated instrumentation systems, such as those developed for continuous monitoring of meteorological variables. Lack of instrumentation for measuring insect flight activity limits understanding of the impact of movement on the population dynamics of pest species (Stinner et al. 1983). Insufficient near real-time information on the abundance of insect pests and associated predators and parasites is often an impediment to development of integrated pest management and biological control programs.

Richards (1955) discovered that a photocell exposed to the sun detected transient waveforms caused by reflections from the wings of individual insects flying in front of the solar disk. Each waveform contained a harmonic series, with the wingbeat frequency as the fundamental. Richards (1955) suggested that his discovery "might be turned to account by workers interested in the flight habits of insects, and particularly in variations in insect density during daylight hours."

Moore et al. (1986) digitally recorded two species of mosquitoes with a photosensor based on an optical tachometer developed by Unwin and Ellington (1979). Spectral analysis of the mosquito wingbeat waveforms using a fast Fourier transform revealed a harmonic series with the wingbeat frequency as the fundamental. Moore et al. (1986) suggested that automated instrumentation could be designed to discriminate among species of flying insects by recognizing spectral patterns formed by these harmonics, analogous to the way in which the human ear and brain discriminate among musical instruments by recognizing patterns of harmonics in the sounds they generate. Frequency spectra derived from the mosquito data were used to train an artificial neural network (Moore 1991). The trained network identified the species and sex of mosquitoes with 92% accuracy. When the network was trained and tested using only wingbeat frequencies, the mean absolute error for the network increased by 33%, indicating that the harmonic patterns contributed significant species-specific information.

Moore (1998) developed an optical sensor and data acquisition system for automated monitoring of flying insects in the field. Here, we report on our attempts to use this system to automate identification of alate aphids in the laboratory.

Materials and Methods

Aphids. Colonies of *Aphis craccivora* Koch, *A. gossypii* Glover, *A. nerii* Fonscolombe, *Pentalonia nigron-*

[1] 57 Belmont Avenue, Ottawa, Ontario, Canada K1S 0V2.
[2] College of Agriculture and Life Sciences, University of Guam, Mangilao, Guam 96923 (e-mail: rmiller@uog.edu).

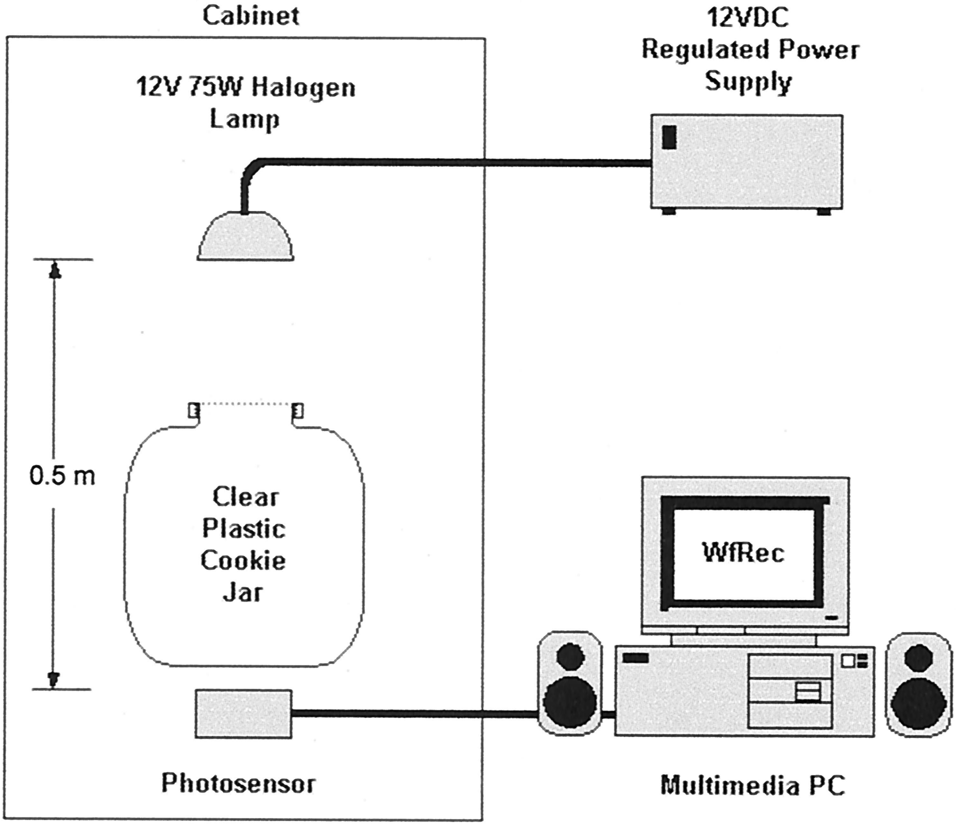


Fig. 1. Apparatus for digital recording of insect wingbeat waveforms.

