Performance of an Algorithm Based on Cross-Correlations that Detects the Number of Infested Wheat Kernels in Grain Samples Tested in ALFID

David K. Weaver[1], Dennis Shuman[1], Richard W. Mankin[1], Robert E. Johnson[2], Mboumda N. Nasah-Lima[2], Robert R. Radford[2], Hok Chia[2], and Betty Weaver[1]

[1]USDA, ARS, Insect Attractants, Behavior, and Basic Biology Research Laboratory, 1700 SW 23rd Drive, P.O. Box 14565, Gainesville, FL 32604

and

[2]University of Florida, Gainesville, FL 32611

ABSTRACT

A custom developed algorithm was used to cluster and match sounds from individual insects in grain samples by cross-correlating signals from an array of acoustic sensors. The algorithm was created with a number of analysis parameters that are empirically optimizable for increased accuracy. The system was tested by placing individual kernels infested with a late-instar rice weevil, *Sitophilus oryzae* (L.), larva at specified locations in an ALFID sample chamber filled with wheat. With the analysis parameters set to facilitate processing of relatively low-level insect sounds, the algorithm successfully counted one insect in 86% of the trials. It miscounted the number as two in 14% of the trials. The algorithm's ability to count multiple insects was assessed by combining signals from data files collected with single infested kernels one or more sensor spacings apart to create a data set that represented sounds from a pair of insects. In these analyses, the algorithm correctly counted two insects in 100% of combinations three sensor spacings apart, 64% of combinations two sensor spacings apart, and 21% of combinations one sensor spacing apart. All of the incorrect analyses gave counts of one. Based on these results and the geometry of the ALFID sampling chamber, the algorithm has an 85% probability of identifying two insects producing sound in kernels located at random in a wheat sample.

INTRODUCTION

Current grain inspection practices focus on the presence of insects in determining the quality of a particular shipment. This is accomplished by sieving and visually inspecting a sample (1 kg) to see if it contains an unacceptable number of insects. The United States Federal Grain Inspection Service (FGIS) guidelines currently classify samples with less than two insects per sample as "clean" grain and samples with greater than one insect per sample as "infested." Most of the severe stored-product pest insects are obligate internal feeders as immatures. Thus infested grain may appear clean if adults are not yet present or if they have been removed by mechanical cleaning. Laboratory methods available for detecting internally feeding larvae involve the measurement of carbon dioxide evolved, resonance spectroscopy, or X-rays. These technologies are not widely implemented because they are costly, time consuming, of limited practicality, or not accurate enough for reliable use. There is a strong need to develop methods to overcome these listed limitations and implement a method for rapid, accurate detection of internally feeding insects in stored grain.

A significant amount of scientific research has been conducted on acoustic phenomena involving insect pests of stored-grain. Most of these involve correlation of the detected acoustical activity to the numbers of insects in discretely collected samples (for example, see Hagstrum et al 1991). The relationship between continuously collected sounds, discrete numbers of large insects in a sample, and the actual population size is not known. Acoustic monitoring of sounds is instructive because increases in acoustic phenomena occur as populations grow. However, until numbers of sounds detected can be directly related to the number of insects producing them, acoustic data are of little use for estimating population size. The ability to make such estimates is essential for grain managers.

The problem of quantification is even more important in grain inspection. When grain is in transportation channels, a very limited time is available to accurately determine if there are any insects in a collected sample. Sample numbers are limited and certainly not representative, but they are diagnostic, so it is very important that information about them be accurate. A system called ALFID (Acoustic Location Fixing Insect Detector) was developed to determine the number of loci from which sounds are originating in a grain sample (Shuman et al 1993). The initial version of ALFID performed well in determining when grain was "clean" by FGIS standards, and performed adequately when determining if the grain was "infested." However, the exact counting of insects in "infested" grain was less precise, which was in part due to irresolvable data from closely spaced insects (Shuman et al 1993).

A new ALFID (now called Acoustic Location "Fingerprinting" Insect Detector) has been developed using a novel design and more sensitive detectors (Shuman et al, these Proceedings). A recently completed algorithm uses waveforms from adjacent sensors to produce cross-correlations that are analyzed by a clustering procedure. A single array of output cluster data is linked into groups of matched sounds. The number of groups of linked sounds is interpreted as the number of insects. Here we report the results of tests using single late-instar larvae of the rice weevil, *Sitophilus oryzae* (L.), to evaluate the accuracy of this new ALFID and the efficiency of the algorithm.

