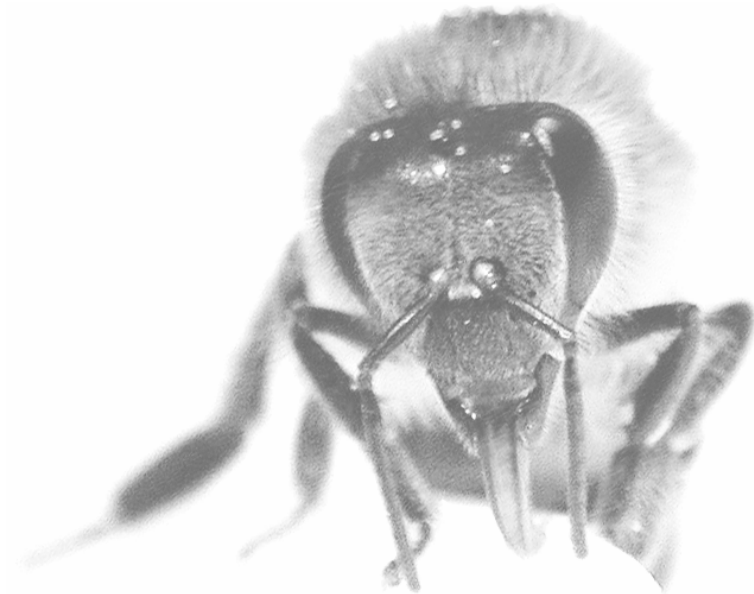


JULIUS-MAXIMILIANS-UNIVERSITÄT
LEHRSTUHL FÜR VERHALTENSPHYSIOLOGIE UND SOZIOBIOLOGIE

**AUTOMATISCHE IDENTIFIZIERUNG BEI SOZIALEN INSEKTEN:
DESIGN UND PRAXISTEST**

PROMOTIONSSCHRIFT

SEBASTIAN STREIT



DR. JÜRGEN TAUTZ (ERSTGUTACHTER)
DR. STEFAN FUCHS (ZWEITGUTACHTER)

I. Inhaltsverzeichnis

Automatische Identifizierung bei sozialen Insekten: Design und Praxistest	1
I. Inhaltsverzeichnis	2
II. Verzeichnis der Abbildungen	4
III. Verzeichnis der Tabellen	4
IV. Zusammenfassung	5
V. Summary	6
1 Einleitung.....	7
1.1 Individuelle Identifizierung bei Insekten.....	7
1.2 Identifizierung bei Insekten.....	8
1.3 Angestrebte Ziele	9
1.3.1 Zuverlässige Identifizierung (Problematik Bienen, Ameisen)	9
1.3.2 Richtungserkennung	10
1.3.3 Mobiler Einsatz.....	10
1.3.4 Masseneinsatz	11
1.3.5 Einfache Auswertungsmöglichkeit.....	11
1.3.6 Kostenfaktor	12
2 Technik: Theorie.....	13
2.1 Stand der Technik – prinzipielle Möglichkeiten	13
2.2 Vergleich Barcode und RFID.....	14
2.3 Geschichte des Barcode	15
2.3.1 Technik.....	17
2.3.2 Vorteile	18
2.3.3 Nachteile	19
2.4 Geschichte der RFID.....	20
2.4.1 Technik.....	20
2.4.2 Vorteile	24
2.4.3 Nachteile	24
2.5 Entscheidung für RFID	25
3 Technik: Praxis.....	26
3.1 Hardware.....	26
3.1.1 Auswahl und Evaluierung der RFID Tags	26
3.1.2 Bau und Evaluierung der Scanner	28
3.1.3 Bau und Evaluierung des Tunnels	31
4 Software	35
4.1 Prinzipielle Forderungen	35
4.1.1 Portabilität	36
4.1.2 Simplizität.....	36
4.1.3 Skalierbarkeit	37
4.2 Realisierung	37
4.2.1 Schichtenmodell.....	37
4.2.2 Konkrete Auswahl der Komponenten.....	38
4.2.3 Verwendete Datenbank.....	39
4.2.4 Datenerfassung mit <i>Data2DB</i> : Präparation	42
4.2.5 Versuchsdurchführung mit <i>Data2SQL</i> : Erfassung.....	44
4.2.6 Versuchsauswertung mit <i>Data2Web</i> : Auswertung	46
4.3 Alternative Auswertungsmöglichkeiten.....	52
5 Anwendung: Proof of concept mit <i>Bombus terrestris</i>	55
5.1 Fragestellung.....	55

5.2	Setup.....	56
5.3	Auswertung	57
5.4	Diskussion.....	58
6	Ausblick.....	60
6.1	Software	60
6.2	Technische Weiterentwicklung und Ausblicke.....	60
7	Danksagung	62
8	Literaturverzeichnis	63
9	Internetquellen.....	64

II. Verzeichnis der Abbildungen

Abbildung 1: Karl von Frisch beim Markieren individueller Bienen mit Farbe	8
Abbildung 2: REM Bild einer Biene mit einem RFID Tag.....	25
Abbildung 3: Aufkleben eines RFID Chips mittels Schellack	27
Abbildung 4: Diagramm der Scanlogik	29
Abbildung 5: Spezielle RFID Scanner integriert in den benötigten Tunnel	30
Abbildung 6: Reichweite und Position des Erkennungsfeld der Scanner	31
Abbildung 7: Schematische Positionierung eines Insekts im Tunnel	32
Abbildung 8: Gesamtansicht eines Tunnels.....	33
Abbildung 9: Durchsicht durch den Tunnel (2 Gänge, von außen nach innen).....	34
Abbildung 10: Abstrahiertes Schichtenmodell des verwendeten Software Designs ..	37
Abbildung 11: Scan-Abfrage in SQL.....	40
Abbildung 12: Data2DB - schematische Übersicht	42
Abbildung 13: Data2DB Programmfenster.....	43
Abbildung 14: Data2SQL - schematische Übersicht.....	44
Abbildung 15: Data2SQL Programmfenster	45
Abbildung 16: Data2Web - schematische Übersicht.....	48
Abbildung 17: Unveränderte Daten des Experimentes	49
Abbildung 18: Daten nach dem ersten Verarbeitungsschritt.....	51
Abbildung 19: Setup des „Proof of Concept“- Experiment	56
Abbildung 20: Abfragebeispiele der Aktivitäten	58
Abbildung 21: <i>Apis mellifera</i> mit RFID Tag	59

III. Verzeichnis der Tabellen

Tabelle 1: Systeme zur automatischen Identifikation.....	13
Tabelle 2: Vergleich Barcode und RFID	14
Tabelle 3: Vergleich ein- und zweidimensionaler Barcode Standards	16
Tabelle 4: Tabellarische Aktivitätsübersicht der ID 2306128607359016917	51

IV. Zusammenfassung

Die Faszination der Beschäftigung mit dem Verhalten sozialer Insekten liegt eben in dem *sozialen* Aspekt dieser Gemeinschaften - genau diesen Aspekt zu beschreiben und vollständig zu erfassen ist aber schwierig.

Komplexe Verhaltensweisen wie beispielsweise die Nestklimatisierung und Thermoregulation von Honigbienen erklären wir heute mit dem Begriff der Emergenz. Emergenz definiert sich dabei als das Entstehen neuer Strukturen aus dem Zusammenwirken vieler Elemente in einem komplexen System. Wenn aber nur dieses Zusammenwirken die Struktur eines Systems wie beispielsweise eines Bienenstockes bestimmt, dann muss jeder Forschungsansatz, der diese Strukturen beleuchten will, die Quelle dieser Emergenz, also die *Masse* der mitwirkenden Individuen, mit einbeziehen.

Hierzu ist ein Mittel die Markierung und Beobachtung einer Auswahl von Individuen. Traditionelle Markierungssysteme wie Farben oder Ringe leisten exzellente Dienste bei der Unterscheidung einzelner Tiere - diese einzelnen Tiere werden aber immer nur eine im Promillebereich liegende Auswahl aus einem Volk mit oft mehreren 10.000 Individuen darstellen, über deren repräsentative Aspekte man nur spekulieren kann. Versucht man, die Anzahl der zu beobachtenden Tiere zu erhöhen, dann stößt man schnell an Grenzen, die vor allem im personellen Aufwand der Beobachtung liegen. Hier bietet die automatische Identifikation einen möglichen Ausweg. Längst genutzt bei Maut, Logistik oder Personenidentifikation, könnte die Umsetzung der aktuell verfügbaren Techniken auf die Insektenforschung eine Vielzahl von Beobachtungen und Versuchsansätzen ermöglichen, die bis heute scheitern. Die Vision, mehrere 1000 Tiere in ihrem Verhalten über Wochen und Monate zu erfassen, wäre Realität.

Die vorliegende Arbeit prüft zunächst verschiedene Systeme zur automatischen Identifikation auf die Nutzbarkeit für die Verhaltensforschung an sozialen Insekten. Aus den verfügbaren Systemen wird mit der RFID (Radio Frequency Identification) Technik ein passendes ausgewählt und daraus eine Lösung für den praktischen Einsatz entworfen und getestet.

Diese Lösung umfasst zum einen die Entwicklung der Hardware, die sich zusammensetzt aus

- den Markierungen
- den Erkennungsgeräten
- und einem funktionierenden Versuchsaufbau an einer Kolonie von Honigbienen.

Zum anderen muss die benötigte Software entwickelt werden, die

- das experimentelle Setup in Form der Markierungserfassung in Gruppen unterstützt
- die automatische Identifizierung und Erfassung der Daten in Datenbanken leistet
- und eine Schnittstelle zur Sichtung und Auswertung der entstehenden Datenmassen bietet.

Die entstehende Lösung wird anhand der Markierung und automatischen Verfolgung von einer Hummelkolonie im Labor getestet.

V. Summary

The fascination of dealing with the behavior of social insects often lies in the *social* aspect of these communities - to exactly describe and completely measure this aspect is however difficult. Looking at behavior like air conditioning or thermal regulation of honey bee hives we explain these complex behaviors with the term of the emergence. Emergence defines itself thereby as the development of new structures from the cooperation of many elements in a complex system.

But if only this cooperation determines the structure of a system like a beehive then each research that wants to light up these structures, must include the source of this emergence, thus the mass of the participating individuals.

To achieve that a common mean is the marking and observation of a selection of individuals. Traditional marking systems such as colors or rings do excellent services in the distinction of a few individual animals - these individual animals will represent but a small selection from a society counting often several 10,000 of individuals - thereby lying in the parts per thousand range, upon whose representative aspects one can only speculate.

If one tries to increase the number of animals which are to be observed then boundaries are reached fast, which consist particularly in the personnel expenditure of the observation.

Here the automatic identification offers a possible way out. Today already used with tolls, logistics or person identification, the conversion of the available up-to-date techniques could make a multiplicity of observations possible, which would be impossible today.

The vision to monitor several 1,000 animals in their behavior over weeks and months would be reality.

This text first examines different systems for automatic identification for the applicability for the behavior research at social insects. From different systems we select the RFID (Radio Frequency Identification) technology as suitable and a solution for the field use is sketched and tested. This solution on the one hand comprises the development of the hardware with

- the markings
- the recognition devices
- and a functioning experimental setup realized on a colony of honey bees

On the other hand the necessary software must be developed, that

- supports the experimental setup in form marking and group definition
- guarantees the automatic identification and collection of the data in data bases
- and offers an interface for the sifting and evaluation of the developing data masses offers

The developing solution is tested on the basis of the marking and automatic monitoring of a bumblebee colony in the laboratory.

1 Einleitung

Die folgende Arbeit dokumentiert das Design, die Entwicklung und den Praxistest eines Systems zur automatischen Identifizierung und Überwachung sozialer Insekten, im Besonderen Bienen und Hummeln.

Aus den aktuell verfügbaren Techniken soll die Technik identifiziert werden, die möglichst optimale Resultate mit praxisorientierter Realisierung in Einklang bringt. Aspekte wie Zeit-, Geld - und Personalaufwand sind zu berücksichtigen: Realisiert werden muss eine praxistaugliche Lösung, die der Arbeit im biologischen Alltag entgegenkommt.

1.1 Individuelle Identifizierung bei Insekten

Für viele Studien ist die individuelle Identifikation unverzichtbar. Auf der einen Seite kann hier der oft sehr breiten Streuung intraspezifischen Verhaltens Rechnung getragen werden, indem man auffällige Individuen eindeutig identifiziert und diese - je nach Schwerpunkt der Untersuchungen - in den Mittelpunkt stellen oder aber sie aus dem Normalbereich des Verhaltens ausschließen kann.

Auf der anderen Seite fordern methodische Ansätze wie das *focal sampling* (ALTMANN 1974) die Möglichkeit zur individuellen Identifizierung, ohne die diese sinnlos werden.

Bei dieser Technik wird eine Teilmenge einer zu untersuchenden Gruppe kontinuierlich beobachtet, um so auf das Verhalten der Gesamtgruppe schließen zu können. Die Größe dieser Gruppe variiert nach Tierart und Verhalten - in jedem Fall aber muss dabei eine eindeutige, individuelle Identifizierung möglich sein.

Dem entgegen stehen Techniken wie das *scan sampling*, in denen in verschiedenen Zeitabschnitten das aktuelle Verhalten der Tiere erfasst wird. Hier ist eine individuelle Erkennung nicht unbedingt erforderlich - allerdings ist diese Technik nur bei einer begrenzten Anzahl Individuen nutzbar und kann - methodisch bedingt - keine komplexeren Verhaltenssysteme wie z.B. Verhaltensabfolgen erfassen.

Die methodischen Schwierigkeiten der individuellen Markierung sind heute für die verschiedensten Tierarten weitgehend gelöst: Je nach Größe und Tierart werden Farben, Tattoos, Ringe, Halsbänder, Fellrasuren usw. genutzt. Die verhaltensspezifischen Implikationen der Markierung dagegen werfen immer wieder Probleme aller Art auf, da die Markierung durchaus das Verhalten des markierten Tieres ändern kann. Erwähnt sei hier als Beispiel eine Studie von NANCY T. BURLEY (BURLEY 1986) an Zebrafinken, die einen starken Einfluss von farbigen Markierungsbändern auf das Geschlechtsverhalten der Vögel nachwies: Bestimmte Farben der Markierungsbänder steigerten die Attraktivität des markierten Tieres, während andere diese senken ließen.

Bei sozialen Insekten als Schwerpunkt des hier zu entwickelnden Systems kommen meist Farbmarkierungen, bunte Drahringe oder Opalithplättchen zum Einsatz, die eine individuelle Unterscheidung sehr großer Anzahlen ermöglichen.

1.2 Identifizierung bei Insekten

Die eindeutige Identifizierung sozialer Insekten bildet die Grundlage vieler Insektenforschung und wurde in der Neuzeit von Karl von Frisch an Honigbienen eingeführt. Dessen System von farblich unterschiedlichen Markierungen auf Thorax und Abdomen der Tiere wird bis heute genutzt:



Abbildung 1: Karl von Frisch beim Markieren individueller Bienen mit Farbe

Eine automatische Erkennung dieses Farbsystems ist heute durchaus möglich - die Schwierigkeit der Skalierung auf viele tausend Individuen sowie die momentane Unmöglichkeit, die Erkennung von bewegten Insekten in Echtzeit zu leisten, lassen es aber nicht als geeignete Grundlage erscheinen. Zu viele Bienen werden gar nicht oder falsch eingeordnet.

Für einige Experimente, in denen es eher um eine einfache Markierung und schnelle - aber eben *nicht* automatische - Erkennung weniger Insekten geht, stellen solche Systeme nach wie vor die beste, praktikabelste und billigste Lösung dar.

1.3 Angestrebte Ziele

Das hier zu entwickelnde System muss langfristig geeignet sein, das Leben einer einzelnen Biene über eine beliebige Zeitspanne zu verfolgen.

Zu dem Leben der Biene gehört dabei natürlich aber jede Änderung seiner Welt - weit über bloße Bewegungsanalyse hinaus. Daher sollten langfristig möglichst die Speicherung unterschiedlicher physikalischer Faktoren (Temperatur, CO₂-Gehalt, Lichtstärke usw.) in das System integriert werden, um die Überwachung aller abiotischen Faktoren, die das Leben der Biene beeinflussen, zu ermöglichen.

Mit einem solchen System wäre der nächste Schritt schnell ausgemacht: Soziale Insekten wie z.B. Honigbienen sind erst durch die Kooperation mit dem Stock in ihrem Leben definiert - ein Überwachungssystem muss also entsprechend diesen Aspekten Rechnung tragen oder letztlich scheitern: Die ideale Ausbaustufe eines Systems muss Qualität und Quantität jeder Interaktion mit anderen Insekten als weit wichtigere Parameter als die rein Abiotischen demnach auch erfassen können.

Leider geht die Realisierung dieser Systeme noch über den aktuellen Stand der Technik hinaus, ist aber letztlich nur eine Frage der Zeit. In der vorliegenden Arbeit möchte ich den Weg zur vollkommenen Überwachung eines Einzelinsektes, der vor rund hundert Jahren eingeschlagen wurde, einen Schritt weitergehen.

Konkret entwickelt und getestet werden soll ein System, das eine beliebige Anzahl von Insekten mit einer möglichst hohen Wahrscheinlichkeit erkennt, deren Bewegungen an definierten Punkten (Stocheingang, Futterstelle) protokolliert und eine Schnittstelle zu den so entstehenden Daten bildet, die eine einfache Sichtung und Auswertung dieser ermöglicht.

1.3.1 Zuverlässige Identifizierung (Problematik Bienen, Ameisen)

Die Verfolgung eines einzelnen Insektes setzt zunächst die Identifikation dieses Insektes voraus. Wo die Verhaltensforschung von Säugern oft auf die natürlichen Unterschiede zwischen den Individuen zurückgreifen kann, stellen sich dem Entomologen die besonderen Bedingungen der Insekten entgegen, die durch Lebensweise, Größe und besondere genetische Verteilung im Falle der sozialen Insekten eine individuelle Identifikation alleine mit dem menschlichen Unterscheidungsvermögen unmöglich machen.

Die Identifizierung, die hier letztlich eingesetzt werden soll, muss eine beliebige Anzahl an Markierungen zulassen und sollte durch ihre Art keinen Einfluss auf das intraspezifische oder individuelle Verhalten des Insektes haben. Da wiederholt nachgewiesen wurde, wie vor allem Farb-Markierungen starken Einfluss auf das markierte Versuchstier haben, muss dieser Punkt besonders beachtet werden, um unverfälschte Ergebnisse erzielen zu können (BURLEY 1986).

Eine zuverlässige Identifizierung ist mit reinen Farbmarkierungen zwar sehr gut möglich, jedoch meist wie erwähnt nicht maschinell auswertbar und durch die fehlende Redundanz der gespeicherten Information fehleranfällig: Die verwendete Farbmarkierung kann durch Feuchtigkeit, aktives Putzen der Biene oder ihrer Nestgenossinnen oder durch Verschmutzung unleserlich gemacht werden.

Auch das Gewicht, die Form und Größe sind wichtige Faktoren: Eine Markierung, die immer wieder zu starkem Putzverhalten veranlasst, die beim Fliegen behindert oder deren Luftwiderstand den Energieaufwand der Bewegungen unverhältnismäßig anwachsen lässt, ist nicht geeignet.

1.3.2 Richtungserkennung

Die Fähigkeit, Richtungen zu erkennen wäre bei einem perfekten System nicht notwendig, solange man nur den Anfangspunkt jedes Tieres kennt. Danach kann man leicht abzählen, wo sich das Tier gerade befinden muss.

In der Realität werden vor allem zwei Faktoren eine zuverlässige Richtungserkennung unumgänglich machen:

- Nicht-Erkennung durch technische Unzulänglichkeiten
- Nicht-Erkennung durch das Verhalten des Versuchstieres

Technische Unzulänglichkeiten sind nicht vermeidbar - seien es echte Fehlfunktionen, schlechte Setups oder ganz einfach ein Stromausfall zum falschen Zeitpunkt. In solch einem Fall könnten alle folgenden automatischen Beobachtungen nicht mehr eindeutig zugeordnet werden - das Experiment müsste an dieser Stelle scheitern.

Die Nicht-Erkennung durch das Verhalten der Versuchstiere darf auch nicht vernachlässigt werden: Biologie definiert sich über komplexes, individuell oft weit differentes Verhalten und hat damit eine weite Spannbreite, die in Technik heute nicht alle erfassbar sind. Schon ein schneller Blick auf den Eingang eines Bienenstockes zeigt Verhaltensweisen, die nur schwer einer Richtung zuordenbar sind: minutenlanges Stehen bleiben auf einer Stelle zum Ventilieren des Stockes oder schnelles Pendeln innerhalb weniger Zentimeter bei Putz- oder Bautätigkeiten sind solche Beispiele.

1.3.3 Mobiler Einsatz

Das System sollte dem Trend der Miniaturisierung im technischen Bereich so weit wie möglich folgen. Zunächst muss dies aus rein praktischen Gründen geschehen: Die Logistik und der Aufwand für den Transport eines Handheld oder Notebook sind ungleich einfacher als der Transport des kühlschrankgroßen optischen Identifizierungssystems, das an jeder Ladenkasse zu sehen ist.

Auch die Energieversorgung, die bei der Beobachtung der Tiere in natürlichen Lebensräumen immer zusätzliche Probleme bereitet, sinkt meist analog zur Größe.

Zum anderen wird der Einsatz eines solchen Systems immer von der Möglichkeit, es an beliebige Orte zu transportieren und zu montieren abhängen: Mit zunehmender Größe der Einheit zur Identifikation der Versuchstieren können auch natürliche Lebensräume wie enge Höhlen oder Baumstämme als Einsatzgebiet genutzt werden.

1.3.4 Masseneinsatz

Die Schwierigkeiten eines Masseneinsatz können sich ergeben aus

- dem Preis (dessen besondere Aspekte unter 1.3.6 behandelt werden) jedes einzelnen Identifizierungsobjektes und dessen Aufnahmegerät
- der Möglichkeit, eindeutige Identifizierungen in dem Objekt abzulegen: So haben bis heute übliche Farbkombinationen oft eine recht begrenzte Anzahl von Kombinationsmöglichkeiten

Vor allem die oft hochmobilen soziale Insekten bringen es auf der einen Seite auf eine Vielzahl von Individuen pro Kolonie, die eventuell markiert werden müssen und kombinieren dies aber mit einer Mobilität, die sich über mehrere Kilometer erstrecken kann. Ein sinnvolles System muss also eine eindeutige Identifizierung aller markierter Individuen einer ganzen Region garantieren, da sich an einem möglichen zweiten Versuch ein paar Kilometer weiter mit hoher Wahrscheinlichkeit Tiere aus weit verstreuten Kolonien treffen werden und trotzdem unterscheidbar bleiben müssen.

