Entwicklung statistischer Auswertungsmethoden und Diskussion von Buridanlaufexperimenten

Bachelorarbeit

Jedrzej Blaszkiewicz

Betreuer: Dr. Björn Brembs, Neurobiolgie FU-Berlin
Zweitkorrektor: Prof. Dr. Hans-Joachim Pflüger

Selbstständigkeitserklärung:
Hiermit bestätige ich, dass die vorliegende Arbeit selbstständig verfasst wurde und keine weiteren Hilfsmittel als die angegebenen verwendet worden sind.

Berlin, 15.11.2010 Jedrzej Blaszkiewicz

Einführung

Moderne Experimente ermöglichen das Beobachten einer Vielzahl von Variablen zur selben Zeit und ermöglichen auf diese Weise eine breite Diskussionsgrundlage. Eine große Menge an vorhandenen Daten birgt jedoch auch Probleme, so ist es sehr schwierig die gesammelten Daten aufeinander zu Beziehen und Korrelationen zwischen Ihnen zu finden.

Um diesem Problem entgegen zu treten können die Daten mit Hilfe statistischer Methoden und moderner Computerprogramme vereinfacht werden und somit auch einfacher visualisiert werden. Im Folgenden werde ich mich mit der Analyse von Buridanlaufexperimenten, welche mit unterschiedlichen genetischen Stämmen von Drosophila melanogaster durchgeführt wurden mit Hilfe einer Hauptkomponentenanalyse auswerten und diskutieren.

Die Hauptkomponentenanalyse ist ein mächtiges mathematisches Werkzeug zur Reduktion von N-Dimensionalen Datenräumen in einen kleineren Mengenraum, welcher sich bis auf drei Dimensionen beschränkt, ohne dabei allzu viele Informationen zu verlieren.

Material Methoden

Alle benutzten Daten wurden wie in der Projektarbeit [1] beschrieben in einer Laufarena, der Buridanarena gesammelt. Diese Arena basiert auf Experimenten von Götz K.G.[2] und ermöglicht es Fruchtfliegen bestimmten Reizen auszusetzen. Im Folgenden wird sowohl die Arena, als auch die Methoden der Datenerhebung kurz erklärt und vorgestellt, da Dieses Hauptbestandteil der vorhergehenden Arbeit war, wird nicht näher auf Details eingegangen.

Bei der Arena handelt es sich um eine mittig angeordnete Plattform mit einem Durchmesser von 115mm, diese ist von einem eingesenktem Ring mit 295mm Durchmesser umgeben, welcher für die Experimentierphasen mit Wasser gefüllt wird, sodass die Fliegen keine Fluchtmöglichkeiten haben (Abb.2)

Umgeben wird die gesamte Arena von einer weißen Wand, diese wird durch mehrere hochfrequente (1kHz) Leuchtstoffröhren weiß beleuchtet. Oberhalb der Arena befindet sich an einem beweglichen Metallarm eine Webcam (Abb. 2), welche mit einem Computer verbunden ist.

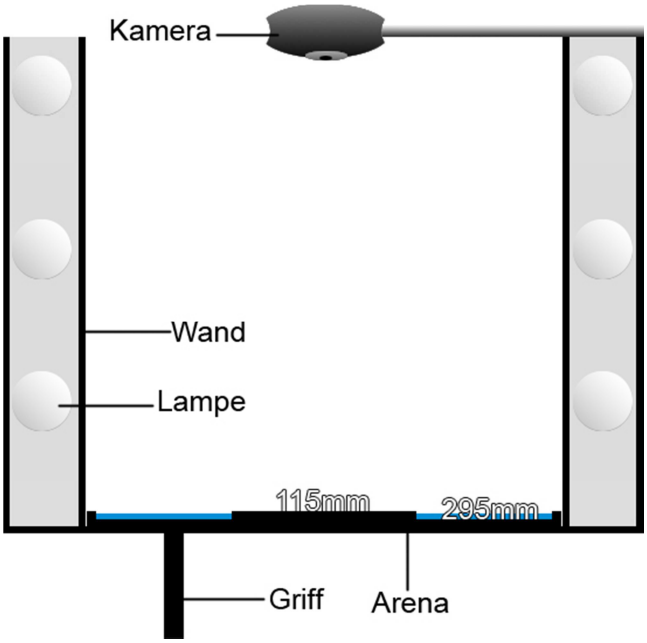


Abbildung 1 skizzierte Buridanarena im Querschnitt. Die mittige Plattform ist von Wasser und eine weiß illuminierten Wand umgeben. Oberhalb der Arena ist eine Kamera zum Beobachten der Fliege angebracht.

Für die Experimente wurden fünf genetisch unterschiedliche Stämme verwendet. Als Vergleichsstamm diente hierbei der Wildtypstamm Canton S TP (cstp). Außerdem wurden

die Stämme *radish* (rsh), *dunce* (dnc), *rutabaga* 1 und 2080 (rut1 und rut2080) genutzt. Für die endgültige Auswertung der Experimente wurden die Daten von insgesamt 300 Fliegen genutzt. Für jeden Experimentiermodus und je Stamm wurden jeweils 20 Fliegen benutzt. Um Laufexperimente durchzuführen mussten den Fliegen die Flügel um 2/3 gestutzt werden.

Das Buridanexperiment baut auf der Beobachtung von Individualverhalten auf, jeder Durchlauf wurde mit einer Fliege durchgeführt.

Die Experimente wurden in drei Modi durchgeführt. Hierbei wurden an der Arenawand zwei schwarze, exakt gegenüberliegende Streifen aus Pappe angebracht. Es wurden breite Streifen mit einer Breite von 20° und schmale mit einer Breite von 11° benutzt, der dritte Modus war komplett Streifenlos und somit ohne weitere Reize. Die Streifenbreite bezieht sich auf die Breite der gesamten Arenawand (360°) gemessen vom Mittelpunkt der Arena. 15 Minuten.

Software

Die Daten wurden mit Hilfe der speziell für diese Zwecke an der FU-Berlin entwickelten Buridantrackingsoftware aufgezeichnet und verarbeitet. Hierbei wurden alle Laufbewegungen in einzelne Dateien aufgezeichnet. Zusätzlich zur Trackingsoftware liegen viele verschiedene Skripte, welche in der Sprache R geschrieben worden sind zum Auswerten der Daten vor. Auf die Ausgaben dieser Skripte baut auch die Hauptkomponentenanalyse auf. Da aufgrund von fehlerhaften Skripten nicht alle Rohdaten benutzt werden konnte, werden im Folgenden nur die später in der HKA vorkommenden Variablen näher erläutert.

Sum_act
Die Variable sum_act beinhaltet die Zeit in Sekunden, während welcher die Fliege aktiv war. Dabei zählt als Aktivität die Zeit in welcher sich die Fliege bewegt und nicht in Pause befindet.

Median_act
Die Variable median_act gibt den Median einer durchschnittlichen Aktivität der Fliege in Sekunden an. Als Aktivitäten zählen dabei alle Bewegungen zwischen den Pausen.

Median_pause
Die Variable median_pause ist das Äquivalent zur Variable median_act, diese gibt den Median der Pausen einer Fliege in Sekunden an, wobei eine Pause mit einem Stillstand von mindestens 1/10s definiert ist.

Number_pause
Die Variable number_pause gibt die Gesamtzahl der Pausen einer Fliege an, wobei die Pause auch hier mit einem Stillstand von mindestens 1/10s definiert ist.

Speeds
Die Variable speeds gibt die durchschnittliche Geschwindigkeit einer Fliege in mm/s an.

Distance
Die Variable distance gibt die zurückgelegte Strecke einer Fliege in mm an.

Walks
Die Variable walks gibt die Anzahl der Läufe einer Fliege zwischen den Streifen an.

Die Rohdaten für alle Variablen lagen in Textdateien vor und wurden in „Microsoft Excel 2010„ importiert und in eine gemeinsame Tabelle mit 300 Datensätzen eingearbeitet. Hierbei wurden je Geno- und Streifentyp 20 Datensätze benutzt.

Hauptkomponentenanalyse

Für die Statistik wurde das Programm „XLSTAT“ in der Version 2010.5.04 von Addinsoft genutzt. Bei XLSTAT handelt es sich um ein Statistikprogramm für „Microsoft Excel 2010“, das Programm enthält auch ein Visualisierungsprogramm „XLSTAT-3DPLOT“ in der Version 7.2, mit welchem die Visualisierung der Hauptkomponentenanalyse erstellt wurde.

Die Hauptkomponentenanalyse (HKA) ist eine effiziente statistische Methode, welche es erlaubt Korrelationen zwischen einer großen Zahl von Variablen zu Analysieren und vor allem auch zu Visualisieren. Hierbei werden multidimensionale Datensätze (mehr als 3 Variablen) so kompensiert, dass diese ohne großen Varianzverlust grafisch in zwei oder drei Dimensionen dargestellt werden können. Ausgehend von einer großen Datenmenge mit sehr vielen verschiedenen Variablen besteht das Problem, dass man für die Visualisierung dieser Daten N-Dimensionen (N=Variablenzahl) mit M Datenpunkte hätte (M=Messpunkte), da man jedoch nur bis zu drei Dimensionen auf konventionelle Art und Weise visualisieren kann stößt man schnell an die Grenzen des Machbaren. In diesem Fall kann die HKA helfen, durch die Berechnung von Redundanzen und Korrelationen kann die Dimensionszahl verringert werden und die Werte neuen, synthetischen Achsen, den Hauptkomponenten zugeordnet werden. Dadurch, dass viele Variablen häufig miteinander verknüpft und somit redundant sind, werden neue Variablen mit verringerter Redundanz geschaffen[3].

Da man mit vielen unterschiedlichen Variablen und somit häufig auch verschiedenen Einheiten arbeitet, werden die Variablen zum Anfang standardisiert auf eine Varianz von Eins, damit man eine Korrelationsmatrix erstellen kann. Die Ähnlichkeit wird durch den Korrelationskoeffizienten nach Pearson hergestellt – es handelt sich um ein lineares Modell. Die Korrelation einzelner Variablen lässt sich auch grafisch darstellen, wenn man annimmt, dass die Variablen Vektoren gleichen Ursprungs sind, welche durch die Standardisierung eine vergleichbare länge haben, kann man mit Hilfe der Richtung und Entfernungen die Korrelation berechnen. Variablen, welche ein schmales Dreieck bilden und somit nah

einander Verlaufen (Abbildung 2 A zu B) sind positiv korreliert, solche die orthogonal (Abbildung 2 A zu C) zueinander verlaufen sin korreliert und Variablen die in entgegengesetzte Richtungen (Abbildung 2 A zu D) verlaufen negativ korreliert. Somit lässt sich aus dem Winkel zwischen den Vektoren, mit Hilfe von trigonometrischer Methoden auf die Korrelation schließen. Hierzu bildet man aus zwei Vektoren ein rechtwinkliges Dreieck, indem man sie durch in Lot verbindet. Die Strecke, welche durch den Schnittpunkt des Lotes (Abbildung 2, x) und einem der Vektoren entsteht ist über den Kosinus mit dem Winkel verknüpft. Die Länge dieser Strecke ist abhängig von dem Winkel (Abbildung 2 alpha) zwischen den Vektoren, durch die Standardisierung haben alle Vektoren dieselbe Länge und die durch das Lot abgetrennte Strecke von -1 bis +1 einnehmen. Bei dieser Strecke handelt es sich um den Pearson-Korrelationskoeffizienten, welcher das Korrelationsmaß zwischen den Variablen angibt.

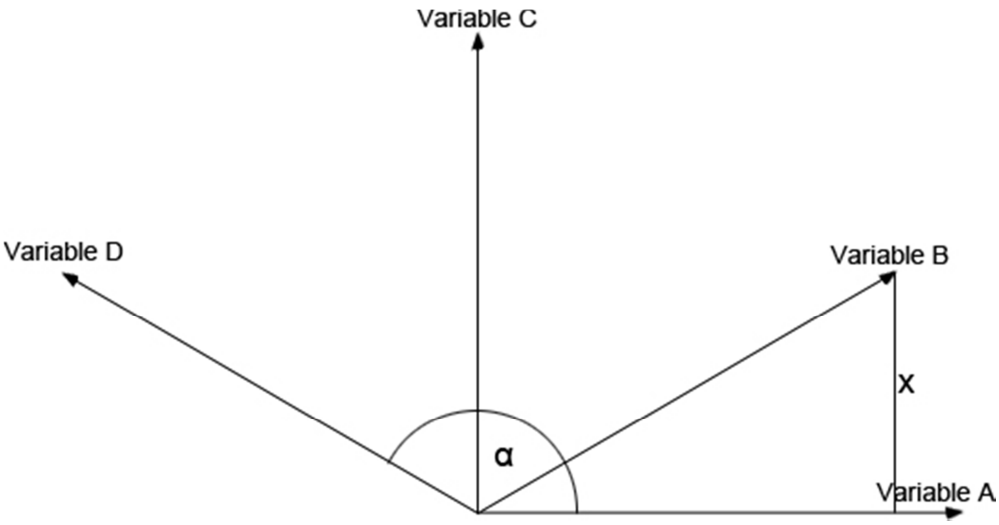


Abbildung 2 Modell der Korrelation von Variablen anhand geometrischer Methoden. Jede Variable entspricht einem standardisierten Vektor gleichen Ursprungs, durch das schlagen eines Lots (x) kann man die Korrelation zwischen zwei Variablen bestimmen.

Auf diese Weise kann man jedem Variablenpaar einen Korrelationskoeffizienten zuordnen. Diese Methode funktioniert auch für beliebig viele Variablen, bei der Hauptkomponentenanalyse wird nun versucht neue Achsen zu finden, bei welchen die Korrelation zwischen den Variablen möglichst hoch ist, bzw die aufsummierten Strecken der Vektoren möglichst groß sind. Diese Streckensumme nennt sich Eigenwert und wird aus den quadrierten Strecken summiert. Die Strecken müssen quadriert werden, da sie auch negative Werte annehmen können, bzw. es zu negativen Korrelationen kommen kann. Die Statistikprogramme schwenken hierfür die Achsen und durchlaufen je nach Variablenmenge viele der möglichen Kombinationen (in den Programmen wählbar) und ermitteln die idealen neuen Achsen, diese neuen optimalen Achsen sind die Hauptkomponenten. Es handelt sich dabei um abstrakte Achsen, welche ebenfalls mit den Variablenvektoren durch die beschriebene Methode korreliert werden können. Die Korrelation zwischen den Variablen und Hauptkomponenten werden Ladungen genannt. Je größer die Ladung einer Variable, desto höher der Einfluss dieser Variable auf die Hauptkomponente. Die Summe aller

Ladungen, welche auf einer Hauptkomponente zu finden sind, werden Eigenwerte genannt. Die Summe aller Eigenwerte zusammen entspricht der Summe der Hauptkomponenten, je größer der Eigenwert einer Hauptkomponente, desto größer ist der Anteil dieser Hauptkomponente an der Gesamtvarianz, die Eigenwerte können zu einfacheren Interpretation auch in Prozentangaben umgerechnet werden. Die Eigenwerte nehmen mit der Zahl der Hauptkomponenten von Hauptkomponenten zu Hauptkomponente ab, da die Erste bereits die bestmögliche Reduktion erreicht hat.
Auf die beschriebene Weise können nun weitere Hauptkomponenten berechnet werden, diese müssen jedoch unkorreliert, also orthogonal zu den anderen Hauptkomponenten sein um unabhängige Informationen darzustellen. Sodass am Ende jede neue Achse einen Satz von redundanten Informationen beschreibt, wobei man davon ausgeht, dass mit der steigenden Zahl der Komponenten nur noch Rauschen übrig bleibt. [4]

Das Ziel ist es also Achsen zu extrahieren, welche möglichst viel der Varianz enthalten, jede weitere Achse enthält den möglichst größten Teil der Restvarianz.

Um nun die Werte in ein Koordinatensystem übertragen zu können und somit eine grafische Auswertung zu ermöglichen werden die Werte jeder Variable mit ihrer Ladung auf der zugehörigen Hauptkomponente multipliziert.

Die vorliegenden Berechnungen der Korrelationsmatrix wurden mit der Methode nach Pearson durchgeführt, welcher die Korrelationen auch einem Signifikanztest mit einem Signifikanzniveau von 5% unterzogen werden.

Alle Grafiken, welche aus der HKA entstanden sind wurden mit XLSTAT-3DPLOT ausgehend von den einzelnen Koordinaten erstellt, hierbei wurden jeweils zwei Versionen eine in zwei (Hauptkomponenten 1 und 2) und eine in drei (Hauptkomponenten 1 bis 3) Dimensionen erstellt. Um eine bessere Vergleichbarkeit und Übersicht zu schaffen wurden des Weiteren gruppierte Grafiken erstellt, hierfür wurde der Median aller Einzelkoordinaten berechnet und als Koordinate benutzt.

Resultate

Um ein vollständiges Bild liefern zu können werden in diesem Abschnitt zuerst die Resultate der vorhergehenden Projektarbeit angeschnitten um diese später im Diskussionsteil besser miteinander vergleichen zu können, hierbei wurden die Ausgaben nach Aktivität, Pausen und Laufparameter unterteilt. Die Grafiken zeigen alle drei Streifentypen eines Genotypen im direkten Vergleich.

<u>Aktivität</u>

Die Aktivität wird auf zwei Weisen gemessen, als komplette Aktivitätszeit, diese beschreibt die Gesamtaktivitätszeit in Sekunden während des Versuches und als mittlere Aktivitätszeit, diese beschreibt die Dauer einer einzelnen Aktivität, zwischen zwei Pausen. Die Werte sind Mittelwerte aller gemessenen Fliegen. Aufgrund fehlender Optionen zur Einstellung der Skala, unterscheiden sich die Dimensionen in den Grafiken in Abhängigkeit zu den Werten.

Total activity time

Abbildung 3 zeigt die Gesamtaktivität von cstp in s, die meiste Aktivität wurde im schmalstreifigen Modus mit ca. 500s gemessen.

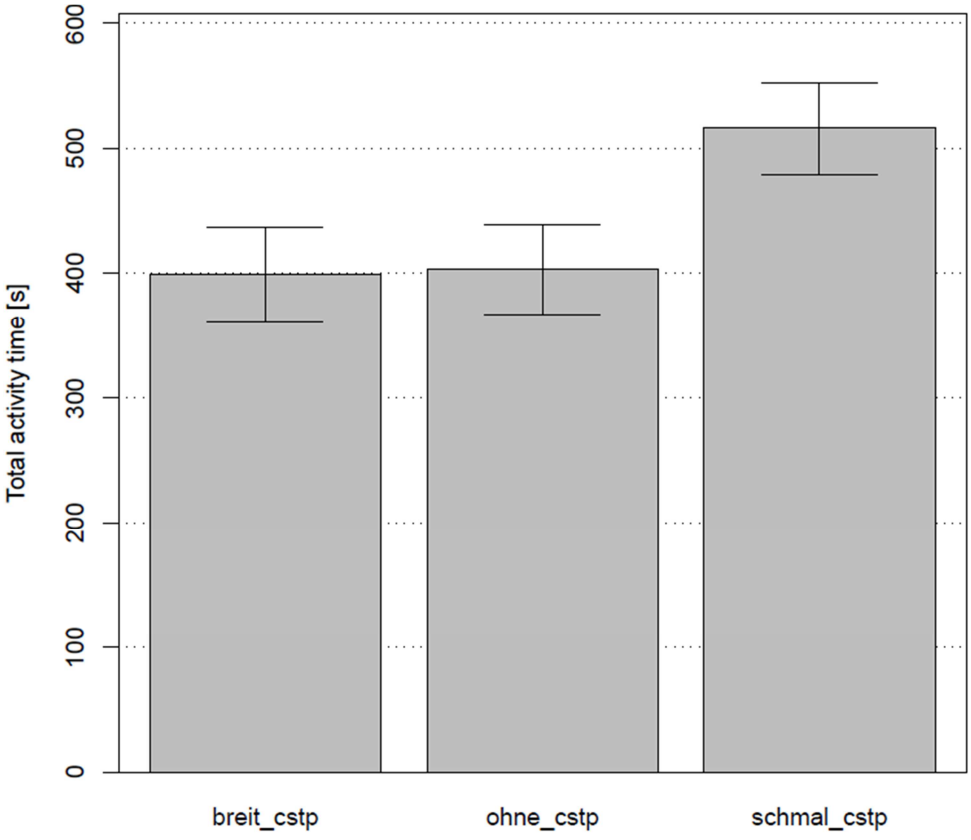


Abbildung 3 grafische Darstellung der Gesamtaktivität von cstp in s.. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Die Aktivität bei schmalen Streifen ist mit ca. 500s am höchsten.