MATERIALS AND METHODS

Active late-instar (20- to 25-day-old) larvae from a laboratory stock culture were selected for testing by individually placing putative infested kernels on a piezoelectric microphone and verifying that they were producing sounds. The culture was maintained on soft white wheat in an incubator that was operating at 26 ± 1° C and 65 ± 5% relative humidity with a 14:10 (L:D) h photoperiod. Kernels selected for testing were placed at one of several positions in the ALFID system (Shuman et al., this issue) as the unit was filled with one kg of grain. Positions tested were: A) at the midpoint between directly opposing sensors, B) at the midpoint between two sensors that were orthogonally opposed, C) offset from a sensor but within 1 cm of it, D) directly in front of a sensor, within 1 cm, and E) halfway between position A) and one of the directly opposing sensors, effectively one-quarter of the distance between the two facing sensors and displaced at the midpoint between two adjacent sensors (Figure 1). Different sensors in the grain container were used with each of these positons. Data were recorded from tests run for 30 minutes in an anechoic chamber at 20 - 23° C and 50 - 65% ambient relative humidity. Eighteen trials were conducted without insects in the ALFID chamber, while twenty-one were conducted with a single insect at one of the locations described above. Evaluations of the accuracy of the algorithm in scoring two insects were conducted using data sets created by combining two data files obtained with individual insects that were spaced at locations equivalent to one or more sensor positions apart. This facilitated determining whether analysis runs correctly discriminated between sounds made by the two insects. Forty-two such files were tested, fourteen of which were approximately one sensor spacing apart (Figure 1 - 1x), fourteen of which were approximately two sensor spacings apart (Figure 1 - 2x), and the last fourteen were approximately three sensor spacings apart (Figure 1 - 3x). The algorithm (Shuman et al, this issue) contains several analysis parameters which are optimizable, but are currently set at levels which enhance detection of low-level signals. First, the RMS ratio is set at a threshold value which retains low-level sounds. Increasing the threshold value discards most of very weak insect sounds and most of the sensor channels outputing only random noises. In addition, the cross-correlation peak absolute and relative thresholds both are set at empirically determined default levels that include all but the weakest insect sounds. At these settings, a considerable number of random noises are also retained and these may confound the analyses. These parameters will be empirically optimized in the near future to separate noises from insect sounds. Other parameters are set at selected default levels that encompass the dimensions of the sampling chamber, sensor location within the sampling chamber, and the performance of the sensor units with characteristic insect sounds. Clustering and grouping parameters are also optimizable, but are used at chosen default settings, as well. The default threshold value for cluster width is nine time samples and three sounds are required for the cluster to be considered valid. A default threshold value for a valid match for a pair of sounds is also three, meaning that both sounds have share membership in at least three clusters. The threshold value for minimum group size (criteria for interpretation as an insect) is once again three, which means that if each sound can only be a member of one group, any group must contain three sounds to be interpreted as an insect for these tests

RESULTS AND DISCUSSION

The algorithm counted single insects correctly in 86% of the tests (n = 21), with the remainder being counted as two insects (Figure 2.). The initial version of ALFID (Shuman et al 1993) was accurate in 70% of the tests, with 11% counted as two (Figure 2), and 19% counted as zero. In tests using data combined from two individual insects, the system counted correctly in 62% of the combinations overall (n = 42), with all miscounts being counted as one (Figure 3). The initial version counted correctly in 55% of the trials with two insects and undercounted in 42% of the tests (Figure 3). This enhanced performance is significant because the initial version could not resolve data for insects less than 1.75 sensor spacings apart, while the current version is 21% accurate at approximately 1 sensor spacing interval for combined data (n = 14), with 79% undercounting (Figure 4). The new ALFID scored correctly in 64% of the combinations with two insects approximately two sensor spacings apart (n = 14) and counted two insects correctly in all combinations with the insects approximately three sensor spacings apart (n = 14) (Figure 5). When sensor spacings were two or greater, the initial version ALFID counted correctly only 72% of the time. Therefore, this new ALFID with default tunable parameters is approximately 10 - 15% more accurate than its predecessor when counting one or two insects in a grain sample. The accuracy relative to the initial version increases primarily as the distance between pairs of insects decreases.