Eine Standardisierung der verwendeten Markierung zum weltweiten Einsatz oder zumindest die Verwendung solch variabler Codes, dass eine Doppelbelegung sehr unwahrscheinlich ist, wäre wünschenswert.

1.3.5 Einfache Auswertungsmöglichkeit

Automatische Identifizierung soll vor allem eines sein - „automatisch“. Damit verliert aber der Experimentator die Kontrolle über die Datensammlung - im Bereich sozialer Insekten mit vielen markierten Tieren werden so schnell 100.000 und mehr Datensätze in recht kurzen Zeiträumen gespeichert werden.

100.000 Datensätze sind schon nicht mehr in einem Standardprogramm auch nur darstellbar - dort verkräftet jedes Tabellenblatt zum aktuellen Zeitpunkt gerade mal 65.000 Zeilen. Auf solche Datenmengen spezialisierte Anwendungen gibt es durchaus - die Gefahr aber, dass die interessanten Datenreihen in der Masse der Daten untergehen, ist hoch.

Gerade in der Entwicklung von Techniken, die nur einem beschränkten Personenkreis letztlich dienen werden - und eigentlich immer in der Entwicklung von Software - wird oft vergessen, einen gewissen „Human Factor“ in die Entwicklung mit einzubeziehen. Hier wird schnell eine zu mathematische oder EDV lastige Lösung konzipiert, die in der Personengruppe der Anwender - hier nämlich oft eher mit ganzheitlichen Ansätzen beschäftigte Entomologen - zu keinem weit verbreiteten Erfolg führen wird.

1.3.6 Kostenfaktor

Die Kosten sind heute immer ein begrenzender Faktor. Gerade in einer Zeit, wo hierzulande ein Großteil der „wissenschaftlichen“ Arbeit im Ausfüllen von Antragsformularen für Drittmittel besteht, kann dieser Faktor nicht ignoriert werden. Dabei muss zwischen zwei Kostenarten unterschieden werden:

- Investitionskosten (Scanner, Daten verarbeitende Hardware, Software)
- Laufende Kosten (Markierungen)

Die laufenden Kosten in Form von Markierungen sind im Bereich der Insekten außerhalb der Laborbedingungen nicht recyclebar (sie gehen meist verloren mit dem Tod des Tieres) und müssen daher besonders beachtet werden.

Zu erwarten ist eine unterschiedliche Rentabilität abhängig von der Menge der zu markierenden Tiere: Spielen bei geringen Mengen eher die Investitionskosten die tragende Rolle, so werden bei sehr großen Mengen an benötigten Markierungen die laufenden Kosten zunehmend Gewicht erlangen.

2 Technik: Theorie

Im den folgenden Abschnitten werden die prinzipiellen Techniken und Möglichkeiten der automatischen Identifizierung gezeigt und daraus die für die aktuelle Anwendung passenden diskutiert und ausgewählt.

2.1 Stand der Technik – prinzipielle Möglichkeiten

Automatische Identifikation wird heute in einer Vielzahl von prinzipiell unterschiedlichen Techniken implementiert. Die folgende Tabelle stellt einige wichtige Zweige der daraus resultierenden Einsatzgebiete vor:

Verwendete Technik	Häufiges Einsatzgebiet
Barcode	Produktkennzeichnung (Logistik, Lebensmittel)
Optical Character Recognition	Bankenwesen, Nummernschilder von KFZ
Sprachidentifizierung	Zugangskontrollen
Chipkarten	Bankenwesen
RFID Systeme	Logistik, Identifizierung

Tabelle 1: Systeme zur automatischen Identifikation

Die allgemeine Forderung an die Technik bezieht sich vor allem auf Größe, automatische Detektierbarkeit und Kosten pro Markierung.

Hier ist zunächst die Größe ein wesentlicher Aspekt: Alle anderen Aspekte sind schon heute zufriedenstellend gelöst. Als Beispiel sei die Verwendung des Argos¹ Tracking Systems angeführt. Hier werden zwei Satelliten genutzt, um telemetrisch weltweit die Position von entsprechend markierten Objekten festzulegen. Genutzt wird hierzu der Doppler-Effekt, der sich aus der Bewegung der Satelliten zum Objekt ergibt, um die Position auf ca. 300 m genau festzulegen. Wird dieses System noch mit einem GPS System gekoppelt, kann die Genauigkeit auf unter 100 m verbessert werden. Mit der Entwicklung von Sendern, deren Gewicht unter 160 Gramm liegt, wurde so die Beobachtung von beispielsweise Walen (MATE et. al. 2000), Vögeln (MEYBURG, 2003) oder Fischen weltweit für alle Arbeitsgruppen möglich.

160 Gramm Gewicht bedeuten eine beeindruckende Miniaturisierung der benötigten Techniken (und sind nur möglich, weil Solartechniken die benötigte Energie ganz oder teilweise speisen) - sind aber für die Identifizierung von Insekten noch immer entschieden zu viel.

Noch vor den radio-telemetrischen Techniken sind optische Techniken mit großem Erfolg zur Identifizierung genutzt worden - hier sind vor allem die Barcodes zu nennen, die aus heutigen Geschäftsprozessen nicht mehr wegzudenken sind - zunehmend aber von RFID Techniken verdrängt werden. Nach Tabelle 1 kommen für ein System zur Identifizierung von Insekten am ehesten Techniken in Frage, die auf

¹ <http://www.argosinc.com/biology.htm>

Barcodes oder RFID basieren. Einzig mit OCR (Optical Character Recognition) steht ein System zur Verfügung, das am ehesten sich nahtlos in die bestehenden Markierungssysteme auf Farbbasis einfügen würde, da es diese Farben auch automatisch erkennen und zuordnen könnte. Hier sprechen aber der extreme Aufwand dieser Systeme und die damit verbundenen Kosten gegen eine sinnvolle Implementierung einer solchen Lösung.

2.2 Vergleich Barcode und RFID

Im Folgenden ein kurzer Vergleich einiger wichtiger Parameter von Barcode- und RFID - basierten Systemen leicht modifiziert nach FINKENZELLER (2002):

Parameter	Barcode	RFID Systeme
Datenmenge (Byte)	1 ~ 100	16 ~ 64 k
Datendichte	gering	sehr hoch
Maschinenlesbarkeit	gut	gut
Einfluss v. Schmutz / Nässe	sehr stark	sehr gering
Einfluss von opt. Abdeckung	totaler Ausfall	kein Einfluss
Einfluss von Lage u. Richtung	gering	kein Einfluss
Abnutzung	bedingt	kein Einfluss
Anschaffungskosten Scanner	gering bis sehr hoch	hoch
Lesegeschwindigkeit	gering bis mittel	sehr hoch
Kosten Markierungen	sehr gering	mittel
Größe Markierungen	sehr klein (bis 1 mm ²)	sehr klein (bis 1 mm ²)

Tabelle 2: Vergleich Barcode und RFID

Der Vergleich zeigt beide Systeme als prinzipiell geeignet. Die Datenmengen und -Dichten sind in beiden Fällen ausreichend. Zwar überwiegt die Datenmenge eines RFID Systems zunächst einmal bei weitem die von Barcode Systemen, da hier aber im ersten Schritt nur eine Identifikationsnummer genutzt werden soll, reicht eine Speichermenge von 4 Byte ($= 2^{32} = 4.294.967.296$) in jedem Fall aus.

Wichtiger ist die vergleichbar gute Maschinenlesbarkeit bei guten Bedingungen, wobei hier RFID im Alltag des Experimentes Vorteile für sich verbuchen kann, da diese Systeme weniger anfällig auf Verschmutzung, Abnutzung, schlechte Sichtverhältnisse und die Lage des zu scannenden Objektes reagieren.

Bei den Kosten ergibt sich kein einheitliches Bild: Bei Barcodes stehen der starken Schwankung der Investitionskosten sehr geringe laufende Kosten gegenüber, während bei RFID diese Kosten erst einmal höher anzusetzen sind.

Die Größe beider Markierungen scheint in jedem Fall geeignet zu sein - daher werden die folgenden Kapitel den Aufbau und die prinzipielle Nutzbarkeit für die Identifikation von Insekten beleuchten und daraus schließlich eine Entscheidung für das hier zu implementierende System ableiten.

2.3 Geschichte des Barcode

Barcodes zur Identifizierung von Objekten vielerlei Arten sind eine sehr alte Technik. Im Folgenden soll ein kurzer Überblick über die Entwicklung der Barcodes bis heute gegeben werden, der sich orientiert an PALMER (1995).

Das erste Patent aus diesem Bereich stammt aus dem Jahr 1934 und wurde J.T. KERMODE zugesprochen. Dies beschreibt einen Kartensortierer, der mit vier parallelen Linien als Identifikationsschema arbeitete. 1935 erhielt D. A. YOUNG ein Patent für eine ähnliche Sortiermaschine, die über eine Anordnung verschiedener optischer Markierungen die Karten identifizierte.

Die Beschreibung der Vorteile und die Konzeption eines darauf basierenden Gesamtablaufs, die automatische Identifizierung in den Bereichen vor allem der Warenlogistik bringen können, wurde auch in den 30er Jahren des vorherigen Jahrhunderts zum ersten Mal entwickelt. Die Vision eines vollautomatischen Auslieferungssystems mit Abrechnung und automatisierter Lagerhaltung war damals noch technisch nicht realisierbar – bewies aber die visionäre Kraft von WALLACE FLINT, der später aktiv bei den Normierungsbestrebungen mitwirkte, die dann zum sog. UPC Code führten.

Die späten 40er Jahre führten dann zum ersten Patent eines Barcodes, wie er prinzipiell auch heute noch in Gebrauch ist. NORMAN JOSEPH WOODLAND und BERNARD SILBER entwickelten den - auch als Bull's Eye Code bezeichneten - Barcode, der konzentrische Ringe um einen zentralen Mittelpunkt zur Identifikation von zunächst 7 Artikeln (bei 4 genutzten Linien) und dann später bis zu 1023 Artikel (bei 10 Linien).

Noch war die technologische Ausstattung des Einzelhandels nicht zu einer so massiven Umrüstung der Logistik bereit: Die ersten Tests des *Bull's Eye Codes* im Groß-einsatz scheiterten noch an verschiedenen technischen Unzulänglichkeiten. In Folge dessen wurden die Anstrengungen intensiviert, einen einheitlichen Standard im Zusammenspiel aller in die späteren Produktions-, Logistik- und Verkaufsketten involvierten Unternehmen zu schaffen.

Diese **UGPCC** Gruppe (**U**niform **G**rocery **P**roduct **C**ode **C**ouncil) präsentierte als Ergebnis am 3. April 1973 den U.P.C. (Universal Product Code) als gemeinsamen Industriestandard. Hauptentwickler dieses Codes war IBM² – sein Entwickler war GEORGE J. LAURER. Der enorme Erfolg des UPC Codes in den amerikanischen Supermärkten zog schnell das Interesse des europäischen Marktes auf sich. Im Dezember 1976 wurde ein sehr ähnliches System - von demselben Entwickler - als **EAN** (**E**uropean **A**rticle **N**umbering) etabliert. Die weitere Vereinheitlichung sieht vor, dass bis zum Jahr 2005 alle Lesesysteme in Amerika neben UPC auch den EAN-13 Code erkennen können.

UPC und EAN Codes sind dabei optimiert auf gute Sichtbarkeit auf einer Vielzahl von Oberflächen - und damit entsprechend ungeeignet für Systeme, bei denen extremer Platzmangel die Größe des verwendeten Codes maßgeblich mitbestimmt.

Tatsächlich handelt es sich bei diesen Strichcodes um eindimensionale Codes, deren zweite Dimension - die Höhe - letztlich nur zur Redundanz der Information dient und so Verknitterung oder Verschmutzung des Untergrunds (also z.B. der Lebensmittel-

² <http://www.ibm.com/>

verpackung) entgegenwirkt. Kommen hierzu auch noch das Bedürfnis, eine möglichst hohe Anzahl individuell verschiedener Codes unterscheiden zu können, reicht diese eine Dimension, die zur Kodierung der Information verwendet werden kann, nicht mehr. Im Jahre 1988 wurde daher als erster zweidimensionaler Barcode der *Code 49* eingeführt. Seitdem sind noch zahlreiche weitere zweidimensionale Barcodes - auf verschiedene Anwendungen optimiert - standardisiert worden. Im folgenden Beispiel wird als Zwischenform mit *Codablock* eine sog. „Gestapelter Code“ gezeigt; des Weiteren mit *MaxiCode* und *DataMatrix* echte, zweidimensionale Barcode Standards.

Die folgende Abbildung zeigt den Begriff „Automatic Insect Identification“ kodiert in verschiedenen Barcode Techniken:

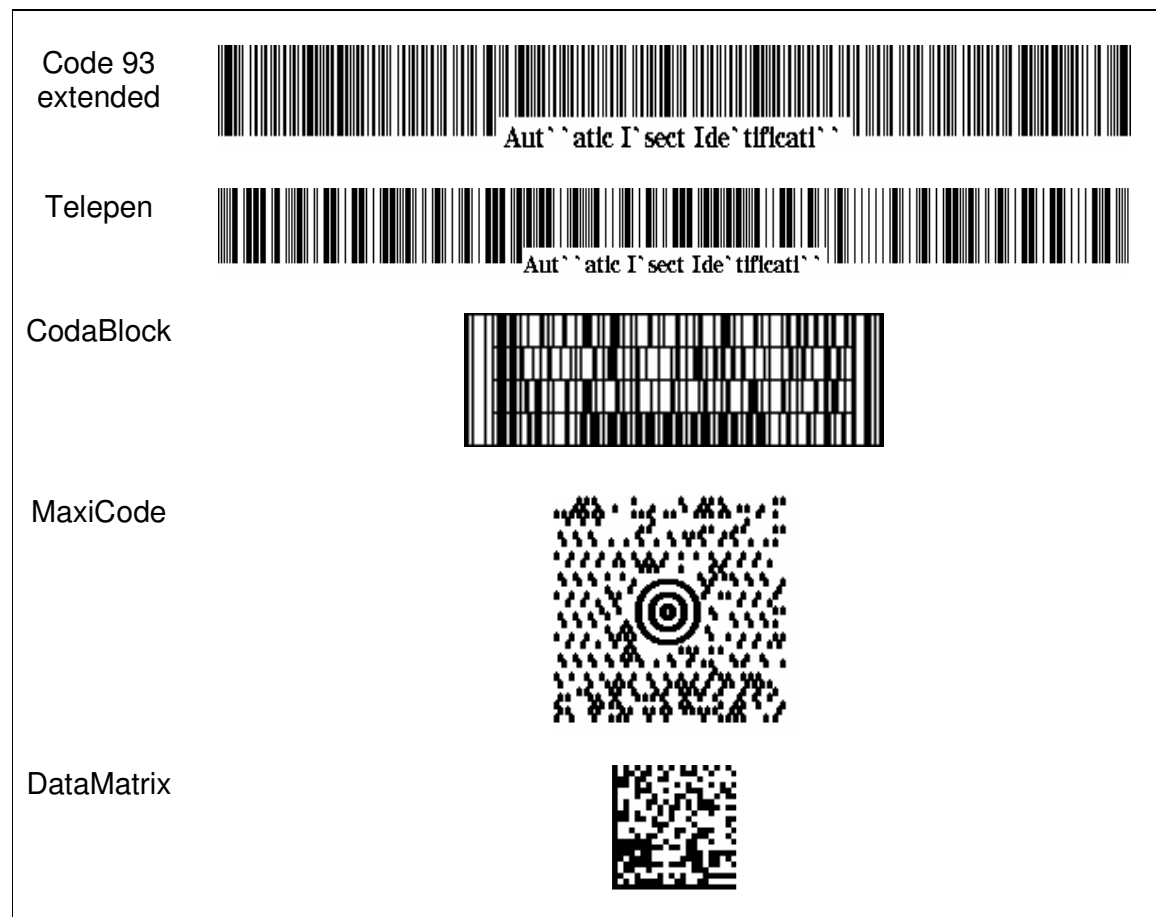


Tabelle 3: Vergleich ein- und zweidimensionaler Barcode Standards

2D Codes setzen sich aus zwei Untergruppen zusammen: Den sog. „Gestapelten Codes“, zu denen auch der erwähnte *Code 49* gehört, sowie den Matrix Codes. Gestapelte Codes sind letztlich mehrere eindimensionale Strichcodes mit einer sehr kurzen Länge in der nicht genutzten zweiten Dimension übereinander. Hier wird der Zuwachs an Information mit einem Abfall der Redundanz und damit der Fehlertoleranz erkauft.

Matrix Codes dagegen führen ein neues Prinzip ein und codieren die Informationen in einer zweidimensionalen Matrize, in der die Position der Einzelpunkte zueinander den Informationsgehalt definiert. In dieses Prinzip integriert ist eine Fehlerkorrektur und mögliche vielfache Redundanz der Informationen, was die Fehlertoleranz dieser Codes stark erhöht. Gleichzeitig steigt durch die Verwendung der zweiten Dimension die Informationsdichte stark an: Die maximale Datenkapazität eines sog. DataMatrix ECC-200 Symbols beträgt 3116 numerische oder 1558 alphanumerische Zeichen (8bit) in einem Symbol bestehend aus 144 Zeichen pro Symbolseite.

Ein weiterer, in der Verhaltensforschung ebenso essentieller Punkt ist die möglichst maximale Unabhängigkeit der Lage des Codes zum Lesegerät. Strichcodes sind hier besonders anfällig, wogegen Matrizen basierte Codes omni-direktional aus jedem Winkel heraus auslesbar sind.

Nutzt man den DataMatrix Code ausschließlich zur Bereitstellung einer eindeutigen Informationsnummer, dann reichen in der Regel 5 - 6 Zahlen zur Darstellung der Identifikationsnummer: So muss nur ein kleiner Teil der verfügbaren Kapazität für die eindeutige Nummer genutzt werden - der Rest steht für redundante Informationen zur Verfügung. Bei einer Identifizierung von Insekten liegt es nah, einer hohen Informationsredundanz den Vorzug zu geben, um möglichen Fehlfunktionen bei z.B. Verschmutzung entgegenzuwirken. Nutzt man nur einige Prozent der maximalen Kapazität, ermöglicht die entstehende Redundanz das Auslesen der Information bei reziproker Verschmutzung oder Beschädigung des Matrix Codes: Selbst eine Beschädigung, die auf 90% der Fläche des Codes einwirkt, verhindert das Auslesen in der Regel nicht.

Die geringe Größe dieser 2-D Codes, verbunden mit der hohen Anzahl der möglichen Codes lässt diese Produktklasse prinzipiell als eine geeignete für den Einsatz bei der Identifizierung sozialer Insekten erscheinen.

2.3.1 Technik

2.3.1.1 Barcode Generierung

Durch die extreme Verbreitung von Barcodetechnologien in verschiedenen Industrien existiert eine Fülle von fertigen Softwareprodukten oder Entwicklungsumgebungen, die die Erstellung der Barcodes und die Kopplung des Herstellungsprozesses mit Datenbanken ermöglichen.

2.3.1.2 Label Erstellung

Barcodes können prinzipiell mit jedem Drucker ausgedruckt werden. Die Anforderung an die Qualität des Druckers steigt parallel mit der Informationsdichte des Barcodes. Eine anvisierte Größe des 2D-Barcodes von maximal 1,5 * 1,5 mm erfordert hochwertige Laser- oder Thermotransferdrucker - sind dann aber mit diesen zuverlässig herzustellen. Die letztlich erreichte Qualität der Barcodes ist eine Funktion der Drucker- und Papierqualität auf der einen und der Konzentration der Information auf der anderen Seite und muss entsprechend angepasst werden.

Als Träger der Label kommen verschiedene beschichtete Papier- oder Kunststoffmaterialien in Frage, die die entsprechenden Anforderungen an Wasserbeständigkeit und Verträglichkeit mit Klebmaterialien mit sich bringen.

Die Generierung der Barcodes ist auf der Ebene der Software kein Problem. Das Ausdrucken und Anbringen der Codes erfordert hochwertige Drucker und Materialien, ist aber mit den verfügbaren Geräten nicht problematisch.

2.3.1.3 Barcode Detektion

Als Scanner kommen bei zweidimensionalen Barcodes ausschließlich CCD (**charge-coupled device**) Scanner zum Einsatz. Diese arbeiten mit LEDs, die den Barcode mit Licht bestrahlen und das reflektierte Licht auf einen speziellen optischen Chip übertragen. Die Photonen des einfallenden Licht können dort Elektronen auf einen höheren Energielevel befördern und befreien diese dabei teilweise vom Atom. Diese freien Elektronen werden in Kondensatoren aufgefangen und geben in ihrer Summe ein sehr klares Bild der Menge des einfallenden Lichtes pro Fläche. Dabei spielt die Wellenlänge des Lichtes eine wichtige Rolle: Licht zu langer Länge durchdringt den Halbleiter bis zu einer gewissen Tiefe, ohne mit der sensitiven Schicht zu interagieren, während Licht zu kurzer Länge von der Oberfläche reflektiert werden kann.

Die Photozellen des Chips übersetzen das einfallende Bild in ein elektronisch dekodierbares Signal und dekodieren damit die Information des Matrixcodes.

Die Dichte des CCD Sensors bestimmt also das maximale Auflösungsvermögen des Scanners. In den Größenbereichen die in der hier untersuchten Anwendung benötigt werden, muss hier eine maximale Dichte angestrebt werden bzw. entsprechende optische Vergrößerungen verfügbar sein.

2.3.2 Vorteile

Der Vorteil der Barcodes liegt klar auf der Seite der Erstellung.

Für den Einsatz bei sozialen Insekten ist auch immer die Massenproduktion von Labels interessant: Bei Barcodes spielen die Kosten der Erstellung der Codes eine untergeordnete Rolle. Bei der Verwendung von hochwertigen Standarddruckern, wie sie in jedem Büro heute zu finden sind, kommt der einzelne Barcode selbst auf beschichtetem Spezialpapier auf Kosten von unter einem € - Cent. Die weithin verfügbaren Softwarebibliotheken reduzieren Zeit und Kosten der Entwicklung weiterhin. Die entstehenden Label sind leicht und können durch Versiegelung mit Lacken wasserbeständig gemacht werden.

Der Aufwand beim Aufbringen der Label auf dem Insektenthorax ist mit der Markierung mittels Opalithplättchen zu vergleichen und liegt damit unter einer Minute pro Insekt.