Abbildung 4 zeigt die Gesamtaktivität von rsh in s, die meiste Aktivität wurde im schmalstreifigen Modus mit ca. 350s gemessen.

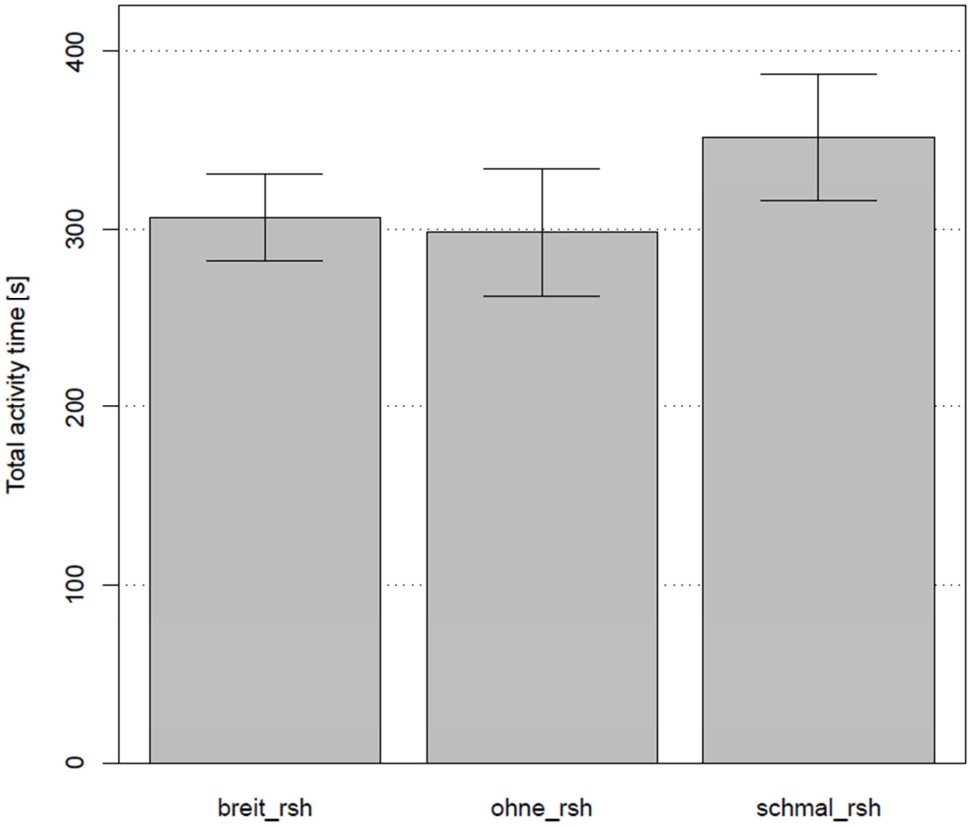


Abbildung 4 grafische Darstellung der Gesamtaktivität von rsh in s. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Die Aktivität bei schmalen Streifen ist mit ca. 350s am höchsten.

Abbildung 5 zeigt die Gesamtaktivität von rut2080 in s, alle Streifemodi haben mit ca. 450s dieselbe Aktivität.

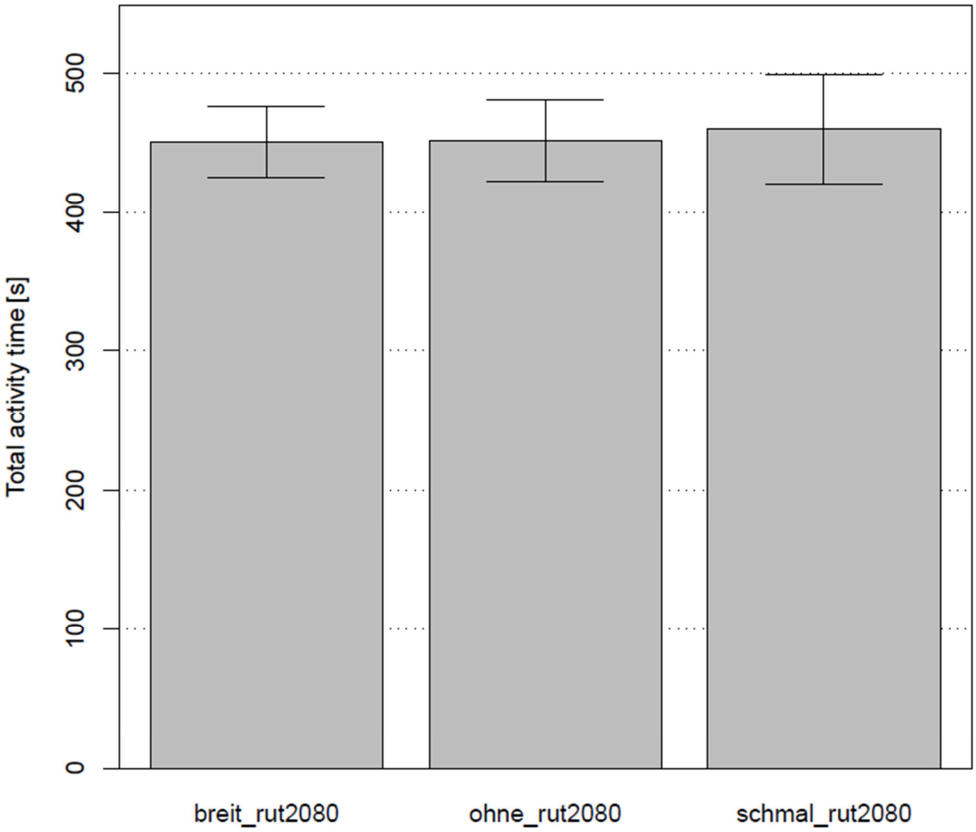


Abbildung 5 grafische Darstellung der Gesamtaktivität von rut^{2080} in s. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Alle Streifenmodi haben eine ähnliche Aktivität.

Abbildung 6 zeigt die Gesamtaktivität von dnc in s, alle Streifemodi haben mit ca. 330s dieselbe Aktivität.

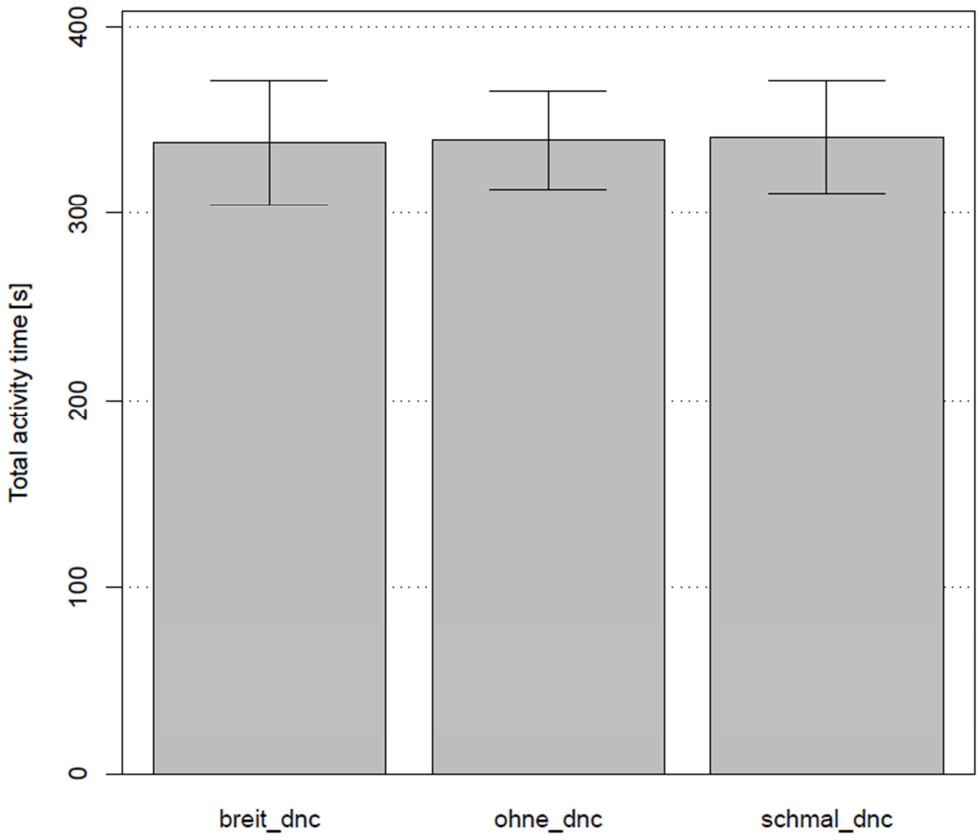


Abbildung 6 grafische Darstellung der Gesamtaktivität von dnc in s. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Alle Streifenmodi haben eine ähnliche Aktivität.

Abbildung 7 zeigt die Gesamtaktivität von rut1 in s, alle Streifemodi haben mit ca. 270s dieselbe Aktivität, es ist eine Tendenz zur höchsten Aktivität beim streifenlosen Modus zu erkennen, durch die hohe Standardabweichung kann jedoch keine sichere Aussage getroffen werden.

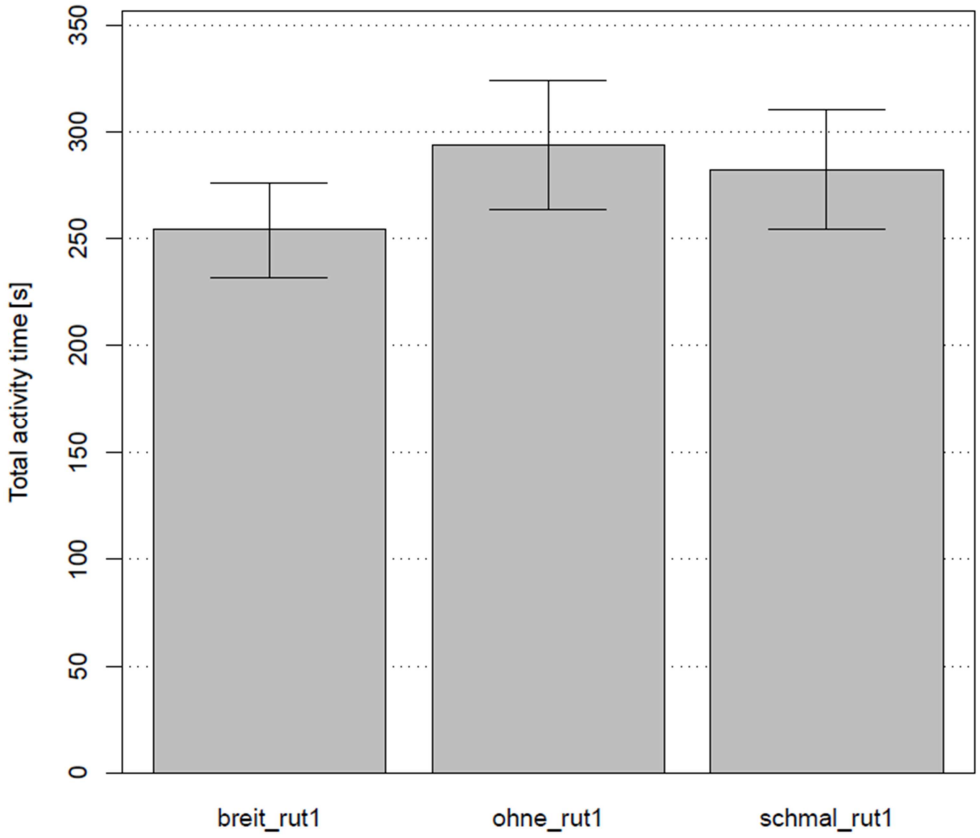


Abbildung 7 grafische Darstellung der Gesamtaktivität von rut[1] in s. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Durch die Standardabweichung ist kein aktiverer Streifenmodus erkennbar.

Mean activity time

Die folgenden Grafiken zeigen die mittlere Aktivitätszeit je Aktivität.

Abbildung 8 zeigt die mittlere Aktivität von cstp in s, der schmale Streifenmodus zeigt mit ca. 1s die meiste Aktivität.

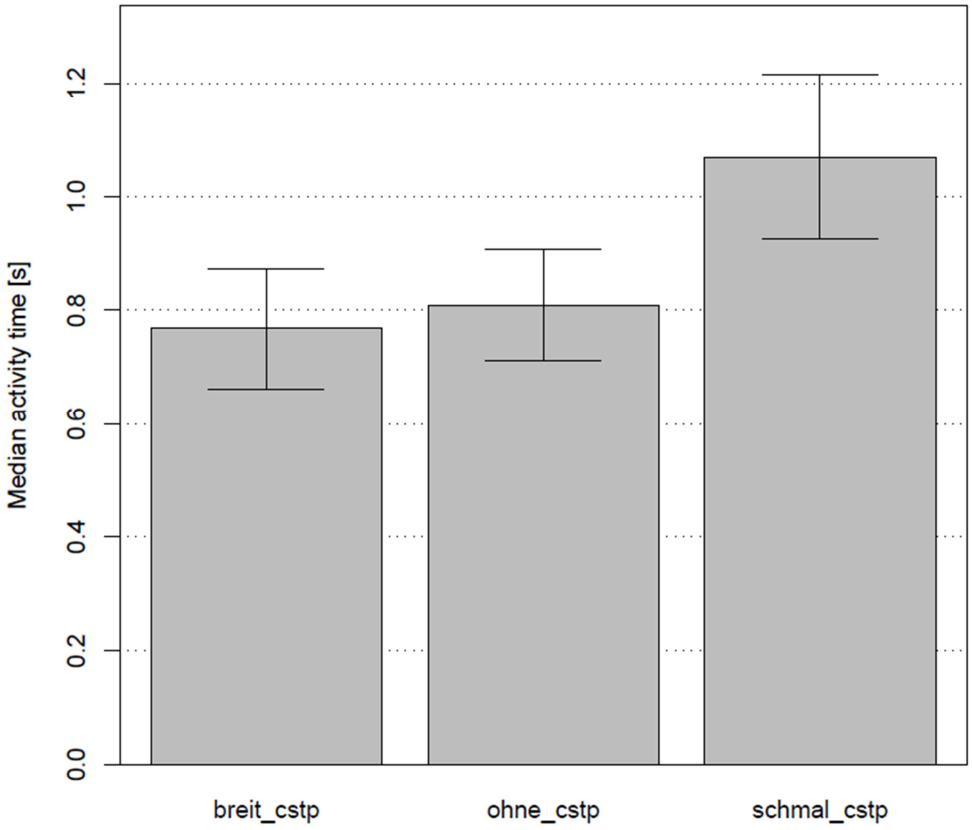


Abbildung 8 grafische Darstellung der Einzelaktivitäten von cstp in s. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Median inklusive Standardabweichung. Im schmalen Streifenmodus ist mit ca. 1s die meiste Aktivität zu beobachten.

Abbildung 9 zeigt die mittlere Aktivität von rsh in s, der schmale Streifenmodus zeigt die mit 0,8s meiste Aktivität, wobei diese durch die hohe Standardabweichung des streifenlosen Modus nicht ohne Zweifel ist, da auch der streifenlose Modus in ähnliche höhen hineinragt.

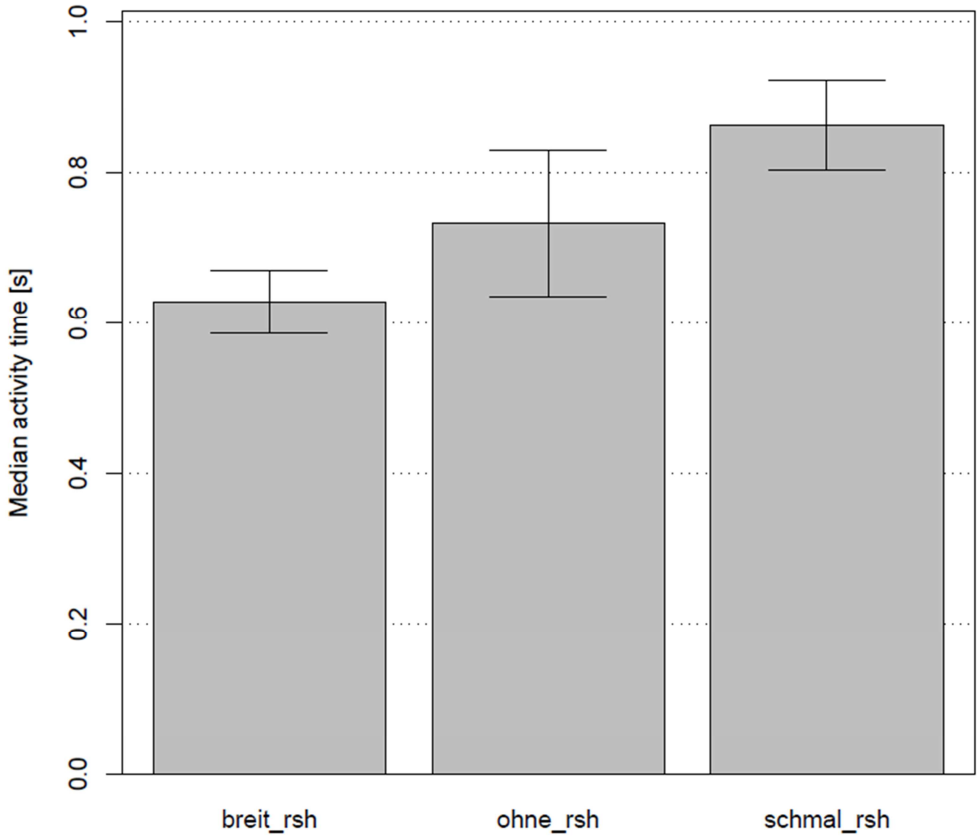


Abbildung 9 grafische Darstellung der Einzelaktivitäten von rsh in s. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Im schmalen Streifenmodus ist mit ca. 0.8s die höchste Aktivität zu beobachten, wobei durch die hohe Standardabweichung kein großer Unterschied zum streifenlosen Modus erkennbar ist.

Abbildung 10 zeigt die mittlere Aktivität von rut1 in s, durch die hohe Standardabweichung kann man keine höchste Aktivität festlegen, diese liegt bei allen Modi um 0,55s.

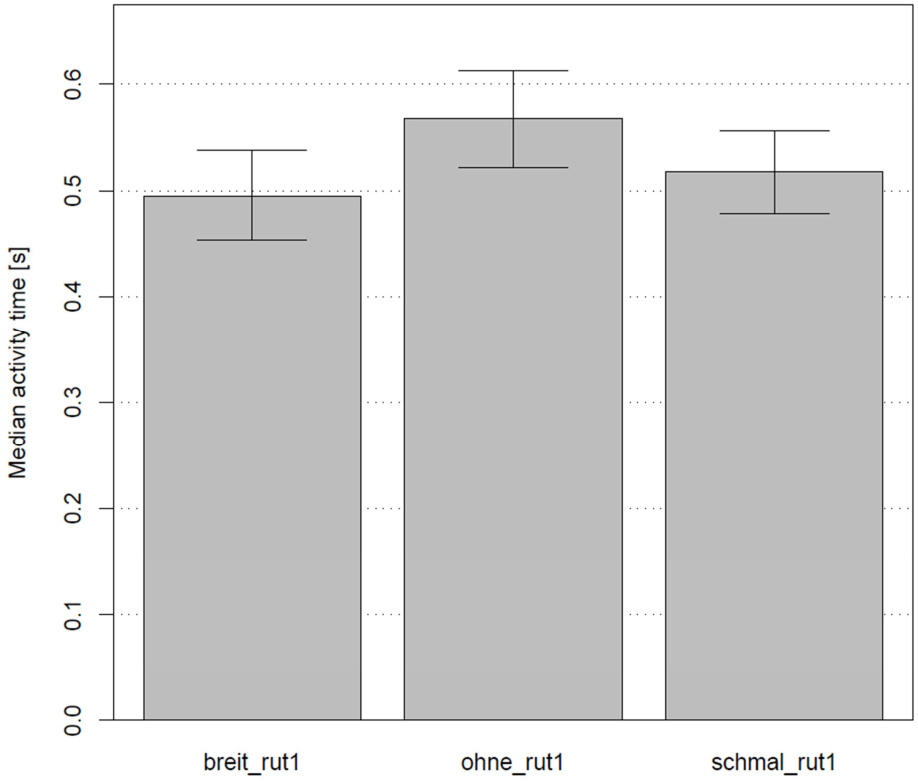


Abbildung 10 grafische Darstellung der Einzelaktivitäten von rut[1] in s. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Durch die Standardabweichung sind alle Streifenmodi mit ähnlichen Aktivitäten besetzt.