The algorithm became available only recently, so it was not possible to optimize any of the parameters designed to tune the system. The increased sensitivity of the new detectors, combined with default parameters designed to retain every potential insect sound, including those that were marginal, results initially in a lower accuracy of the new system for scoring grain containing no insects. The new ALFID had an accuracy of 56% (n = 18), compared to 90% for the earlier system (Figure 6). However, with FGIS standards, 89% of the trials (n = 18) are already considered "clean" with no tuning, which is more comparable to the 100% scored by the initial version for this FGIS criteria.

The observed fraction of correct counts of insect pairs at different spacings allows for approximation of the overall probability of the current ALFID correctly counting a random sample of two insects in the test chamber.

The probability of correctly counting two insects decreases from 1 at 3 sensor spacings, to 0.64 at 2 sensor spacings, to 0.21 at 1 sensor spacing, and to 0 at 0 sensor spacings. This relationship can be expressed approximately as $\rho(r) = r/R$, where $\rho(r)$ is the probability of a correct count when a pair is separated by distance r, and, R is the critical distance at which the probability of a correct count becomes 1 (in this case, 3 sensor spacings). The overall probability of a correct count within a sphere of radius R is the normalized integral of $\rho(r)$ over the total volume, or $P = 3/4$ (see Appendix). So, we estimate that any two insects will be scored correctly in 75% of trials where they are within a sphere of $R = 2$ sensor spacings, and in 100% of trials where their spacing is greater than R, for a cumulative probability of 85% for any trial with two insects (See Appendix). Note that the number is actually higher because the actual critical sphere was probably achieved before reaching three sensor spacings (this was the distance at which we first recorded 100% accuracy, and is probably not the distance at which it first occurred).

In summary, with ongoing parameter optimization, we expect at least the same level of accuracy (or better) with uninfested grain for the new ALFID system as was possible with the initial ALFID. It was our intent to develop the algorithm with default parameters that ensured that every possible insect sound is retained for analysis and this caused some overlap with random noises, thus causing them to be scored as insects when enough such overlaps occurred. Noise "culling" parameters can be readily optimized to reduce analysis of such sounds and subsequent tuning of "clustering" and "matching" parameters will also serve to better separate insects that are located within two sensor spacing intervals of each other. Thus, we can expect that this system will also soon become more accurate at counting low numbers of insects in grain samples, where it already clearly performed better than its predecessor. We plan to proceed next with parameter optimization, followed by more extensive laboratory tests and field trials in the near future.

ACKNOWLEDGEMENTS

The support and interest of Federal Grain Inspection Service, Kansas City, MO is greatly appreciated. This research was supported in part by a specific cooperative agreement between the United States Department of Agriculture and the Agricultural and Biological Engineering Department at the University of Florida and a research support agreement between the United States Department of Agriculture and the Department of Entomology at the University of Florida.

REFERENCES

Hagstrum, D. W., K. W. Vick and P. W. Flinn. 1991. Automated acoustical monitoring of *Tribolium castaneum* (Coleoptera: Tenebrionidae) populations in stored wheat. Journal of Economic Entomology 84: 1604-1608.

Mankin, R. W., J. S. Sun, D. Shuman and D. K. Weaver. 1995. Acoustical shielding for the Acoustic Location Fingerprinting Insect Detector. Proceedings of the Second Symposium on Agroacoustics, National Center for Physical Acoustics, University of Mississippi, Oxford, Mississippi, September 6-7, 1995: xxx-xxx.

Shuman, D., J. A. Coffelt, K. W. Vick and R. W. Mankin. 1993. Quantitative acoustical detection of larvae feeding inside kernels of grain. Journal of Economic Entomology 86: 933-938.

Shuman, D., D. K. Weaver, R. W. Mankin, H. Chia, R. E. Johnson, M. N. Nasah-Lima, R. R. Radford, J. S. Sun and B. A. Weaver. 1995. A cross-correlation implementation of the Acoustic Location Fingerprinting Insect Detector (ALFID) system. Proceedings of the Second Symposium on Agroacoustics, National Center for Physical Acoustics, University of Mississippi, Oxford, Mississippi, September 6-7, 1995: xxx-xxx.

FIGURE CAPTIONS

Figure 1. Placement of infested kernels in ALFID (A - E) and illustration of sensor spacing intervals (1x - 3x) for the combined data sets.

Figure 2. Performance of the current and initial versions of ALFID in counting a one kilogram sample of grain infested with a single late-instar larva of the rice weevil, *Sitophilus oryzae* (L.), developing within a wheat kernel.