ervosa Coquerel, and *Toxoptera citricida* (Kirkaldy) were collected from various locations on the island of Guam. Aphids on cuttings of host plants were collected in the field and maintained in the laboratory at 22°C for a maximum of 3 d. Aphids used in waveform analysis tests were preserved in 70% ethanol and identified using published taxonomic keys (Blackman and Eastop 1984).

Digital Recording of Wingbeat Waveforms. For each recording session, ≈12 alate aphids from a single colony were transferred from a rearing container into a transparent plastic cookie jar (SnapWare, FloTool International, Tustin, CA) measuring 20 by 17 by 13 cm and fitted with a fine muslin top (Fig. 1). We chose 12 as the number of aphids to use for each recording session to ensure that we would record a large number of waveforms during a recording session (1 to 3 h duration), while minimizing the probability that more than one aphid would fly simultaneously. The container was placed in a cabinet between a halogen lamp (Type EYJ, 12V 75W, General Electric, Cleveland, OH) and a photosensor (Qubit Systems, Kingston, Ontario, Canada). The cabinet shielded the sensor from optical noise caused by AC-powered room lighting, which flashes at 120 Hz. The halogen lamp was powered by a regulated DC power supply (7W 13.8V regulated power supply, TrippLite, Chicago, OH). The photosensor was connected to the microphone input jack of a 16-bit sound card installed in a personal computer (Pentium 200 MMX, 48 MB RAM). Air temperature within the cookie jar ranged from 24–30°C during the recording sessions. A transient waveform recorder program (WfRec, Qubit Systems 1999) was used to sample the sound card input at 8,000 samples per s. The waveform recorder's adaptive trigger was set to record 512-sample waveforms whenever the input signal exceeded six standard deviations above the mean ambient noise level. WfRec's cepstrum filter was set to reject waveforms that were low in harmonic content, i.e., those with a peak cepstrum amplitude of <0.06 within the 100–200 Hz frequency band.

Wingbeat waveform recordings were separated into two groups. Group 1 contained recordings of aphids collected and recorded between 11 March and 13 May 1999. This group was used for training nearest neighbor and artificial neural network classifiers. Group 2, from aphids collected and recorded between 19 May and 25 May 1999, was used only for evaluating the classifiers.

Data Reduction. The transient waveform recorder program, WfRec, represented recorded waveforms at three levels of abstraction. First, the raw data were saved as a wingbeat waveform time series, X1, . . ., X512, which contains the signal amplitude at discrete