Abbildung 11 zeigt die mittlere Aktivität von dnc in s, durch die hohe Standardabweichung kann man keine höchste Aktivität festlegen, diese liegt bei allen Modi um 0,75s.

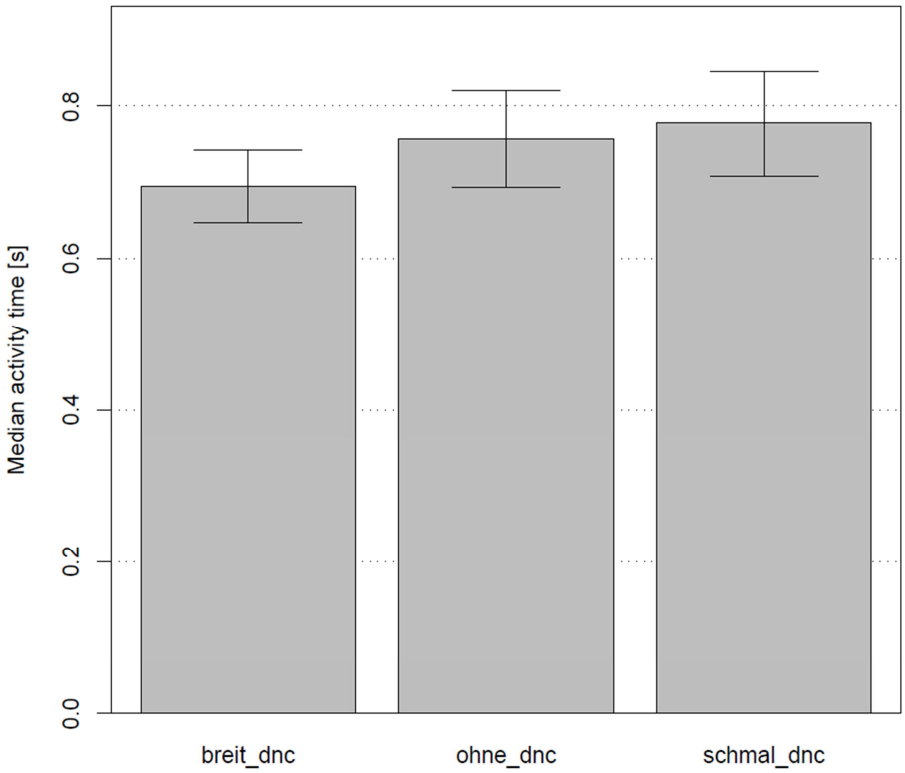


Abbildung 11 grafische Darstellung der Einzelaktivitäten von dnc in s. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Durch die Standardabweichung sind alle Streifenmodi mit ähnlichen Aktivitäten besetzt.

Abbildung 12 zeigt die mittlere Aktivität von rut2080 in s, durch die hohe Standardabweichung kann man keine höchste Aktivität festlegen, diese liegt bei allen Modi um 1,1s.

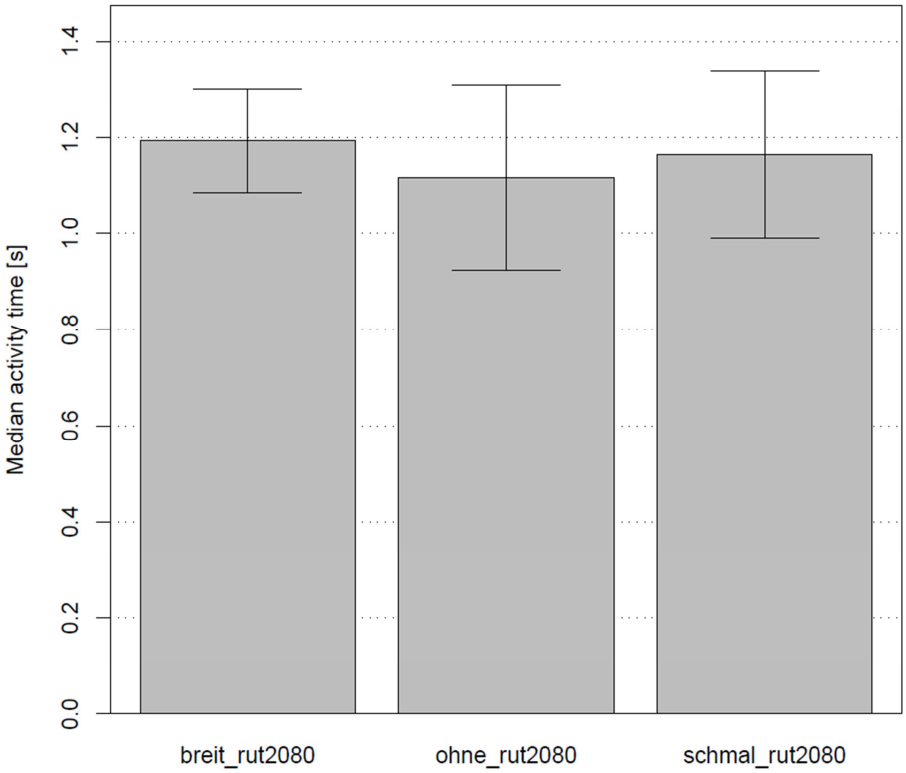


Abbildung 12 grafische Darstellung der Einzelaktivitäten von rut^{2080} in s. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Durch die Standardabweichung sind alle Streifenmodi mit ähnlichen Aktivitäten besetzt.

<u>Pausen</u>

Bei den Pausen wird die durchschnittliche Zeit einer Pause in s und die Gesamtzahl der Pausen je Durchgang gemessen. Die Pausenlänge wurde mir 1/10s definiert.

Mean pause time

Abbildung 13 zeigt die mittlere aller Stämme in s, die Pausenzeit ist bei allen Fliegen sehr ähnlich und liegt bei ca. 1,4s.

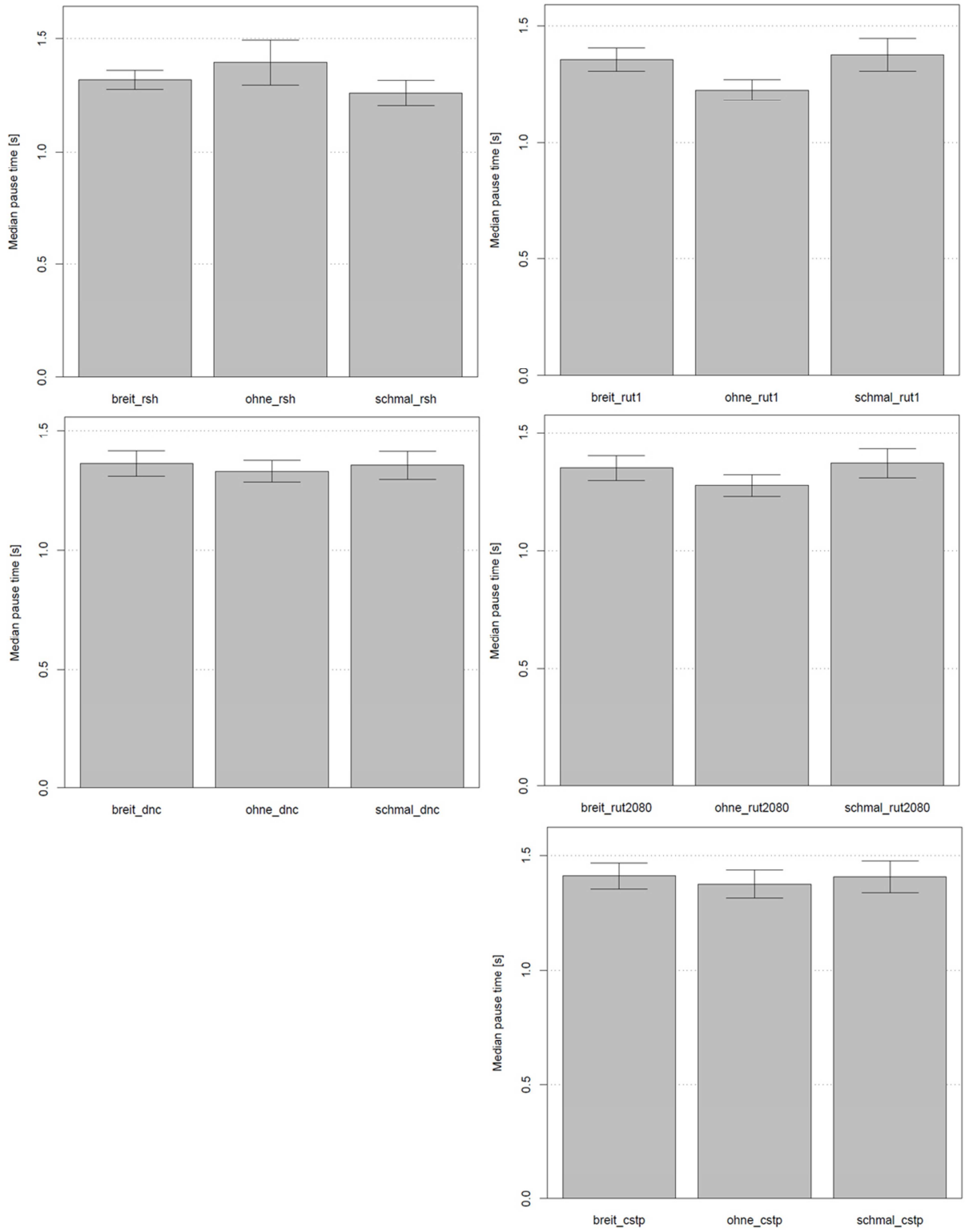


Abbildung 13 durchschnittliche mittlere Pausenzeit in s für alle Stämme. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Die Pausenzeit ist bei allen Genotypen und Streifenmodi bei ca. 1,4s

Number of pauses

Im Folgenden wird die gemittelte Zahl an Pausen je Durchgang dargestellt, mit eingezeichneter Standardabweichung.

Abbildung 14 zeigt die mittlere Pausenzahl von cstp in , diese ist im streifenlosen Modus mit ca. 175 Pausen am höchsten.

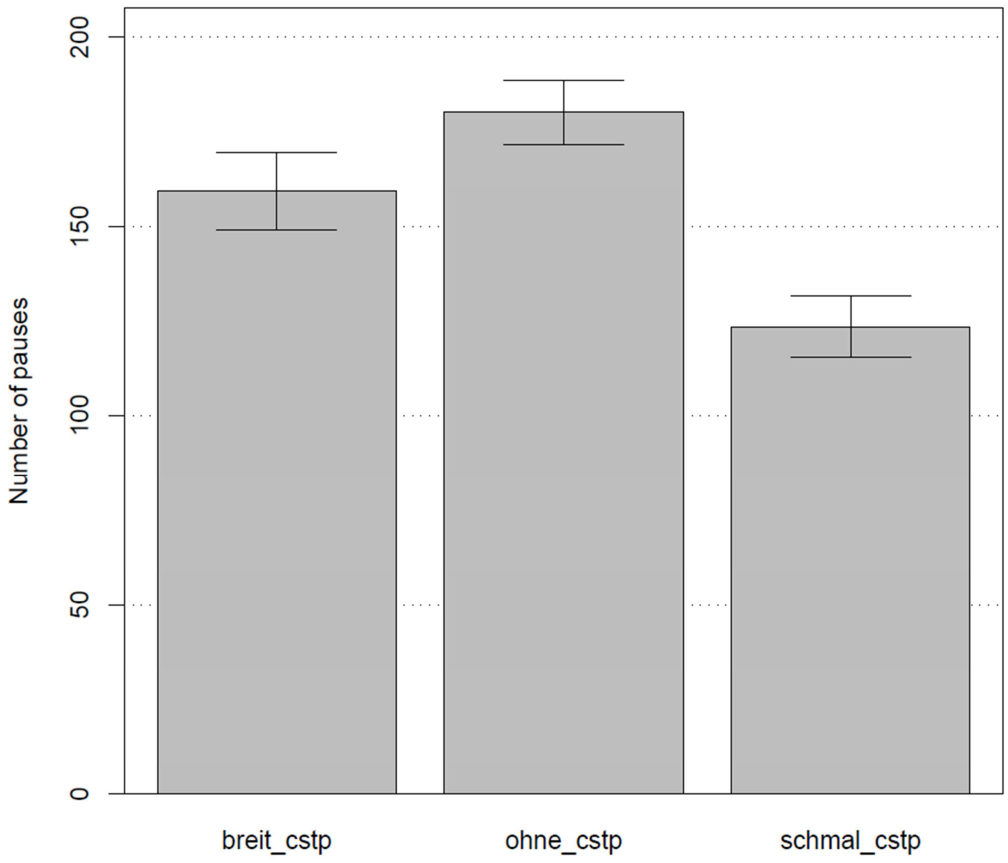


Abbildung 14 durchschnittliche Anzahl der Pausen je Durchgang für cstp. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Die Pasuenzahl ist im streifenlosen Modus mit ca. 175 Pausen am höchsten.

Abbildung 15 zeigt die mittlere Pausenzahl von rut2080 und rut1. Durch die hohe Standardabweichung ist kaum ein Unterschied feststellbar, die Pausenzahl liegt bei ca. 200 bei rut1 und bei rut2080 etwas darunter.

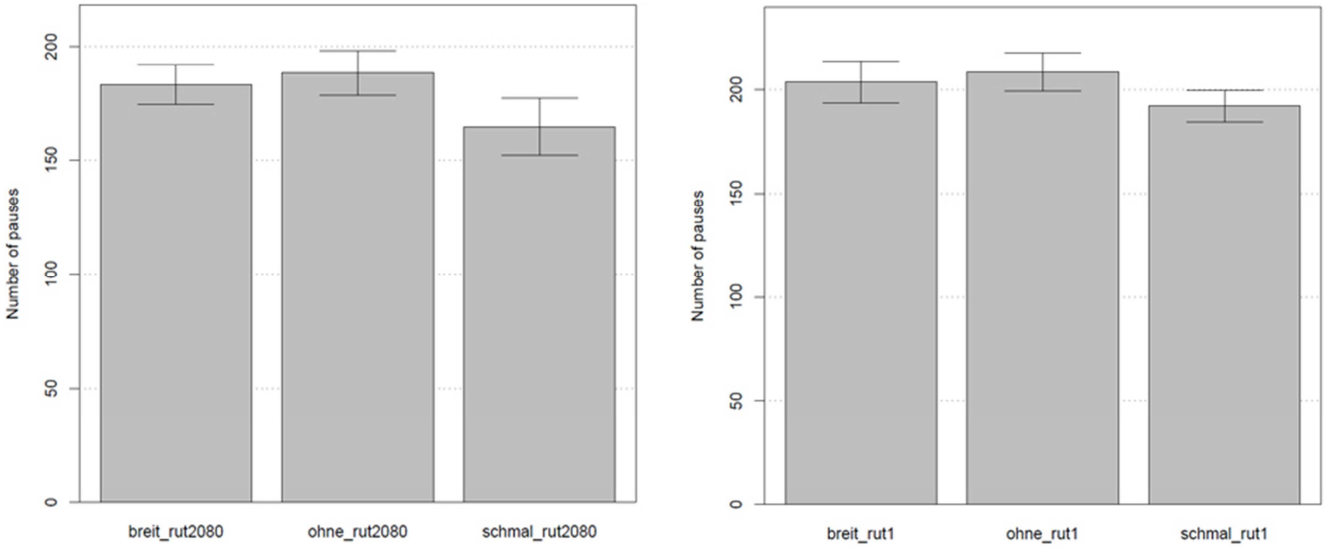


Abbildung 15 durchschnittliche Anzahl der Pausen je Durchgang für rut[1] und rut[2080]. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Die Pausenzahl ist bei beiden Genotypen zwischen den Streifenmodi kaum zu unterscheiden, auch wenn bei beiden eine Tendenz von ca. 200 Pausen beim streifenlosen Typ erkennbar ist.

Abbildung 16 zeigt die mittlere Pausenzahl von dnc und rsh. Bei dnc zeigt der streifenlose Modus mit ca. 180 Pausen die meisten, bei rsh ist eine genaue Aussage aufgrund der Standardabweichung schwierig, breiter und streifenloser Modus streuen um 220 Pausen.

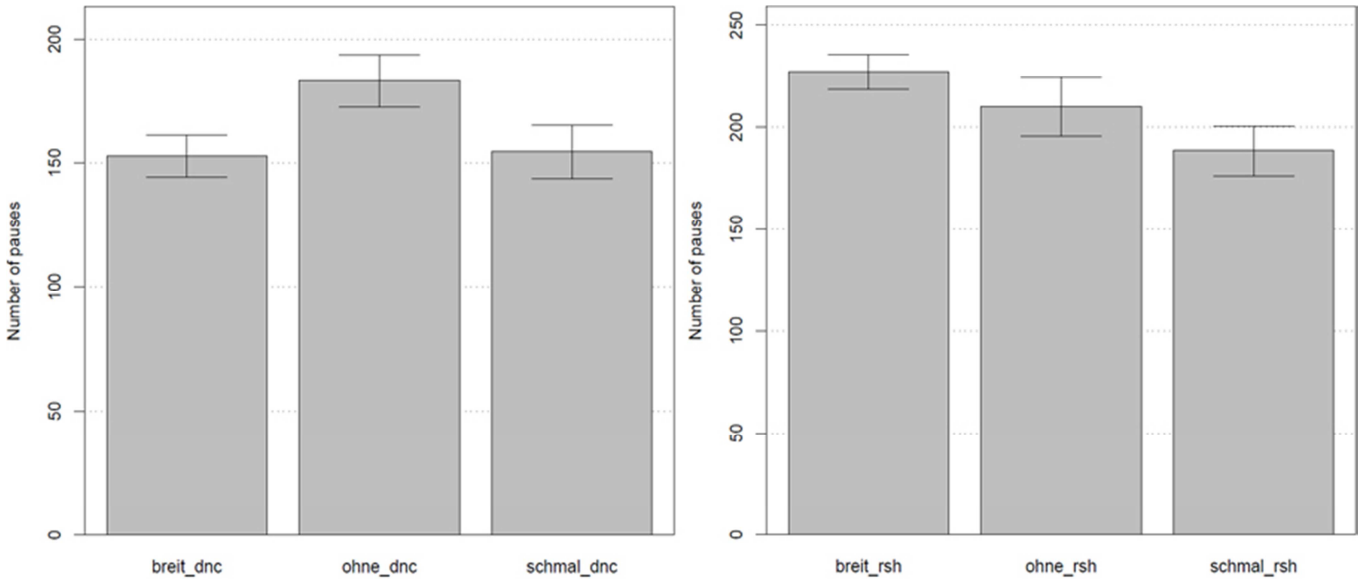


Abbildung 16 durchschnittliche Anzahl der Pausen je Durchgang für dnc und rsh. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. DNC zeigt mit ca. 180 Pausen im streifenlosen die meisten Pausen. Bei rsh ist eine Tendenz für die meisten Pausen beim breiten Modus erkennbar, welche jedoch durch die Standardabweichung in Frage gestellt werden kann.