Figure 3. Overall performance of the current version of ALFID in counting paired combinations of sound data from single insects as two insects and the performance of the initial version of ALFID in counting a one kilogram sample of grain infested with a two late-instar larvae of the rice weevil, *Sitophilus oryzae* (L.), developing within wheat kernels.

Figure 4. Performance of the current version of ALFID in counting paired combinations of sound data from single insects as two insects and the performance of the initial version of ALFID in counting a one kilogram sample of grain infested with a two late-instar larvae of the rice weevil, *Sitophilus oryzae* (L.), developing within wheat kernels when they are located at one sensor spacing interval of each other.

Figure 5. Performance of the current version of ALFID in counting paired combinations of sound data from single insects (late-instar larvae of the rice weevil, *Sitophilus oryzae* (L.) developing within wheat kernels) as two insects in a one kilogram sample of grain when they are located within two or three sensor spacing intervals of each other.

Figure 6. Performance of the current and initial versions of ALFID in counting uninfested samples.

Appendix

Calculations for the probability of correctly counting two late-instar larvae in a one kilogram sample of grain in ALFID.

The observed probability of correctly counting two insects separated by distance, r, is approximately

$$\rho(r) = r/R \qquad | \; r < R \qquad \text{(I)}$$

where $\rho(r)$ is the probability of a correct count and R is the critical distance at which the probability of a correct count becomes 1 (about 3 sensor spacings). The overall probability of a correct count within a sphere of radius R is the normalized integral of $\rho(r)$ over the total volume

$$P = (4\pi R^3/3)^{-1} {}_0\int^R \rho(r)dv, \qquad \text{(II)}$$

where $dv = 4.\pi.r^2dr$ is an infinitestimal volume element, and dr is an infinitestimal length element. Combining equations I - II yields

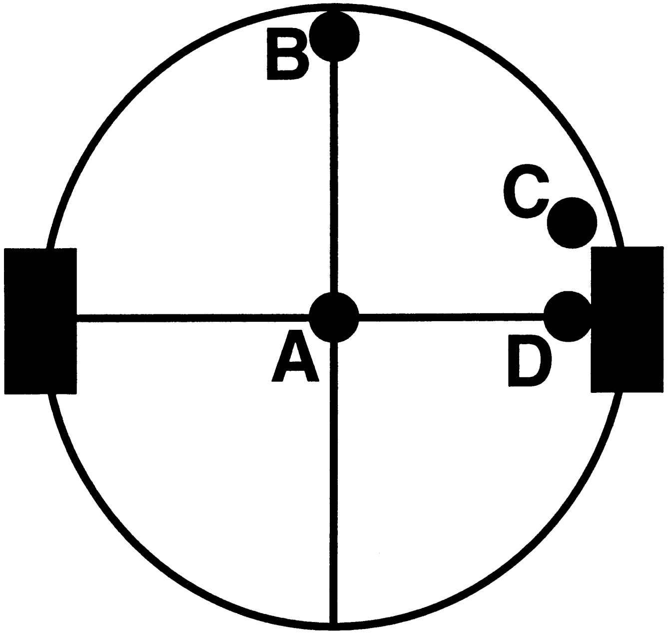
$$P = (4\pi R^3/3)^{-1} {}_0\int^R (4\pi/R)r^3dr, \qquad \text{(III)}$$
$$= (4\pi R^3/3)^{-1}\pi R^3 = 3/4$$

Thus we have an overall probability of accuracy of 0.75 for counting two insects correctly within a critical sphere.