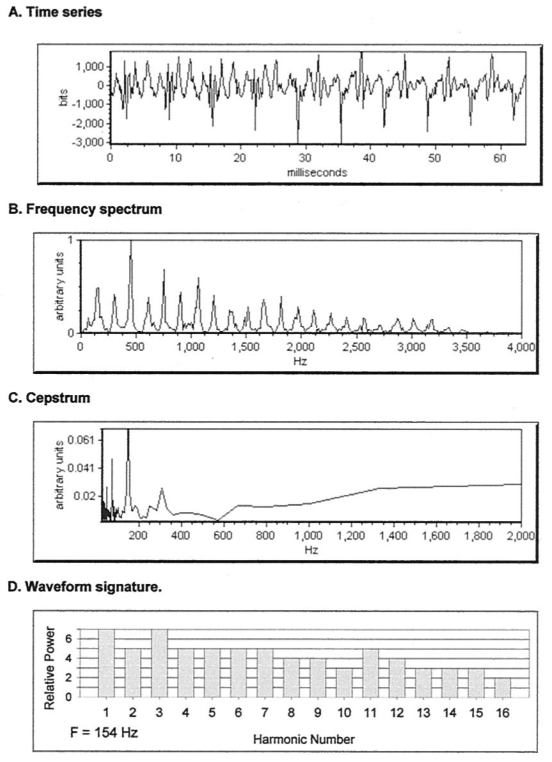


Fig. 2. Digital processing for calculation of a waveform signature starts with a time series of 512 data points sampled at 8,000 samples per second (A). A fast Fourier transform produces a frequency spectrum (B) which contains a harmonic series. A second fast Fourier transform, applied to the frequency spectrum produces a cepstrum (C) that has a peak at the fundamental frequency of the waveform. The amplitude of the peak is a measure of the total harmonic content of the waveform. Areas area under the spectrum within frequency bands centered on the first 16 harmonics are calculated and scaled to produce a waveform signature (D).

time steps of 1/8000 s. Second, a fast Fourier transform generated a wingbeat waveform spectrum, S1, [ellipsis], S256, which contained the relative energy in each of 256 frequency bins ranging from 0 Hz to 4 kHz. Third, the waveform recorder encoded spectral information as a series of 17 numbers, F, H1, [ellipsis], H16, referred to as the wingbeat waveform signature (Fig. 2). The first number in the signature, F, is the fundamental frequency (=wingbeat frequency) measured in Hz. This is followed by a series of 16 integers, H1 to H16, which have values between 0 and 7. These integers represent a histogram of the relative power contained within frequency bands centered on the first through the 16th harmonic of each waveform.

Cluster Analysis. Hierarchical agglomerative clustering using Euclidean distances and Ward's method for linkages (Virtual Institute of Analytical Sciences 1999) was applied to wingbeat waveform signatures for the group 1 data. Each variable was standardized to zero mean and unit variance before analysis. Wingbeat waveform data from recordings of two aphid parasitoids: an introduced aphidiid, *Diaeretiella rapae* (McIntosh), and an unidentified, locally collected aphidiid species were included in this analysis as outclasses for comparison of scale in cluster separation. Results are presented as dendrograms.

Nearest Neighbor Classification. Nearest neighbor classification models (Virtual Institute of Analytical Sciences 1999) were built using wingbeat waveform time series, frequency spectra, and signatures for the group 1 data (6,228 waveforms) and evaluated using the group 2 data (2,273 waveforms). Each input variable was standardized to zero mean and unit variance before analysis. Five nearest neighbor classifiers were built. Each used aphid wingbeat waveform data presented at a different level of abstraction: middle half of the time series (X129, . . ., X384); frequency spectrum (S1, . . ., S256); signature (F, H1, . . ., H16); harmonics alone (H1, . . ., H16); and wingbeat frequency alone (F).

Artificial Neural Network Classification. A commercial artificial neural network package, NeuroShell 2 (Ward Systems Group 1996), was used to construct, train, and test probabilistic networks to classify aphid wingbeat waveforms. We used the same data sets as in the nearest neighbor analysis. Five networks were built to accept input data in the form of a time series, frequency spectrum, signature, harmonics alone, and wingbeat frequency alone. These networks had one neuron for each input variable: 256, 256, 17, 16, and one input neurons, respectively. All networks had five output neurons, one output for each aphid species. There was also a single hidden layer with 4,983 neurons. Waveform data from group 1 were normalized so that the values of each variable ranged between zero and 1. Waveforms were then selected at random to be members of a training set ($n = 4983$) and a test set ($n = 1,245$). The test set was used for internal calibration of the network to prevent overtraining. During training, the network was presented with a waveform data as input, and a five-digit code consisting of a one and four zeros to indicate the species of aphid that produced the signature (e.g., 0 1 0 0 0 represented *Aphis gossypii*). After training, the group 2 data were classified by presenting waveform data to the network. Each waveform was classified as the aphid species corresponding to the output neuron having the highest excitation (Fig. 3).

Results

Cluster Analysis. When the complete wingbeat waveform signature were included in the analysis (Fig. 4A), recordings of aphids and aphid parasitoids fell into two very distinct clusters. The aphid branch contained many mixed-species clusters. Note, however, that all the aphid clusters containing only two aphid colonies are conspecific. The analysis based only upon the relative amplitudes of harmonics (Fig. 4B) produced almost identical results to those from the complete signatures, whereas analysis based only on wingbeat frequency (Fig. 4C) showed little separation among clusters.