Laufparameter

Ausgegeben werden Die Laufgeschwindigkeit, Laufentfernung und die Anzahl der Läufe zwischen den Streifen.

Speed plot

Der Speedplot stellt die durchschnittliche Geschwindigkeit je Lauf in mm/s inklusive Standardabweichung dar.

Abbildung 17 zeigt die mittlere Laufgeschindigkeit von rut2080 ,rut1, cstp und rsh. Die genauen Daten können Tabelle 1 entnommen werden, rut2080 zeigt in allen Modi die schnellste Geschwindigkeit.

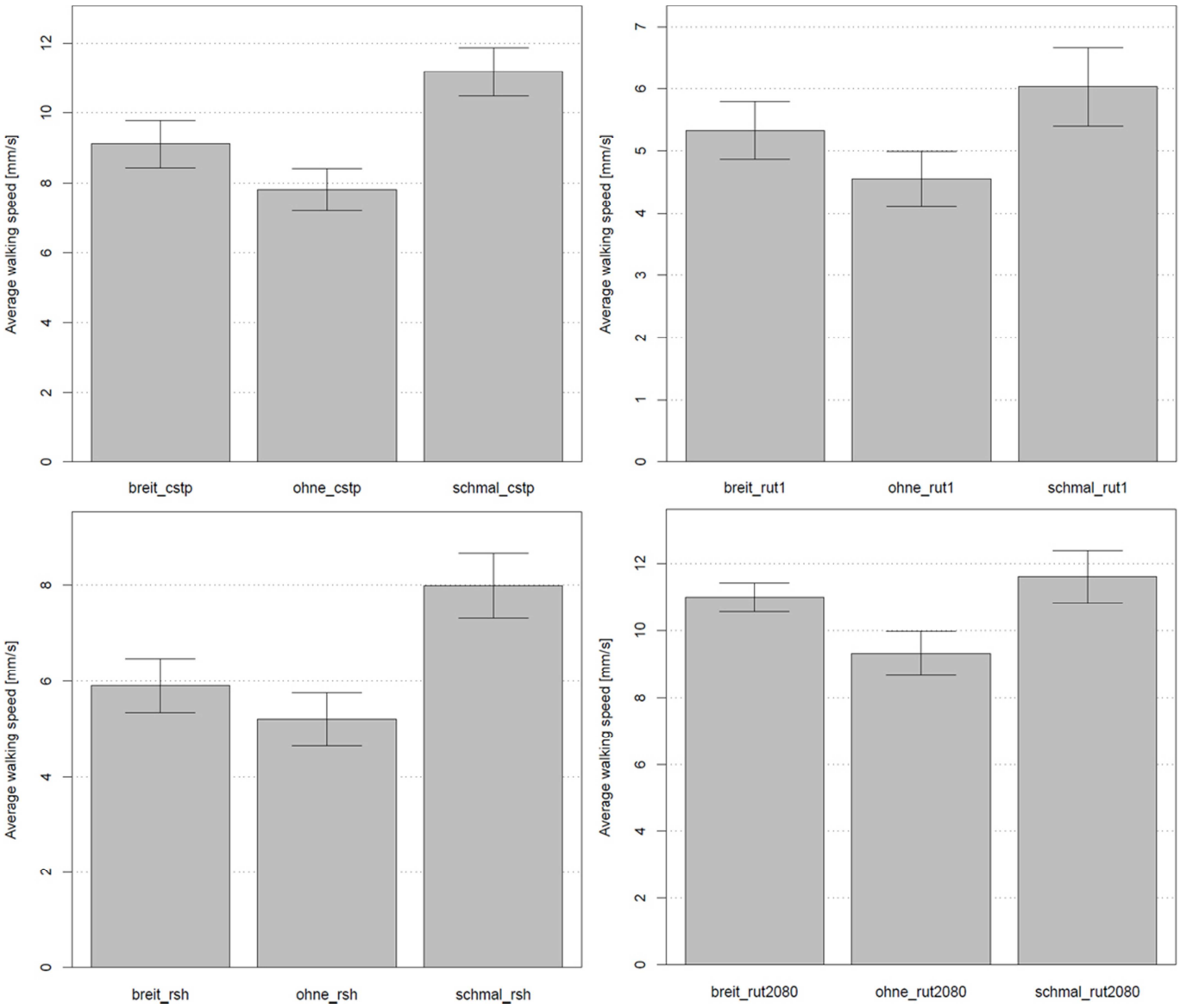


Abbildung 17 durchschnittliche Laufgeschwindigkeit je Durchgang für cstp, rut[1], rsh, und rut[2080]. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Bei allen Genotypen ist zu erkennen, dass bei schmalen Streifen die schnellste Laufgeschwindigkeit vorhanden war.

Die folgende Tabelle zeigt die Laufgeschwindigkeiten der zuvor genannten Stämme

Stamm	breit	ohne	schmal
rut[1]	5,2	4,5	6
cstp	9	7,8	11
rsh	6	5,2	8
rut[2080]	11	9,3	11,7

Tabelle 1 Mittlere Laufgeschwindigkeit der Stämme rut[1], cstp, rsh und rut[2080]

Abbildung 18 zeigt die mittlere Laufgeschwindigkeit von dnc, alle Modi zeigen eine Geschwindigkeit von ca. 6mm/s.

Abbildung 19 zeigt die mittlere Laufstrecke von rut2080 ,rut1, cstp und rsh. Die genauen Daten können Tabelle 2 entnommen werden, rut2080 läuft im breiten und streifenlosen, cstp im schmalen Modus am weitesten von allen Stämmen.

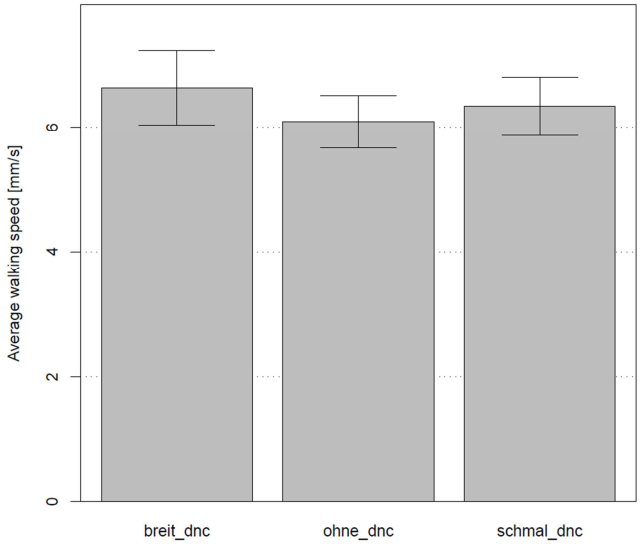


Abbildung 18 durchschnittliche Laufgeschwindigkeit je Durchgang für dnc. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Alle Streifenmodi zeigen eine ähnliche Laufgeschwindigkeit.

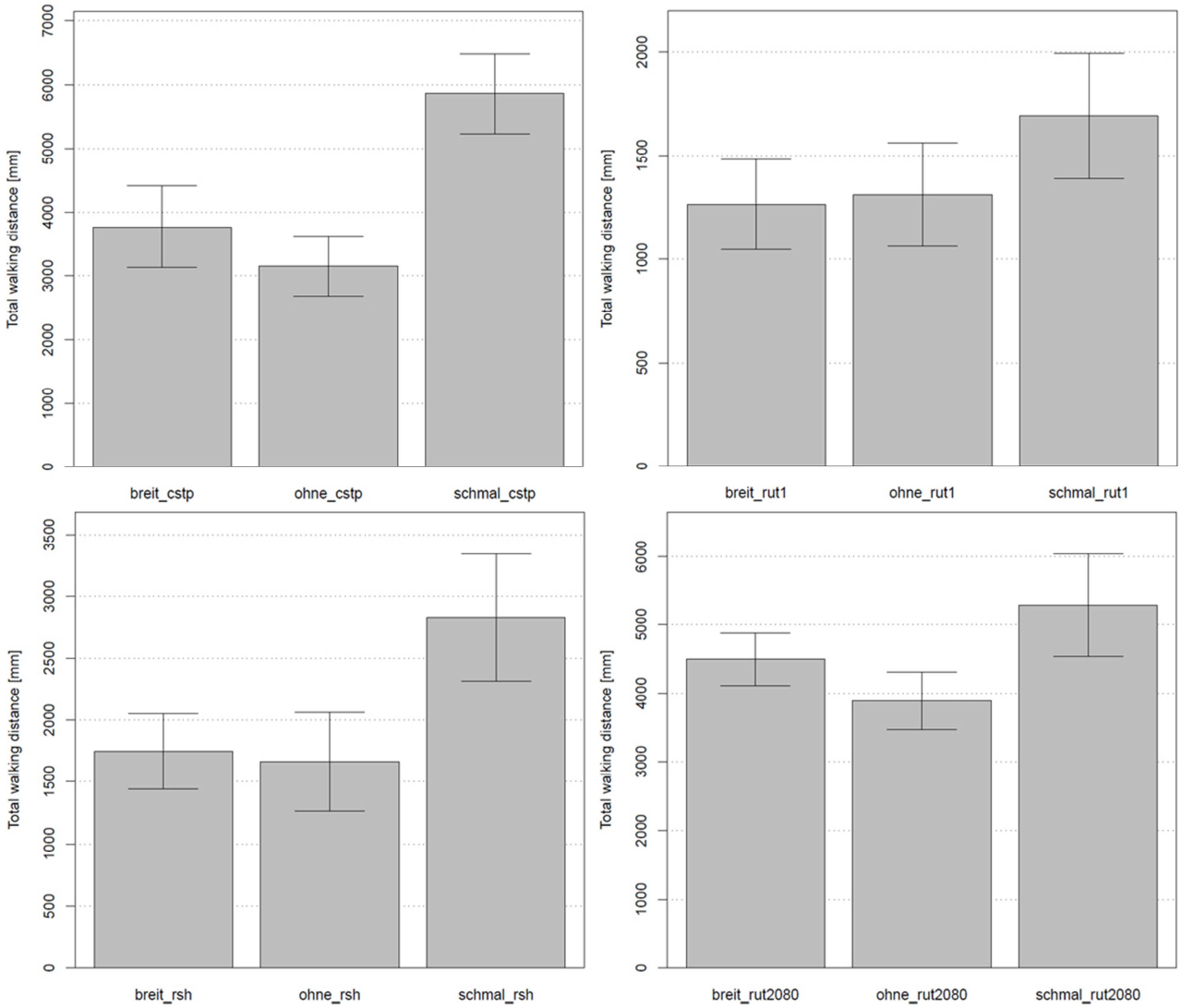


Abbildung 19 durchschnittliche Laufdistanz in mm je Durchgang für cstp, rut[1], rsh, und rut[2080]. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Bei allen Streifenmodi ist im schmalen Streifenmodus eine deutlich größere Laufstrecke zu erkennen.

Die folgende Tabelle zeigt die Laufentfernung der ersten Gruppe:

Stamm	breit	ohne	schmal
rut[1]	1250	1250	1650
cstp	3800	3100	5800
rsh	1700	1650	2800
rut[2080]	4500	3950	5200

Tabelle 2 Mittlere Laufentfernung in mm je Durchgang der Stämme rut[1], cstp, rsh und rut[2080]

Abbildung 20 zeigt die mittlere Laufstrecke von dnc, alle Streifenmodi haben eine Entfernung von ca. 2000mm zurückgelegt.

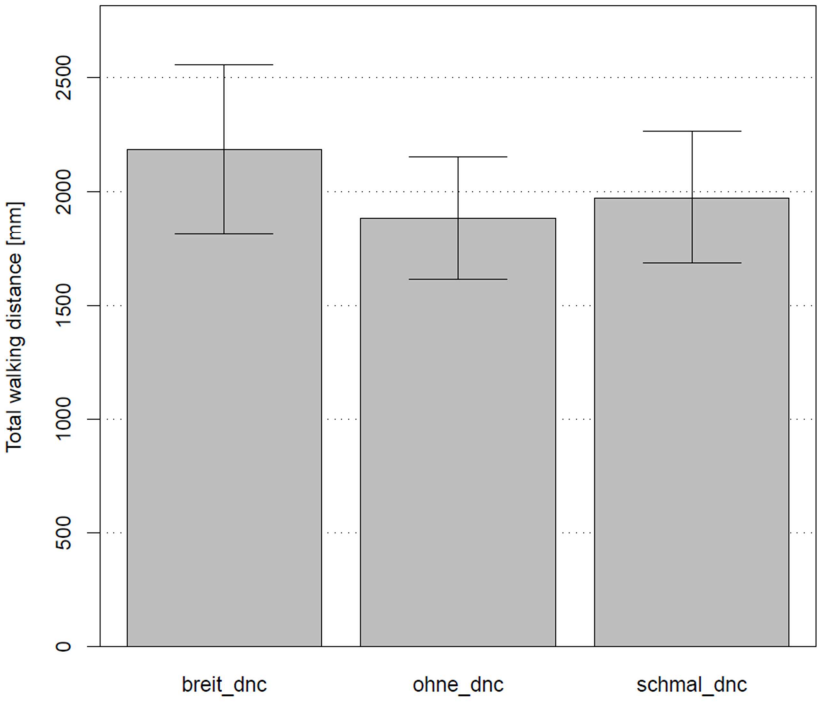


Abbildung 20 durchschnittliche Laufdistanz in mm je Durchgang für dnc. Jeder Balken stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge im Mittelwert inklusive Standardabweichung. Die Laufstrecken unterscheiden sich kaum zwischen den Modi.

Walks between stripes

Zuletzt wird die Anzahl der Läufe zwischen den Streifen kalkuliert, auch hier im Durchschnitt je Durchgang mit Standardabweichungen (in den Grafiken).

Stamm	breit	schmal
rut[1]	3	5
cstp	10	18
rsh	5	10
rut[2080]	19	23

Tabelle 3 Anzahl der Läufe zwischen den Streifen je Durchgang für die Stämme rut[1], cstp, rsh und rut[2080]

Um alle gemessenen Daten aufeinander beziehen zu können im Folgenden noch die Occupancy- und Transitionplots, diese werden zwar auch in der Diskussion Platz finden, wurden jedoch nicht in der Hauptkomponentenanalyse behandelt, da es sich um eine grafische Interpretation der Rohdaten handelt.

Occupancy- und Transitionplots

Occupancyplot
In dieser Darstellungsmöglichkeit werden die Laufbewegungen aller Fliegen grafisch dargestellt, je länger eine Fliege an einer Stelle war, desto weiter im roten liegt diese, je kürzer desto weiter im blauen. Orte an denen die Fliegen nicht waren sind weiß. Es handelt sich um eine Aufenthaltsgrafik.

Transitionplot
Als zweite Ausgabe kommt es zu einer ähnlichen Darstellung, die Farbbedeutung gleicht der aus den Occupancyplots, nur dass es sich hierbei um Durchlaufwahrscheinlichkeiten handelt. Es wird also festgestellt wie häufig die Fliege einen Punkt durchlaufen hat. Auch hierzu finden sich Details in der Dokumentation.

Als letztes werden die Läufe einzelner Fliegen dargestellt, hierbei ändert sich Farbe der Linie bei jedem Lauf zwischen den Streifen.

cstp

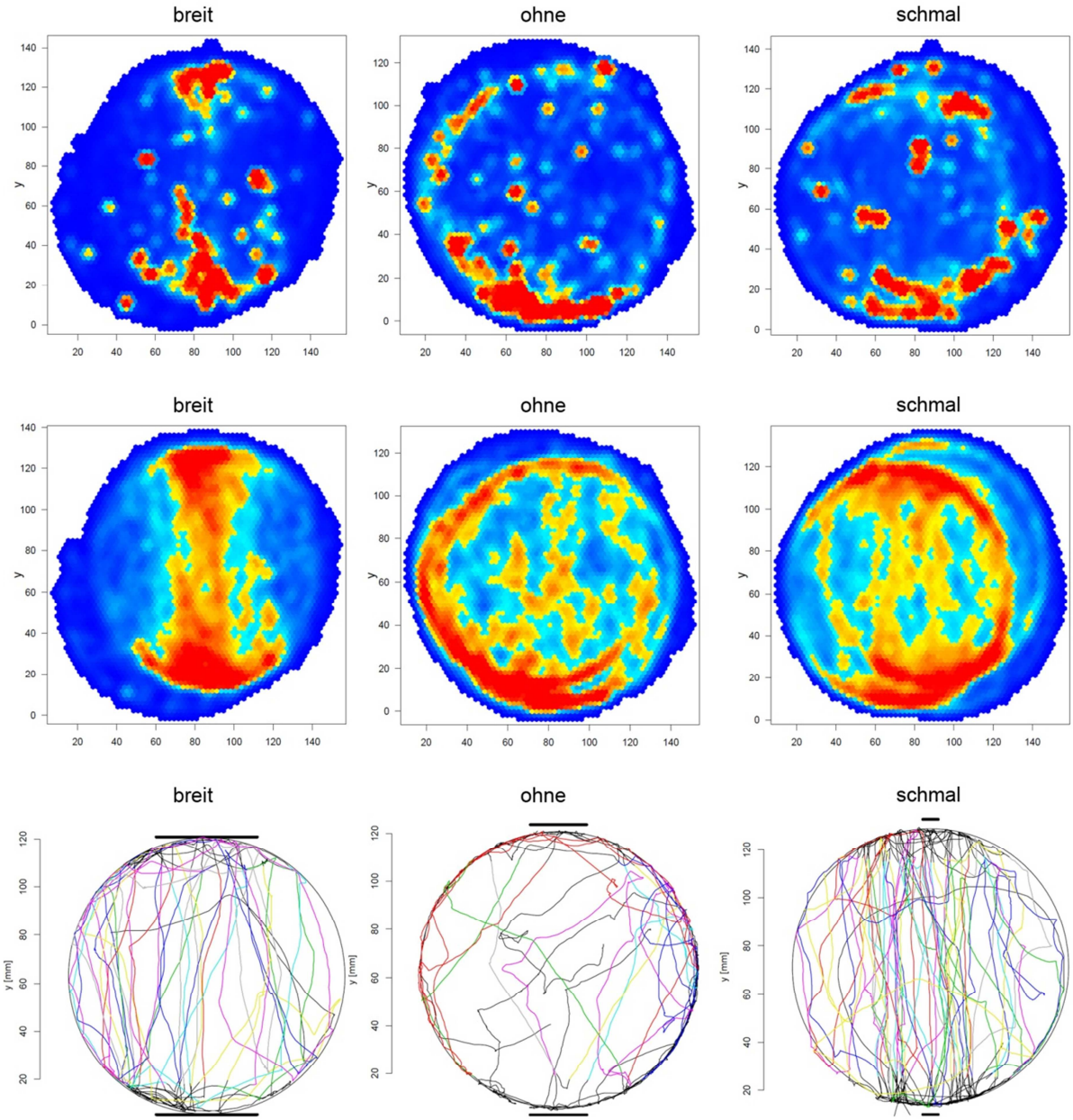


Abbildung 21 Transition-, Occupancy- und Einzelplots für die drei Experimentiermodi des Stammes cstp. Jeder Plot stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge (der Einzelplot zeigt das Laufmuster einer Fliege). Die X- und Y- Achsen stellen die Dimensionen der Arena in mm dar. Die Aktivität wird durch die Farbe ausgedrückt, von wenig (blau) bis viel (rot) Aktivität. Beim Einzelplot sind die exakten Laufbewegungen einer einzelnen Fliege zu sehen, die Farbänderungen stehen für Läufe zwischen den Streifen.