2.3.3 Nachteile

Den Vorteilen der Erstellung der Barcodes stehen Nachteile im Bereich der Detektion entgegen. Das Licht ist dabei naturgemäß die wichtigste Größe optischer Systeme und muss hier sowohl in Wellenlänge als auch in Lichtstärke auf den jeweiligen Scanner optimiert werden.

Eine Detektion solch kleiner Barcodes bei hoher Geschwindigkeit erfordert eine sehr gute Ausleuchtung des zu scannenden Bereiches und kann damit bestimmten experimentale Bedingungen zuwider laufen: Lichtscheue Insekten werden so ganz ausgeschlossen, während bei allen anderen Insekten der Einfluss des Lichtes während der Dunkelheitsperioden einen starken artifiziellen Einfluss auf das zu beobachtende Verhalten darstellen kann. Wie stark dieser Einfluss sein kann, zeigt recht anschaulich allnächtlich jede Straßenlaterne und jede Kerze...

Ein weiterer Punkt ist weniger konzeptioneller denn monetärer Struktur: CCD-Scanner solcher Präzision sind sehr teuer. Einer der führenden Anbieter für 2D Scanner im industriellen Einsatz ist die 1982 gegründete Firma Microscan³, die mit ihrer „Quadrus EZ“ Produktionslinie⁴ einen entsprechenden Scanner anbietet - zu einem Preis von etwa 1700 € pro Scanner zum diesem Zeitpunkt. Jeder zu überwachende Tunnel des experimentellen Setups braucht mindestens einen solchen Scanner, so dass diese Kosten hier sehr schnell wachsen.

Der starke Preisverfall bei CCD Geräten, der sich ja z.B. bei digitalen Kameras heute zeigt, wird hier ohne Zweifel die Anschaffungskosten im Laufe der Zeit erheblich reduzieren - die hohe Spezialisierung dieser Scanner mit speziellen Optiken und extrem kurzen Schaltzeiten wird dem auf der anderen Seite entgegen wirken.

³ <http://www.microscan.com/deutsch/>

⁴ <http://www.quadrus-ez.com/>

2.4 Geschichte der RFID

Die Verbreitung von RFID (Radio Frequency Identification) Technologien begann in den 60er Jahren des letzten Jahrhunderts mit der Einrichtung von elektronischen Warensicherungssystemen (EAS, Electronic Article Surveillance). Diese auf elektromagnetischer Basis arbeitenden Systeme sind bis heute an den Ein- und Ausgängen der meisten größeren Kaufhallen zu finden.

Der Fortschritt in Elektronik und Informationstechnologie der 70er erlaubte den Einsatz in der Automatisierung, der Kennzeichnung von Großvieh und zunehmend der Kontrolle von Mautsystemen im Verkehrsbereich. Vor allem dieser letzte Punkt wurde in den 80ern stark ausgebaut, als mehrere amerikanische Bundesstaaten sowie Norwegen flächendeckend RFID basierte Systeme in diesem Bereich einführten.

In den letzten 15 Jahren ist die Miniaturisierung und günstige Massenproduktion prägendes Merkmal der RFID Entwicklung und stellt damit auch den Durchbruch für den breiten Einzug von RFID Markierungen in unser tägliches Leben dar. Im weiteren Sinn zu den RFID Systemen gezählt werden dabei auch Chipkarten, wie sie im Zahlungsverkehr heute beinahe ausschließlich eingesetzt werden. Zwei aktuelle Beispiele aus einer Fülle von Ankündigungen jede Woche sind

- der Umstieg zu RFID im Fahrkartensystem der London Underground bis 2005
- der Ersatz aller Hundemarken mit RFID Transponder in Wien innerhalb 2004

Dies zeigt, dass aktuelle Schätzungen, die von bis zu 5 jederzeit getragenen RFID Transponder pro Person in Deutschland für das Jahr 2010 sprechen, als eher konservativ zu bewerten sind.

2.4.1 Technik

2.4.1.1 Einführung

RFID beschreibt letztlich ein Bündel unterschiedlicher Technologien. Zum Einsatz kommen ebenso eher einfache elektro-magnetischen Verfahren wie auch komplexe Mikrochip basierte Technologien.

Prinzipiell besteht ein RFID System dabei immer aus zwei Komponenten:

- dem Transponder oder „Tag“, der auf dem zu identifizierenden Objekt angebracht wird
- dem Lesegerät oder Lese-/Schreibgerät (Scanner), das den Transponder auslesen kann

Der Transponder besteht aus einem Koppellement sowie aus einem Mikrochip. Das Lesegerät (auch als Antenne oder Scanner bezeichnet) besteht aus einem Hochfrequenzmodul sowie dem Koppellement zum Transponder.

Eine erste Orientierung bietet die Einteilung der genutzten Markierungen (Tags) in aktive und passive RFID Tags, der auch prinzipiell unterschiedliche Technologien als Grundlage dienen.

Aktive Tags verfügen über eine eigene Energiequelle und können meist sowohl gelesen, als auch beschrieben werden. Sie befinden sich als Standard in einem Ruhezustand, in dem sie keine Informationen aussenden. Sobald ein spezielles Aktivierungssignal empfangen wird, reagiert der Sender. Aktive RFID - Tags sind durch Ihre Energiequelle im Vergleich zu passiven Tags meist größer und schwerer. Wegen ihrer höheren Sendereichweite und größeren Speicherkapazität der verwendeten Mikrochips werden sie beispielsweise in Mautsystemen gerne eingesetzt.

Passive Tags haben keine eigene Energiequelle und müssen daher ihre Energie zur Übertragung der Informationen aus den empfangenen Funkwellen beziehen. Der Aufbau der verwendeten Chips erlaubt einzig das Lesen der gespeicherten Informationen. Dieser Speicher wird meist benutzt, um eine eindeutige Identifikationsnummer zu hinterlegen.

Im Vergleich der beiden Ansätze ergeben sich die folgenden Vorteile:

aktive Tags:

- hohe Reichweite
- große Datenmengen speicherbar

passive Tags:

- sehr kleine Baugrößen realisierbar (bis 1 mm^3)
- nahezu unbegrenzte Lebensdauer
- sehr günstig in Massenproduktion realisierbar

Die Anforderung an das hier vorgestellte System würde natürlich von den Vorteilen der aktiven Tags sehr profitieren - allein die Größe und das Gewicht, das durch eine Energiequelle der aktiven Tags unvermeidbar ist, lässt momentan nur passive Tags als die geeignete Wahl für Insektenidentifikation erscheinen.

2.4.1.2 Frequenz, Kopplung

Wichtigste Unterscheidung zwischen verschiedenen passiven RFID Systemen sind Betriebsfrequenz, das physikalische Kopplungsverfahren und die resultierende Reichweite.

- Die Frequenzen variieren dabei zwischen 135 kHz und 5,8 GHz
- Die Kopplung von Sender und Tag kann elektrisch, magnetisch oder elektromagnetisch erfolgen
- Die Reichweite der Systeme variiert zwischen einem Millimeter und bis zu 15 Metern

Die maximale Entfernung, die der Tag zu dem Scanner haben darf, ist im praktischen Einsatz ein bestimmendes Element. Sie hängt ab von

- a) der Positioniergenauigkeit der Transponder (Tag)
- b) dem minimalen Abstand mehrerer aufeinander folgender Transponder
- c) der Geschwindigkeit der bewegten Transponder relativ zum Scanner
- d) der Größe der Antenne des Tags
- e) der Leistung der verwendeten Scanner

Die Punkte *a* - *c* sind abhängig vor allem vom Versuchstier, welches den RFID - Tag trägt. Hier ist eine artspezifische Differenzierung zu erwarten, da die Geschwindigkeit der Tiere und die Möglichkeiten, diese exakt zum Scanner zu positionieren, stark differieren.

Die Punkte *d* und *e* sind rein technischer Natur. Dabei stellt die Größe der Antenne auf den Tags (*c*) einen Punkt dar, mit dessen Maximierung eine größere Leistung des Systems (Reichweite) zu erreichen wäre - dessen Minimierung aber erklärtes Ziel des Setups ist, um mögliche Einflüsse durch abstehende Antennen auf das Verhalten der Tiere auszuschließen. Hier bietet sich eine integrierte Lösung an, die Reichweit mit minimalem Flächenbedarf vereint, die sog. „*coil-on-chip*“ Technologie.

Bei den meisten Bauformen besteht die Antenne der Tags aus einer Transponder-spule, die mit dem Transponderchip gekoppelt wird (hybride Technologie). Dabei liegt die Größe des Chips meist im Bereich $1-5 \text{ mm}^3$, die der Antenne im Bereich einiger dutzend Quadratzentimeter.

Die „*coil-on-chip*“ Technologie dagegen integriert die Spule auf der Oberseite des Chips. Die Spule wird als planare Spirale unmittelbar auf dem abschließenden Isolator des Chips platziert und mit den darunter liegenden Leiterbahnen gekoppelt. Darüber folgt eine Polyamidversiegelung, die die Abschlusspassivierung leistet.

Der Punkt *e*, die Leistung der verwendeten Scanner, verdient besondere Betrachtung. Wie jede Strahlung nimmt die die Leistung der Scanner mit dem Quadrat der Entfernung ab - man muss also mit sehr starken Energien arbeiten, um die Reichweiten nennenswert zu erhöhen.

Dabei läuft man zunehmend in Gefahr, dass Streustrahlung aus dem Setup heraus auf die umgebende Umwelt wirkt. Hier werden zunächst einmal Aspekte der Arbeitsplatzsicherheit berührt. Hochfrequenzstrahlung wird von Wasser - und damit vom Körpergewebe absorbiert, was mit einem Anstieg der Gewebetemperatur verbunden ist. Davon profitiert z.B. die Hochfrequenzverleimung von Polykondensationsklebstoffen und Polyvinylacetatleimen, bei der durch eine hochfrequente Strahlung (13.56 MHz) Hölzer bis zu 10 mal schneller als mit konventionellen Methoden verleimt und getrocknet werden können. Eine Erhöhung der Körpertemperatur des Versuchstieres oder des Experimentators erscheint dagegen eher unerwünscht und begrenzt daher die beliebige Leistungserhöhung der Antennen.

Auf der anderen Seite ist die Verwendung von Sendern elektrischer Wellen jeder Art weltweit streng reglementiert, da zu viele Aspekte der Infrastruktur von Störungen dieses Bereiches berührt werden können. Die Feldstärke eines Senders mit einer Frequenz von 13.56 MHz darf dabei in Deutschland im Abstand von 10 Metern noch eine magnetische Feldstärke von $42 \text{ dB}\mu\text{A/m}$ betragen.

2.4.1.3 Gesetzlicher Rahmen

RFID Systeme erzeugen elektromagnetische Wellen und strahlen diese ab, sind also rechtlich als Funkanlage einzuordnen. Ohne ausreichende Regelung könnte eine Störung so wichtiger Dienste wie Polizeifunk, GSM Netze oder Flugdienste nicht ausgeschlossen werden. Daher wurden bestimmte Bereiche der Funkfrequenz für die Verwendung bei industriellen, wissenschaftlichen oder medizinischen Anwendungen reserviert - man spricht weltweit auch von den ISM-Frequenzbereichen (**I**ndustrial/**S**cientific/**M**edical).

Für die vorliegende Anwendung wird der Bereich 13,553 bis 13,567 MHz betrachtet werden, da dieser Hochfrequenzbereich bei der Verwendung von sehr kleinen Tags die meiste Anwendung findet. Neben RFID Systemen werden in diesem Bereich auch beispielsweise Modellsteuerungen oder Personenrufanlagen betrieben.

Mit der „ERC Recommendation 70-03 relating to the use of *short range devices* (SRD)“ liegt seit 1997 eine Grundlage für nationale Regulierungsvorschriften vor, die die Verwendung dieser Strahlungen regelt. Herausgeber ist die CEPT⁵, die „European Conference of Postal and Telecommunications Administrations“, genauer deren damalige Untergruppe ERC („European Radiocommunications Committee“⁶).

Unterschieden wird in der REC 70-03 zwischen 13 verschiedenen Anwendungen, die in sog. Annexen definiert sind. RFID Anwendungen fallen dabei unter Annex 9 („Inductive applications“) und 11 („RFID applications“).

Weiterhin bildet die Norm *EN 300 200*⁷ die Grundlage nationaler europäischer Regulierungsvorschriften für Funkanlagen kleiner Leistung.

Diesen Empfehlungen und Richtlinien folgen in Deutschland die Verfügungen 61/2000⁸ sowie 73/2000⁹ die im „Amtsblatt der Regulierungsbehörde für Telekommunikation und Post (RegTP)“¹⁰ im Jahr 2000 veröffentlicht wurden.

Weltweit sind in diesem Bereich noch stark unterschiedliche Regelungen zu finden, bei denen es aber aktuell Bemühungen gibt, diese zu vereinheitlichen. So planen die USA und Japan eine Angleichung der nationalen Regelungen im Sinne der REC 70-03.

⁵ <http://www.cept.org/>

⁶ heute: European Radiocommunications Office - <http://www.ero.dk/>

⁷ „Radio Equipment and Systems (RES); Short Range devices, Technical characteristics and test methods for radio equipment to be used in the 25 MHz to 1000 MHz frequency range with power levels ranging up to 500 mW“

⁸ „Allgemeinzuteilung von Frequenzen für die Benutzung durch die Allgemeinheit für induktive Funkanlagen des nicht-öffentlichen mobilen Landfunks (nömL)“, Amtsblatt 12

⁹ „Allgemeinzuteilung von Frequenzen für die Benutzung durch die Allgemeinheit für Funkanlagen geringer Leistung des nicht-öffentlichen mobilen Landfunks (nömL) in ISM Frequenzbereichen; SRD (Short Range Devices)“, Amtsblatt 18

¹⁰ <http://www.regtp.de/>

2.4.2 Vorteile

Die Vorteile von RFID Systemen liegen unter anderem in der hohen Plastizität der möglichen Technologien. Speziell geformte Transponder in allen denkbaren Bauformen sind ebenso verfügbar wie Scanner verschiedener Reichweite und Frequenz. Das Bündel prinzipiell unterschiedlicher Technologien, die als RFID zusammengefasst werden, bietet für viele Anwendungen eine Lösung. Eine berührungslose Identifikation, die durch Verschmutzungen und viele Materialien hindurch arbeitet, also nicht auf irgendeine Ausleuchtung oder auch nur direkte Sichtlinie angewiesen ist, erscheint ideal. Die Verwendung von Mikrochips bietet eine zusätzliche Perspektive, da hier ja ein nutzbarer Datenspeicher ein ständiger Begleiter des zu beobachtenden Tieres wird. Schon heute werden Sensoren mit RFID Systemen gekoppelt, um diese berührungsfrei auslesen zu können - vor allem Temperatursensoren werden so realisiert. Mit der fortschreitenden Miniaturisierung sind mittelfristig hier weitere Optionen zu erwarten, die letztlich zu einem Miniaturlabor auf dem Rücken eines einzelnen Insektes führen können - das ja dann in jedem Fall Speicher zum Ablegen der Messwerte benötigt.

Die Lebenszeit von passiven RFID Tags ist dabei zunächst einmal unbeschränkt und wird in der Praxis durch ihre relative Unempfindlichkeit gegen Wasser oder Hitze tatsächlich ausreichend sein.

2.4.3 Nachteile

Durch die Verwendung von Mikrokontrollern als Bestandteil des Tags sind die Kosten pro Tag sehr hoch. Aktuell ist mit einem Euro pro Tag zu rechnen. Hier kann aber von einer weiteren Verbesserung ausgegangen werden: Aktuell werden Drucker präsentiert, mit denen sich die benötigten Leiterbahnen direkt drucken lassen - die entstehenden Produkte werden schließlich in den Bereich von 5 - 10 Euro Cent pro Tag sinken.

Der Flexibilität der Technik steht aber auch ein gewisser Zwang in der Produktion gegenüber: Wo bei Barcodes die Produktion der Codes vollkommen in der Hand des Experimentators liegt, ist dies bei RFID Systemen kaum möglich und im Gegenteil auch nicht so schnell vorstellbar. Die Kosten für Design, Produktion und Verarbeitung eines Wafer mit RFID fähigen Mikrochips liegen im Bereich von mehreren 100.000 € und damit außerhalb der meisten Budgets. Hier wird man also nehmen müssen, was der Markt anbietet - ohne Chance zu Eigenentwicklung.

RFID Scanner sind in vielen Bauformen erhältlich und preislich in Regionen zwischen 300 und 1.000 € angesiedelt. Die Entwicklung eigener Lösungen erscheint möglich, da die verwendeten Komponenten aus dem Elektronikbereich verbreitet und standardisiert sind.

Die Positionierung der Tags ist ein wichtiges Kriterium für die Detektierbarkeit. Daher wird auch ein RFID System die Positionierung der Versuchstiere zum Scanner nicht ignorieren können.

2.5 Entscheidung für RFID

Die Abwägung zwischen Barcode und RFID System brachte die Entscheidung für das RFID System. In der aktuell möglichen Ausbaustufe sind beide Systeme allerdings ähnlich leistungsfähig und geeignet. Die Kostenvorteile liegen bei einer kleineren Stückzahl von benötigten Markierungen bei RFID, bei einer sehr großen Stückzahl aber bei Barcodes.

Die Positionierung der Tiere ist in beiden Systemen notwendig, wobei RFID Systeme toleranter gegen Verschmutzungen oder Lichtmangel sind.

Mittel- und langfristig aber erscheint das RFID System das interessantere:

- die Preise für die Tags werden weiter sinken und die Kosteneffizienz damit steigen
- die Miniaturisierung von Sensoren wird einen bis jetzt noch kaum vorhersehbaren Ausbau der Identifikations-Tags zulassen, bei dem der heute schon vorhandene Speicher effizient genutzt werden kann
- die Flexibilität der Technik lässt ein Eigendesign der Scanner zu, die damit auch Aufgaben wie Richtungsdetektion erledigen könnten

Aus diesen Gründen soll im Folgenden die genaue Bauweise und eingesetzte Technik eines bei sozialen funktionsfähigen Insekten beschreiben werden, sowie der Test dieses Systems.

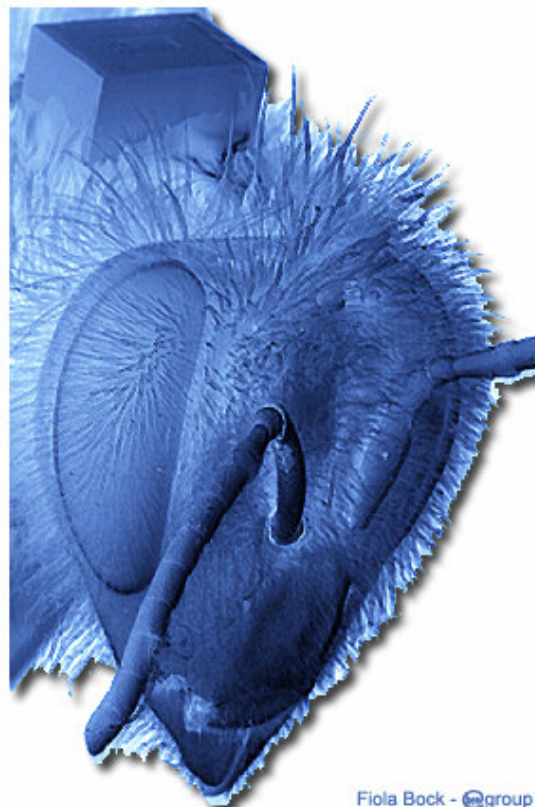


Abbildung 2: REM Bild einer Biene mit einem RFID Tag.

3 Technik: Praxis

Kapitel 3 und 4 dokumentieren die letztlich erreichten Lösungen für eine Reihe von Problemen, die in den Jahren 2002 und 2003 erkannt und gelöst wurden. Die Notwendigkeit eines speziellen Tunnels war in der Anfangsphase des Projektes nicht erkannt; ergab sich aber aus dem konsequenten Leitsatz der Minimalisierung von artifiziellen Einflüssen auf das Experiment - das eben auch mit einer „Miniaturisierung“ der Reichweite der Scanner einhergeht.

3.1 Hardware

Die Entwicklung und Evaluierung der Hardware wurde in einer Reihe kleinerer Tests durchgeführt. Es wurde in vielen Bereichen ein Bauen/Testen/Bauen Rhythmus notwendig, der nur eine langsame Annäherung an einen Idealwert erlaubte.

Den wenigsten Einfluss habe ich in dieser initialen Phase der Entwicklung auf die RFID Tags gehabt. Hier kommen momentan nur fertige Produkte des Marktes zum Einsatz, die ganz klar nicht auf die Identifizierung von Insekten und die spätere Kopplung mit Sensoren optimiert sind.

Erst massive Investitionen in diesem Bereich werden hier noch eine weitere Verbesserung erwirken können; sind aber langfristig notwendig. Das hierfür voraussichtliche Kapital wird im Bereich von € 10^6 liegen und ist damit nur in einer internationalen Anstrengung zu schultern.

3.1.1 Auswahl und Evaluierung der RFID Tags

Nach dem Grundsatz „Je kleiner und leichter desto besser“ wurde nach einem geeigneten RFID Tag gesucht. Neben der zu minimalisierenden Größe musste auf die Bestimmungen der Telekommunikationsgesetze genauso geachtet werden wie auf eine Unempfindlichkeit der Tags gegenüber den beim Aufkleben verwendeten Wirkstoffen.

Der Einsatz von RFID Tags im Bereich sozialer Insekten wurde schon 1999 von Ron Gilbert an dem Pacific Northwest National Laboratory versucht (GILBERT 2000). Mit einem Gewicht von 27 mg und einer Länge von etwa einem Zentimeter waren die dort verwendeten Tags aber nie geeignet, ein ungestörtes Verhalten der Insekten zu gewährleisten - dementsprechend wurde dieses Projekt auch nie für den breiten Einsatz im Bereich der Forschung an sozialen Insekten ausgebaut.

Die Größe und das Gewicht der verwendeten Chips ist dabei aber nicht das limitierende Element, da dieser schon heute in Größen von $0,4 * 0,4$ mm verfügbar ist („ μ -chip“ von Hitachi¹¹). Das Problem entwickelt sich mit der Anbindung des Chips an eine geeignete Antenne, die aufgrund der dort anliegenden physikalischen Effekte eben durch ihre Größe auch ihre Leistung steigert. Ein Kompromiss aus diesem Di-

¹¹ <http://www.hitachi.co.jp/Prod/mu-chip/>

lemma ist die Produktion von Antennenspulen auf der Chipoberfläche, den sog. „coil-on-chip“ (JURISCH 1994).