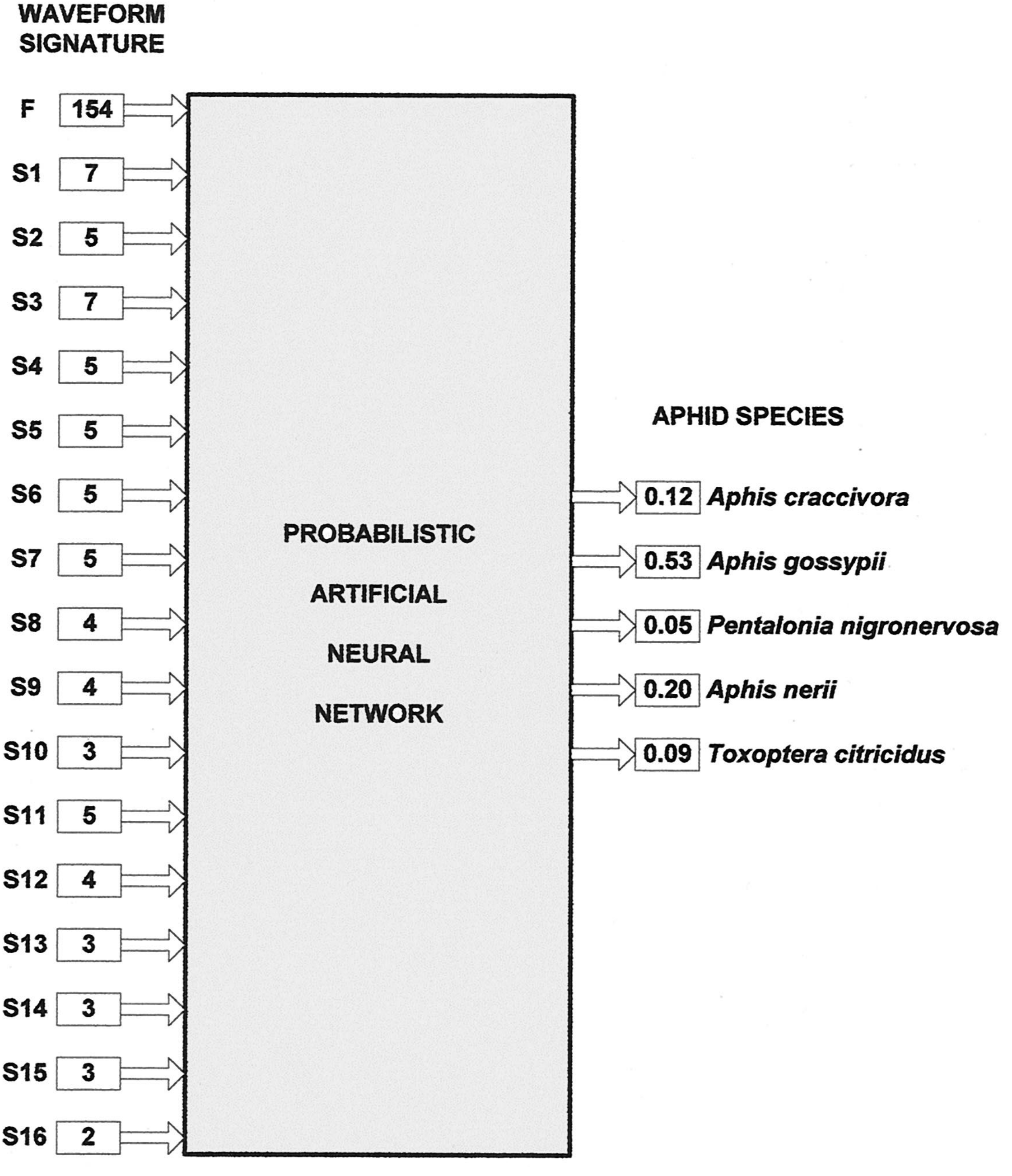


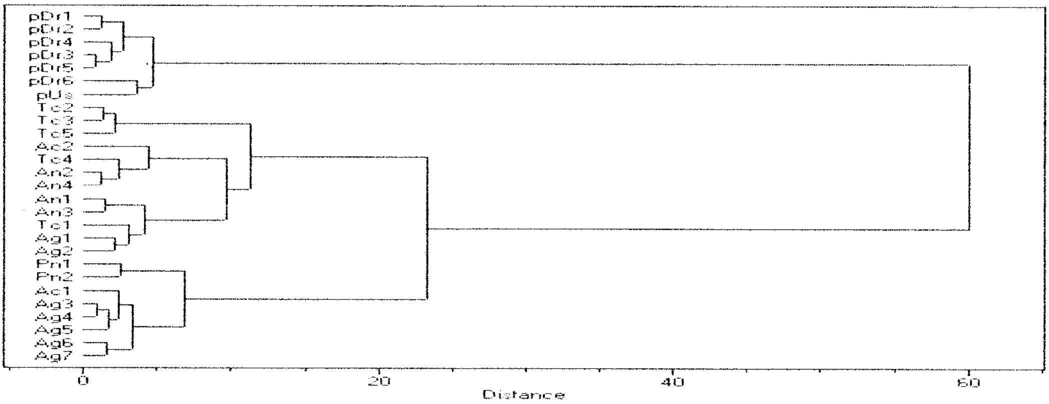
Fig. 3. Probabilistic artificial neural network for classifying aphids based on wingbeat waveform signatures. Inputs are fundamental frequency, F, and relative power in frequency bands centered on the first 16 harmonics, S1 through S16. Values shown are for recording number 679, *Aphis gossypii*, 21 IV 1999.

Automated Classification. Mean wingbeat frequencies were significantly different for all five species of aphids (Table 1). However, wingbeat frequency alone was not very useful