dnc

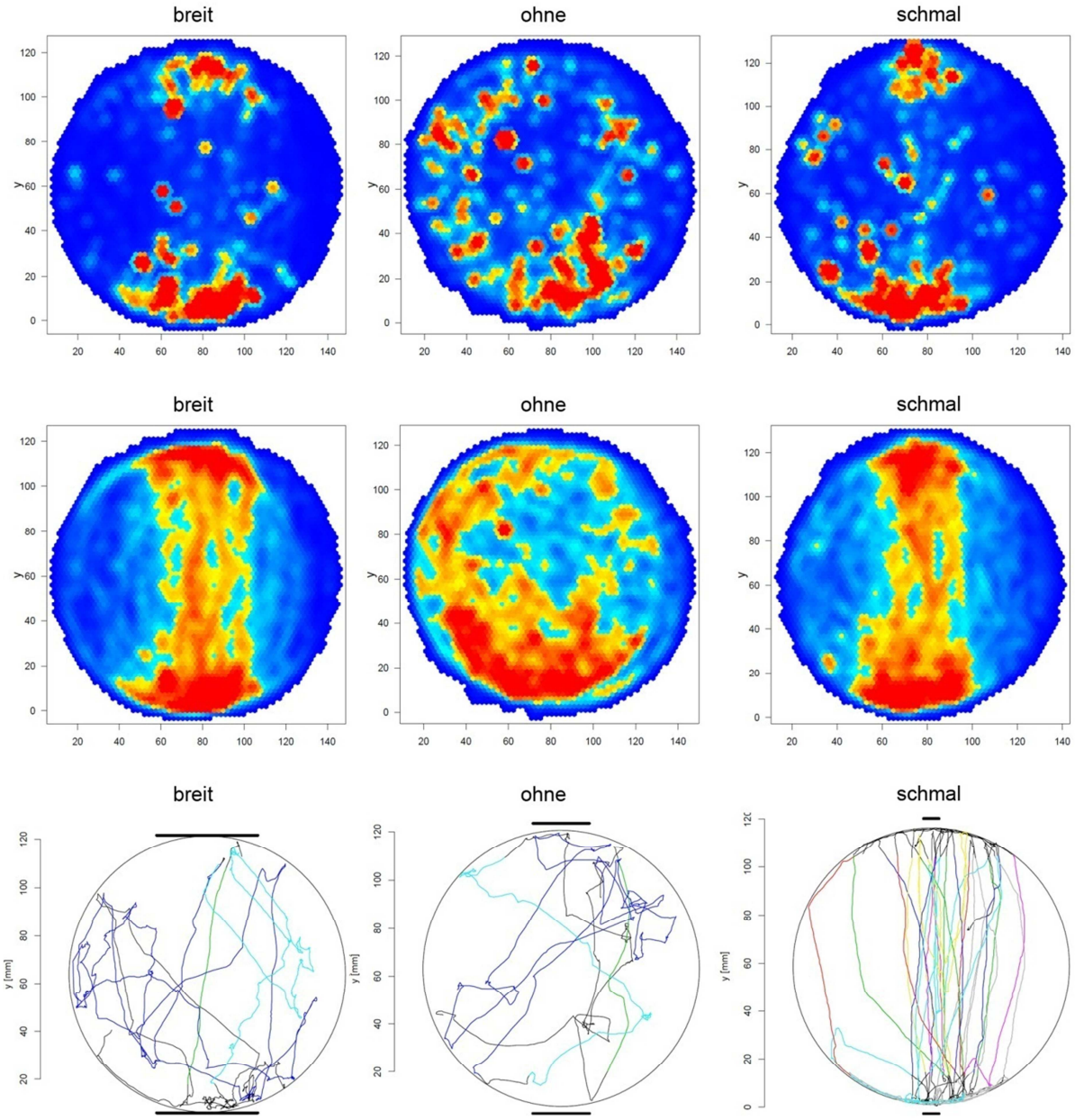


Abbildung 22 Transition-, Occupancy- und Einzelplots für die drei Experimentiermodi des Stammes dnc. Jeder Plot stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge (der Einzelplot zeigt das Laufmuster einer Fliege). Die X- und Y- Achsen stelle

rsh

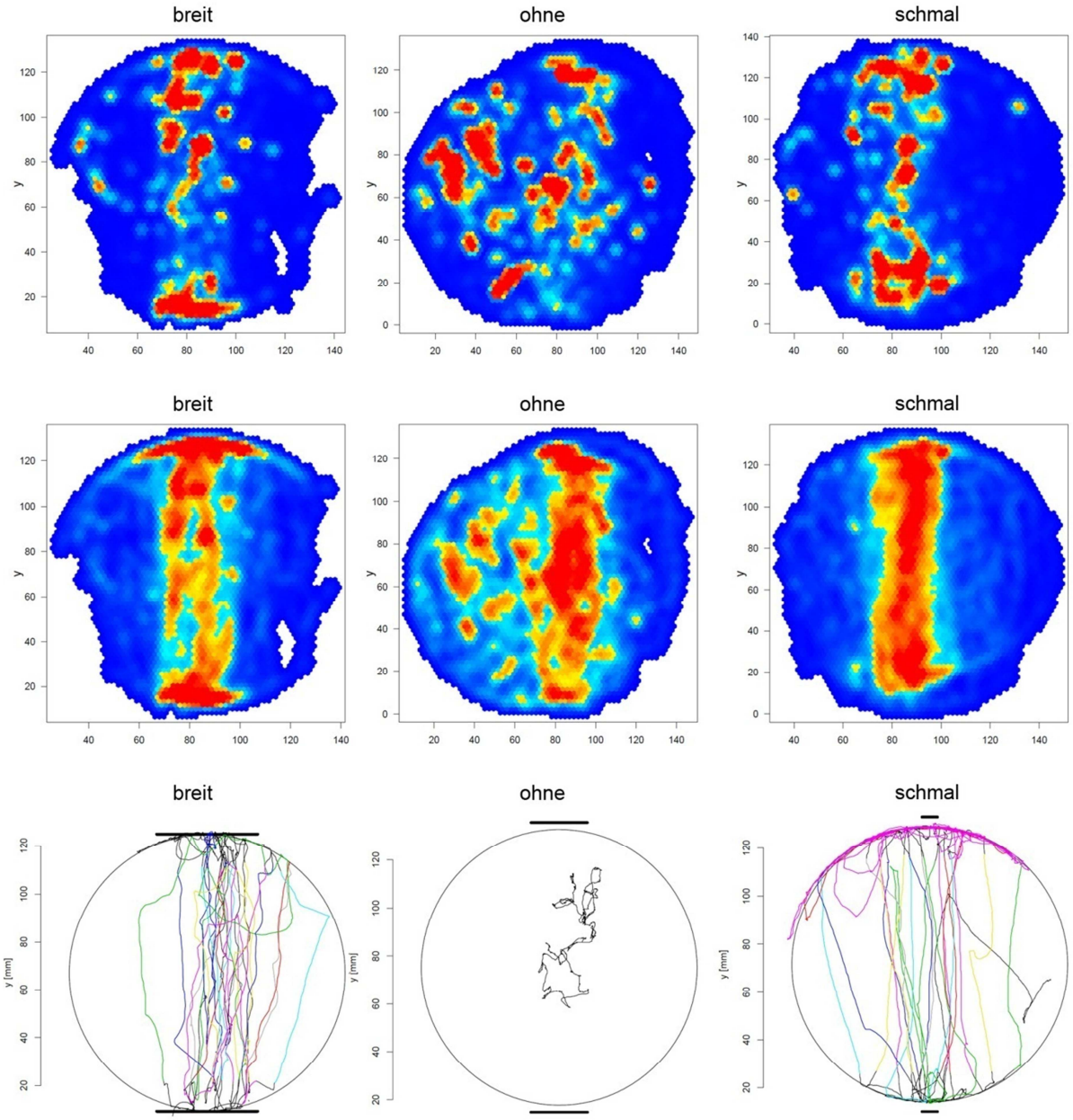


Abbildung 23 Transition-, Occupancy- und Einzelplots für die drei Experimentiermodi des Stammes rsh. Jeder Plot stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge (der Einzelplot zeigt das Laufmuster einer Fliege). Die X- und Y- Achsen stelle

rut[1]

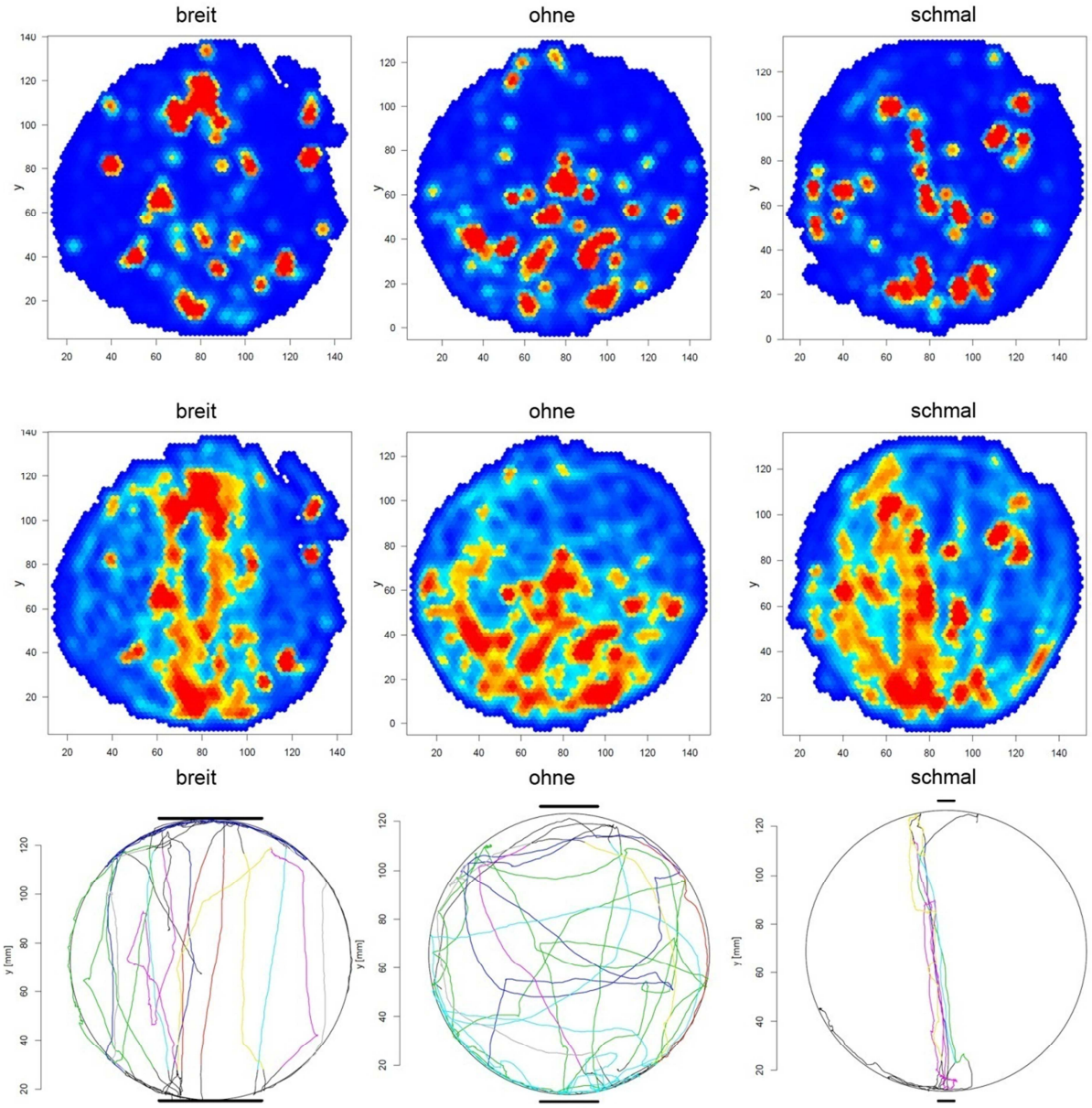


Abbildung 24 Transition-, Occupancy- und Einzelplots für die drei Experimentiermodi des Stammes rut[1]. Jeder Plot stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge (der Einzelplot zeigt das Laufmuster einer Fliege). Die X- und Y- Achsen stelle

rut^{2080}

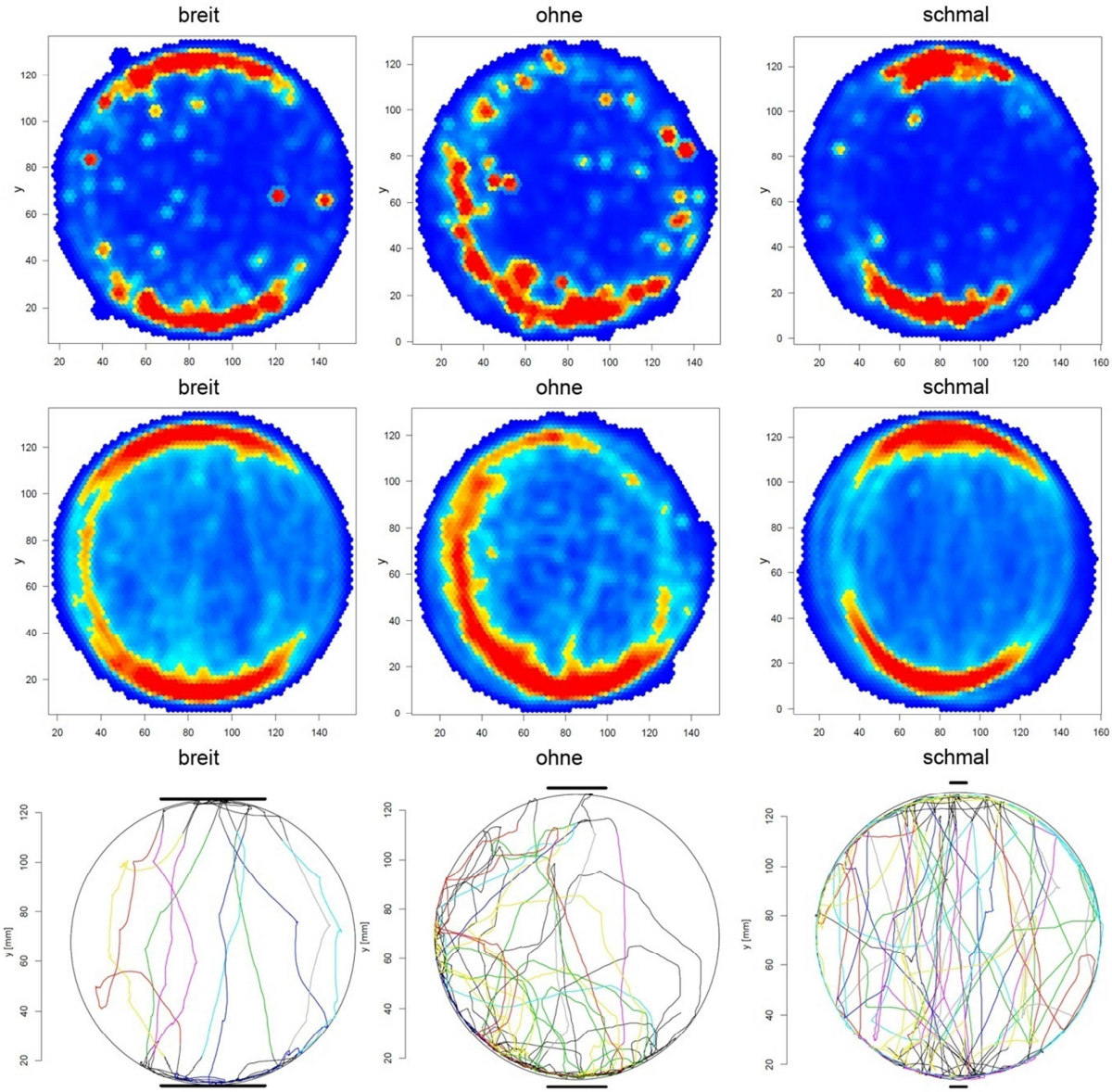


Abbildung 25 Transition-, Occupancy- und Einzelplots für die drei Experimentiermodi des Stammes rut^{2080}. Jeder Plot stellt einen Streifenmodus dar, es handelt sich um jeweils 20 Durchgänge (der Einzelplot zeigt das Laufmuster einer Fliege). Die X- und Y- Achsen stelle

Hauptkomponentenanalyse

Die Hauptkomponentenanalyse wurde in mehreren Gruppen durchgeführt, alle 300 Datensätze wurden verarbeitet und nach Streifen- und Genotyp zugleich sortiert. Da diese Sortierung die Unterschiede innerhalb der genetisch unterschiedlichen Stämme nicht zeigen kann wurden diese separat durchgeführt.

Hauptkomponentenanalyse gesamt

Die Korrelationsmatrix zeigt wie gut Werte korreliert sind. Wenn r=1 ist, so ist das Wertepaar zu 100% positiv korreliert, r=-1 wiederum zeigt eine exakte negative Korrelation, r=0 bescheinigt keinen Zusammenhang des Wertepaares.

Es ist eine durchgehend hohe positive Korrelation zwischen den Aktivitätsvariablen (sum_act; median_act) und den Laufvariablen (speeds; distance; walks) erkennbar, wobei die Korrelation von walks etwas geringer ist, als bei den anderen. Beide Pausenvariablen sind durchgehend negativ Korreliert, auch untereinander, wobei diese Korrelation in den meisten Fällen nicht so stark ausfällt wie bei den Lauf- und Aktivitätsvariablen, da r durchgehend kleiner als -0.5 ist.

Korrelationsmatrix (Pearson (n)):

Variablen	sum_act	median_act	median_pause	number_pause	speeds	distance	walks
sum_act	**1**	**0,730**	**-0,425**	**-0,302**	**0,722**	**0,900**	**0,592**
median_act	**0,730**	**1**	**-0,263**	**-0,381**	**0,540**	**0,694**	**0,449**
median_pause	**-0,425**	**-0,263**	**1**	**-0,195**	-0,082	**-0,238**	-0,080
number_pause	**-0,302**	**-0,381**	**-0,195**	**1**	**-0,438**	**-0,461**	**-0,406**
speeds	**0,722**	**0,540**	-0,082	**-0,438**	**1**	**0,894**	**0,723**
distance	**0,900**	**0,694**	**-0,238**	**-0,461**	**0,894**	**1**	**0,771**
walks	**0,592**	**0,449**	-0,080	**-0,406**	**0,723**	**0,771**	**1**

Werte in Fettdruck sind von Null verschieden mit einen Signifikanzniveau von alpha=0,05

Tabelle 4 Korrelationsmatrix nach Pearson für die GesamtHKA. R=1 ist vollständig positiv, r=-1 vollständig negativ und r=0 gar nicht korreliert.

Die Eigenwerte sind einer der wichtigsten Bestandteile der HKA, so lässt sich mit diesen die Varianzverteilung zwischen den einzelnen Hauptkomponenten ermitteln. Alle Eigenwerte in der Summe entsprechen der Zahl der Hauptkomponenten, um diese Werte einfacher deuten zu können werden sie zu der entsprechenden Prozentzahl umgerechnet, zusätzlich steht die summierte Variabilität der vorhergehenden Hauptkomponenten in der letzten Zeile.

Die Tabelle 5 zeigt, dass die ersten drei Hauptkomponenten bereits 86,63% der Gesamtvarianz. Die erste Hauptkomponente deckt bereits 59,12% der Gesamtvarianz ab, diese könnte in einer eindimensionalen Grafik dargestellt werden.

Eigenwerte:

	F1	F2	F3	F4	F5	F6	F7
Eigenwert	4,143	1,283	0,637	0,449	0,296	0,161	0,031
Variabilität (%)	59,189	18,334	9,102	6,415	4,222	2,298	0,440
Kummulierter %	59,189	77,523	86,625	93,040	97,262	99,560	100,000

Tabelle 5 Eigenwerte für die berechneten Hauptkomponenten aller Fliegendaten, Spalten zwei und drei enthalten die prozentuale Umrechnung im Bezug auf Gesamtvarianz und den kummulierten Wert dieser Prozentwerte.

Der „Scree plot“ ist eine grafische Aufarbeitung der Eigenwerte und zeigt diese als Säulen an, zusätzlich ist der Anteil an Variabilität in % aufgetragen (Abb. 26).