Der letztlich verwendete Tag ist der - mit Antenne - aktuell kleinste auf dem deutschen Markt. Es handelt sich um den „mic3@-64-TAG“ der Firma Microsensys¹².

Die technischen Daten sind dabei nach Herstellerangaben:

- Closed Coupling Read Only Transponder 64bit in mic3 Technologie
- Trägerfrequenz: 13.56 MHz
- Speicher: 64 bit laser programmed
- Größe: 1.0 * 1.6 * 0.5 mm³
- Mögliche Scanner: PEN-USB, PEN-PDA, mic3-CARDreader
- Mögliche Antennen: M12, K3

Aufgeklebt wird der Chip auf dem Thorax einer Biene mittels Schellack - eine Prozedur ganz ähnlich dem Aufbringen von Opalithplättchen. Die Chips sind durch eine Lackversiegelung gegen Wasser geschützt und recht unempfindlich gegen mechanische Einwirkungen. Einzig spitze metallene Gegenstände wie Pinzetten können die auf die Oberseite der Chips aufgebraute Antenne unter Umständen beschädigen.

Im folgenden Bild das Aufkleben des Chips auf einer Honigbiene:



Abbildung 3: Aufkleben eines RFID Chips mittels Schellack

¹² <http://www.micosensys.de/all-transponder.htm>

3.1.2 Bau und Evaluierung der Scanner

Der Bau der Scanner wurde notwendig, weil mit auf dem Markt verfügbaren Scannern zwar ein reines Auslesen der Chips möglich ist - die besonderen Ansprüche des Setups, wie Richtungserkennung und Vorverarbeitung der gescannten Daten, aber nicht realisiert sind. Handelsübliche Scanner werden daher nur in Verbindung mit „Data2DB“ (siehe Kapitel 4.2.4) eingesetzt, da hier mögliche kleinste Verbesserungen den immensen finanziellen Aufwand einer Eigenentwicklung kaum rechtfertigen würden.

Die im Zusammenhang mit der Software zum Vorverarbeiten und Speichern und Auslesen der Daten, „Data2SQL“ (siehe Kapitel 4.2.5) genutzten Scanner aber fordern eine Reihe von speziellen Anpassungen, die eine Eigenentwicklung unumgänglich machen.

Die Richtungserkennung ist dabei essentiell: Bei 100% Erkennung aller Tiere wäre sie eigentlich nicht notwendig, da man ja die Anfangsposition kennt und danach nur noch abzählen müsste. Tatsächlich aber werden das Verhalten der Tiere und unser Bestreben, mit der kleinstmöglichen Tag Größe zu operieren immer einen gewissen Fehler bedingen. Ein einziger Fehler beim Scan im gesamten Leben einer Biene aber würde jede weitere Messung im Bezug auf den augenblicklichen Aufenthaltsort des Tieres täuschen - daher muss ein Mechanismus geschaffen werden, der immer wieder ein definiertes Rücksetzen des aktuellen Aufenthaltsorts erlaubt.

Zu diesem Zweck wurde ein Scanner konstruiert, der aus zwei Teilscannern besteht, die sich gemeinsam Datenlogik und Speicherbereich teilen. So kann aus der Abfolge, in der eine Biene die beiden Teilscanner durchquert, die Richtung erschlossen werden. Die erste Generation der Scanner bietet nur die zum Auslesen der ID Nummern und deren Speicherung notwendige Logik. Sie muss daher mehrmals pro Sekunde abgefragt werden, um so die genauen Ereigniszeiten abspeichern zu können. Die Abfrage der Scanner erfolgt über die serielle Schnittstelle eine angeschlossenen PCs und liefert eine Reihe von Tag IDs, die dann mit genauem Zeitpunkt des Scans in die Datenbank geschrieben werden können.

Die interne Logik der Scanner reduziert den Mehrfachscan eines einzigen Tags. Dies tritt auf, wenn Insekten sich zu langsam unter dem Scanner hindurch bewegen oder dort ganz stehen bleiben. Jeder Teilscanner vergleicht intern die ID Nummer mit der letzten von ihm selbst ausgelesenen und verweigert eine Speicherung, wenn diese beiden IDs gleich sind.

Der Einsatz ab dem Sommer 2003 in einem Versuch mit 1000 Bienen zeigte dabei die Verbesserungsmöglichkeiten, die in den Entwurf der zweiten Generation der Scanner mündete.

Diese Generation unterscheidet sich von der ersten durch

- Verbesserung der internen Verarbeitungslogik
- Ausbau des Speichers auf 16 MBit
- Wegfall eines angeschlossenen PCs

Das Design dieser Scanner lief Anfang des Jahres 2004 an und wird ab Sommer 2004 einsatzbereit sein.

Das folgende Diagramm verdeutlicht schematisch die interne Logik der Scanner, die aktuell zum Einsatz kommt. Es geht von Teilscannern A und B aus, die sich einen gemeinsamen Speicher zur Ablage der Daten teilen:

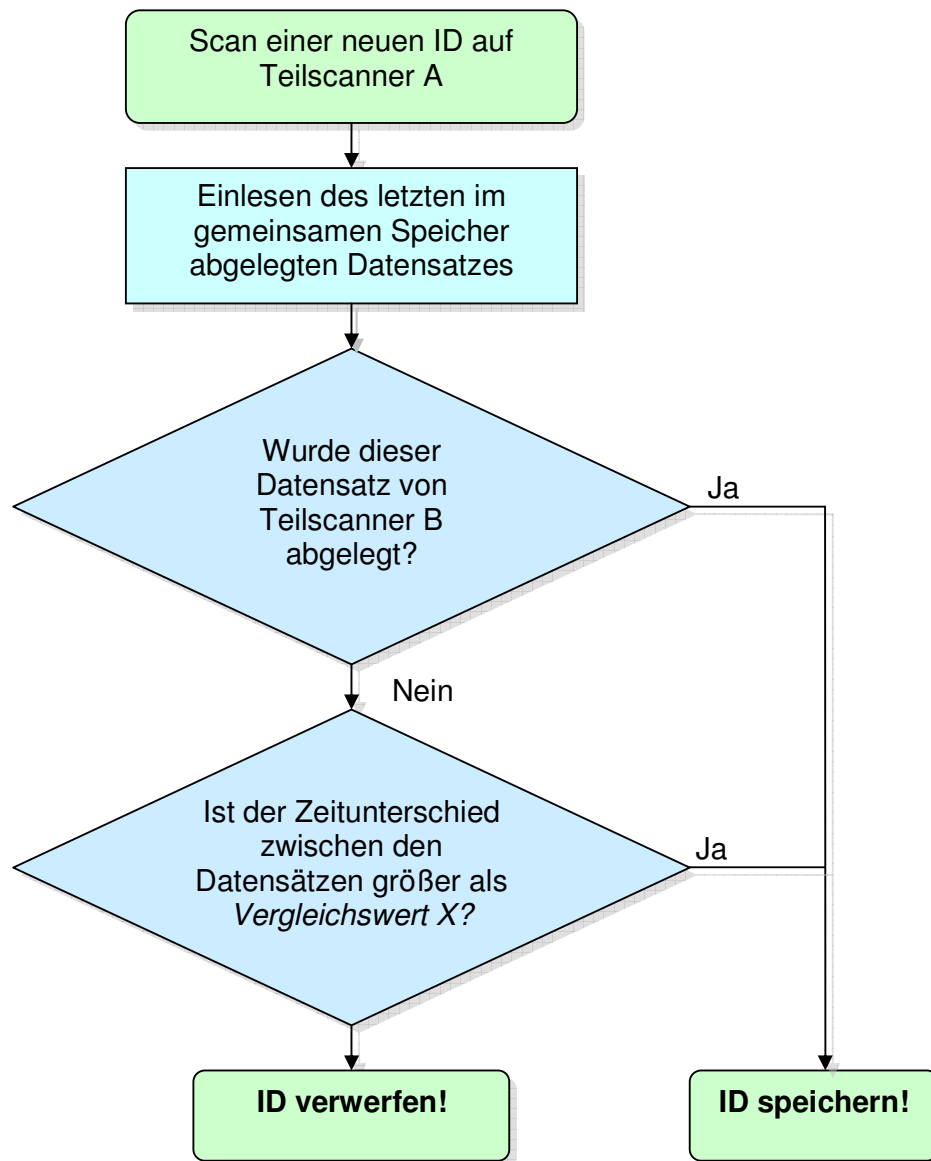


Abbildung 4: Diagramm der Scanlogik

Der *Vergleichswert X* ist der Wert, nach dem erst der neue Scan gespeichert werden soll, wenn er vom selben Teilreader kommt. Dies soll Mehrfachscans reduzieren, die oft entstehen, wenn die Tiere unter dem Scanner stehen bleiben. Auf der anderen Seite darf *X* aber nicht so groß gewählt werden, dass ein schnelles Durchlaufen, Umdrehen und Zurücklaufen nicht mehr als solches erkannt wird. *X* ist damit vom

Experiment und zu untersuchenden Tier abhängig und muss erst empirisch ermittelt werden (Werte zwischen einer und vier Sekunden sind hier realistisch). Damit muss aber schon klar sein, dass der Fehler, den diese Methode minimal haben wird, niemals Null sein kann: Ein Versuchstier, das bis exakt unter den Scanner läuft um dann rückwärts wieder zurückzukrabbeln, kann nicht in jedem Falls als ein solches erkannt werden. Solche Fehler müssen entsprechend in der nachfolgenden Bearbeitung und Sichtung der Daten erkannt und behoben werden. Diese und andere Schritte leistet die Verarbeitungssoftware Data2Web (Kapitel 4.2.6).

Warum nun wird überhaupt eine Vorverarbeitung der gescannten Daten vorgenommen und nicht einfach alle Daten zur späteren Sichtung weitergegeben? Zunächst einmal ist der Scanner in der Lage, etwa 20-mal pro Sekunde einen Tag einzulesen - das heißt, die große Mehrzahl der Scans besteht in jedem Fall aus Doppelscans, da ein Insekt in der Regel mehr als eine zwanzigstel Sekunde benötigen wird, um einen der beiden Teils Scanner zu unterqueren. Werden solche Daten aber nicht im Vorfeld verworfen, dann läuft der interne Speicher der Scanner zu schnell voll, um von großem Nutzen zu sein. Die Trennung von Scanner vom PC war immer erklärtes Ziel der Entwicklung und konnte nur gelingen, wenn nicht jeder Scan gespeichert wird. Die Scanner der zweiten Generation werden etwa 100.000 Datensätze speichern können: Genug für etwa einen Monat autarker Datenerfassung bei Vorverarbeitung der Scans - oder für 90 Minuten ohne diese.

Das folgende Bild zeigt die Position zweier Scanner an dem speziellen Tunnel - das Design, die Konzeption und der Bau des Tunnels werden im folgenden Kapitel 3.1.3 beschrieben werden:

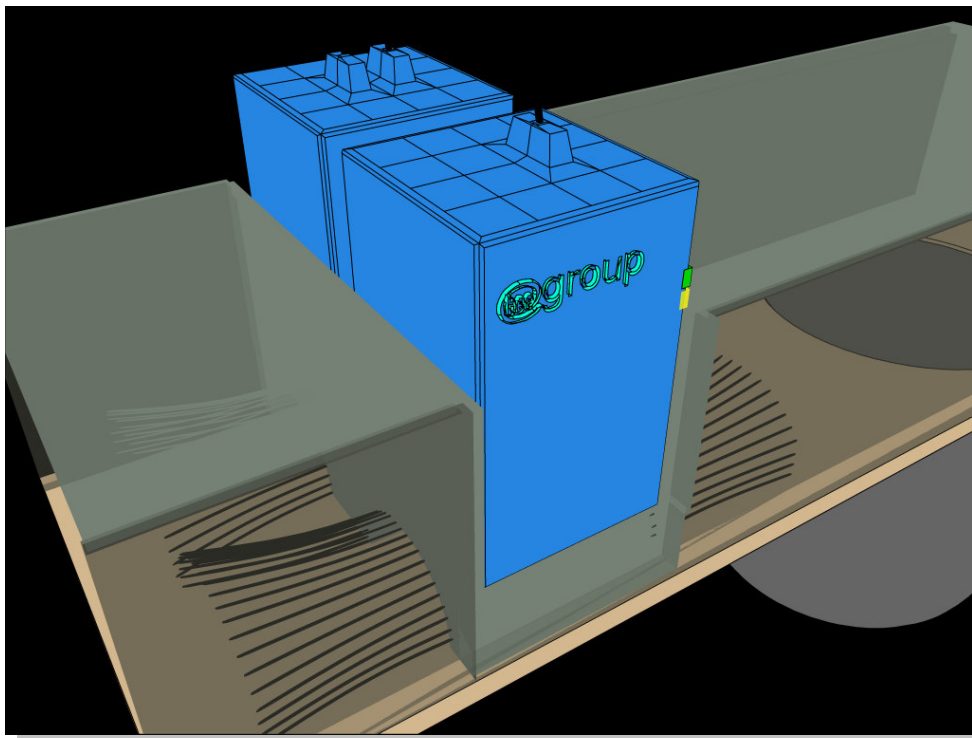


Abbildung 5: Spezielle RFID Scanner integriert in den benötigten Tunnel

3.1.3 Bau und Evaluierung des Tunnels

Die Notwendigkeit eines speziellen Tunnels entsteht aus der Forderung, den biologisch relevanten Einfluss der RFID Tags zu minimieren. Würde man an dieser Stelle eine ca. 8 mm lange Antenne tolerieren, dann würde die Reichweite, bei der dieser RFID Tag zuverlässig erkannt wird, auf etwa 10 cm steigen und damit die exakte Ausrichtung der Bienen überflüssig machen.

Als primär für Verhaltensbeobachtungen ausgelegtes System wurde aber die Minimierung der Tags als *conditio sine qua non* erkannt. Deren Reichweite liegt bei etwa 2 mm und erlaubt eine Winkelabweichung von maximal etwa 25° in beide Richtungen zur Senkrechten des Tags:

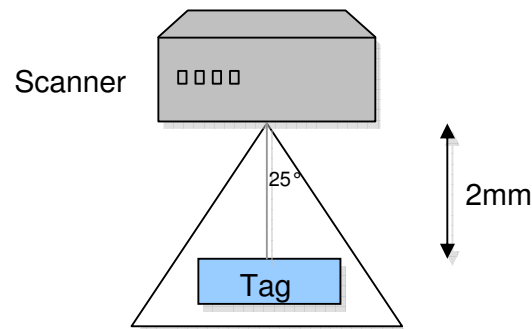


Abbildung 6: Reichweite und Position des Erkennungsfeld der Scanner

Damit wird die Positionierung der Tags zum Scanner aber eine die wichtigste Größe für die Effizienz des Systems: Nur wenn es gelingt, einen Großteil der Insekten zu einem Unterqueren der Scanner mit dem Tag möglichst senkrecht zum Scanner zu bewegen, wird man eine ausreichende Erkennungsleistung erreichen.

Nun befähigen die Haftorgane der Insekten diese auf beinahe beliebig glatten Flächen zu laufen: Die meisten leichteren Insekten wie Honigbienen oder Ameisen werden ohne echte Präferenz für den Boden eines Ganges sowohl die Wände als auch die Gangdecke nutzen.

Versuche, mittels extra glatten Materialien dieses zu unterbinden, scheiterten an dem erstaunlichen Haftvermögen der Tiere: Spätestens nach ein paar Stunden wurden die ehemals glatten Flächen mit einer dünnen Wachsschicht überzogen und als Lauffläche genutzt.

Der intensive Test verschiedener Geometrien und Materialien führte schließlich zur aktuell verwendeten Lösung, die bei etwa 299 von 300 Tieren zu einer ausreichend genauen Positionierung führt.

Zunächst musste der Eintritt in den Gang so manipuliert werden, dass er nur noch von der Bodenfläche her stattfindet. Dazu werden in einem Halbkreis um den Gang Borsten mit einer Länge von ca. 2 cm angebracht, die einen stacheligen Kranz bilden. Die Bienen können den etwas unregelmäßigen Saum dieses Kranzes nicht umlaufen um den Tunnel mit einer nicht gewünschten Ausrichtung zu betreten: Der gesamte Eintritt in den Tunnel erfolgt vom Boden des Ganges.

Mit der Vorliebe der hier getesteten Honigbienen, senkrecht durch enge Spalten zu laufen, ergab sich das nächste Problem. Anstatt mit dem Tag nach oben den Spalt unter dem Scanner zu durchqueren, drehten sich etwa 50% der beobachteten Tiere um 90 Grad und entgingen damit einer erfolgreichen Detektion. Hier war die Lösung geometrischer Natur: Bienen sind mit ihren Beinen - von vorne betrachtet - im Querschnitt breiter als hoch. Baut man den Gang nun entsprechend nach und zwingt die Tiere, in ihm leicht in die „Knie“ zu gehen, also den Körper beinahe auf den Boden zu drücken, dann wird ein Abkippen des Körpers um 90° aus Platzgründen vermieden:

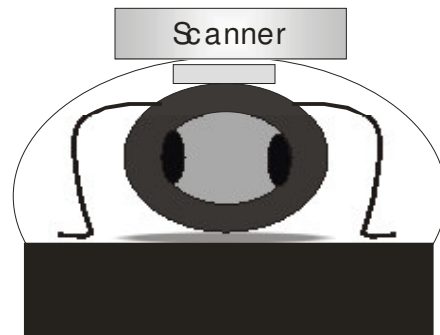


Abbildung 7: Schematische Positionierung eines Insekts im Tunnel

Dieses Design behindert natürlich bei sehr hoher Aktivität den Fluss der Tiere durch den Tunnel hindurch. Versuche mit verschiedenen Stock- und Tunnelgrößen zeigten eine Notwendigkeit von etwa einem mit Scannern bestückten Durchgang pro 1500 Bienen einer Kolonie, um einen insgesamt ungehinderten Verkehrsfluss zu gewährleisten.

Werden diese Werte stark überschritten, dann sinkt die Attraktivität des Stockes, was unweigerlich bei Bienen ein Verlassen der Nisthöhle nach sich zieht.

Selbst wenn diese Werte beachtet werden, ergeben sich weitere Probleme im Falle von Größenpolymorphismus der zu beobachtenden Insektenart: Die Tunnelhöhe als einziger Ausgang des Stockes kann je nach Setup beschränkend für andere Größen der Kolonie wirken: Bei Bienen sind Drohnen zum Beispiel nicht in der Lage, diese Tunnel zu durchqueren. Auch tote Bienen können nur sehr schlecht durch solche Tunnel transportiert werden - benötigt wurde also eine Vorrichtung, die tote Bienen sammelt - seien es natürlich im Stock gestorbene oder aber Drohnen, die im Stock nach einiger Zeit getötet wurden.

Hierzu wurde im Anfangsbereich des Tunnelbodens ein Glas mit einem Durchmesser beinahe der gesamten Tunnelfläche angebracht. Arbeiterinnen, die mit dem Abtransport aus dem Stock beschäftigt sind, fallen dort hinein und lassen ihre Last schnell dort liegen. So wird auf der einen Seite eine Möglichkeit zur Entsorgung von toten Bienen aus dem Stock geschaffen, die aus hygienischer Sicht notwendig ist und auf der anderen Seite eine weitere Möglichkeit für den Experimentator, Vorgänge des Stockes zu überwachen.

Die folgenden Bilder zeigen verschiedene Konzeptionen solcher Tunnel - die genauen Größen der Gänge müssen im Vorfeld jedes Setups an die Größe der Versuchstiere angepasst werden.

Zunächst die Gesamtansicht, die im linken Bereich den gläsernen Auffangbehälter und im rechten zwei Tunnel mit jeweils einem (hier blauen) Scanner zeigt. Die Borsten um die Eingänge ermutigen die Versuchstiere, nur vom Boden des Tunnels her den Gang unter dem Scanner zu betreten.

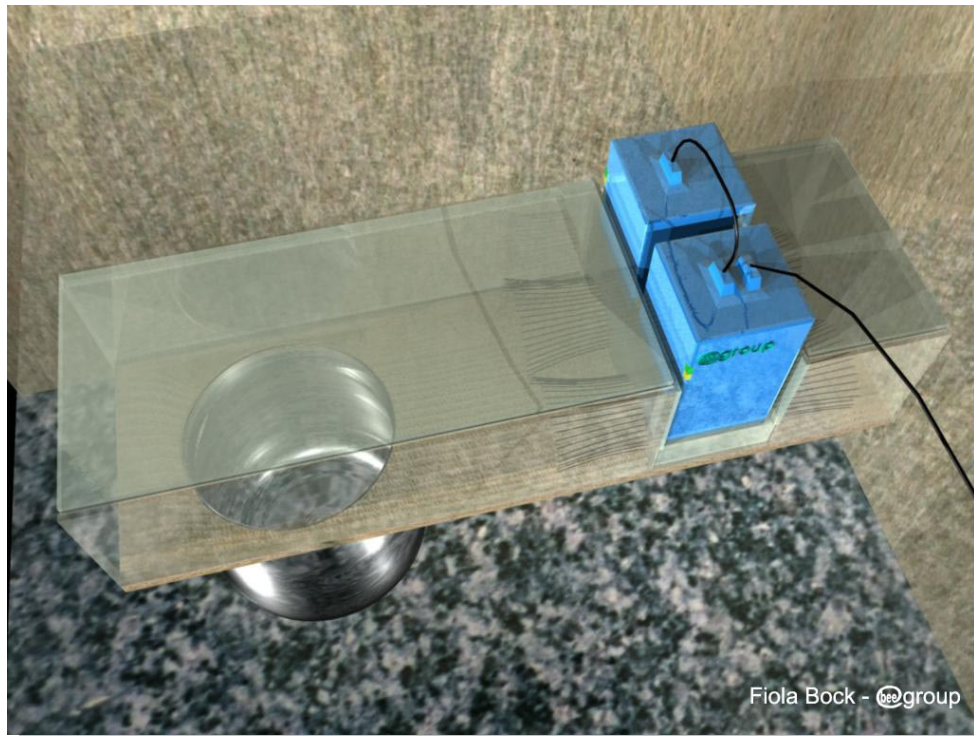


Abbildung 8: Gesamtansicht eines Tunnels

Zu beachten ist die Minimierung der hier rechts dargestellten Außenseite des Tunnels: Wird dieser Abschnitt zu lang gewählt, dann kann er schnell für die Stocktiere einen „Stockinneren“ Charakter bekommen und ist damit zunehmend irrelevant als Mechanismus zum Zählen echter Ausflüge. Auch hier kann noch immer ein Tier den Tunnel zwar durchqueren, danach aber sofort umdrehen und zurück in den Stock gelangen. Solche Verhaltensweisen werden durchaus beobachtet und müssen durch geeignete Methoden in Rechnung genommen werden: Sei es durch prozentuale Abschätzung und Addition zur Fehlerrate oder durch eine Begrenzung der Zeit, die erst als echter Ausflug gewertet wird.