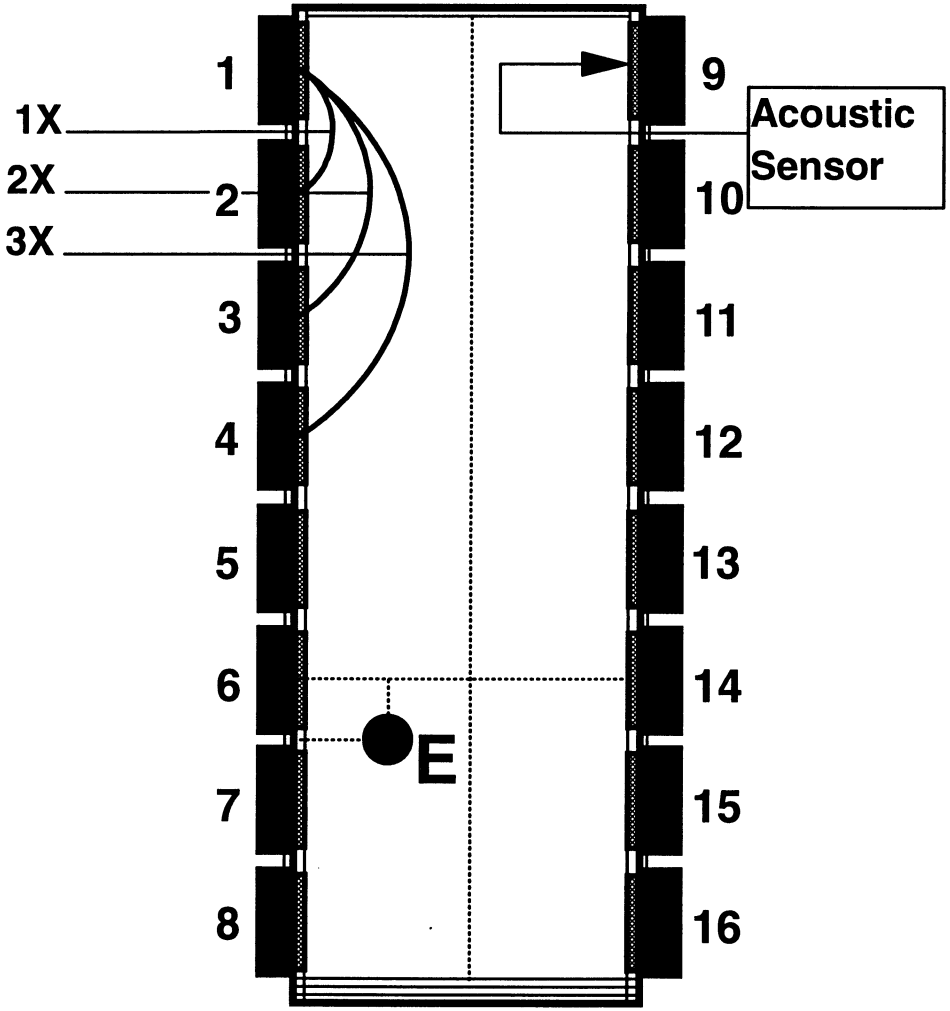
To consider the probability of correct counting over the entire sampling chamber, we used Monte-Carlo methods to determine the probability of two insects that were randomly placed within the ALFID chamber being within the critical radius, R, or three sensor spacing intervals of each other. The simulations were run 100,000 times. Based on these simulations, there is a 60.5% probability that the insects will be located at or within the length of the critical sphere (three sensor spacings) of each other. Therefore, the overall probability of accuracy within the chamber is (0.75 X 0.605) + (1.0 X 0.395) = 0.849 ≈ 85%.

NOTE: The actual value is probably higher than 85% because the critical radius is probably shorter than when first detected at three sensor spacings.