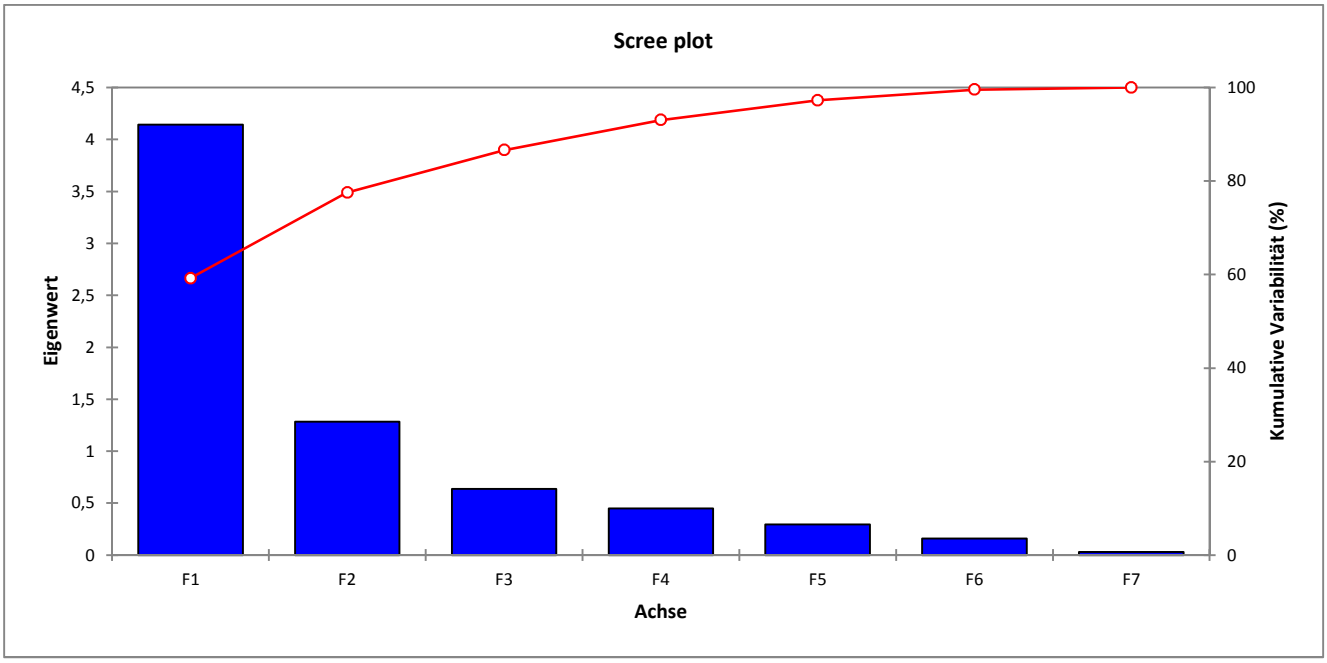


Abbildung 26 Screeplot der Gesamthauptkomponentenanalyse, eingezeichnet sind die Eigenwerte der Hauptkomponenten und der Anteil an der Gesamtvarianz in Prozent

Die Faktorladung (Tabelle 6) hilft den Einfluss der einzelnen Variablen auf die Hauptkomponenten festzustellen. Die Werte können hierbei zwischen -1 und 1 annehmen, wobei die Korrelation bei 1 am höchsten ist. Es lassen sich somit die Variablen sum_act, median_act, speeds, distance und walks der ersten Hauptkomponente zuordnen, median_pause und number_pause der Zweiten.

Faktorladungen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,901**	-0,278	-0,035	0,109	-0,137	0,271	-0,078
median_act	**0,778**	-0,144	-0,425	0,356	0,228	-0,125	-0,002
median_pause	-0,271	**0,874**	0,114	0,377	-0,006	0,090	-0,004
number_pause	**-0,533**	**-0,604**	0,465	0,364	0,053	-0,017	0,009
speeds	**0,883**	0,152	0,260	0,039	-0,257	-0,243	-0,054
distance	**0,975**	-0,005	0,106	0,034	-0,107	0,061	0,145
walks	**0,799**	0,185	0,385	-0,183	0,380	0,029	-0,023

Tabelle 6 Faktorenladung für die Gesamthauptkomponentenanalyse, gibt die Korrelation zwischen Variablen und Hauptkomponenten an, wobei r=1 vollständig positiv; r=-1 vollständig negativ und r=0 gar nicht korreliert sind.

Die Faktorenladung lässt sich auch grafisch in einem Korrelationskreis (Abbildung 27) darstellen, hierbei werden die beiden ersten Hauptkomponenten als Achsen gewählt. Die einzelnen Variablen werden ausgehend von dem Achsenursprung als Geraden aufgetragen, je weiter die Variablen vom Zentrum entfernt sind, desto höher ist die Korrelation. Des Weiteren gilt, dass die Korrelation zwischen Variablen, welche nah einander liegen positiv ist. Variablen, welche orthogonal zueinander angeordnet sind sind nicht korreliert und Variablen, welche sich gegenüber liegen sind negativ korreliert. Das Bild aus Tabelle 6 wird durch die Abbildung 27 vervollständigt und die beiden Hauptkomponenten lassen sich den Variablen zuordnen.

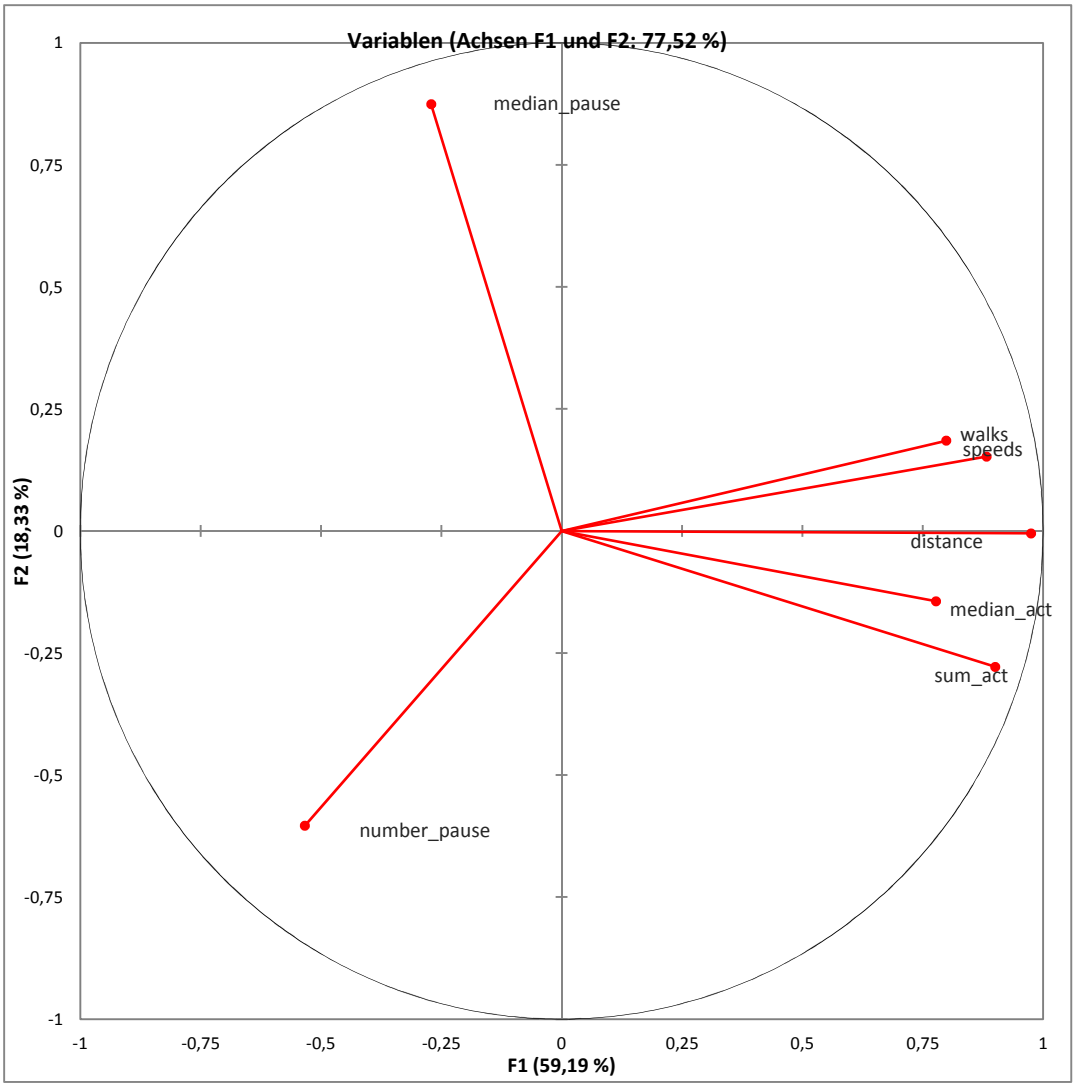


Abbildung 27 Korrelationskreis der GesamtHKA, mit den Achsen F1 (erste Hauptkomponente) und F2 (zweite Hauptkomponente), in Klammern hinter den Achsennamen steht die Variablität, welche durch diese Achse abgedeckt wird. Eingezeichnet sind die Faktorladungen der einzelnen Variablen, je näher die Variablen der Achse oder anderen Variablen sind, desto höher ist die Korrelation.

Um den Einfluss der einzelnen Variablen auf die Hauptkomponenten zu ermitteln quadriert man den Cosinuswert der Variablen, wie der folgenden Tabelle 7 entnommen werden kann. Hierbei entsprechen die fettgedruckten Werte dem Faktor, für den der quadrierte Kosinus am größten ist und somit den größten Einfluss auf die Hauptkomponente hat. Die Ergebnisse ähneln der vorhergehenden Tabellen 6. Man kann wieder die Einteilung von Aktivitäts- und Laufvariablen bei der ersten und Pausenvariablen bei der zweiten Hauptkomponenten machen. Die erste HK wird am stärksten durch distance, die Zweite am stärksten durch median_pause beeinflusst. Aus den vorliegenden Werten lassen sich auch die Beiträge der Variablen in Prozent berechnen, wie in Tabelle 8 zu sehen ist.

Quadrierte Cosinuswerte der Variablen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,812**	0,077	0,001	0,012	0,019	0,073	0,006
median_act	**0,605**	0,021	0,180	0,127	0,052	0,016	0,000
median_pause	0,073	**0,764**	0,013	0,142	0,000	0,008	0,000
number_pause	0,284	**0,364**	0,216	0,132	0,003	0,000	0,000
speeds	**0,780**	0,023	0,068	0,002	0,066	0,059	0,003
distance	**0,951**	0,000	0,011	0,001	0,012	0,004	0,021
walks	**0,638**	0,034	0,148	0,034	0,145	0,001	0,001

Tabelle 7 quadrierte Cosinuswerte der Gesamthauptkomponentenanalyse, die fett gedruckten Werte für jede Variable entsprechen dem Faktor, für den der quadrierte Kosinus am größten ist und somit den meisten Einfluss einer Variable darstellt

Beiträge der Variablen (%):

	F1	F2	F3	F4	F5	F6	F7
sum_act	19,587	6,031	0,197	2,635	6,327	45,490	19,732
median_act	14,595	1,613	28,301	28,246	17,561	9,672	0,013
median_pause	1,772	59,502	2,026	31,594	0,012	5,053	0,042
number_pause	6,864	28,386	33,869	29,471	0,941	0,187	0,282
speeds	18,819	1,807	10,628	0,335	22,305	36,727	9,380
distance	22,955	0,002	1,772	0,257	3,896	2,347	68,772
walks	15,408	2,660	23,209	7,462	48,958	0,524	1,779

Tabelle 8 Beiträge der einzelnen Variablen in Prozenten an den Hauptkomponenten der Gesamthauptkomponentenanalyse

Abschließend einer der größten Vorteile der HKA, die grafische Darstellung. Diese kann je nach der Zahl der verwendeten Hauptkomponenten in zwei oder drei Dimensionen dargestellt werden.

Die Darstellung beider Varianten bringt einige Vorteile in der Übersichtlichkeit, Abbildung 28 zeigt eine 3D Darstellung, bei welcher die erste Hauptkomponente der F1-Achse, die Zweite der F2-Achse und die Dritte der F3-Achse entsprechen. Angezeigt werden die Datensätze aller 300 Fliegen, wobei diese nach Geno- und Streifentyp farbliche Zuordnungen erhalten habe, welche der Legende innerhalb der Grafik entnommen werden können. Wie die Grafik zeigt, ist das Gesamtbild recht stark gestreut, auffällig ist hierbei, dass die Streuung um die F3-Achse gering ausfällt, unabhängig von Streifenbreite und Genotyp. Auf der F1-Achse, welche hauptsächlich die Aktivitäts- und Laufvariablen repräsentiert kann man eine starke Konzentration der Punkte um den Achsenursprung feststellen, einzig cstp_breit; cstp_schmal; rut2080_breit und rut2080_schmal streuen über eine weite Strecke von ca. -3 bis +8, wobei bei rut2080 ein die meisten Punkte im positiven Bereich zu finden sind. Die restlichen Variablen befinden sich knapp im negativen Bereich, mit ein paar Ausreisern. Um die F2-Achse, welche die Pausenvariablen widerspiegelt findet unter allen Geno- und

Streifentypen eine ziemlich gleichmäßige Streuung zwischen -2 und +3 statt, Ausnahmen bilden hierbei rsh_ohne und rut1_breit welche deutlich stärker ins positive gehen (bis zu +5) und rut2080 in allen Streifentypen, welches einen viel schmaleren Bereich von -2 bis ca. +1 (bei rut2080_breit sind 5 von 60 rut2080-Werten zwischen +1 und +3 angesiedelt). Um ein übersichtlicheres Bild zu erzeugen finden sich in Abbildung 29 dieselben Werte, jedoch in zwei Dimensionen aufgetragen, das heißt, dass die F3-Achse, welche die dritte Hauptkomponente dargestellt hat wegfällt. Die Ergebnisse sind bis auf diese Achse aber identisch.

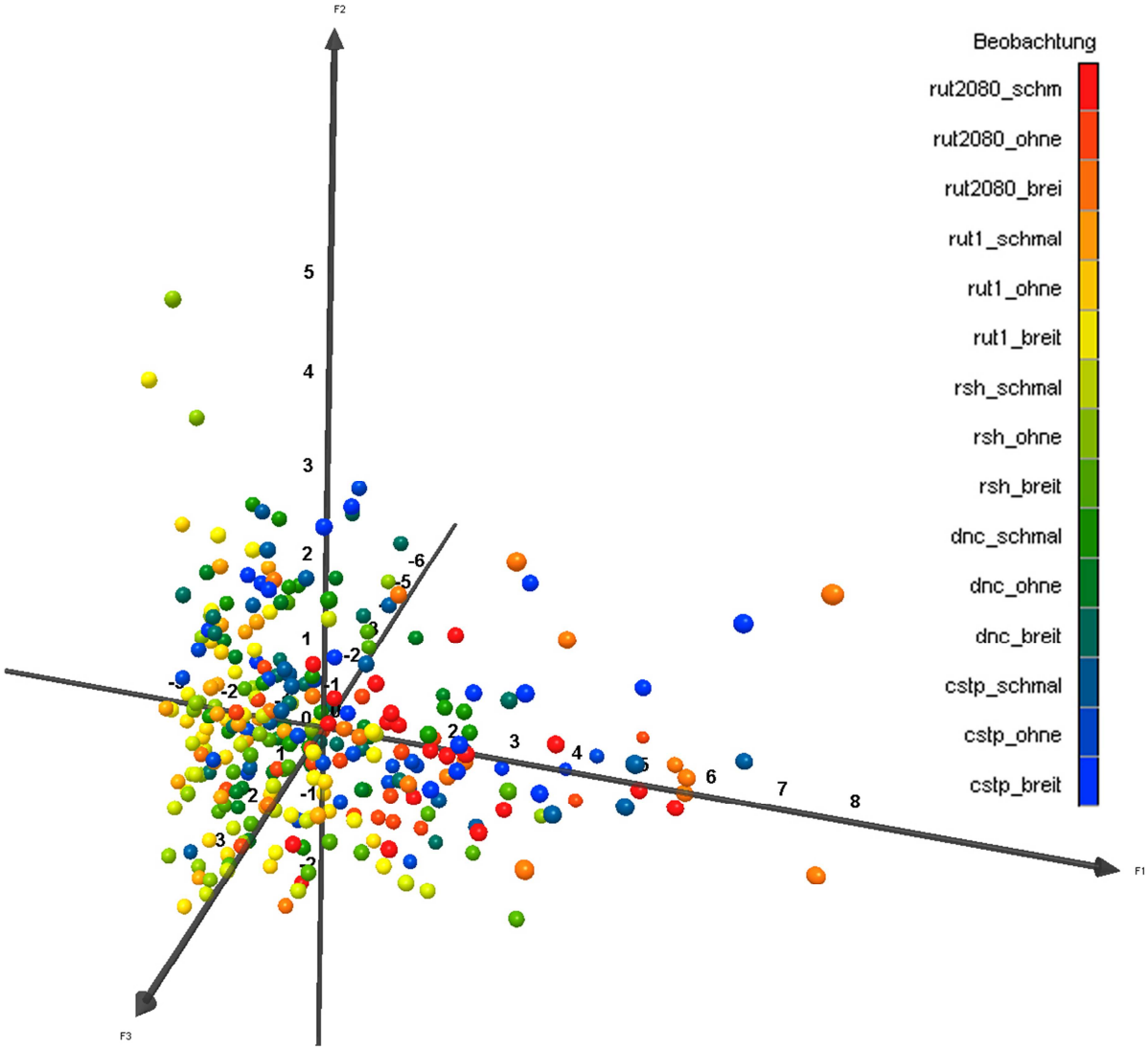


Abbildung 28 Dreidimensionale Darstellung der HKA mit allen Daten, Einteilung nach Geno- und Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

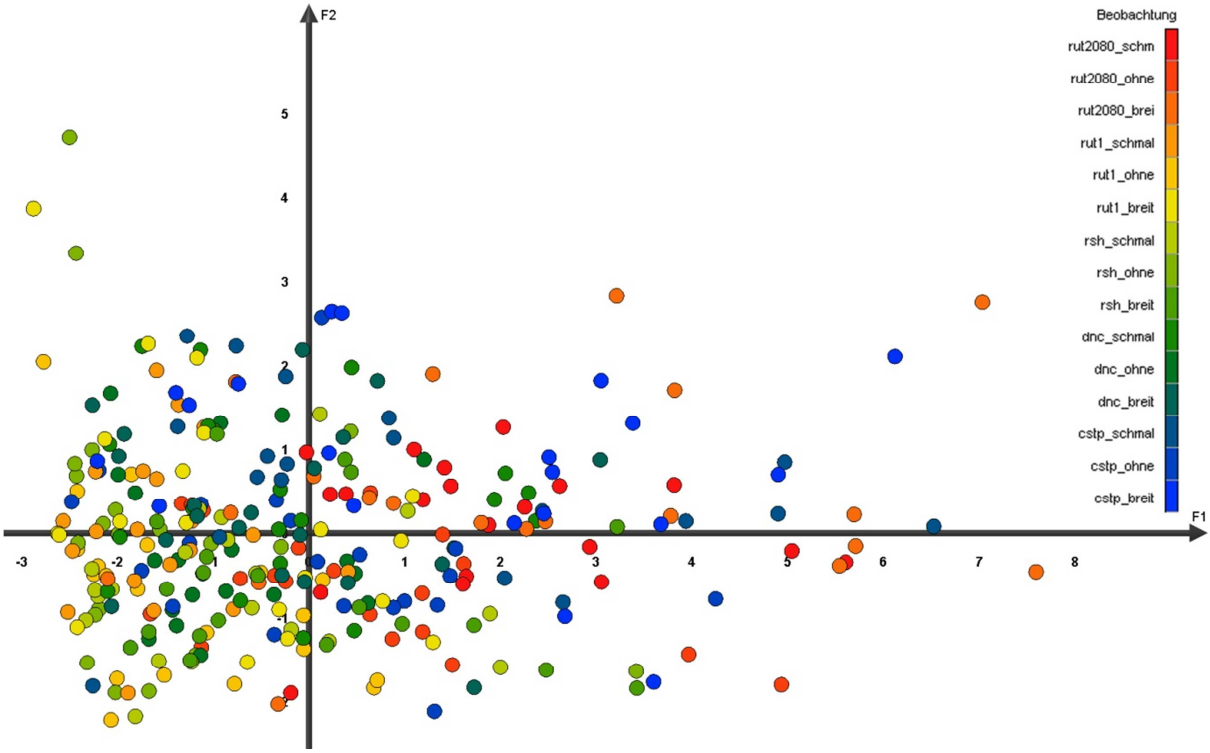


Abbildung 29 Zweidimensionale Darstellung der HKA mit allen Daten, Einteilung nach Geno- und Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Um eine bessere Vergleichbarkeit zwischen den Geno- und Streifentypen zu ermöglichen wurden in den Abbildungen 30 und 31, die aus den Abbildungen 28 und 29 bekannten Werte mit Hilfe des Medians zusammengefasst und als einzelne Punkte in das Koordinatensystem gezeichnet. Auch hier gilt, dass die Abbildung 30 eine Achse (dritte Hauptkomponente) mehr enthält, ansonsten aber mit denselben Werten und Koordinaten arbeitet.