In der folgenden Abbildung dargestellt ist eine Einsicht in den Tunnel von der Außenseite durch die beiden Gänge hindurch auf die spätere Innenseite des Tunnels:

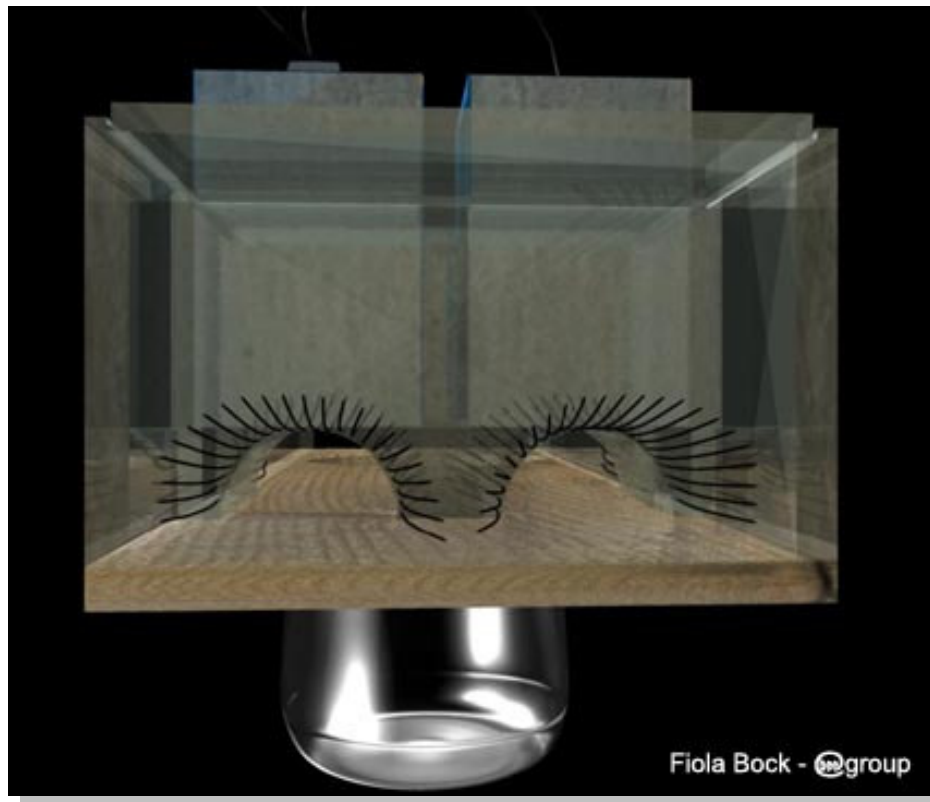


Abbildung 9: Durchsicht durch den Tunnel (2 Gänge, von außen nach innen)

4 Software

Eine wesentliche Bedeutung fällt der verwendeten Software zu. RFID generierte Experimente können Datenmengen produzieren, die nur noch mit spezialisierter Software zu bewältigen sind.

Nehmen wir eine Gruppe von nur 10 Individuen, deren Futtersuche in einem Zeitraum von 2 Monaten untersucht werden soll. Wenn jeweils nur 20 Bewegungen pro Tag erfasst werden, dann können die nach 2 Monaten entstehenden $10 * 20 * 60$ (Tage) = 12.000 Datenbankeinträge schon nur noch schwerlich mit konventionellen Hilfsmittel bewältigt werden. Schon eine Erweiterung auf 50 Individuen würde die Grenzen einer MS Excel ($\sim 2^{16} = 65.000$ Zeilen) Tabelle sprengen – von Hand auf Plausibilität und innere Zusammenhänge zu sichten sind diese Daten in keinem Fall.

Die Softwareschnittstelle macht ein RFID System für den Einsatz bei sozialen Insekten erst sinnvoll – ohne einfachen Zugang zu den gewonnenen Daten ist eine Nutzung nicht möglich.

4.1 Prinzipielle Forderungen

Die Forderungen an die Software ergeben sich aus dem Einsatzgebiet des Gesamtsystems. Gegliedert werden kann dabei der Einsatz des Systems in drei Schritte, deren jeweilige Software unterschiedlichen Ansprüchen gerecht werden muss:

- Erfassung der Individuen und Gruppen
- Datenerfassung während des Experimentes
- Datenauswertung

Die Erfassung der Individuen und Gruppen muss mobil und einfach möglich sein. Verschiedene Betriebssysteme oder Sprachversionen dürfen nicht im Wege stehen, da ja diese Tätigkeit vor Ort - das heißt auch ohne Probleme in Gebieten beschränkter Infrastruktur - durchgeführt werden muss.

Die laufende Datenerfassung muss vor allem schnell und stabil sein. Ausfälle dieser Software ziehen Ausfälle in der Erfassung mit sich, die das laufende Experiment schnell ruinieren könnten. Der hierbei benötigte Rechneinsatz sollte minimiert werden, da die Datenerhebung ja vor allem auch an den Nist- und Brutstätten der zu untersuchenden Spezies möglich sein soll.

Die Datenauswertung muss auf der einen Seite einfach genug sein, um ohne aufwendige Kenntnisse von Datenbanksprachen Tendenzen und Ergebnisse zu erzielen – muss aber auf der anderen Seite alle Möglichkeiten bieten, spezialisierte Werkzeuge zur Datenanalyse einsetzen zu können.

4.1.1 Portabilität

Unter Portabilität von Software werden die Möglichkeit zum Einsatz auf verschiedenen Betriebssystemen und Betriebssystemversionen sowie die Anbindung an einen möglichst verbreiteten Standard zur Abfrage der Ergebnisse verstanden. Als Betriebssysteme für Desktop Computer und Notebooks kommen heute überwiegend drei Betriebssysteme zum Einsatz:

- Microsoft Windows¹³
- Apple OS X¹⁴
- Linux¹⁵

Die entwickelte Software auf allen drei Systemen lauffähig zu gestalten wäre ideal, da so der Einsatz auf beinahe allen Rechnern denkbar wäre. Diese eigentlich einfache Forderung gestaltet sich überraschend schwierig, da die verwendeten Grundlagen von Microsoft Windows auf der einen und Apple OS und Linux auf der anderen Seite sich so stark unterscheiden, dass ein Einsatz derselben Software zugleich auf beiden Systemen zunächst unmöglich ist.

Die Portabilität der Ergebnisse definiert sich in der implementierten Schnittstelle zu der verwendeten Datenbank. Hier hat sich durch den weit verbreiteten Einsatz von Datenbankstrukturen z.B. in der Geschäftswelt eine Reihe von konkurrierenden Standards entwickelt, die einer weiteren Verwendung der Daten im Wege stehen könnten. Statistikpakete wie SPSS¹⁶ oder Statistica¹⁷ bieten Schnittstellen zu standardisierten Datenbanken und können auch riesige Datenmengen bewältigen. Mit einer genauen Kenntnis der verwendeten Schnittstellensprachen können so große Datenmengen statistisch ausgewertet werden oder noch nicht erkannte Zusammenhänge über DataMining Funktionen erkannt werden.

4.1.2 Simplizität

Die Einfachheit der Bedienung ist immer elementar: Zunächst einmal werden Fehler bei der Erfassung der Parameter und Eingabe der Grunddaten vermieden, die das weitere Experiment negativ beeinflussen würden. Zum anderen erlaubt erst die Abstraktion der Datenaufbereitung von dem darunter liegenden Datensystem und dessen Sprachen den breiten Einsatz solcher Systeme.

Die Simplizität der Benutzerschnittstellen darf aber in keinem Fall die darunter liegenden Schichten der Datenbanken und RFID Scannersteuerung negativ beeinflussen - hier muss prinzipiell jede Möglichkeit vorhanden sein, mit beliebigen Auswertungsansätzen auf die entstehenden Daten zuzugreifen.

¹³ <http://www.microsoft.com/windows/>

¹⁴ <http://www.apple.com/macosex/>

¹⁵ Beispielsweise: <http://www.kernel.org/>

¹⁶ <http://www.spss.com/>

¹⁷ <http://www.statsoft.com/>

4.1.3 Skalierbarkeit

Unter Skalierbarkeit wird die Möglichkeit zur Nutzung bestehender Strukturen in mengenmäßig größeren Zusammenhängen. Am Beispiel des realisierten RFID Systems bedeutet dies die Spannweite der Möglichkeiten von der Überwachung von 10 Individuen einer Gruppe bis hin zur parallelen Überwachung beliebig vieler Kolonien staatenbildender Insekten und damit einer Erfassung von Millionen von Bewegungen pro Tag. Die Grenze des möglichen Ausbaues darf hier im Vorfeld nicht zu niedrig angelegt werden: Noch hindert die recht aufwendige Markierungsarbeit einen millionenfachen Einsatz – sobald diese aber auch automatisiert geschehen kann, stünde der Überwachung ganzer Populationen nichts mehr im Wege.

4.2 Realisierung

Im Folgenden soll ein Überblick über einen der gangbaren Wege gegeben werden, die zu den insgesamt 3 Softwaremodulen führte. Dabei sind die verwendeten Schnittstellen und Umgebungen durchaus variierbar - tatsächlich wurde auf eine möglichst gute Portierbarkeit geachtet, um Weiterentwicklungen auch mit anderen Systemen zu ermutigen.

4.2.1 Schichtenmodell

Die aufgeführten Forderungen bildeten das Rückgrat für die Realisierung des Systems.

Die folgende Graphik zeigt das Zusammenspiel der Schichten, in denen die Funktionen des RFID Systems angelegt sind:

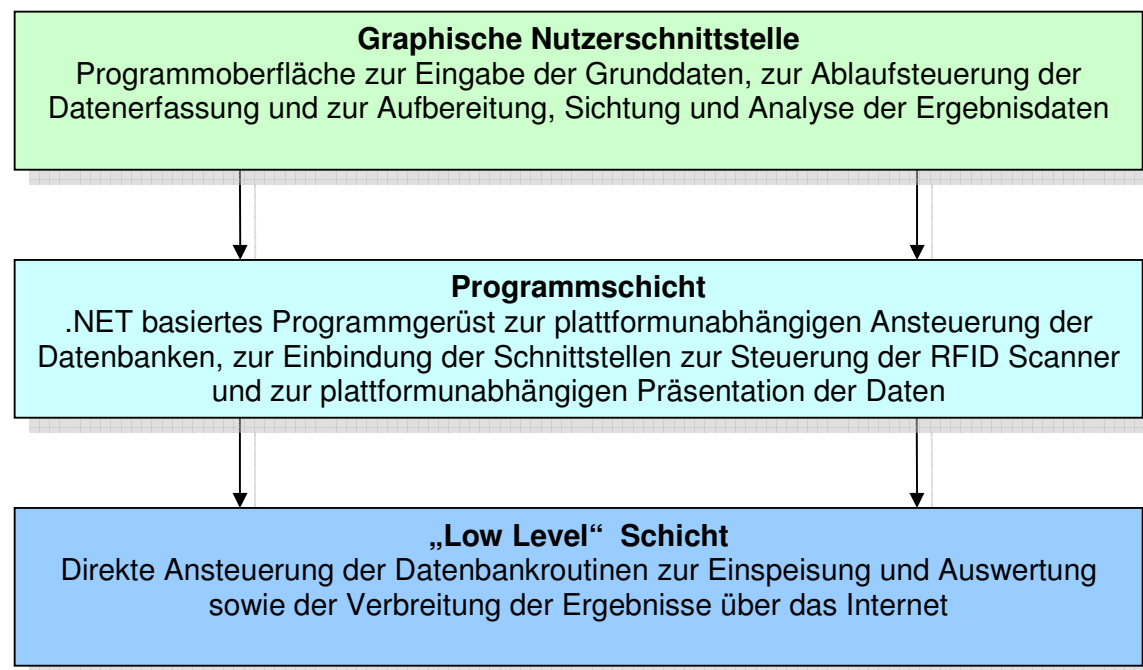


Abbildung 10: Abstrahiertes Schichtenmodell des verwendeten Software Designs

Durch die Abstraktion der Schichten untereinander kann eine weitgehende Plattformunabhängigkeit erreicht werden. Je weiter aber direkt auf Funktionen des Betriebssystems wie z.B. dem Erstellen und Modifizieren von Daten (also auch Datenbanken) zugegriffen werden muss, desto mehr sinkt auch die Plattformunabhängigkeit der verwendeten Strukturen: Daher also wird diese von der obersten zur untersten Schicht sinken.

4.2.2 Konkrete Auswahl der Komponenten

Graphische Nutzerschnittstelle:

Wo immer es möglich war, wird die Nutzerschnittstelle über die etablierten Mechanismen des „World Wide Web“ und damit über das sog. HTTP Protokoll¹⁸ abgewickelt. Dem Nachteil des Geschwindigkeitsverlustes eines dezentral organisierten Systems stehen massive Vorteile gegenüber, die ja das WWW so erfolgreich gemacht haben:

- Zentrale Informationsspeicherung – und damit die Möglichkeit des gleichzeitigen Zugriffs beinahe unbeschränkt vieler Nutzer darauf
- Vollkommene Plattformunabhängigkeit durch Nutzung der WWW Browser als Eingabeschnittstelle, die es für jedes der angesprochenen Betriebssysteme gibt
- Unbegrenzte Skalierbarkeit durch etablierte Techniken der Datenverarbeitung und –Speicherung („Cluster“), wie sie in den verschiedenen Breichen des Internets etabliert wurden

Programmschicht:

Die Programmschicht wurde im Microsoft .NET Framework¹⁹ entwickelt. Das .NET Framework ist eine Umgebung für die Entwicklung, Bereitstellung und Ausführung von sog. Services und Anwendungen und hat vor allem die Portabilität und Skalierbarkeit als Hauptmerkmal bei seiner Entwicklung verfolgt.

Der Austausch von Daten zwischen den drei verschiedenen Anwendungen erfolgt über ein „XML“ (eXtensible Markup Language²⁰) genanntes Format.

Dabei handelt es sich um eine standardisierte Datenbeschreibungssprache - also eine Hersteller und Plattform unabhängige Methode für Softwareprogramme, Daten für andere Softwareprogramme in einer verständlichen Weise zu beschreiben. Die Speicherung aller relevanten Daten in diesem Format ermöglicht den Datenaustausch mit Auswertungssoftware wie Statistica²¹ oder MS Excel²², die dieses Format in ihren aktuellen Versionen ohne Probleme importieren können.

Die Programmierung selber wurde in C#²³ (sprich: „C-Sharp“) durchgeführt. Diese ebenfalls in den letzten Jahren neu entwickelte Sprache ist ECMA²⁴ standardisiert

¹⁸ <ftp://ftp.isi.edu/in-notes/rfc2616.txt>

¹⁹ <http://msdn.microsoft.com/netframework/>

²⁰ <http://www.w3.org/XML/>

²¹ <http://www.statsoft.com/>

²² <http://www.microsoft.com/office/excel/default.asp>

²³ <http://msdn.microsoft.com/vcsharp/>

²⁴ <http://www.ecma-international.org/>

und für alle angesprochenen Betriebssysteme verfügbar. Durch den Einsatz des .NET Frameworks entfällt aber eine Begrenzung auf eine Sprache: Die Weiterentwicklung der Software ist innerhalb des .NET Frameworks auch in so unterschiedlichen Sprachen wie Visual Basic²⁵ oder Borland Delphi²⁶ möglich.

Durch die Verfügbarkeit des .NET Frameworks unter Microsoft Windows™ sowie unter Linux²⁷ (Mono²⁸, DotGNU²⁹), konnte hier eine weitgehende Plattformunabhängigkeit erreicht werden – eine Neukompilierung des Programmcodes reicht in der Regel, um die Softwaremodule an andere Betriebssysteme anzupassen.

„Low Level“ Schicht:

In diese Schicht fallen Schnittstellen für Gerätetreiber wie die der RFID Scanner, sowie Schnittstellen für die Manipulation der Datenbanken. Die Gerätetreiber der Scanner sind zum aktuellen Zeitpunkt nur als Windows DLL (Dynamic Link Library) verfügbar – sollen aber im Laufe des Jahres 2004 auf den sog. „managed code“ – also das .NET Framework umgestellt werden. Daher kann zum aktuellen Zeitpunkt die tatsächliche Ansteuerung der Scanner nur unter MS Windows erfolgen. Als Gerätetreiber zu der verwendeten Datenbank wurde der so genannte ODBC Treiber (**O**pen **D**ata**B**ase **C**onnectivity³⁰) im Level 0-3.5.1 verwendet.

Dabei handelt es sich um einen Standard für Datenbankzugriffe, der ursprünglich von Microsoft stammt - mittlerweile aber für die meisten Betriebssysteme vorliegt und so beinahe beliebige Datenbankschnittstellen miteinander verbinden kann. Mit Verwendung dieses Standards anstatt dem direkten Zugriff auf die Datenstrukturen (der aber nie für alle möglichen Betriebssysteme zu realisieren ist), geht auch hier wieder ein Geschwindigkeitsverlust von ca. 20% einher – der aber aus den genannten Gründen in Kauf genommen wird.

4.2.3 Verwendete Datenbank

Bei der Auswahl der verwendeten Datenstrukturen spielt eine zentrale Rolle zunächst die verwendete Abfragesprache. Hier hat sich SQL, die „**S**tructured **Q**uery **L**anguage“³¹ als weltweiter ANSI (American National Standards Institute³²) Standard etabliert. SQL ist eine standardisierte Sprache zur Kommunikation mit Datenbank Management Systemen (DBMS). Sie ermöglicht den Zugriff auf und die Manipulation von relational strukturierten Daten, also miteinander verknüpfte Tabellen. Obwohl eigentlich als Standard definiert, liegt SQL in mehreren Dialekten und Entwicklungsstufen vor, die sich Datenbank-abhängig leicht unterscheiden.

Die verwendete Datenbankabfragesprache definiert im Wesentlichen die Abfragefähigkeiten des Programms: SQL erlaubt ineinander verschränkte Abfragen mit beinahe beliebig vielen Parametern, in die verschiedene Tabellen mit einbezogen werden können.

²⁵ <http://msdn.microsoft.com/vbasic/>

²⁶ <http://www.borland.com/delphi/>

²⁷ Beispielsweise: <http://www.kernel.org/>

²⁸ <http://www.go-mono.com/>

²⁹ <http://www.dotgnu.org/>

³⁰ <http://www.iodbc.org/>

³¹ ISO/IEC 9075-1:2003 bis -13:2003 unter <http://www.iso.ch/>

³² <http://www.ansi.org/>

Die Struktur von Relationalen Datenbanken sieht die Gliederung der Daten in mehrere Tabellen vor. In unserer Anwendung enthält

- eine Tabelle dabei die vom Scanner gelieferten Daten
- eine andere die Definition der Gruppen
- eine dritte die Zuordnung der einzelnen, eindeutigen RFID Tags zu den Gruppeneinträgen

Genutzt wird diese Struktur, um Mehrfacheinträge zu vermeiden und damit drohende Inkonsistenzen zu minimieren: Anstatt in jede Zeile eines Scans auch die genaue Gruppe des jeweiligen Tags mit einzutragen, reicht ein Verweis in der hier genannten dritte Tabelle, um diesen Eintrag eindeutig einer Gruppe zuordnen zu können. Diese Struktur wird in der Abfrage genutzt, um komplexe Beziehungen zwischen Tabellen herzustellen. So können Abfragen gestaltet werden, die ganz unterschiedliche Parameter in einem Ergebnis bündeln.

Eine Abfrage, die alle Scans

- am 01. September 2003 und 01. Oktober 2003
- zwischen 12:00 und 16:00
- ausschließlich aus dem Volk 101

sucht, lautet in SQL (Dialekt MySQL v4.1) und bei den aktuell verwendeten Tabellennamen:

```
SELECT * FROM tags.tag_processed INNER JOIN tags.tag_group_rel on
tags.tag_processed.t_ID = tags.tag_group_rel.rel_TagID INNER JOIN tags
.gruppen on tags.tag_group_rel.rel_groupID = tags.gruppen.g_ID WHERE
((t_Time >= '2003-9-1 00:00:00.0' AND t_Time < '2003-9-2 00:00:00.0') OR
(t_Time >= '2003-10-1 00:00:00.0' AND t_Time < '2003-10-2 00:00:00.0') )
AND ((HOUR(t_Time) >= 12 AND HOUR(t_Time) < 16)) AND
(rel_groupID=2) GROUP BY t_auto
```

Abbildung 11: Scan-Abfrage in SQL

Eine ebenso mächtige wie zunächst etwas verwirrende Abfragesprache. Ein wichtiger Punkt wird daher sein, dem Nutzer zu ermöglichen, die direkte Berührung mit dieser Abfragesprache zu umgehen. Soweit sich die Abfragen in eine eingängige Oberfläche umsetzen lassen, sollte dies geschehen: Anders ist eine weite Verbreitung des Systems schwer vorstellbar.

Eine weitere zentrale Rolle spielt die tatsächlich verwendete Datenbank. Datenbanken waren lange im professionellen Bereich extrem kostspielige Komponenten. Heute liegt mit MySQL eine Datenbank vor, die unseren Ansprüchen an Skalierbarkeit und Portabilität voll genügt und dabei kostenlos ist. MySQL unterstützt SQL weitgehend im Standard „SQL-92“, was für unsere Bedürfnisse voll ausreicht. Die maximale Größe der Datentabellen beträgt hier zwischen 2 und 8 Tera Byte (abhängig vom verwendeten Betriebssystem) und die optimale Anzahl der gescannten Datenreihen jeder Tabelle zwischen 10 und 100 Millionen Einträgen: Momentan

genug für die aktuell denkbaren Versuche und jederzeit erweiterbar, falls mehr notwendig sein sollten.

Die Struktur aller sog. Relationalen Datenbanken ist dabei ähnlich: Hier handelt es sich um Tabellen, die miteinander in einer Beziehung stehen - also eine definierte Relation haben.

So existiert in der hier verwendeten Datenbank eine Tabelle, die alle Scandaten direkt aus Data2SQL beinhaltet. Eine andere enthält alle Gruppen, die bei der Erfassung mit Data2DB erstellt wurden. Nun gehört zu der Tabelle der Scandaten natürlich immer ein Feld, das die ID Nummer des gescannten Tags beinhaltet. Das Wissen über die Gruppe, zu der diese ID gehört, liegt aber in der Gruppentabelle der Datenbank.