for identifying individuals due to a high degree of overlap among species (Table 1). The most accurate classification model of several that were tested (Table 2) was a probabilistic neural network using waveform data presented as a 256-bin frequency spectrum (Table 3). The proportion of waveforms that were correctly classified was 69%, ranging from 94% for *Aphis gossypii* down to 41% for *A. craccivora.*

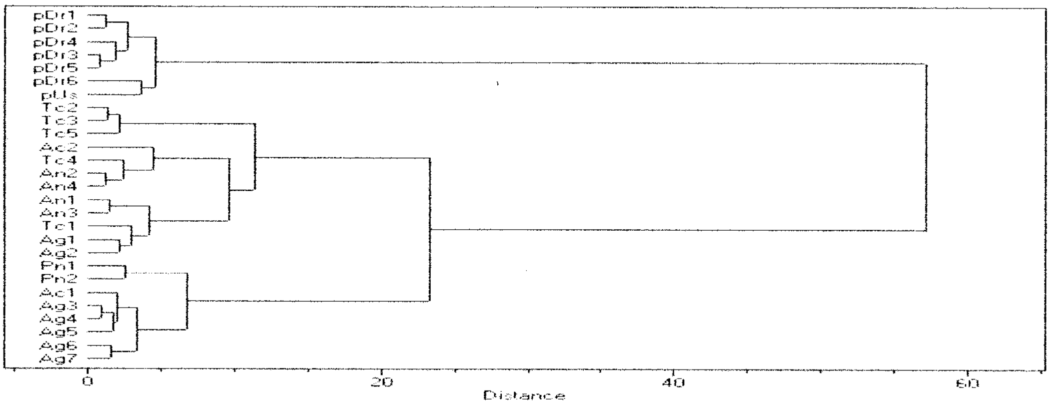
Discussion

The literature contains little information on aphid wingbeat frequencies. Sotavalta (1947) listed only one