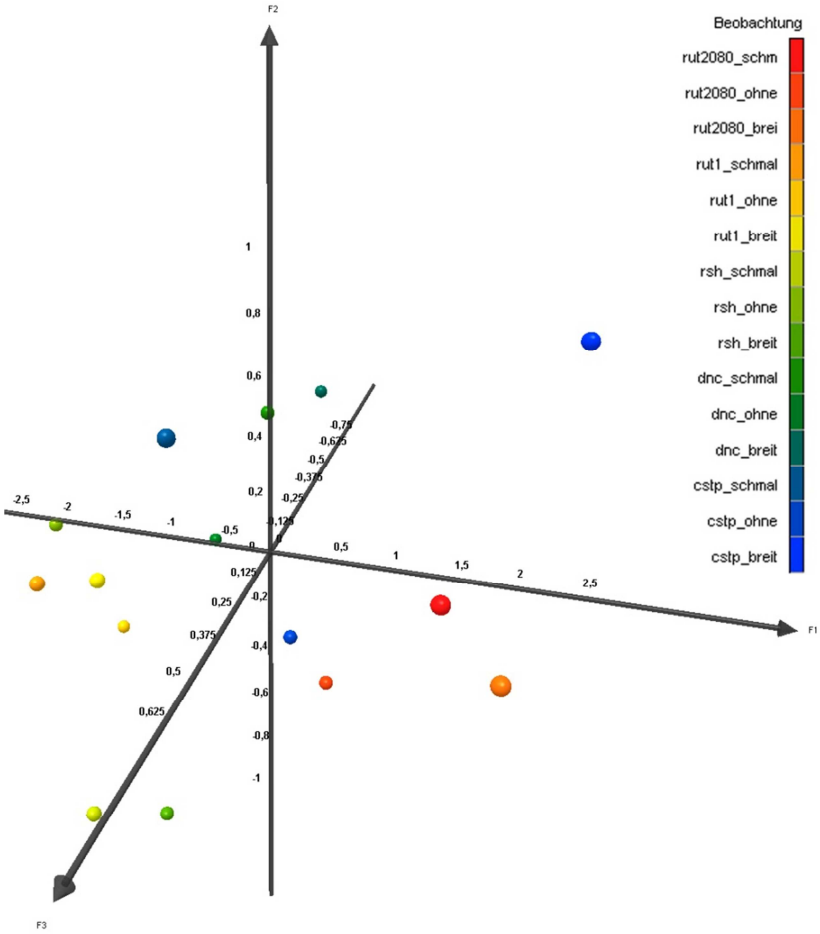


Abbildung 30 Dreidimensionale Darstellung der gruppierten HKA mit allen Daten, Einteilung nach Geno- und Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2;

F3=HK3. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Geno- und Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

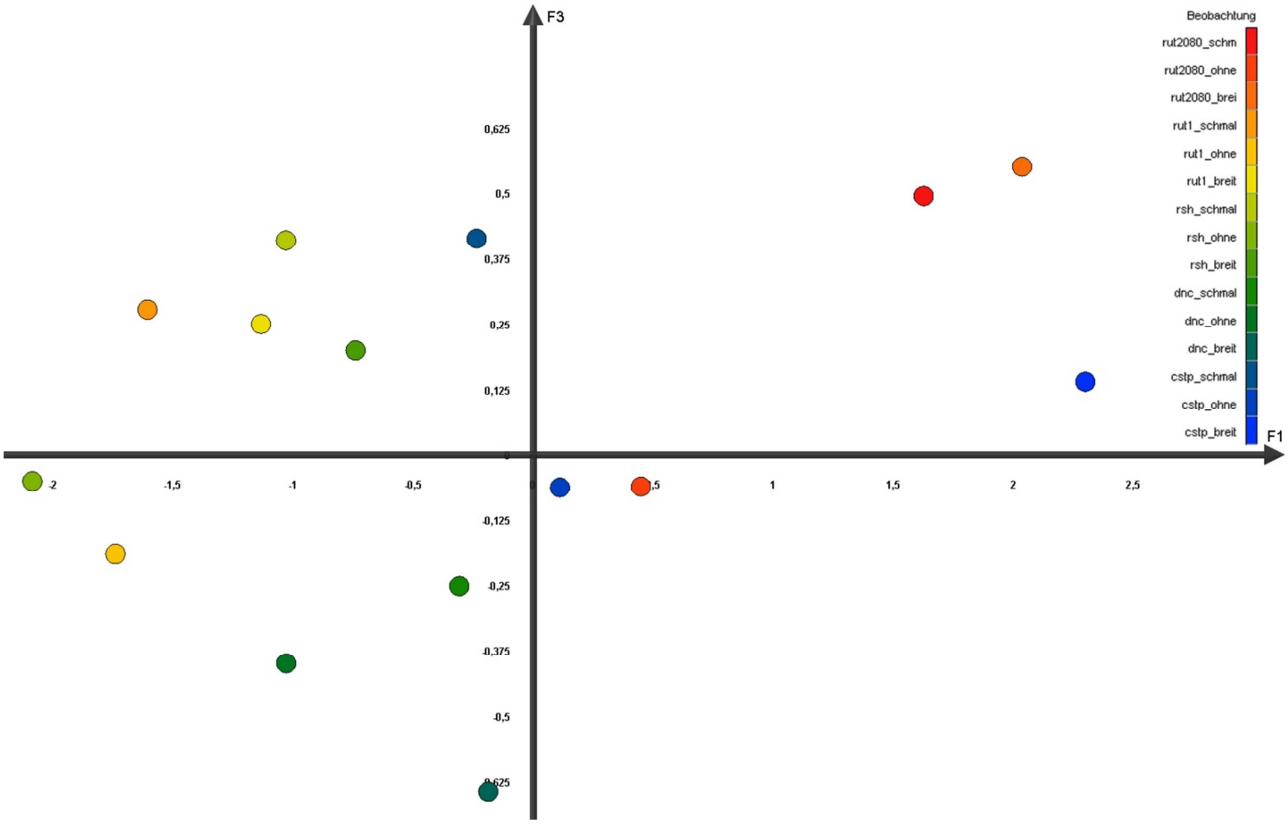


Abbildung 31 Zweidimensionale Darstellung der gruppierten HKA mit allen Daten, Einteilung nach Geno- und Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Geno- und Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Aufgrund der großen Menge an Daten und um eine bessere Vergleichbarkeit zu gewährleisten, werden im Folgenden die Resultate von Hauptkomponentenanalysen für die fünf untersuchten Genotypen gezeigt. Hierbei handelt es sich um dieselben Berechnungen und Grafiktypen wie zuvor, weshalb auf eine detaillierte Erklärung verzichtet wird.

Hauptkomponentenanalyse CSTP

Für die folgenden Resultate wurden Berechnungen mit den Rohdaten aus 60 Datensätzen durchgeführt, diese sind zu jeweils 20 Stück auf die drei Streifenbreiten verteilt. Die Korrelationsmatrix (Tabelle 9) zeigt ein ähnliches Bild wie die Korrelationsmatrix der gesamten HKA, die Aktivitäts- und Laufvariablen zeigen auch hier eine positive Korrelation untereinander, ähnlich stark ausgeprägt wie im Gesamten, die Korrelationsunterschiede zu walks sind hier noch größer. Auffällig ist, dass die negative Korrelation zwischen den Pausenvariablen und den Restlichen deutlich höher ist, r ist hierbei teilweise kleiner als -0,5.

Korrelationsmatrix (Pearson (n)):

Variablen	sum_act	median_act	median_pause	number_pause	speeds	distance	walks
sum_act	**1**	**0,769**	**-0,577**	**-0,540**	**0,725**	**0,907**	**0,522**
median_act	**0,769**	**1**	**-0,422**	**-0,528**	**0,421**	**0,628**	**0,310**
median_pause	**-0,577**	**-0,422**	**1**	-0,069	-0,154	**-0,368**	-0,166
number_pause	**-0,540**	**-0,528**	-0,069	**1**	**-0,633**	**-0,654**	**-0,559**
speeds	**0,725**	**0,421**	-0,154	**-0,633**	**1**	**0,905**	**0,705**
distance	**0,907**	**0,628**	**-0,368**	**-0,654**	**0,905**	**1**	**0,703**
walks	**0,522**	**0,310**	-0,166	**-0,559**	**0,705**	**0,703**	**1**

Werte in Fettdruck sind von Null verschieden mit einen Signifikanzniveau von alpha=0,05

Tabelle 9 Korrelationsmatrix nach Pearson für die HKA von cstp. R=1 ist vollständig positiv, r=-1 vollständig negativ und r=0 gar nicht korreliert.

Die Eigenwerte (Tabelle 10) zeigen, dass knapp über 90% der Gesamtvariabilität auf die ersten drei Hauptkomponenten fällt. Der Anteil ist etwas größer als bei der gesamt HKA (87%), was auch für die ersten beiden Hauptkomponenten gilt. Die Eigenwerte sind in der Abbildung 32 als Scree plot visualisiert.

Eigenwerte:

	F1	F2	F3	F4	F5	F6	F7
Eigenwert	4,352	1,267	0,691	0,355	0,225	0,084	0,025
Variabilität (%)	62,177	18,102	9,877	5,072	3,209	1,200	0,363
Kummulierter %	62,177	80,278	90,156	95,228	98,437	99,637	100,000

Tabelle 10 Eigenwerte für die berechneten Hauptkomponenten von cstp, Spalten zwei und drei enthalten die prozentuale Umrechnung im Bezug auf Gesamtvarianz und den kumulierten Wert dieser Prozentwerte.

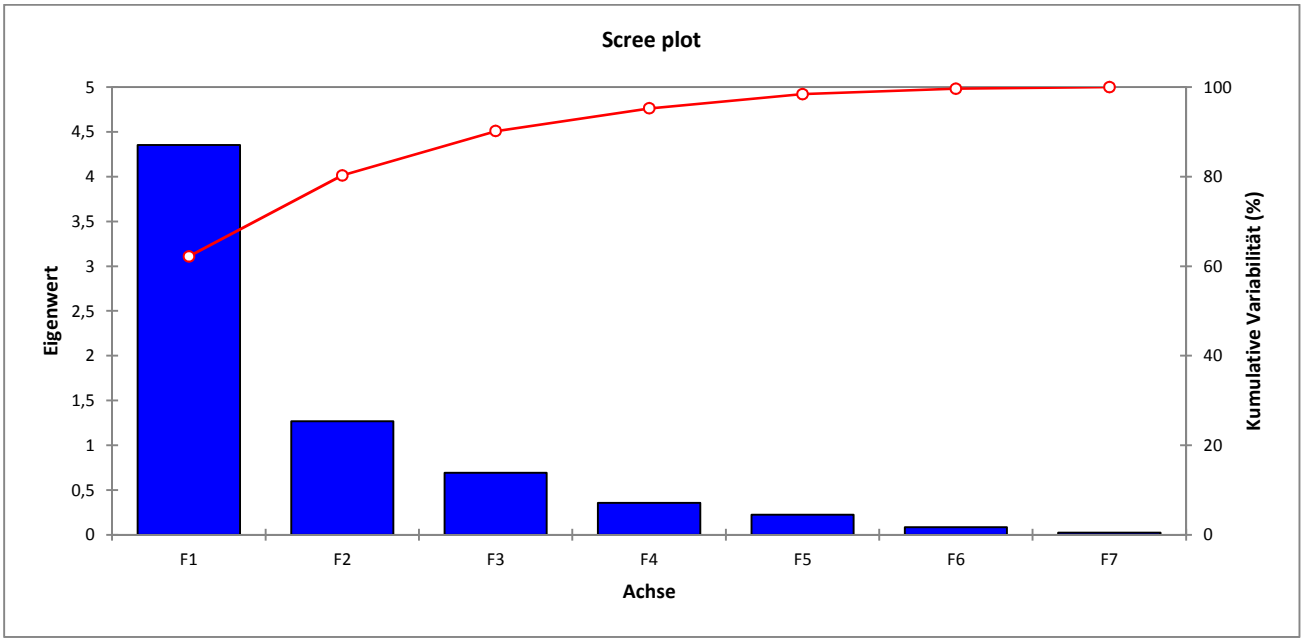


Abbildung 32 Screeplot der HKA von cstp, eingezeichnet sind die Eigenwerte der Hauptkomponenten und der Anteil an der Gesamtvarianz in Prozent

Faktorladungen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,925**	-0,286	0,046	-0,144	0,007	0,181	-0,085
median_act	**0,739**	-0,329	**0,513**	0,115	0,246	-0,091	0,009
median_pause	-0,420	**0,829**	0,261	-0,142	0,206	0,077	-0,009
number_pause	**-0,733**	-0,445	-0,370	-0,171	0,314	0,006	0,003
speeds	**0,868**	0,285	-0,214	-0,295	0,005	-0,176	-0,046
distance	**0,968**	0,039	-0,100	-0,176	0,012	0,073	0,125
walks	**0,741**	0,330	-0,407	0,391	0,152	0,024	-0,014

Tabelle 11 Faktorenladung für die HKA von cstp, gibt die Korrelation zwischen Variablen und Hauptkomponenten an, wobei r=1 vollständig positiv; r=-1 vollständig negativ und r=0 gar nicht korreliert sind.

Die Faktorladung (Tabelle 11) zeigt eine positive Korrelation von sum_act; median_act; speeds; distance; walks und eine negative Korrelation von number_pause zur ersten Hauptkomponenten und eine starke positive von median_pause zur Zweiten. Es zeigt sich ein sehr ähnliches Bild wie bei der Gesamthauptkomponentenanalyse, wie man auch sehr gut im Korrelationskreis (Abbildung 33) erkennen kann.

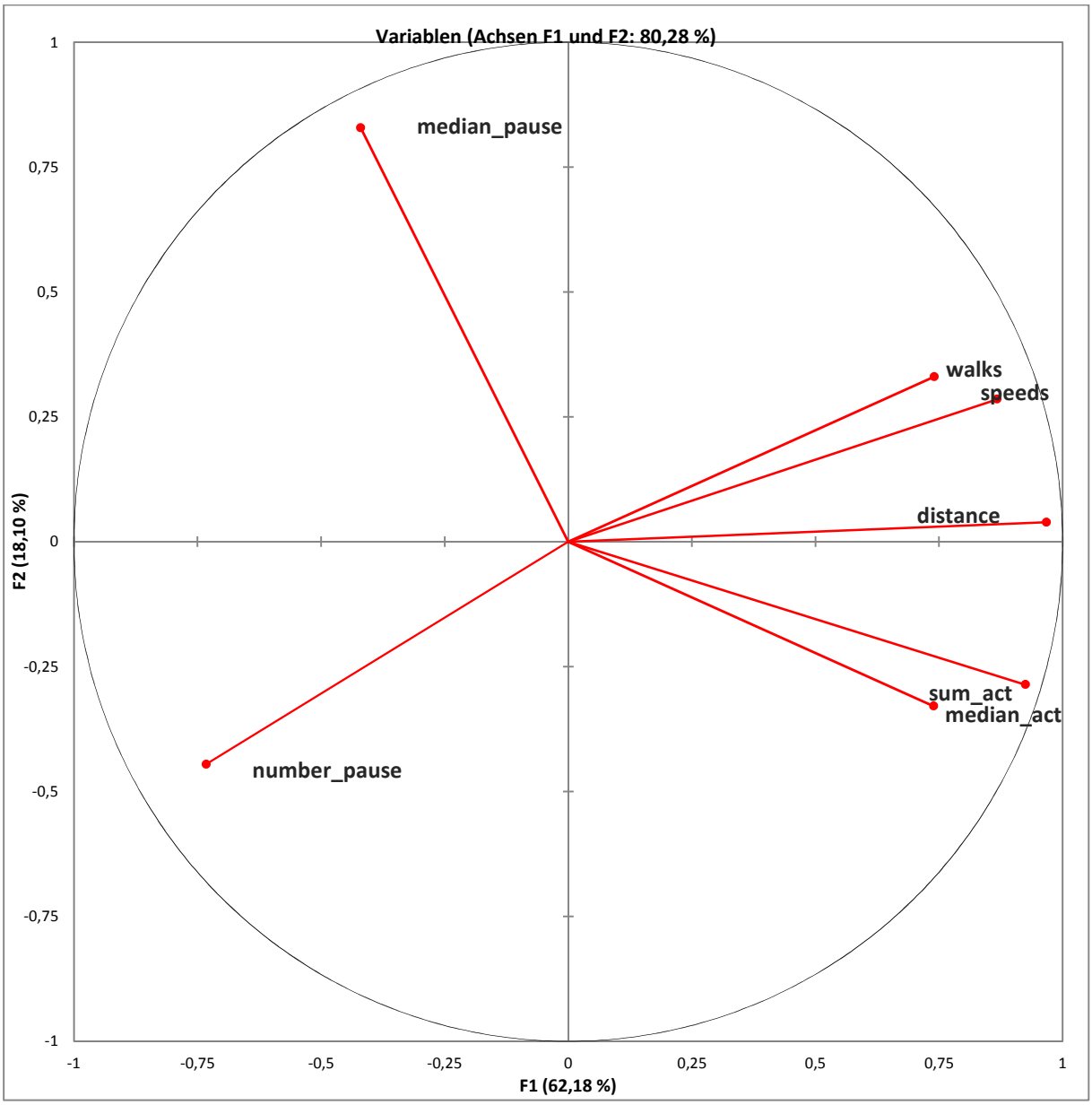


Abbildung 33 Korrelationskreis der HKA von cstp, mit den Achsen F1 (erste Hauptkomponente) und F2 (zweite Hauptkomponente), in Klammern hinter den Achsennamen steht die Variablität, welche durch diese Achse abgedeckt wird. Eingezeichnet sind die Faktorladungen der einzelnen Variablen, je näher die Variablen der Achse oder anderen Variablen sind, desto höher ist die Korrelation.

Die quadrierten Cosinuswerte (Tabelle 12) vervollständigen das Bild der Faktorenladung, vor allem die Variablen sum_act; speeds; und distance beeinflussen die erste HKA und median_pause die Zweite.

Quadrierte Cosinuswerte der Variablen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,855**	0,082	0,002	0,021	0,000	0,033	0,007
median_act	**0,546**	0,108	0,263	0,013	0,060	0,008	0,000
median_pause	0,176	**0,687**	0,068	0,020	0,042	0,006	0,000
number_pause	**0,537**	0,198	0,137	0,029	0,099	0,000	0,000
speeds	**0,753**	0,081	0,046	0,087	0,000	0,031	0,002
distance	**0,936**	0,002	0,010	0,031	0,000	0,005	0,016
walks	**0,548**	0,109	0,165	0,153	0,023	0,001	0,000

Tabelle 12 quadrierte Kosinuswerte der HKA von cstp, die fett gedruckten Werte für jede Variable entsprechen dem Faktor, für den der quadrierte Kosinus am größten ist und somit den meisten Einfluss einer Variable darstellt.