Damit wird eine weitere Tabelle notwendig, die ausschließlich die Relation der beiden Tabellen speichert: Eine Tabelle, die nur eine Liste der gesamten IDs des Versuches mit jeweils einem Verweis auf ihre Gruppe kombiniert. Diese Querbeziehung wird in

```
...INNER JOIN tags.gruppen on tags.tag_group_rel.rel_groupID = tags.  
gruppen.g_ID ...
```

aus obigem Beispiel reflektiert. Eine mögliche Abfrage, die sich nur auf die Mitglieder einer einzigen Gruppe bezieht, kann daher in obiger Abfrage

```
... AND (rel_groupID=2) ...
```

nutzen, um diese zu erhalten und alle anderen auszuschließen.

4.2.4 Datenerfassung mit *Data2DB*: Präparation

Data2DB ist die Schnittstelle zur Kontrolle des ersten Einlesens der Tags der beteiligten Versuchstiere. Im diesem ersten Scandurchgang wird die ID des aufgeklebten Tag erfasst und damit

- in der Datenbank als bekanntes auszuwertendes Tier definiert
- beliebig vielen Gruppen zugeordnet

Die Gruppenzuordnung ist dabei ein entscheidender Schritt, der im Experiment über die späteren Auswertungsmöglichkeiten entscheidet. Vor dem eigentlichen Scandurchgang werden in Data2DB beliebig viele Gruppen definiert, in die dann jede ID eingeordnet wird. So kann eine Gruppe eine genetische Linie, eine Aufzuchtbedingung, einen Gesundheitszustand oder andere Parameter beinhalten: Genau diese Parameter werden es sein, die in der späteren Auswertung bestimmen, welches Verhalten welcher Gruppen voneinander abgegrenzt und verglichen werden kann.

Nachfolgend der Aufbau des Programms:

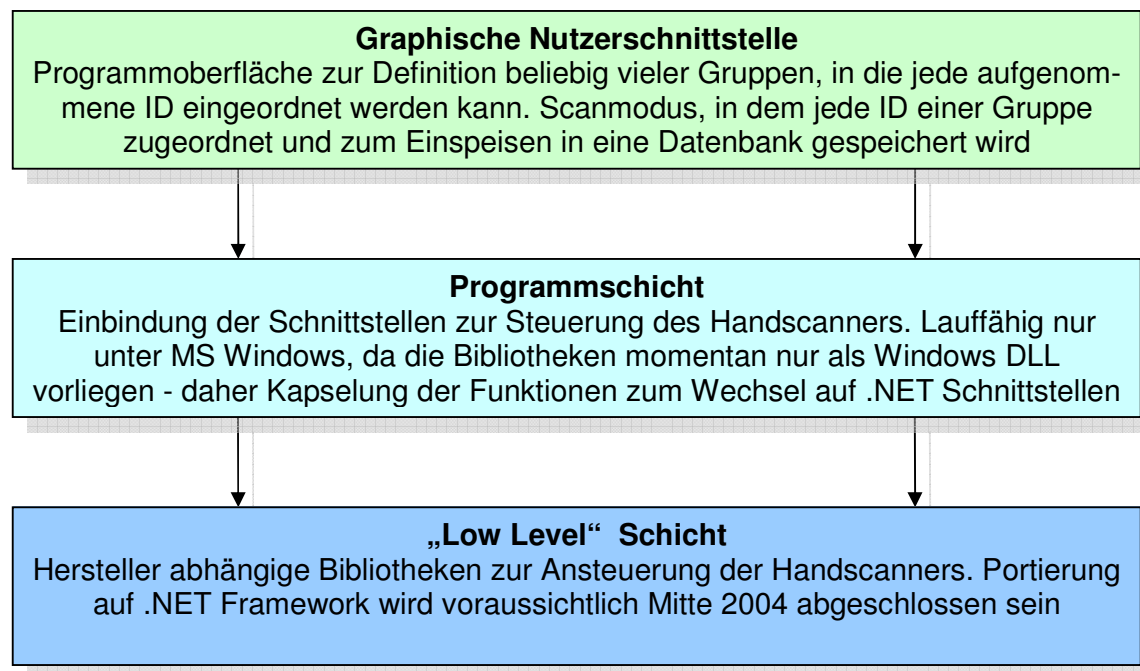


Abbildung 12: Data2DB - schematische Übersicht

Neben der Gruppenzuordnung wird im Scanmodus jede ID gescannt, angezeigt und den entsprechend markierten Gruppen zugeordnet. Diese Gruppeneinteilung - die zwischen keiner und allen Gruppen frei definierbar ist - wird zusammen mit der eindeutigen ID in einer XML Ausgabedatei festgehalten, die in Data2Web in die Auswertung eingespeist werden kann und die Grundlage für alle Gruppenfunktionen von Data2Web bilden wird. Gleichzeitig wird eine Liste der bis dahin im Experiment eingesetzten Tiere erstellt und aktualisiert: Bei der irrtümlichen Verwendung einer nicht eindeutigen RFID Tag Nummer bemerkt das Programm dieses und verweigert

die Annahme dieses Tags. Dies ist besonders bei kleineren Tag Serien wichtig, in denen jede Zahl beliebig oft vorkommen kann: Eine Normierung der Tag IDs nach DIN ist hier meist nicht vorgesehen.

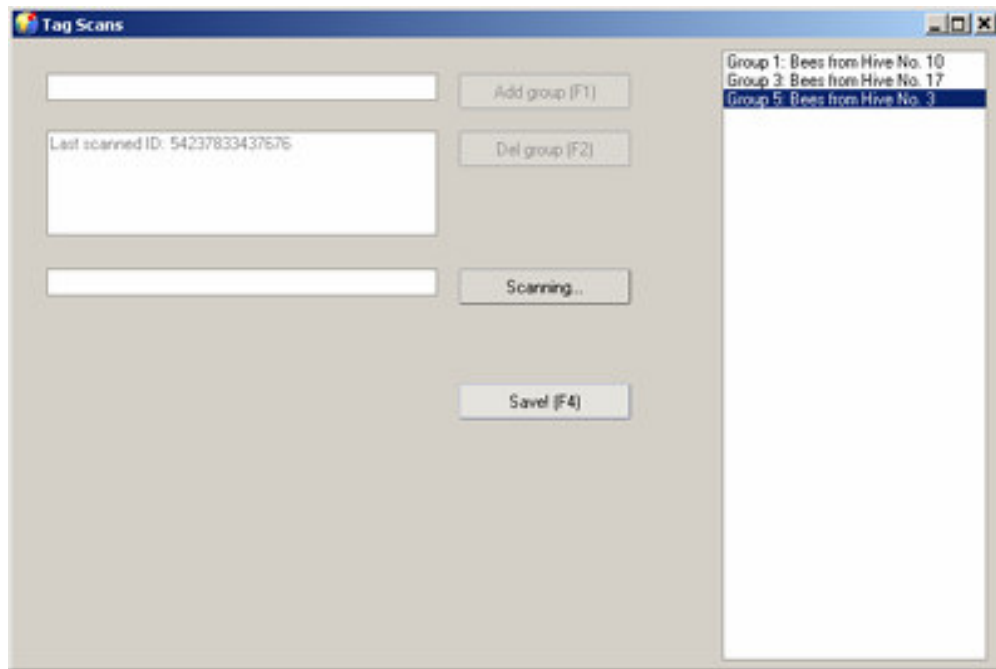


Abbildung 13: Data2DB Programmfenster

Ebenso wie Data2SQL ist dieses Programm stark von den Bibliotheken des Hand-scanners abhängig, den es zur Eingabe der Daten nutzt. Diese liegen zum gegenwärtigen Zeitpunkt nur als Windows DLL vor und begrenzen damit die Portabilität stark - die entsprechende Umsetzung auf das .NET Framework sollte bis zum Sommer 2004 abgeschlossen sein.

Interessant ist auch hier die Portierung auf Kleinst-PCs wie Handhelds, die bei der Markierung im Feld erhebliche Erleichterung gegenüber dem Einsatz von z.B. Notebooks bieten. Da diese Portierung schon von Anfang an geplant und vorgesehen war, wird ihre tatsächliche Ausführung nach Fertigstellung der .NET Bibliotheken der Scannertreiber keine Probleme verursachen.

4.2.5 Versuchsdurchführung mit *Data2SQL*: Erfassung

Data2SQL ist die Schnittstelle zur Steuerung der laufenden Datenerfassung - steuert also letztlich das zeitgenaue Auslesen der Scanner der ersten Generation. Hier bestimmen zwei Punkte das Programm: Maximale Stabilität zur Vermeidung von Fehlfunktionen, die die laufende Datenerfassung unterbrechen könnten und die direkte Abhängigkeit von den Softwarebibliotheken der Scanner, die einige Einschränkungen vor allem der Portabilität mit sich bringen.

Nachfolgend der Aufbau des Programms:

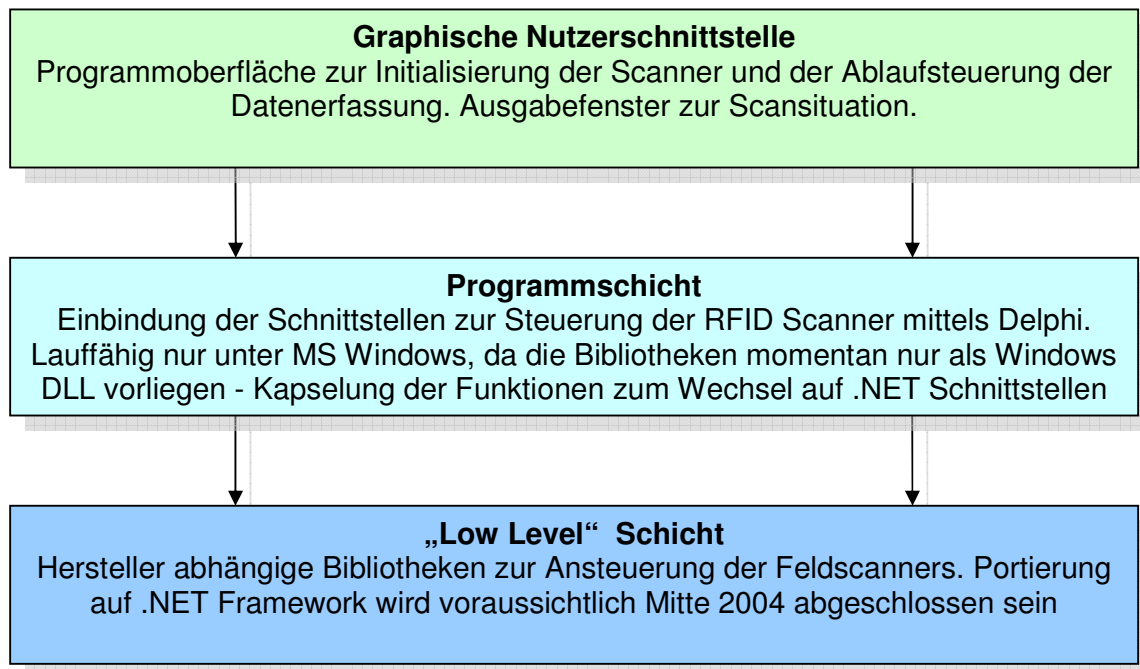


Abbildung 14: Data2SQL - schematische Übersicht

Data2SQL bietet mit nur wenigen Funktionsknöpfen eine Schnittstelle zur Kette der über RS-232 angeschlossenen Scanner und realisiert deren Steuerung. Die eingelesenen Daten werden in eine auf einem beliebigen Rechner eingerichtete Datenbank eingespeist - allerdings empfiehlt sich eher der lokalen Rechner als zumindest temporärer Datenbankserver, um mögliche Fehlerquellen beim Transport der Daten über das Netzwerk auszuschließen.

Die Parameter der anzusteuernenden Datenbank übernimmt eine XML Datei, die mit einem beliebigen Texteditor bearbeitet werden kann.

Das Programm erkennt automatisch alle Scanner, die an die RS-232 Schnittstelle des Computers angeschlossen wurden. Ein Problem stellt zum aktuellen Zeitpunkt das Fehlen von eigenen Timerbausteinen in den Scannern dar, da diese noch in Software realisiert werden müssen. Dies bedeutet aber eine Abfrage mehrmals pro Sekunde, die einen isolierten Versuch mit den Scannern ohne angeschlossenen Computer nicht möglich macht. Ohne Computer würden die Scanner nach wie vor

die gescannten Daten in dem eigenen Speicher ablegen - bei der folgenden Auswertung aber könnte diesen kein genaues Scandatum zugeordnet werden, was eine Quantifizierung z.B. einer Aufenthaltszeit nicht möglich macht. Die schon in Arbeit befindliche Folgegeneration der Scanner wird dieses Problem nicht mehr haben, da alle benötigten Funktionen in Hardware realisiert werden und damit die Notwendigkeit der permanenten Softwareanbindung entfällt: Data2SQL wird sich dann im Auslesemodus auf ein reines Abfragen aller Datensätze auf einmal beschränken. Dies erlaubt aber auf der anderen Seite einen echten Einsatz im Feld, in dem bis zu 100.000 Datensätze ohne erneutes Auslesen in dem Scanner gespeichert werden können.

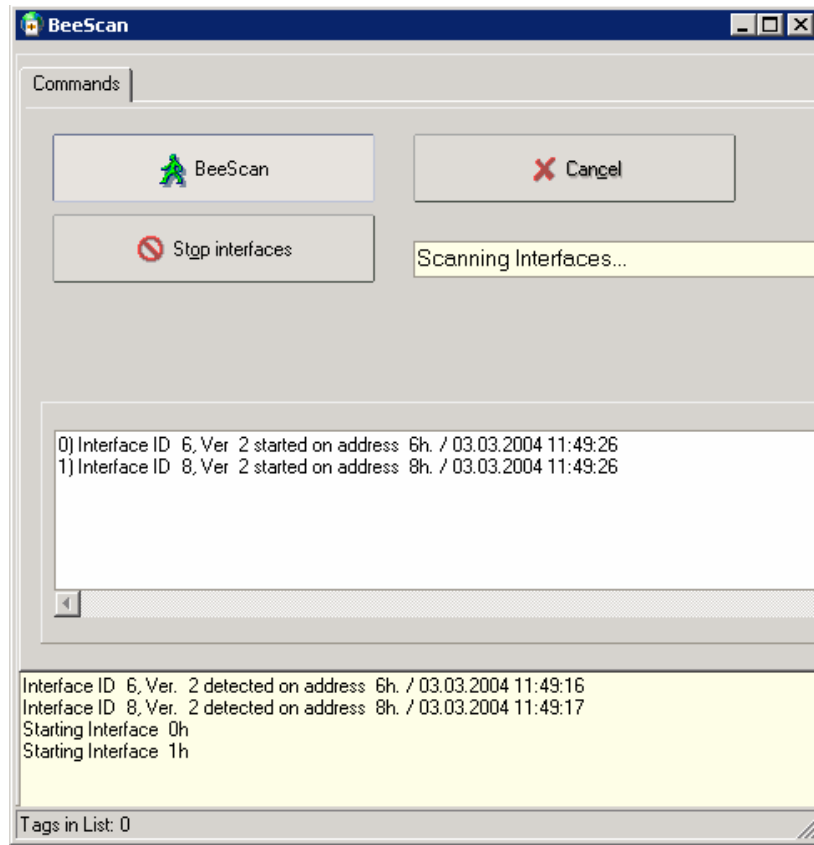


Abbildung 15: Data2SQL Programmfenster

Parallel zur Realisierung der Funktionen in Hardware in der zweiten Scannergeneration läuft die Portierung auch der Low-Level Funktionen auf das .NET Framework. So wird auf der einen Seite eine breitere Hardwareunterstützung gesichert - auf der anderen aber vor allem die aktuelle Generation von Kleinstcomputern (Handhelds) damit unterstützt, so dass ein mobiles Auslesen der Datensätze stark erleichtert wird.

4.2.6 Versuchsauswertung mit *Data2Web*: Auswertung

4.2.6.1 Theorie

Data2Web bildet die Oberfläche zur Manipulation und Abfrage der Datenbank, zur Auswahl der Verarbeitungsschritte, zur graphischen Aufbereitung der entstehenden Datenmenge und zum Export der Daten in externe Dateien zur Weiterverarbeitung.

Dabei bildet die Generierung von Datenbankabfragen via einfacher Menüs den Schwerpunkt der Applikation. Die Hürde, aus den in solchen Experimenten oft entstehenden Datenmassen Teilmengen gezielt herauszufiltern soll hier überwunden werden. Bei den Auswertungs- und Darstellungsmöglichkeiten gilt es, einige Sonderpunkte zu beachten. In mehreren Schritten versucht das Programm aus den gescannten Rohdaten sinnvolle Ketten zu bilden und diese graphisch auf verschiedene Weise aufbereitet, an den Nutzer zurückzugeben.

Dabei kann jeder dieser Schritte unbedingt auch mit Fehlern behaftet sein: Seien es auf der einen Seite Fehler technischer Art, wo eine schlechte Positionierung des RFID Tags zum Scanner für seine Nichterkennung sorgt, oder auf der anderen Seite Fehler, die sich aus dem Verhalten der Versuchstiere ergeben können. Vor allem dieser zweite Punkt spielt eine gewichtige Rolle: Beobachtungen zeigen immer wieder Bienen, die im Bereich des Scanners Bewegungen durchführen, die später leicht als Ausflüge interpretiert werden könnten, wenn man nur die Zahl der gescannten Durchgänge betrachtet. Ein Blick auf die Tabelle der Rohdaten zeigt aber schnell eine „Dauer“ dieser Ausflüge von nur wenigen Sekunden und entlarvt sie damit als nicht relevante Bewegungen. Diese Kontrollen sind durchaus auch schon teilweise in der Software realisiert - können aber in keinem Fall die notwendige Kontrolle der Daten im Rahmen der Versuchsauswertung ersetzen. Das Programm leistet hier massive Hilfestellung in Form von Sortierung und Einschränkung der ungeheuren Datenmengen - die Endkontrolle verbleibt aber immer beim Anwender. Die Graphen und Tabellen, die man hier per Mouseclick generieren kann, führen schnell zu einem Schluss auf Endergebnisse - sind aber nur erste Tendenzen und Hinweise, in welche Richtung sich die Auswertung entwickeln kann. Notwendig aber bleiben sie, um dem Ansatz der möglichst umfassenden Überwachung - ohne unbedingt mit einer Anfangserwartung zu starten - die notwendige Mächtigkeit geben: Ohne eine schnelle Vorsichtung der Daten ist es beinahe ausgeschlossen, aus mehreren 100.000 Datensätzen sinnvolle Ergebnisse mit Mitteln, die sich bis jetzt als ausreichend erwiesen haben („von Hand“ Sichtung, Tabellenkalkulationen), gewinnen zu wollen.

Alle Verarbeitungsroutinen werden kontinuierlich getestet und weiterentwickelt, um schließlich auch die Realisierung von Experimenten mit Datenmengen, deren Verarbeitung nur noch rein maschinell geschehen kann, zu ermöglichen.

4.2.6.2 Aufbau

Realisiert ist Data2Web als reine Intra- und Internetapplikation, was eine parallele Sichtung der Daten zulässt und diese auf jedem Betriebssystem ermöglicht.

Die Voraussetzung für den Einsatz von Data2Web bildet daher ein .NET fähiger Webserver wie der MS IIS 5³³ und höher oder aber Apache³⁴ 2.0 - hier ist darauf zu achten, dass das Modul mod_mono genutzt wird - und höher. Durch die extreme Belastung dieses Servers im Zuge der Auswertungen ist von einer parallelen Nutzung der Webdienste in der Produktion abzuraten - hier sind sonst starke Geschwindigkeitseinbrüche unvermeidbar. Es bietet sich eine kurzfristige Installation und Nutzung einer beliebigen Arbeitsstation (mind. 1 GB RAM) als Server in einem Intranet für diese Applikation am ehesten an - dafür kann unabhängig von aufwendigen Serverinstallationen jede Installation von Windows 2000 oder XP Pro genutzt werden.

Data2Web bildet zunächst im als „Basic“ bezeichneten Bereich grundlegende Funktionen der Datenbankpflege nach: Sicherung der Datenbanken, Initialisieren der Datentabellen durch die Löschung doppelter Scans und die Erstellung von Arbeitsdatentabellen für alle folgenden Verarbeitungsschritte sind hier gesammelt.

Im als „Processing“ benannten Bereich folgt die eigentliche Auswertung. Die Schritte zur Generierung einer neuen Auswertung sind dabei in ihrem Aufbau zunächst immer die gleichen:

- i. Eingrenzung der Zeit (Datum und Tageszeit)
- ii. Eingrenzung der Gruppen
- iii. Eingrenzung der in die Auswertung einfließenden IDs
- iv. Auswahl der Auswertung

Prinzipiell ist bei jeder Auswertung zu beachten, dass der Transfer der Daten von der Ergebnismenge der Datenbank über die ODBC Schnittstelle in den ASP.NET Speicher den die Geschwindigkeit bestimmenden Schritt darstellt. Dabei kann die benötigte Zeit bei einer entsprechend großen Ergebnismenge leicht auf 5 Minuten und mehr (Ergebnismenge im Bereich > 200.000 Datensätze) wachsen - daher sollten nicht benötigte Daten schon im Vorfeld gefiltert werden um die spätere Übersichtlichkeit der Auswertung zu erhöhen und die benötigte Zeit zu minimieren.

Das Eingrenzen der Zeit erlaubt eine Konzentration auf bestimmte Datums- oder Tageszeitabschnitte - kann aber auch helfen, die Qualität der Ergebnismengen zu verbessern: So erwies sich schnell, dass vor allem im Nachtbetrieb viele Tiere offensichtlich im Bereich der Scanner übernachteten und somit dutzende Einträge in der Datenbank generierten, die aber keinen Ausflügen entsprachen. Ein Filter, der nur tagesaktive Versuchstiere einschließt, umgeht diese mögliche Fehlerquelle.

Die Eingrenzung der Gruppen folgt exakt der Definition von Gruppen, die in der ersten Stufe des Experimentes - der Aufnahme aller IDs mittels Data2DB - festgelegt wurde. Diese Definition bestimmt also die Mächtigkeit dieser Stufe und sollte möglichst fein eingeteilt werden. Jedes Tier kann dabei in beliebig viele Gruppen aufgenommen werden.