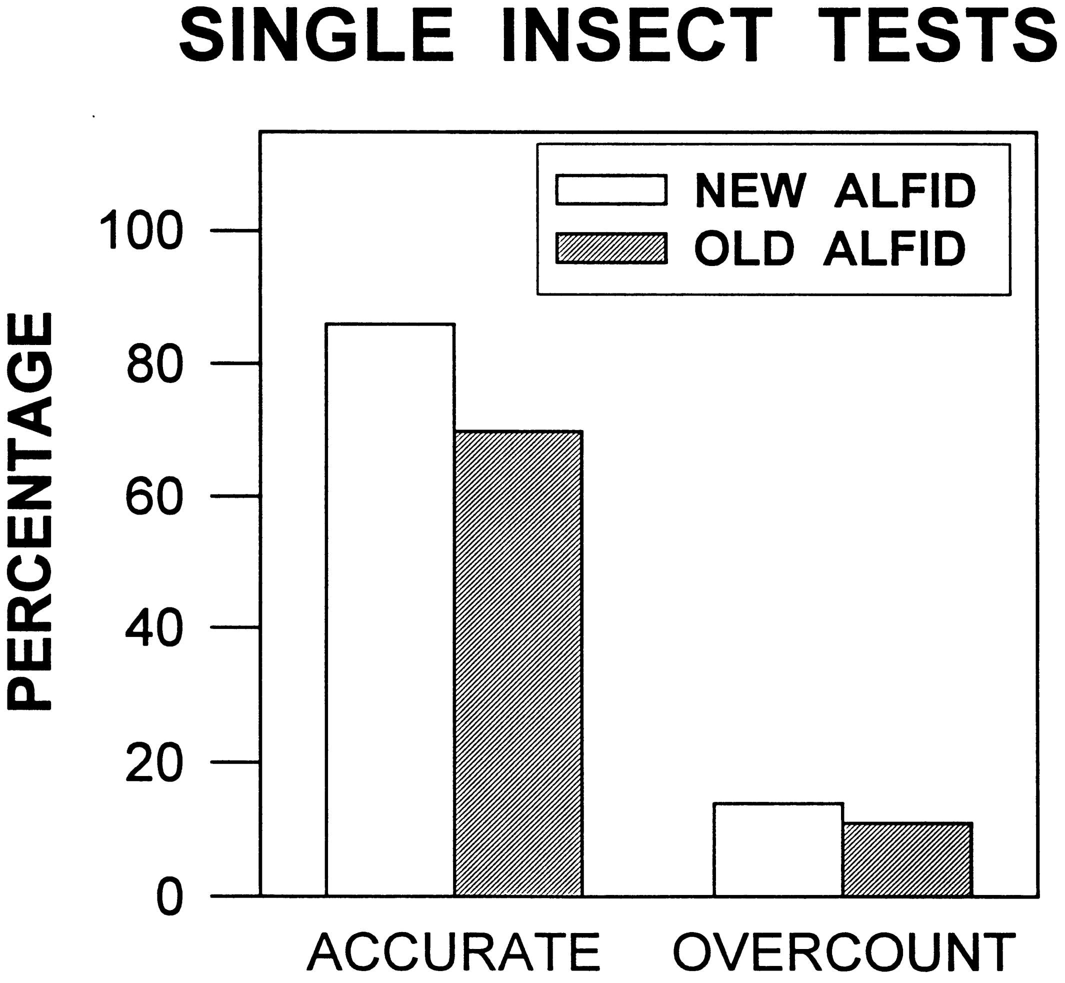
POSITION OF INFESTED KERNELS

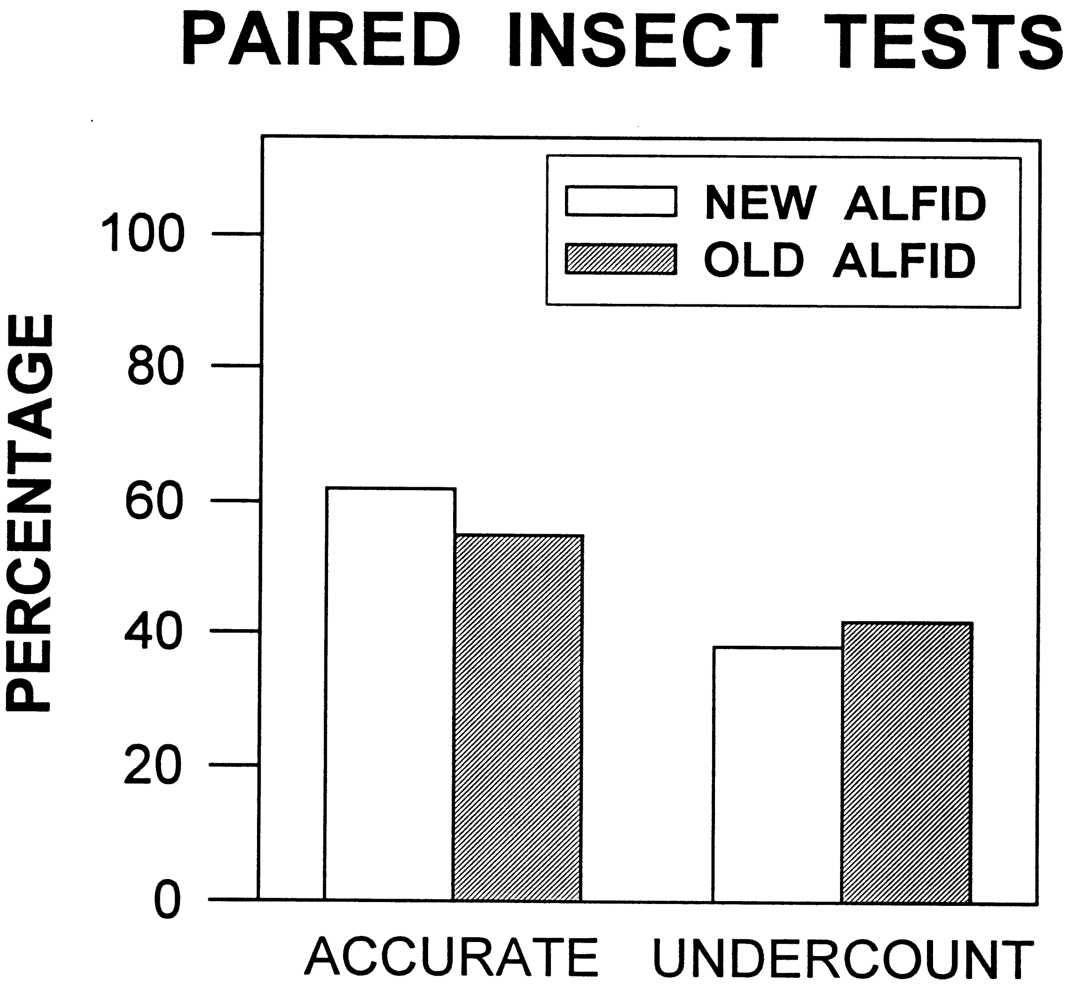


CROSS-SECTION
TOP VIEW

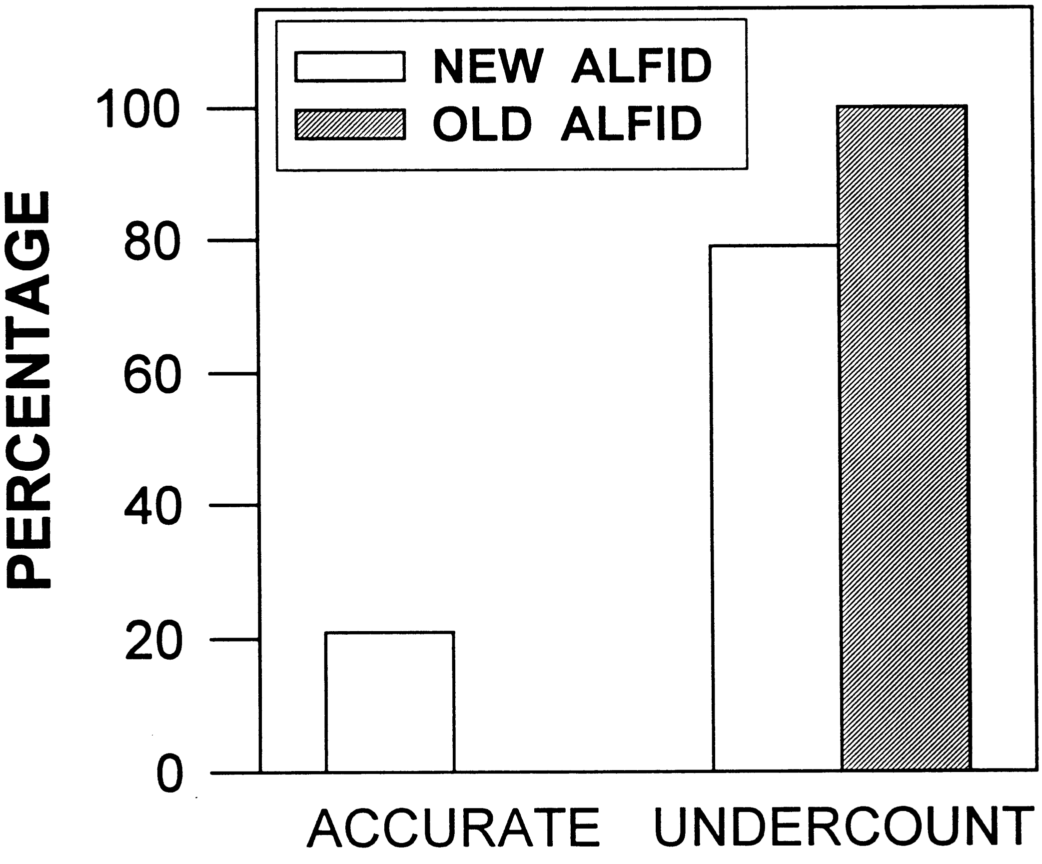