A. Variables: Wingbeat frequency (F) and relative amplitudes of harmonics (S1 .. S16)



B. Variables: Relative amplitudes of harmonics (S1 .. S16)



C. Variable: Wingbeat frequency (F)

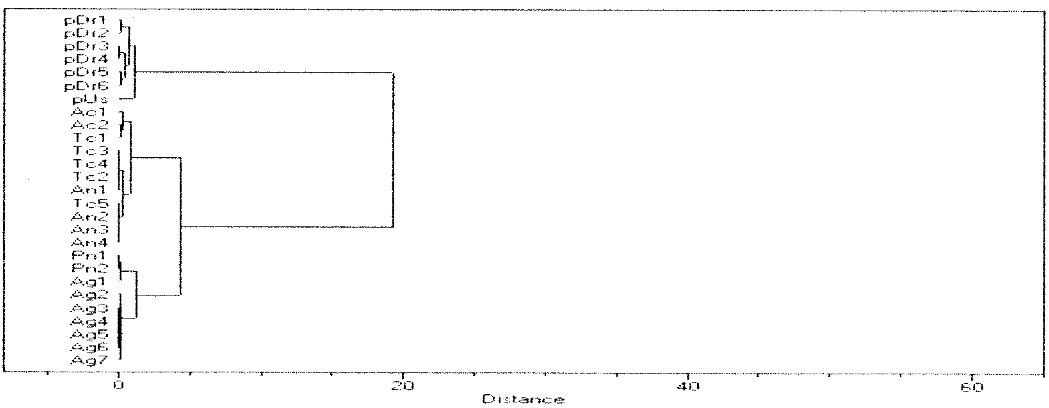


Fig. 4. Cluster analyses for components of wingbeat waveform signatures of aphids and aphid parasitoids. Species codes are: pDr for the aphid parasitoid *Diaeretiella rapae*, pUs for an unidentified aphid parasitoid, Ac for *Aphis craccivora*, Tc for *Toxoptera citricida*, An for *A. nerii*, Pn for *Pentalonia nigronervosa*, and Ag for *A. gossypii*. Numbers indicate separate colonies.

Table 1. Distribution of wingbeat frequencies for aphids recorded on Guam between 11 March and 13 May 1999 (group 1)

Frequency bin (Hz)	Species				
	Aphis craccivora	*Toxoptera citricida*	*Aphis nerii*	*Pentalonia nigronervosa*	*Aphis gossypii*
108	3				
111	3				
114	6				1
118	11				1
121	34	2			6
125	46	2	3		6
129	59	23	7		9
133	46	108	34	4	11
138	41	293	68	16	49
143	47	490	124	83	132
148	31	342	151	162	278
154	11	82	96	195	347
160	2	14	19	108	647
167				91	1,035
174				40	715
182				6	83
190					4
200					1
No. of waveforms	340	1,356	502	705	3,325
Median frequency (Hz)	133a	143b	148c	154d	167e

Medians followed by different letters are significantly different at $P < 0.05$, Dunn's method.

aphid in his table of wingbeat frequencies for over 200 insect species. He reported *Cryptomyzus galeopsidis* (Kaltenbach) as having a wingbeat frequency between 82 and 87 Hz. This measurement was made without instrumentation. Sotavalta relied solely on his keen hearing and musical gift of perfect pitch. Byrne et al. (1988) used an optical tachometer (Unwin and Ellington 1979) to measure wingbeat frequencies for 10 individuals from each of five aphid species flying in a greenhouse at 24 ± 3°C. Mean wingbeat frequencies for *Aphis nerii* and *A. gossypii* were not significantly different (118.1 and 123.4 Hz). In contrast, we measured much higher, significantly different mean wingbeat frequencies for *Aphis nerii* and *A. gossypii* (145.6 and 162.7 Hz). It is likely that our wingbeat frequency measurements are higher because our aphids were flying at higher temperatures, 24–30°C.

The objective of our study was to investigate the feasibility of using optically sensed insect wingbeat waveforms for automated classification of aphids at the species or subspecies level. Our approach is to use the wingbeat waveform as a signature rather than extracting the wingbeat frequency as a single character. The proportion of waveforms that were correctly classified by species using a neural network presented with the whole waveform spectra as input was 69%, ranging from 94% for *Aphis gossypii* down to 41% for *A. craccivora.* The precision of identification was limited because of large variances among waveforms recorded for each species. We suspect that the within-species variance among the waveforms was large due to two reasons. First, many of the recordings were triggered by aphids that where not in flight. We observed aphids buzzing their wings while walking inverted on the test chamber mesh. Second, differences in waveforms may have been due to differences in the orientation of aphids with respect to the sensor and light source. Farmery (1981) showed that, while the