Die Eingrenzung der verwendeten IDs schließlich ermöglicht eine noch wesentlich feinere Einteilung der Versuchstiere. Diese Stufe ist notwendig, um den Ergebnissen des Experimentes Rechnung tragen zu können: Hier werden interessante Gruppen

³³ <http://www.microsoft.com/windowsserver2003/iis/default.aspx>

³⁴ <http://www.apache.org/>

von Tieren in Listen zusammengestellt, und können so gesondert beobachtet und ausgewertet werden. Jede Liste kann dabei als „include“ oder „exclude“ Liste verwendet werden: So ist eine ausschließliche Konzentration auf die Mitglieder der Liste genauso möglich, wie eine Konzentration auf alle Mitglieder, die eben nicht auf der aktuellen Liste stehen.

Die letzte Stufe - die Auswahl der Auswertung - wird gebildet von den Auswertungsmodulen, deren Entwicklung nicht mit dieser Arbeit stocken soll, sondern bei jeder neuen Auswertung entsprechend entwickelt und eingebunden wird. So entsteht eine Bibliothek von Auswertungen, die ganz unterschiedlichen experimentellen Anforderungen entsprechen werden.

Der schematische Aufbau der Software ist geprägt von dem Bestreben nach maximaler Plattformunabhängigkeit, die dem Zeitaufwand, die während des Experimentes in jedes einzelnes der drei Softwaremodulen Rechnung trägt:

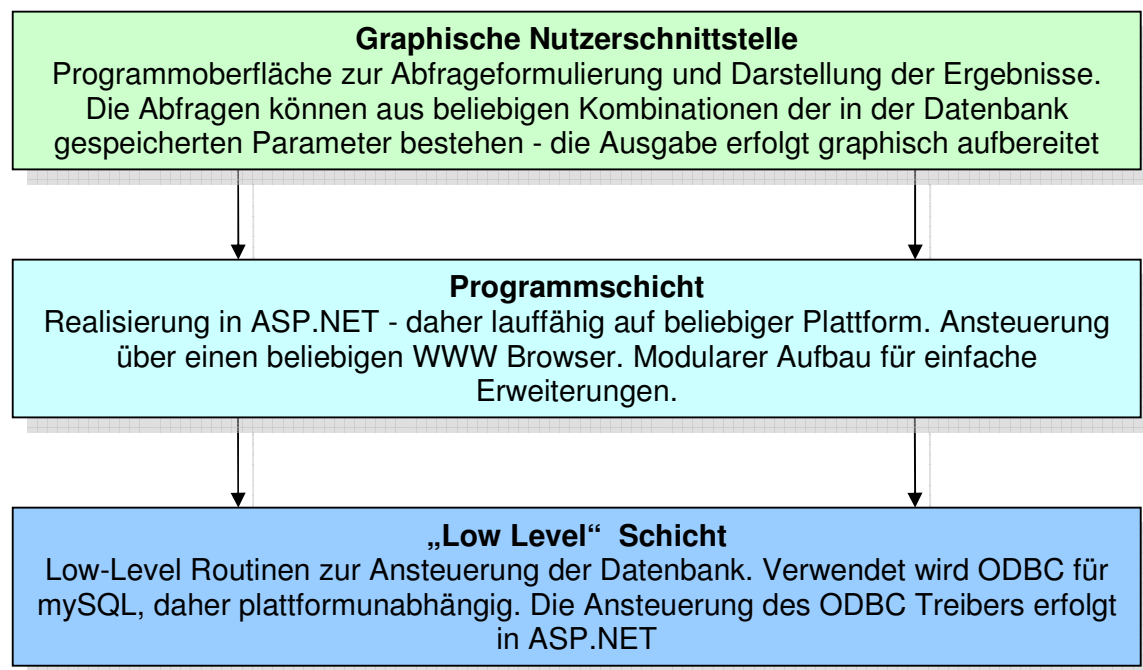


Abbildung 16: Data2Web - schematische Übersicht

Im Experiment wird Data2DB immer in der Anfangsphase des Markierens und ersten Erfassens genutzt werden und hat danach aber keinerlei Funktion mehr. Data2SQL läuft dagegen während des gesamten Experimentes und erfasst die laufenden Daten - dies aber ohne jede Notwendigkeit für einen Benutzer, einzugreifen. Die Auswertung der Daten dagegen dürfte den Schwerpunkt der zeitlichen Auseinandersetzung mit der dargestellten Software bilden.

4.2.6.3 Praxis

Im folgenden soll der dringend zu beachtende Unterschied zwischen den aus dem Experiment gewonnenen Rohdaten, den nach dem ersten Schritt zur Bereinigung der Datenbank gewonnenen Arbeitsdaten und der Darstellung in der Ergebnisstabelle und dem dazu gehörenden Graphen dargestellt werden. Schnell werden anhand dieses Beispiels die Stärken und Schwächen des Systems wie es zum aktuellen Zeitpunkt besteht, verdeutlicht werden - aber auch die Richtung, die ein weiterer Ausbau nehmen muss, um diese Fehlerquellen zu minimieren und mittelfristig alle verarbeitenden Routinen soweit zu optimieren, dass eine vollkommen computergestützte Auswertung möglich sein wird. Das gezeigte Beispiel ist dabei nicht repräsentativ für die Gesamtergebnisse der gewonnenen Daten, sondern wurde ausgewählt, weil es alle möglichen Fehler zusammenfasst.

Zunächst folgt die Tabelle der Rohdaten, die alle erfassten Bewegungen für die ID 2306128607359016917 für den Zeitraum vom 15. bis zum 19. August des im Sommer 2003 angesetzten Experimentes aufzeigt:

[burli] Query Window

File Edit View Query Options HotKeys

SELECT * FROM tag_raw WHERE ((t_Time >= '2003-8-15 00:00:00.0' AND t_Time <= '2003-8-20 00:00:00.0')) AND (t_ID=2306128607359016917) GROUP BY t_auto

Query 1

	t_auto	t_ID	t_Time	t_Interface_ID	t_tmp	t_Antenna
1	2018	2306128607359016917	2003-08-15 15:53:56	6	0	1
2	2048	2306128607359016917	2003-08-15 16:02:42	6	0	1
3	4121	2306128607359016917	2003-08-17 16:23:21	8	0	0
4	4122	2306128607359016917	2003-08-17 16:23:22	8	0	0
5	4187	2306128607359016917	2003-08-17 16:51:25	6	0	0
6	4192	2306128607359016917	2003-08-17 16:54:56	6	0	1
7	4193	2306128607359016917	2003-08-17 16:54:58	6	0	0

Result 1

7 rows in set (1.30) sec

Messages History Explain

Executing Query Read Only

Abbildung 17: Unveränderte Daten des Experimentes

Die für diese Betrachtung wichtigen Spalten der Tabelle werden von der Zeit (t_Time), dem Scanner (t_Interface_ID) und der Teilantenne des Scanners (t_Antenna) gebildet.

Das Beispiel eines perfekten Scans bilden die Tabellenzeilen 6 und 7: Hier erfolgte eine Bewegung innerhalb von 2 Sekunden (16:54:56 bis 16:54:58 am 17.08.2003) unter derselben Antenne (Nr. 6) hindurch: Die Bewegung innerhalb der Unterantennen erfolgt von 1 (Zeile 6) nach 0 (Zeile 7) und zeigt somit eine Richtung von

draußen nach drinnen an. Dazu passt der Eintrag aus Zeile 5, der etwa 3,5 Minuten vorher eine Bewegung anzeigt, bei der aber der zweite Scan fehlt: Die Richtung dieser Bewegung ist nicht klar aus den Daten herauszulesen, sondern ergibt erst mit der eindeutigen Bewegung aus den Zeilen 6 und 7 Sinn.

Die Zeilen 3 und 4 zeigen einen typischen Doppelscan innerhalb einer Sekunde an: Die Unterantenne ist dabei immer 0 - das Versuchstier hat offensichtlich die Scanner recht langsam unterquert. Solche Einträge werden in dem ersten Verarbeitungsschritt gelöscht. Damit bliebe einer der beiden Einträge zurück, der aber als einzelner Eintrag nicht ganz eindeutig ist: An diesem Tag (17.08.2003) gibt es keinen Eintrag mehr - man kann also nur vermuten, dass das Versuchstier bei einem vorherigen oder nachfolgenden Aus- oder Einflug nicht vom Scanner erfasst wurde.

Ähnlich stellt sich die Situation am 15.08. aus den Zeilen 1 und 2 dar: Es ist nicht ganz klar, ob die beiden Einträge das Hin und Zurück eines ganzen Fluges zeigen, oder aber jeweils für sich eine Flug darstellen, bei dem entsprechend eine Passage durch den Tunnel nicht aufgezeichnet wurde.

Jetzt folgt die Tabelle mit Daten (Abbildung 18), die innerhalb von Data2Web einem ersten Verarbeitungsschritt unterzogen wurden: Alle doppelten Scans wurden aus der Tabelle entfernt.

Diese entstehen, wenn ein Versuchstier zu langsam unter dem Scanner hindurchkrabbelt und damit mehrfach gescannt wird. Dieser Verarbeitungsschritt soll in der zweiten Generation der Scanner in Hardware realisiert werden. Im Experiment des Sommers 2003 bestanden etwa 20 % der gewonnenen Erstdaten aus doppelten Scans und wurden daher in diesem Schritt entfernt.

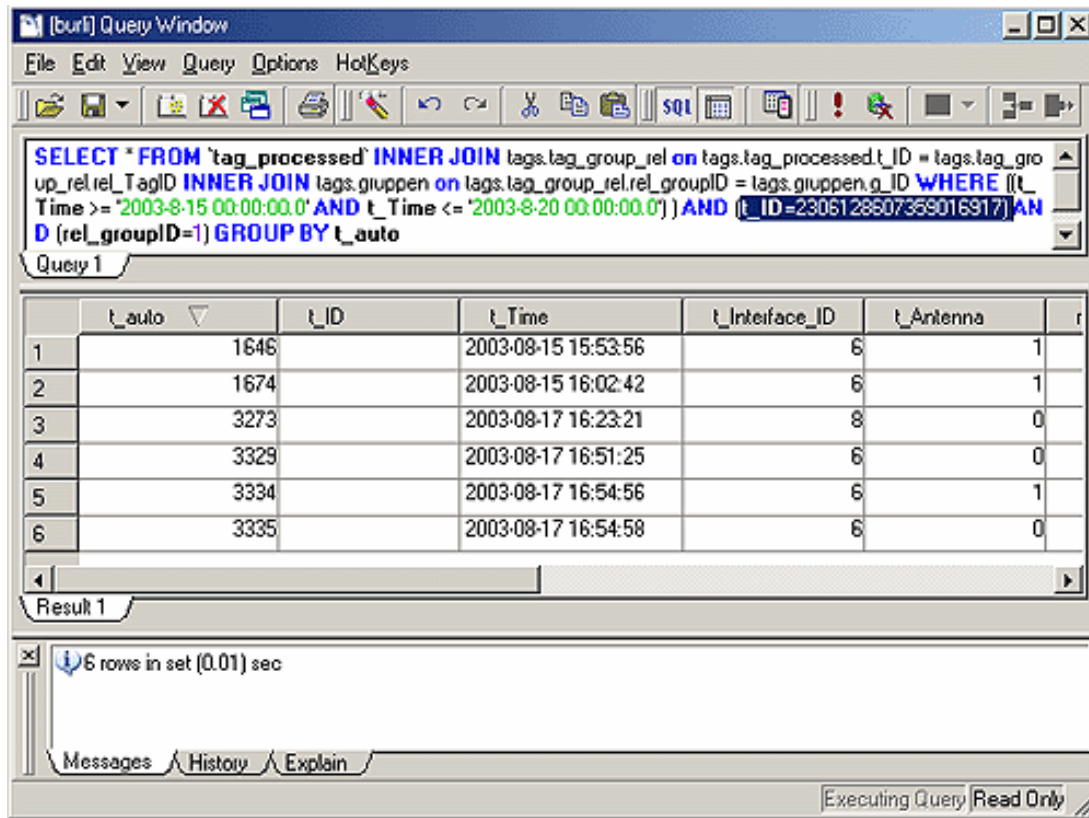


Abbildung 18: Daten nach dem ersten Verarbeitungsschritt

Gelöscht wurde die Zeile 7 aus Abbildung 17, die einen doppelten Scan zur vorhergehenden Zeile darstellt.

Diese Tabelle wird jetzt durch die Auswertung „Aktivität pro Tag“ in eine entsprechende Tabelle sowie den dazu gehörenden Graphen umgewandelt. Diese erste Auswertung des Programms stellt nur erkannte Durchgänge unter dem Scanner dar und erlaubt damit vor allem den Vergleich von Gruppen, ohne eine Aussage über die absolute Anzahl der tatsächlichen Flüge pro Tag zu treffen.

Day	Counter	Unique IDs	Average	Deviation
15.08.2003	2	1	2	0
16.08.2003	0	0	0	0
17.08.2003	3	1	3	0
18.08.2003	0	0	0	0
19.03.2003	0	0	0	0

Tabelle 4: Tabellarische Aktivitätsübersicht der ID 2306128607359016917

Der Graph, der diese Tabelle entsprechend übersichtlich darstellt, wird im Rahmen dieser Auswertung ebenfalls generiert.

Das Beispiel zeigt Stärken und Schwächen der aktuellen Version des Systems. Die zweite Generation der Scanner, die viele Funktionen in Hardware realisieren, werden den Großteil der aufgezeigten möglichen Fehlerquellen abstellen - ohne die Notwendigkeit zur zumindest teilweisen manuellen Kontrolle zu negieren.

4.3 Alternative Auswertungsmöglichkeiten

Die Datenpräparation und -erfassung ist - da abhängig von vorgegebenen Schnittstellen - momentan ohne Alternative. Durch die weitgehende Unabhängigkeit von den verwendeten Plattformen sollte dies aber in keinem Fall zu Problemen führen, die sich nicht mit einer Neukompilierung des Quellcodes beheben lassen.

Ein anderer Punkt bildet die Auswertung: Data2Web entspricht der Realität und weniger einem eher abstrakten Ideal, wenn hier Simplizität vor Komplexität gesetzt wurde. Die Arbeit an sozialen Insekten mit ihrem stark interdisziplinär geprägten Ansätzen und ihrer ungeheuren Komplexität bei der Untersuchung von Phänomenen, die die Totalität aus tausenden von einzeln agierenden Individuen bilden, zieht heute in erster Linie ganzheitlich orientierte Forscher an.

Bei allen Vorteilen dieses Forschertyps bildet hier aber doch die Arbeit an Programmiersprachen und die Beschäftigung mit den eher abstrakten Bereichen der EDV bei der Mehrzahl dieser Forscher eher ein Randgebiet. So wurde Data2Web als einfachstes Frontend zu einer sehr umfangreichen wie komplexen Datenmenge eingeführt, das vor allem schnelle, erste Eindrücke und Tendenzen für jedermann zugänglich darstellen soll. Dabei ist die Auswertung der Daten aber rein deskriptiv: Neben Parametern wie Mittelwert und Standardabweichung bildet der automatisch generierte Graph ein Kernstück der Auswertungen. Auch dieser ist immer nur ein Kompromiss, der auf eine Übersicht der Ergebnisse ausgelegt ist und nicht weiter konfiguriert werden kann, um beispielsweise Beschriftungen zu lokalisieren oder Ausschnitte hervorzuheben.

Jede weitergehende Auswertung und Visualisierung muss demnach gesondert geschehen. Hier bieten sich drei prinzipielle Wege an:

1. der Export über externe Dateien und deren Weiterverarbeitung
2. das Ergänzen von Data2Web um weitere Module
3. die direkte Ansteuerung der Datenbank mittels anderer Programme

4.3.1.1 Export von Daten

Der Export von Dateien erfolgt über die direkte Generierung der ausgewählten Daten im MS Excel 2003 eigenen Format. Hier bildet XML die Grundlage für ein Dokumentenformat, das auch andere Tabellenkalkulationen oder Statistikprogramme ohne weitere Probleme importieren.

Der Vorteil dieser Methode liegt in der einfachen Definitionsmöglichkeit der Daten, die durch Data2Web ermöglicht wird: Die gesuchte Grundmenge der gesamten Daten (beispielsweise Zeitabschnitt oder Gruppeneinschränkung) wird per Mausklick vorgeneriert und kann direkt in ein beliebiges Statistikprogramm zur weiteren Analyse eingespeist werden.

4.3.1.2 Ergänzen weiterer Module

Der modulare Aufbau von Data2Web mit seinen klaren Schnittstellen erlaubt das Ergänzen von Modulen. Hierzu ist die Beherrschung einer Programmiersprache obligatorisch. Durch die Wahl der verwendeten Softwareumgebung aber ist diese (beinahe) beliebig: C#, Visual Basic, Delphi oder Java sind als .NET Dialekte verfügbar und können daher genutzt werden.

Der Aufwand ist nichtsdestotrotz nicht klein und muss im Einzelfall abgewogen werden: Sollte es sich aber um eine häufige Anwendung handeln, dann wird diese einmalige Investition selbst komplexeste Auswertungen einem unbeschränkten Personenkreis verfügbar machen.

Die Entwicklung von Modulen zur Auswertung ist mit der Abgabe der vorliegenden Arbeit nicht abgeschlossen, sondern wird weitergeführt - wobei Kooperationen sehr willkommen sein werden.

4.3.1.3 Direkte Ansteuerung der Daten

Die direkte Ansteuerung der Daten gibt die vollkommene Kontrolle über die erhaltenen Daten und deren Auswertung - bürdet aber auch die Last auf, sich mit mehreren hunderttausend Datensätzen und deren logische Verbindungen auseinander setzen zu müssen.

Schritte wie Bereinigung der Daten oder die zeitliche oder gruppenorientierte Einschränkung der Ergebnisse müssen entsprechend der verwendeten Datenbanksprache in SQL formuliert werden (siehe 4.2.3). Prinzipiell ist diese direkte Ansteuerung aber kein Problem, da ja auf die Verwendung ausschließlich standardisierter Schnittstellen geachtet wurde.

Im Folgenden sei für StatSoft Statistica der mögliche Weg der Anbindung der Daten dargestellt:

- zunächst würde eine neue Instanz von Statistica Query initialisiert werden
- hier muss zunächst die Art der Verbindung angegeben werden, die die Daten bereitstellen wird - ODBC als am weitesten verbreitete Schnittstelle bietet da keinerlei Schwierigkeiten
- nach den Grundlagen der Verbindung kann hier die tatsächliche SQL Syntax definiert werden, die den Umfang der zurück gelieferten Daten bestimmt
- das Ausführen der so definierten Abfrage liefert die gewünschte Untermenge der Daten in Form eines Statistica Spreadsheets zurück

Hier werden immer relativ unbearbeitete Rohdaten zurückgeliefert - der einzige Schritt, der in diesem Stadium schon durchgeführt wurde, ist die Löschung aller Mehrfachscans, die entstehen, wenn ein markiertes Insekt sich zu lange unter dem Scanner aufhält.

Daher muss nach diesem Schritt zunächst eine sinnvolle Ordnung der Daten - nach beispielsweise Ausflugdauer oder -häufigkeit - durchgeführt werden, wozu im Falle

von Statistica mit Statistica BASIC eine auf Statistica optimierte Programmiersprache zur Verfügung steht.

Welche der drei beschriebenen Möglichkeit nachhaltig die günstigste ist, hängt immer von der jeweiligen Auswertung ab. Die für eine möglichst große Anzahl Nutzer sinnvollste Alternative ist dabei die Entwicklung weiterer Module und deren Einbindung in Data2Web - nur auf diesem Weg können mehrfache Parallelentwicklungen minimiert werden.

5 Anwendung: Proof of concept mit *Bombus terrestris*

5.1 Fragestellung

Die Anforderungen an ein Proof of Concept Experiment sollten einerseits die späteren Anforderung in Betracht ziehen, ohne aber andererseits den vollen Umfang dieser zu implementieren: In diesem Stadium ist ein Scheitern noch immer möglich und das Finden noch unbekannter Designmängel eines der erklärten Ziele.

So bot sich eine mittelgroße *Bombus terrestris* Kolonie als Testvolk an: Auf der einen Seite werden hier Probleme, die sich aus dem Leben in einem sozialen Volk ergeben das erste Mal angesprochen:

- die Erkennung und Reaktion der Individuen zueinander, sobald sie mit Fremdkörpern wie den RFID Chips markiert sind
- die räumlichen Verhältnisse vor allem im Tunnelbereich, wo die Lesegeräte auch mehrere Individuen in sehr kurzen Zeitabständen sicher detektieren müssen
- der Futtereintrag der markierten Bienen, der auch mit Markierung, Tunnelsystem und Scannern nicht in einem Maß sinken darf, das die Kolonie beeinträchtigt oder bedroht wird

Gleichzeitig bieten Hummeln aber gegenüber den meisten Bienen- und Ameisenkolonien einige Vorteile, die den Ersteinsatz erleichtern - prinzipielle Schwierigkeiten aber schon andeuten, ohne das Experiment schon im Vorfeld undurchführbar zu machen:

- Hummeln sind im Vergleich zu Bienen recht groß und entsprechend weniger agil: Vor allem das Laufen an Wänden und Decken des Gangsystems erfordert spezielle Lösungen, um die Funktion der Scanner trotzdem zu gewährleisten – Hummeln dagegen bleiben im Tunnel wo möglich in der Waagerechten
- Hummelkolonien sind in ihrer Größe überschaubar: Kolonien von unter 100 Tieren sind leicht zu handhaben und trotzdem stabil über einen längeren Zeitraum
- Hummelkolonien sind unabhängig von der Witterung im Labor nutzbar

Hier soll also gezeigt werden, dass einige ausgewählte Individuen, die aus den aktiven Nahrungssuchern der Kolonie ausgesucht wurden, über einen Zeitraum von einer Woche kontinuierlich beobachtet werden können. Die entstehenden Daten sollen kontinuierlich aufbereitet und präsentiert werden. Parallel zur automatischen Identifizierung werden Zählkontrollen gemacht, um den Wirkungsgrad des Systems zu testen.

Bewusst soll hier im Rahmen dieses Experiments keine neue Fragestellung angegangen werden, da hier der Schwerpunkt auf den praktischen Belangen der Methode liegt - auf deren Erprobung, Dokumentation und Verbesserung. Ganz im Gegenteil hierzu sollten aus der entstehenden Datenbank Beispiele generiert werden, wie im nachhinein Antworten aus diesen Rohdaten destilliert werden können - egal, wie spe-

ziell der Zeitraum der Fragestellung oder wie verschachtelt die Kombination der Parameter ist.