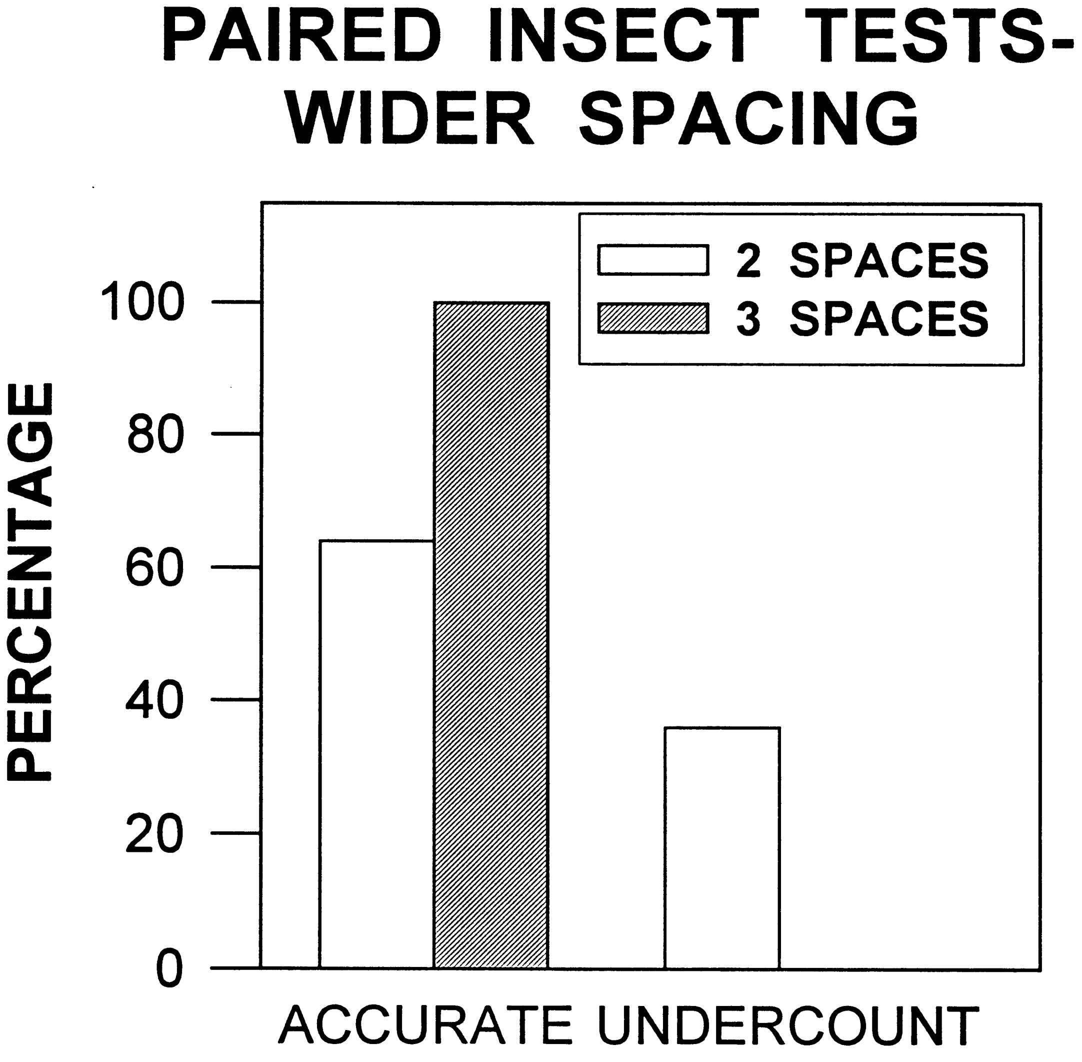
CROSS-SECTION SIDE VIEW


SINGLE INSECT TESTS
NEW ALFID
OLD ALFID
PERCENTAGE
100
80
60
40
20
0
ACCURATE
OVERCOUNT


PAIRED INSECT TESTS
NEW ALFID
OLD ALFID
PERCENTAGE
100
80
60
40
20
0
ACCURATE
UNDERCOUNT

PAIRED INSECT TESTS- ONE SENSOR SPACING




PAIRED INSECT TESTS-
WIDER SPACING
2 SPACES
3 SPACES
PERCENTAGE
100
80
60
40
20
0
ACCURATE
UNDERCOUNT

"NO INSECT" TESTS