Table 2. Performance of classification models

Classification model	Input variables	% of waveforms classified correctly					
		Aphis craccivora	*Aphis gossypii*	*Pentalonia nigronervosa*	*Aphis nerii*	*Toxoptera citricida*	All species
Expected values		20	20	20	20	20	20
Nearest neighbor	Time series ($X_{129}, \ldots, X_{384}$)	31	82	50	17	65	45
	Spectrum ($S_1, \ldots, S_{256}$)	25	91	59	26	61	51
	Signature ($F, H_1, \ldots, H_{16}$)	13	92	45	16	45	43
	Harmonics ($H_1, \ldots, H_{16}$)	7	76	39	10	33	34
	Wingbeat frequency (F)	3	98	14	18	14	36
Probabilistic neural network	Time series ($X_{129}, \ldots, X_{384}$)	37	71	68	22	63	47
	Spectrum ($S_1, \ldots, S_{256}$)	41	94	84	54	66	69
	Signature ($F, H_1, \ldots, H_{16}$)	18	84	64	44	31	54
	Harmonics ($H_1, \ldots, H_{16}$)	13	62	62	34	25	43
	Wingbeat frequency (F)	43	96	44	0	86	43

All models were trained using wingbeat waveform signatures for aphids collected and recorded on Guam between 11 March and 13 May 1999 and all were evaluated using data from a separate group of aphids collected and recorded on Guam between 19 May and 25 May 1999. See text for full description of input variables.

Table 3. Probabilistic artificial neural network classification of optically sensed aphid wingbeat waveforms

Actual Species	Network classification (no. of waveforms)					Proportion correct, %	Probability[a]
	Aphis craccivora	*Aphis gossypii*	*Pentalonia nigronervosa*	*Aphis nerii*	*Toxoptera citricida*		
A. craccivora	**59**[b]	0	12	40	32	41	<0.00001
A. gossypii	2	**540**[b]	23	4	3	94	<0.00001
P. nigronervosa	1	15	**285**[b]	14	24	84	<0.00001
A. nerii	110	1	192	**516**[b]	140	54	<0.00001
T. citricida	26	0	3	59	**172**[b]	66	<0.00001

Input data for each waveform were values from a 256-bin frequency spectrum ranging from 0 to 4 kHz. The network was trained with data from aphids collected and recorded on Guam between 11 March and 13 May 1999, and tested with aphids collected and recorded on Guam between 19 May and 25 May 1999.

[a] Probability of observed proportion correct being due to chance. Each probability was estimated by 100,000 Monte Carlo trials in which waveforms were assigned to each of the five species at random.

[b] Correct classifications.

wingbeat frequency stays constant, the harmonic pattern of optically sensed wingbeat waveforms changes radically when the orientation of the insect with respect to the sensor is varied. This is due to changes in angles of reflection and changes in visibility of body parts. It may be possible to reduce the within-species variance in waveforms by eliminating waveforms from aphids buzzing while not in flight, and by restricting the orientation of flying aphids with respect to the sensor and light source by modifying the apparatus so that the photosensor records aphids flying down a length of transparent tube.

Although all aphid species were not identified with high accuracy, the artificial neural network correctly identified *Aphis gossypii* waveforms with 94% accuracy (Table 3). The error rate for automated identification of waveforms made by this species was only 3% for false positives and 6% for false negatives. This indicates that it should be possible to build an automated field flight monitor for this species, which is the most numerous and important aphid pest species on Guam and other Pacific islands.

Our results show that the relative amplitudes of wingbeat waveform harmonics (H1, .., H16) contain species-specific information, in addition to the wingbeat frequency. The cluster analysis clearly shows that patterns of harmonics abstracted from insect wingbeat waveforms contain species-specific information, even when frequency information is removed. Nearest neighbor and neural network classification models worked equally well with wingbeat frequency (F) or harmonic pattern (H1, .., H16) as input variables. Our discovery that of insect wingbeat waveform harmonics contain information useful for automated identification is significant, as this may allow accurate identification of insects in which the wingbeat frequency changes with temperature, as has been reported for many insects including locusts (Foster and Robertson 1992, Robertson et al. 1996, Xu and Robertson 1996), beetles (Oertli 1989), armyworm moths (Farmery 1982), bees (Unwin and Corbet 1984, Harrison et al. 1996, Roberts et al. 1998), and flies (Unwin and Corbet 1984). Although the current study was performed at a constant temperature, we expect the harmonic pattern to remain constant, even if the wingbeat frequency changes.