5.2 Setup

Eine Box mit einer Kolonie *Bombus terrestris* wurde verbunden mit einer Flugarena (siehe Abbildung 19). Als Verbindung diente der erste Entwurf eines speziellen Tunnelsystems. Der eingesetzte Tunnel bildet dabei den Kernpunkt der Installation und erfüllt mehrere Funktionen:

- Ermöglichung des Ein- und Ausgangsverkehrs in die angeschlossene Flugarena
- Vermeidung von Stauknotenpunkten an den Stellen, an denen sich der Tunnel für einen erfolgreichen Scan der Tag ID verengen muss
- Erreichen einer möglichst geringen Anzahl von fehlerhaften Scans durch eine spezielle Geometrie, die die zu scannenden Tiere in der richtigen Position am Scanner vorbei führt

Damit ergab sich als schematische Übersicht folgende Anordnung:

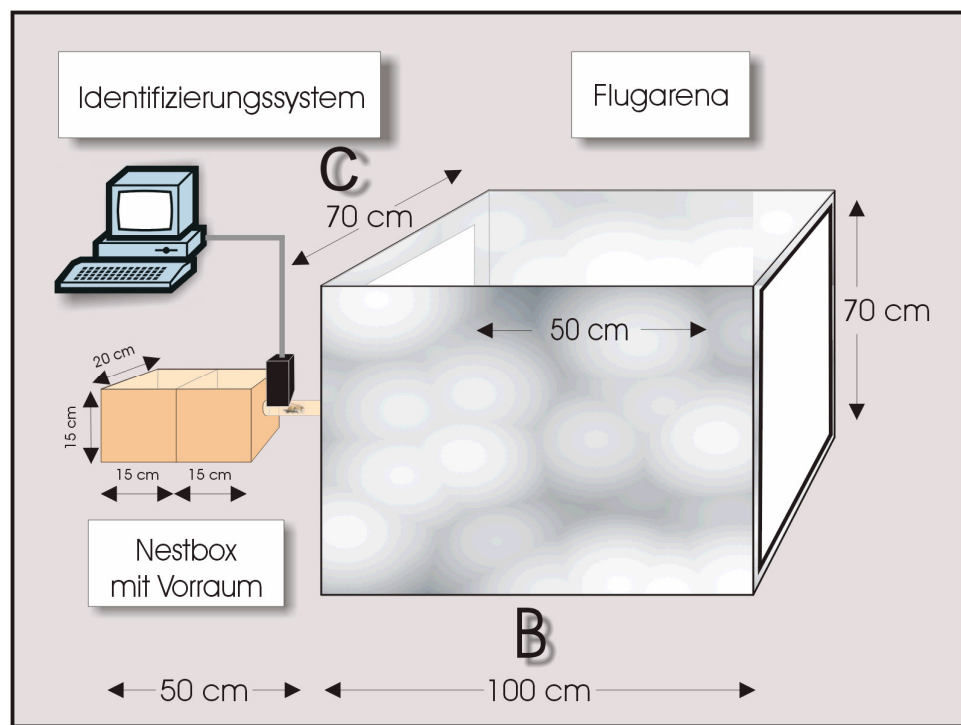


Abbildung 19: Setup des „Proof of Concept“- Experiment

Als Scanner kam eine noch nicht auf die Anwendung optimierte Versionen eines USB Lesestiftes zum Einsatz. Das aufwendige und kostspielige Design und die Produktion einer optimierten Version sollte erst nach dem erfolgreichen Abschluss des Proof of Concept Experiments erfolgen.

Die technischen Daten des verwendeten USB Stiftes:

- Microsensys™ PEN-USB
- Read Write Unit 13.56 MHz im Stiftdesign mit M12 Antenne
- Abmessungen: Breite 12 mm, Länge 170 mm
- HOST Interface: USB 1.1

Als RFID Tags kamen die oben beschriebenen mic3®-64-TAGs zum Einsatz. Deren technische Daten waren:

- Microsensys™ mic3®-64-TAG
- Closed Coupling Read Only Transponder 64bit in mic3 Technologie
- Verwendete Frequenz: 13.56 MHz
- Speicher: 64 Bit Laser programmiert
- Abmessungen: 1.0 * 1.6 * 0.5 mm³
- Benötigte Antenne: M12

Die Tags wurden mit Schellack auf dem Thorax der Versuchstiere, die alle die gleiche Größe hatten, befestigt. Dabei erwies es sich als notwendig, dass die Hummeln zunächst noch einige Minuten in der Fangvorrichtung, die zum Markieren genutzt wurde, verbleiben, da direkt im Anschluss ernsthafte und ausdauernde Versuche von ihnen unternommen wurden, die Tags wieder loszuwerden. Nach ca. einer Stunde war dieses Putzverhalten nicht mehr zu beobachten und das Experiment konnte gestartet werden. Auch im späteren Verlauf konnte kein auffallendes Putzverhalten der markierten Tiere mehr beobachtet werden. Neben den RFID Tags auf dem Thorax erhielt jede Hummel eine farbige Markierung auf dem Abdomen, die die Vergleichsbeobachtungen erleichtern sollte.

Die Dauer des Experimentes betrug 7 Tage, in denen alle Durchquerungen des Scanners von den 5 getagten Tieren protokolliert wurden. Die Fütterung der Tiere erfolgte mit Zuckerwasser, das in Futterschälchen in der Flugarena angeboten wurde sowie durch Pollen, der täglich direkt in das Nest zur Aufzucht der Larven hinzu gegeben wurde.

Neben der automatischen Protokollierung erfolgte als Kontrolle der Effizienz des Systems eine Zählung der Durchquerungen und ein folgender Abgleich mit den gesammelten Scann-Daten. Hierzu wurde 30 Tiere markiert und bei einem starken Futterreiz in der Flugarena die beobachtete resultierende Aktivität mit der automatischen Aufzeichnung derselben verglichen.

5.3 Auswertung

Der Vergleich der automatisierten Zählung der Durchquerungen mit der manuellen Zählung ergab eine einzige nicht gescannte Durchquerung bei insgesamt 300 beobachteten, was einer Fehlerquote von 0,33 % entspricht.

Die Einspeisung in die angeschlossene Datenbank erfolgte ohne Ausfälle oder Probleme - die folgende Abbildung zeigt einige mögliche Fragen, die an das Ergebnis gestellt werden konnten:

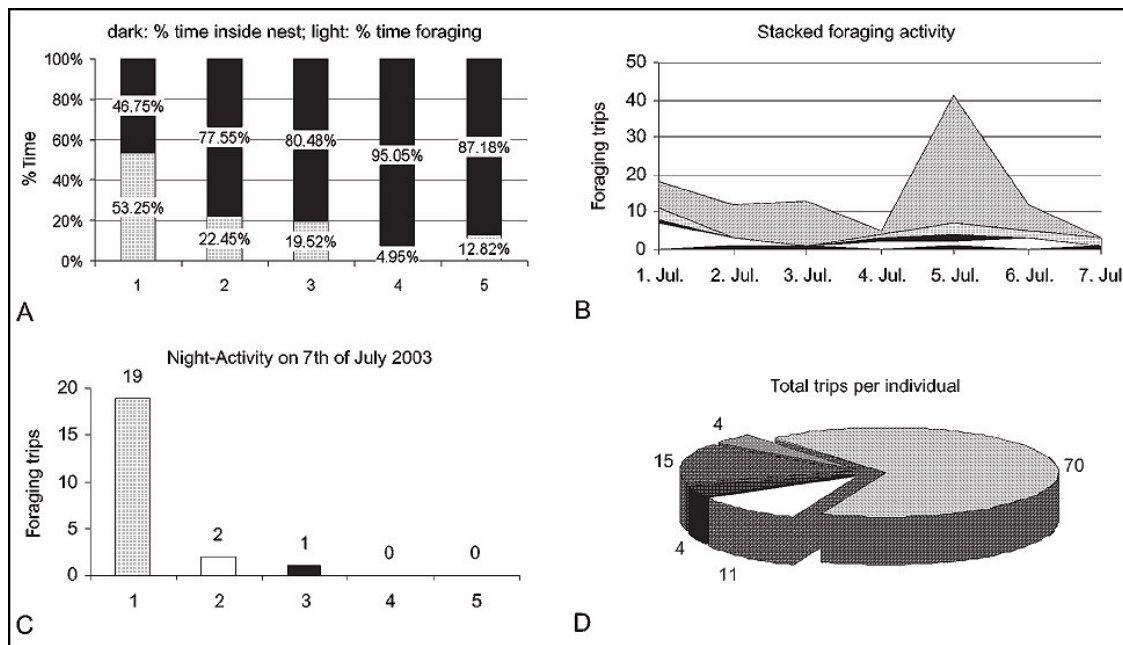


Abbildung 20: Abfragebeispiele der Aktivitäten

Die vier hier zusammengefassten Abbildungen zeigen im Einzelnen:

- Das Verhältnis der Aufenthalte innerhalb der Kolonie sowie außerhalb der Kolonie für alle 5 Tiere im gesamten Zeitraum vom 01. bis zum 07. Juli 2003
- Die akkumulierte Aktivität aller 5 Tiere über den gesamten Zeitraum
- Die absolute Nachtaktivität (22:00 - 06:00) in der Nacht vom 07. auf den 08. Juli 2003
- Die absolute Tagesaktivität aller Individuen im Vergleich während des gesamten Beobachtungszeitraums

5.4 Diskussion

Das Ergebnis zeigt auf der einen Seite das Potential des Systems und auf der anderen die Notwendigkeit zur Weiterentwicklung verschiedener Komponenten. Nahezu alle Durchquerungen wurden korrekt aufgezeichnet und in der angeschlossenen Datenbank gespeichert. Damit wird der wesentliche Anspruch erfüllt, eine zeitlich unbegrenzte Überwachung bestimmter Punkte eines Volkes sozialer Insekten zu ermöglichen.

Andere Punkte erweisen sich noch als Schwachstellen:

So ist die Schnittstelle zu der entstandenen Datenbank noch nicht über die direkte Eingabe von SQL Abfragen hinausgekommen und muss dementsprechend entwickelt werden, um eine möglichst breite Akzeptanz des Systems zu erzielen.

Des Weiteren bildet der Tunnel, der die Tiere für den Scanvorgang in eine optimale Position bringen soll, Raum für Verbesserungen.

Hummeln sind eigentlich ideale Teilnehmer an einem Experiment wie diesem: Durch ihr recht hohes Eigengewicht neigen sie weniger dazu, an Decken und Wänden zu laufen. Dies macht eine Kanalisierung des Verkehrs zu einem einzelnen Strang, in dem die Tiere mit der Oberseite zum Scanner laufen und damit den Scan ja erst ermöglichen, sehr einfach. Hummeln krabbeln auch sehr viel weniger übereinander hinweg, wie man es bei Bienen oder Ameisen regelmäßig beobachtet - auch dies muss konsequent in der Tunnelentwicklung für andere soziale Insekten getestet und eingeplant werden.

Denn schon in diesem Experiment konnten bei höherem Verkehrsaufkommen im Tunnel mehrmals Tendenzen zur Staubildung beobachtet werden, die aber im laufenden Experiment nicht weiter quantifiziert wurden.

Die Entwicklung eines Tunnels, der auf der einen Seite der Notwendigkeit des Scanvorgang zur Kanalisierung und Positionierung der RFID Tags entspricht, und auf der anderen einen ungestörten Verkehr ermöglicht, wurde als erste Notwendigkeit weiterer Entwicklungen hier erkannt.

Die große Resonanz, die auf die Veröffentlichung der ersten Studie des gezeigten Systems erfolgte (STREIT et. al. 2003), zeigt aber auch den Bedarf, den es in der Forschung an sozialen Insekten an einem solchen gibt und ermutigt zur Weiterentwicklung.



Abbildung 21: *Apis mellifera* mit RFID Tag

6 Ausblick

Nach dem erfolgreichen Proof of Concept (Kapitel 5) wurde im August 2003 ein Großversuch mit 1.000 Bienen aufgesetzt, der bis zum laufenden Zeitpunkt weitergeführt wird. Die Bienen wurden innerhalb von 2 Tagen am Tag ihres Schlüpfens markiert, von Data2DB erfasst und in einen Versuchsstock eingebracht, der den gezeigten Tunnel als einzigen Ausgang nutzte und so das Flugverhalten jeder markierten Biene erfasste.

Weniger biologisch relevante Fragen, als technische Aspekte und die Stabilität des Systems wurden dabei getestet und hinterfragt. Bis zum März 2004 sind so 380.000 Datensätze aufgelaufen, wobei das System ohne einen einzigen Absturz funktionierte. Einzig der Ausfall der Hardware im November unterbrach den Test für einige Zeit – durch die Winterpause der Bienen hatte dies aber glücklicherweise keine Auswirkungen auf den Versuch.

Die Erkenntnisse aus diesem Großversuch fließen in die Planungen zur Weiterentwicklung des RFID Systems ein. Neben den reinen Verbesserungen, die die Zuverlässigkeit oder Handhabung betreffen, haben sich dabei auch prinzipiell neue Wege aufgetan, die in den kommenden Jahren weiter verfolgt werden sollen.

6.1 Software

Die Software wird natürlich immer zu den laufenden Scanner und Taggenerationen aktuell gehalten. Hinzu kommt aber als wichtigster Punkt die Modulentwicklung. Jede Fragestellung erfordert andere Auswertungen - diese werden nach und nach zu einer Sammlung möglicher Fragestellungen vereint, die dann auf Knopfdruck verfügbar wird. Trotz der Größe der Datenmengen und den Schwierigkeiten, diese sinnvoll zu katalogisieren und zu verarbeiten wird so ein Weg geschaffen, letztlich einfacher und schneller zu Ergebnissen zu kommen, als dies mit der noch heute üblichen „Pen and Paper“ Methode möglich sein wird.

6.2 Technische Weiterentwicklung und Ausblicke

Die technische Weiterentwicklung umfasst Verbesserungen des Systems auf der einen sowie den prinzipiellen Ausbau auf der anderen Seite.

Zur technischen Seite gehören

- eine Erhöhung der Scanreichweite (bei gleichzeitiger Beibehaltung oder sogar Verringerung der Größe der Tags)
- eine Optimierung der Tunnel, um diese auch bei Größenpolymorphismus einsetzbar zu gestalten
- eine weitere Entwicklung der Scanner hin zu autark einsetzbaren Systemen

Die interessanteren Aspekte bietet aber darüber hinaus die prinzipielle Weiterentwicklung des Systems. Wenn wir schon in der Lage sind, einzelne Speicherbausteine auf dem Rücken einer Biene anzubringen und auszulesen, dann wäre eine weitere

Nutzung dieser Bausteine durch Mikrosensoren und deren Kopplung eine sinnvolle Ausbaustufe.

Transportierbare, miniaturisierte Sensoren, die ein ganzes Insektenleben lang biotische und abiotische Parameter aufzeichnen können würden ein sehr viel genaueres Bild der Einflüsse liefern, die auf das Insekten einwirken.

Diese Informationen sind es, die zu Entscheidungen im Insektenstaat führen - zu einer Entscheidungsfindung, deren Mechanismen wir nur ansatzweise heute verstehen.

Diese Strukturen zu entschlüsseln, nachzubauen und für den Menschen nutzbar zu machen, wäre ein weiterer Aspekt eines Fortschreitens dieses Systems: Die Schaffung von miteinander kommunizierenden Strukturen, die flexible Netze aufbauen und über Kommunikation scheinbar intelligente Entscheidungen treffen, wird uns von jeder sozialen Insektenart vorgelebt und ist bis heute aber nicht vollständig kopierbar.

Solche artifiziellen Systeme wären auf der einen Seite extrem flexibel, da sie sich nach Aufgabenstellung neu organisieren können - auf der anderen Seite aber sehr ausfallsicher, da das System auf Verluste großer Bereiche nicht mit Totalausfall, sondern mit angepasster, reduzierter Leistung reagiert. Nur solche Systeme werden in der Lage sein, eigenverantwortlich Erkundungen durchzuführen, wo Menschen nicht bestehen können. Diese Systeme finden nicht unbedingt den mathematisch optimalen Weg - immer aber einen gangbaren, effektiven und der Situation angemessenen.

Ein aktuelles Beispiel ist die Lastenverteilung von Internetservern, die heute nach einem vom Sammeln der Bienen abgeleiteten Algorithmus funktionieren kann: Ohne die Notwendigkeit einer zentralen Verarbeitungsstelle sind Servernetze mit diesem Algorithmus in der Lage, flexibel und äußerst schnell auf Änderungen der (Netzlast-) Situation zu reagieren. Dies ist nur ein kleiner Schritt, in dem ein eher einfacher Aspekt des Bienenverhaltens mit viel Erfolg auf eine moderne Technologie übertragen wurde. Mit der vollständigen Erfassung aller Einflüsse und Verhaltensweisen der Lebensweise von z.B. Honigbienen werden wir in der Lage sein, die bis heute nur in Ansätzen verstandene Selbstorganisation von bis zu 50.000 Bienen eines Stockes für uns zu nutzen.

Der Hardware Aspekt dieser Entwicklung wird gerade weltweit vorangetrieben. Die Probleme entstehen augenblicklich bei der Integration der Einzelkomponenten des Systems, da es bis jetzt keine Modelle gab, die solch massive Interaktion und Intelligenz schon mitbringen, um dann nach und nach durch künstliche Mikrostrukturen ersetzt zu werden.

Mit dem vorliegenden System ist der erste Schritt einer solchen Entwicklung gegangen. Mit den sozialen Insekten haben wir ein System, das intelligent auf die Umwelt agiert, abhängig von der Anzahl der Individuen flexible Reaktionen generiert etc.

Der nächste Schritt besteht in der Integration von Sensoren in den Chip, um die erprobte Umgebung der Bienen zu protokollieren

Der dritte Schritt bestünde in der Generierung von Aktionen in Abhängigkeit von der erlebten Umwelt und anderen mit Chips ausgestatteten Bienen in der Umgebung

Erst wenn diese Parameter funktionieren und verstanden sind, werden tatsächlich selbst organisierte Systeme von Maschinen möglich, die die Organisationsstrukturen von sozialen Insekten nachbilden.

7 Danksagung

Ich danke an erster Stelle Fiola Bock.

Des Weiteren danke ich meinen beiden Betreuern Prof. Dr. Jürgen Tautz und PD Dr. Stefan Fuchs.

Die Realisierung des Systems konnte nur mit der Förderung des Bayerischen Staatsministeriums für Landwirtschaft und Forsten sowie der Landesanstalt für Wein- und Gartenbau erreicht werden.

8 Literaturverzeichnis

Altmann, J. 1974. Observational study of behavior: Sampling methods. *Behaviour* **49**:227–267.

B.R. Mate, G.K. Krutzikowsky and M.H. Winsor (2000). Satellite-monitored movements of radio-tagged bowhead whales in the Beaufort and Chukchi seas during the late-summer feeding season and fall migration. *Canadian Journal of Zoology*, **78**: 1168–1181

Burley, N. (1986). Sex-ratio manipulation in color-banded populations of zebra finches. *Evolution* **40**:1191-1206.

Finkenzeller, K. (2002). RFID-Handbuch. Hanser Verlag, Deutschland

Gilbert, R. (2000). Using PIT Tags for tracking honeybees, Gladstone, Oregon: PIT Tag Workshop 2000 – Abstracts of Presentations, 43.

Jurisch, R. (1994). Coil on Chip - monolithisch integrierte Spulen für Identifikationssysteme, GME-Fachbericht: Identifikationssysteme und kontaktlose Chipkarten. vde-Verlag, Berlin 1994.

Meyburg, B.-U., P. Paillat and C. Meyburg (2003). Migration routes of Steppe Eagles between Asia and Africa: A study by means of satellite telemetry. *The Condor* **105**: 219-227

Palmer, R.C. (1995). *The Bar Code Book: Reading, Printing, and Specification of Bar Code Symbols*, Helmers Publishing, USA

Streit, S., Bock, F., Pirk, C.W.W., Tautz, J. (2003). Automatic life-long monitoring of individual insect behaviour now possible. *Zoology - analysis of complex systems* **106**, 169-171

Van den Bossche, W., Kaatz, K. and Querner, U. (1998). Satellite tracking of white storks *Ciconia ciconia*. *Journal of African ornithology*. In: Adams, N.J. and Slotow, R.H. (editors) *Proc. 22 Int. Ornithol. Congress, Durban*. *Ostrich* **69**: 152

9 Internetquellen

1. <http://www.argosinc.com/biology.htm>
2. <http://www.ibm.com/>
3. <http://www.microscan.com/deutsch/>
4. <http://www.quadrus-ez.com/>
5. <http://www.cept.org/> heute: European Radiocommunications Office
6. <http://www.ero.dk/>
7. „Radio Equipment and Systems (RES); Short Range devices, Technical characteristics and test methods for radio equipment to be used in the 25 MHz to 1000 MHz frequency range with power levels ranging up to 500 mW“
8. „Allgemeinzuteilung von Frequenzen für die Benutzung durch die Allgemeinheit für induktive Funkanlagen des nicht-öffentlichen mobilen Landfunks (nömL)“, Amtsblatt 12
9. „Allgemeinzuteilung von Frequenzen für die Benutzung durch die Allgemeinheit für Funkanlagen geringer Leistung des nicht-öffentlichen mobilen Landfunks (nömL) in ISM Frequenzbereichen; SRD (Short Range Devices)“, Amtsblatt 18
10. <http://www.regtp.de/>
11. <http://www.hitachi.co.jp/Prod/mu-chip/>
12. <http://www.microsensys.de/all-transponder.htm>
13. <http://www.microsoft.com/windows/>
14. <http://www.apple.com/macosx/>
15. Beispielsweise: <http://www.kernel.org/>
16. <http://www.spss.com/>
17. <http://www.statsoft.com/>
18. <ftp://ftp.isi.edu/in-notes/rfc2616.txt>
19. <http://msdn.microsoft.com/netframework/>
20. <http://www.w3.org/XML/>
21. <http://www.statsoft.com/>
22. <http://www.microsoft.com/office/excel/default.asp>
23. <http://msdn.microsoft.com/vcsharp/>
24. <http://www.ecma-international.org/>
25. <http://msdn.microsoft.com/vbasic/>
26. <http://www.borland.com/delphi/>
27. Beispielsweise: <http://www.kernel.org/>
28. <http://www.go-mono.com/>
29. <http://www.dotgnu.org/>
30. <http://www.iodbc.org/>
31. ISO/IEC 9075-1:2003 bis -13:2003 unter <http://www.iso.ch/>
32. <http://www.ansi.org/>
33. <http://www.microsoft.com/windowsserver2003/iis/default.mspix>
34. <http://www.apache.org/>