This study demonstrates the feasibility of developing an insect flight monitor that automatically counts and identifies individual flying insects. Essential components of the monitoring system are a photosensor, sound board equipped personal computer, and software that identifies wingbeat frequency spectra using an artificial neural network.

Acknowledgments

We thank Nenita Dumaliang, Olivia Idechiil, Hideya Osada, and Bernard Bermudes for technical assistance. Thanks also to John Brown, Mac MacCarter, John Obrycki, and Steve Yaninek for reviewing early drafts of the manuscript, and to Keith Pike and Petr Stary for confirming identification of aphids and aphid parasitoids. We also appreciate useful comments from anonymous reviewers. This work was supported in part by funds for Regional Research Project W-185, Biological Control in Plant Systems.

References Cited

Blackman, R. L., and V. F. Eastop. 1984. Aphids on the world's crops. Wiley, Singapore.

Byrne, D. N., S. L. Buchmann, and H. G. Spangler. 1988. Relationship between wing loading, wingbeat frequency and body mass in homopterous insects. J. Exp. Biol. 135: 9–23.

Farmery, M. J. 1981. Optical studies of insect flight at low altitude. Ph.D. dissertation, York University, York, UK.

Farmery, M. J. 1982. The effect of air temperature on wingbeat frequency of naturally flying armyworm moth (*Spodoptera exempta*). Entomol. Exp. Appl. 32: 193–194.

Foster, J. A., and R. M. Robertson. 1992. Temperature dependency of wing-beat frequency in intact and deafferented locusts. J. Exp. Biol. 162: 295.

Harrison, J. F., J. H. Fewell, S. P. Roberts, and H. G. Hall. 1996. Achievement of thermal stability by varying metabolic heat production in flying honeybees. Science 274: 88–90.

Moore, A. 1991. Artificial neural network trained to identify mosquitoes in flight. J. Insect Behav. 4: 391–395.

Moore, A. 1998. Development of a data acquisition system for long-term outdoor recording of insect flight activity using a photosensor. 13th Conference on Aerobiology and Biometeorology, American Meteorological Society, Albuquerque, NM.

Moore, A., J. R. Miller, B. E. Tabashnik, and S. H. Gage. 1986. Automated identification of flying insects by analysis of wingbeat frequencies. J. Econ. Entomol. 79: 1703–1706.

Oertli, J. J. 1989. Relationship of wing beat frequency and temperature during take-off flight in temperate-zone beetles. J. Exp. Biol. 145: 321.

Qubit Systems. 1999. WfRec transient waveform recorder. Kingston, ON, Canada.

Richards, I. R. 1955. Photoelectric cell observations of insects in flight. Nature (Lond.) 175: 128–129.

Roberts, S. P., J. F. Harrison, and N. F. Hadley. 1998. Mechanisms of thermal balance in flying *Centris pallida* (Hymenoptera: Anthophoridae). J. Exp. Biol. 201: 2321–31.

Robertson, R. M., H. Xu, K. L. Shoemaker, and K. Dawson-Scully. 1996. Exposure to heat shock affects thermosensitivity of the locust flight system. J. Neurobiol. 29: 367–83.

Sotavalta, O. 1947. The flight-tone (wing-stroke frequency) of insects (Contributions to the problem of insect flight 1.). Acta Entomol. Fenn. 4: 1–114.

Stinner, R. E., C. S. Barfield, J. L. Stimac, and L. Dohse. 1983. Dispersal and movement of insect pests. Annu. Rev. Entomol. 9: 115–121.

Unwin, D. M., and S. A. Corbet. 1984. Wingbeat frequency, temperature and body size in bees and flies. Physiol. Entomol. 9: 115–121.

Unwin, D. M., and C. P. Ellington. 1979. An optical tachometer for measurement of the wing-beat frequency of free-flying insects. J. Exp. Biol. 82: 377–378.

Virtual Institute of Analytical Sciences. 1999. Teach/Me data analysis, beta test version. Virtual Institute of Analytical Sciences, Vienna, Austria.

Ward Systems Group. 1996. NeuroShell 2. Ward Systems Group, Frederick, MD.

Xu, H., and R. M. Robertson. 1996. Neural parameters contributing to temperature compensation in the flight CPG of the locust, *Locusta migratoria.* Brain Res. 734: 213–22.

Received for publication 30 September 1999; accepted 9 September 2001.