Beiträge der Variablen (%):

	F1	F2	F3	F4	F5	F6	F7
sum_act	19,648	6,455	0,311	5,840	0,020	39,095	28,631
median_act	12,553	8,556	38,067	3,724	26,884	9,908	0,307
median_pause	4,055	54,187	9,881	5,696	18,863	7,017	0,300
number_pause	12,335	15,659	19,775	8,261	43,890	0,038	0,042
speeds	17,297	6,415	6,599	24,557	0,012	36,843	8,278
distance	21,509	0,119	1,446	8,772	0,065	6,403	61,685
walks	12,602	8,609	23,920	43,150	10,265	0,696	0,756

Tabelle 13 Beiträge der einzelnen Variablen in Prozenten an den Hauptkomponenten von cstp.

Im Folgenden die grafische Darstellung der HKA von cstp, in zwei (Abbildung 34)und drei (Abbildung 35)Dimensionen. Da beide Grafiken dieselben Daten zeigen, bei Abbildung 35, jedoch das Koordinatensystem um eine Achse erweitert wurde, sind die folgenden Feststellungen auf Beide zu beziehen mit Ausnahme der Aussagen für die Achse F3 der Abbildung 35, welche die dritte Hauptkomponente beschreibt. Angefangen mit der Achse F1, welche die erste Hauptkomponente darstellt fällt auf, dass es bei den Streifenmodi schmal und breit eine weite Streuung über den Bereich von -3 bis +5 gibt, wobei der streifenlose Modus (bis auf einen Ausreiser) im Bereich von -3 bis +1 zu finden ist. Die Achse F2, welche die zweite Hauptkomponente Darstellt zeigt eine Verteilung der Punkte von -2,5 bis + 2,5, wobei sich die meisten Punkte des streifenlosen Modus im negativen und die meisten Punkten der anderen beiden Modi im positiven Bereich befinden. Die Achse F3 spiegelt die dritte Hauptkomonente wieder. Um diese fällt die Streuung recht gering aus, so befinden sich fast alle Werte der breiten und schmalen Streifenmodi in einem Bereich von -1 bis +2,

die Punkte des streifenlosen Modus befinden sich hingegen im Bereich von -1 bis +1 (bei allen drei Modi gibt es eine kleine Zahl von Ausreisern, jeweils 1 bis 2 Datensätze).

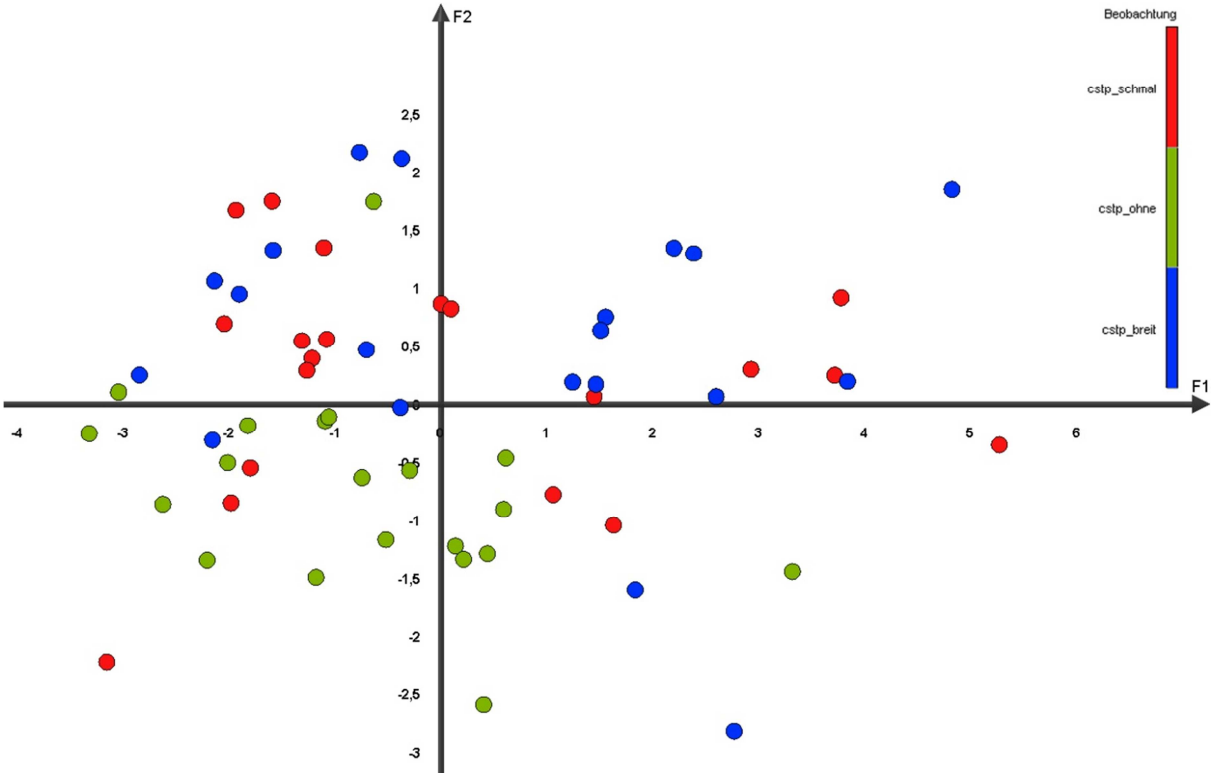


Abbildung 34 Zweidimensionale Darstellung der HKA von cstp, Einteilung nach Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

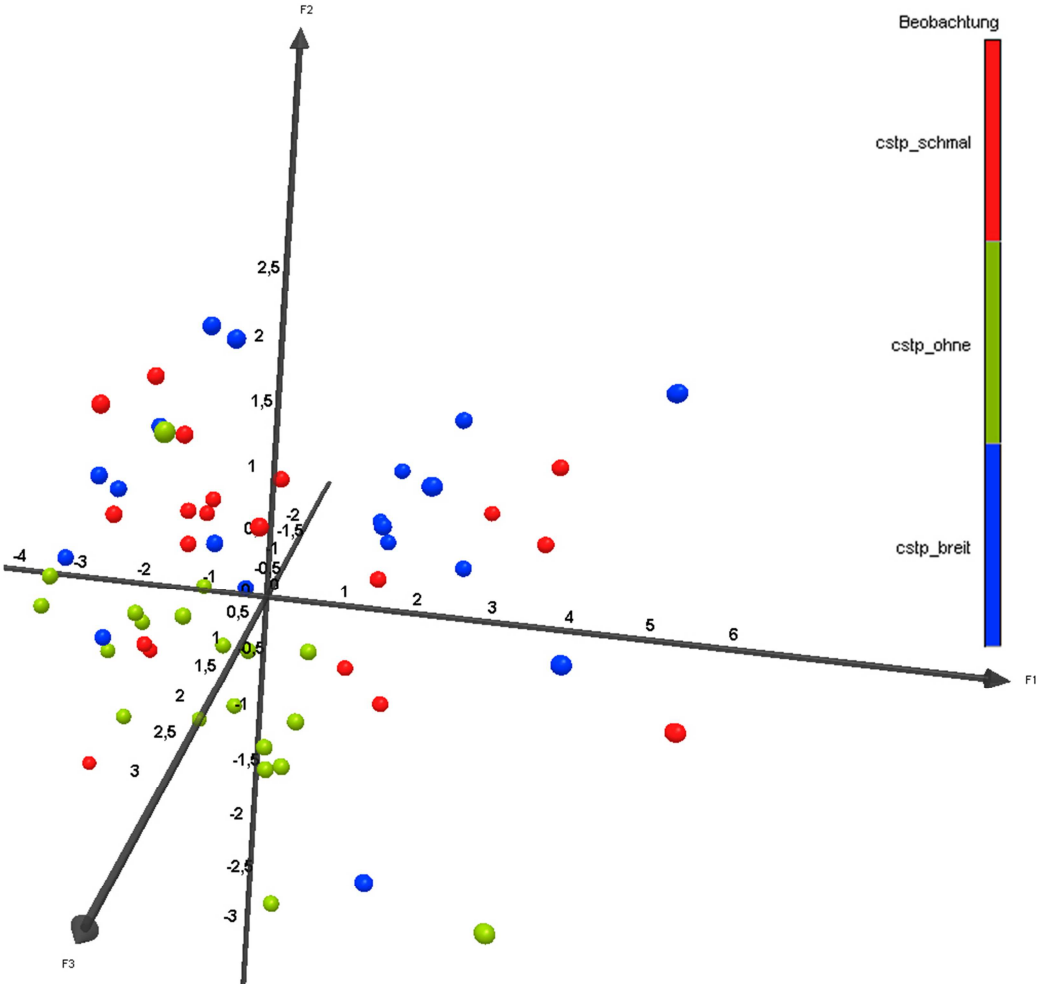


Abbildung 35 Dreidimensionale Darstellung der HKA von cstp, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Zuletzt eine gruppierte Ansicht der Streifentypen (Abb. 36 und 37), hierbei wurden alle Punkte des jeweiligen Streifentyps zu jeweils

einem zusammen-gefasst.

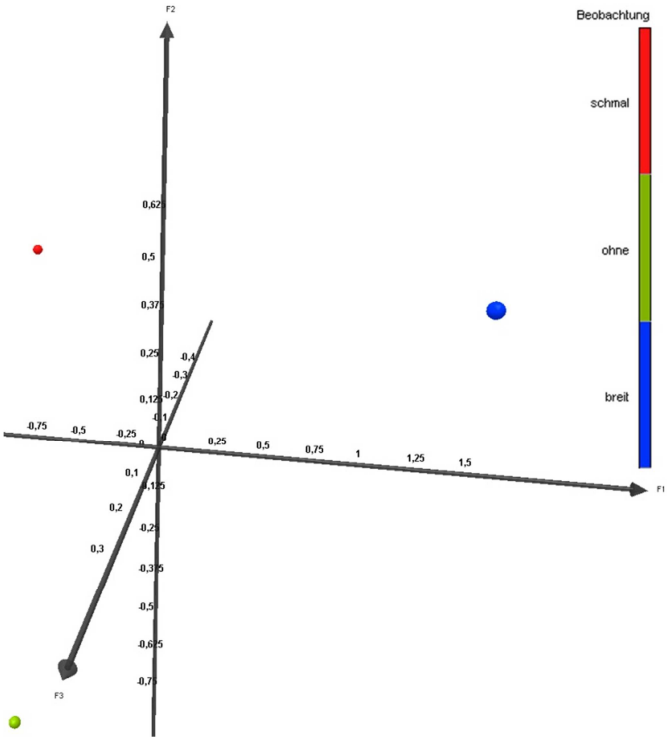


Abbildung 36 Dreidimensionale Darstellung der HKA von cstp, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

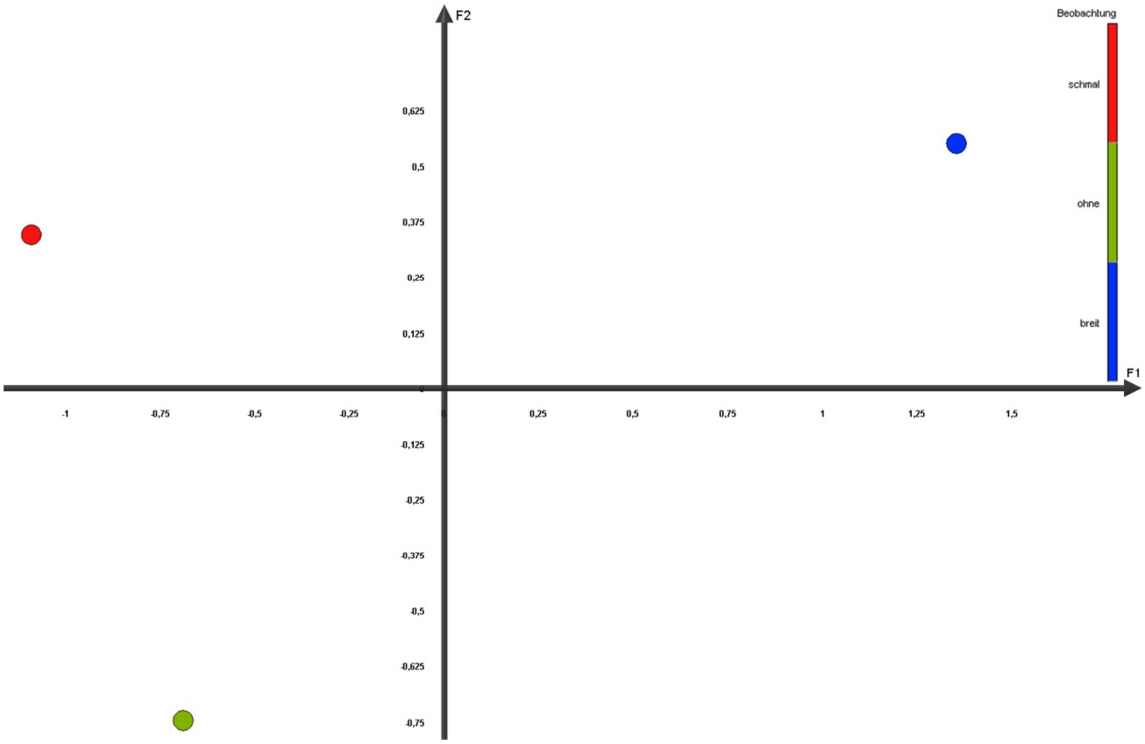


Abbildung 37 Zweidimensionale Darstellung der HKA von cstp, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Hauptkomponentenanalyse DNC

Auch hier (Tabelle 14) ein ähnliches Bild wie bei der Gesamt-HKA, wobei die positiven Korrelationen zwischen den Aktivitäts- und Laufvariablen geringer ausfallen (speeds zu madian_act 0,335, statt 0,540). Auch ist die negative Korrelation zwischen den Aktivitäts- und Laufvariablen und den Pausevariablen nicht so hoch wie im Gesamten.

Korrelationsmatrix (Pearson (n)):

Variablen	sum_act	median_act	median_pause	number_pause	speeds	distance	walks
sum_act	**1**	**0,652**	**-0,364**	-0,084	**0,537**	**0,887**	**0,556**
median_act	**0,652**	**1**	-0,175	-0,230	**0,335**	**0,530**	0,103
median_pause	**-0,364**	-0,175	**1**	-0,113	-0,073	-0,226	-0,139
number_pause	-0,084	-0,230	-0,113	**1**	**-0,438**	**-0,308**	**-0,359**
speeds	**0,537**	**0,335**	-0,073	**-0,438**	**1**	**0,829**	**0,611**
distance	**0,887**	**0,530**	-0,226	**-0,308**	**0,829**	**1**	**0,720**
walks	**0,556**	0,103	-0,139	**-0,359**	**0,611**	**0,720**	**1**

Werte in Fettdruck sind von Null verschieden mit einen Signifikanzniveau von alpha=0,05

Tabelle 14 Korrelationsmatrix nach Pearson für HKA von dnc. R=1 ist vollständig positiv, r=-1 vollständig negativ und r=0 gar nicht korreliert.

Die Eigenwerte (Tabelle 15) zeigen auch eine geringere Varianzverteilung über die ersten drei Hauptkomponenten an, als im Gesamten. So wird bei dnc 82,7% der Gesamtvarianz mit durch die ersten drei Komponenten abgedeckt, im Gegensatz zu 86,3% der Gesamthauptkomponentenanalyse. Die erste Hauptkomponente beschreibt fast 10% weniger Varianz (51,4% gegenüber 59,2%). Die Eigenwerte sind ebenfalls grafisch im Scree plot (Abbildung 38)aufgetragen.

Eigenwerte:

	F1	F2	F3	F4	F5	F6	F7
Eigenwert	3,596	1,296	0,896	0,688	0,364	0,144	0,016
Variabilität (%)	51,368	18,519	12,805	9,834	5,196	2,052	0,227
Kummulierter %	51,368	69,886	82,691	92,525	97,721	99,773	100,000

Tabelle 15 Eigenwerte für die berechneten Hauptkomponenten von dnc, Spalten zwei und drei enthalten die prozentuale Umrechnung im Bezug auf Gesamtvarianz und den kumulierten Wert dieser Prozentwerte.

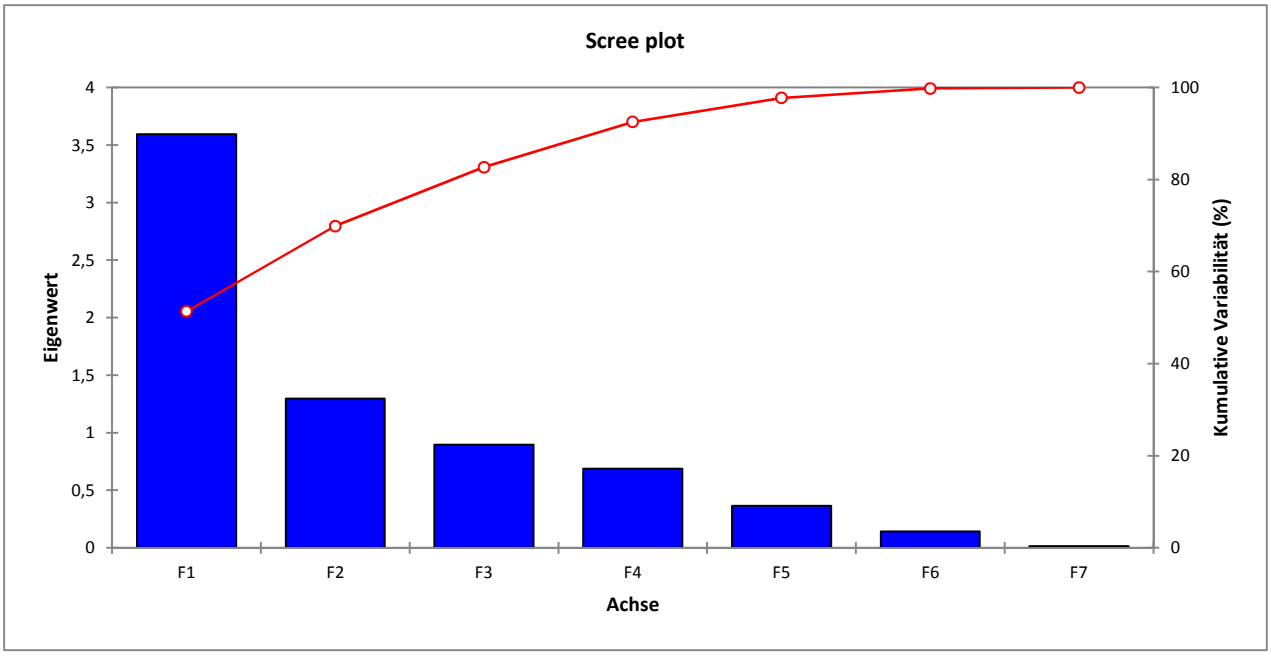


Abbildung 38 Screeplot der HKA von dnc, eingezeichnet sind die Eigenwerte der Hauptkomponenten und der Anteil an der Gesamtvarianz in Prozent

Die Faktorladungen (Tabelle 16) von dnc zeigen ebenfalls ähnliche Ergebnisse wie die Gesamthaupt-komponentenanalyse. Auch hier ist erkennbar, dass die Variablen sum_act; median_act; speeds; distance und walks mit der positiv ersten Hauptkomponente und median_pause (positiv), sowie number_pause (negativ) mit der Zweiten korrelieren. Auffällig bei dnc ist eine starke positive Korrelation von median_act mit der dritten Hauptkomponente. Die Faktorladungen sind im Korrelationskreis (Abbildung 39) grafisch dargestellt.

Faktorladungen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,873**	-0,362	0,063	0,186	-0,154	0,201	-0,067
median_act	**0,612**	-0,303	**0,700**	-0,035	-0,068	-0,193	0,007
median_pause	-0,293	**0,695**	0,332	0,559	-0,086	0,025	-0,005
number_pause	-0,428	**-0,661**	-0,231	0,552	0,102	-0,107	0,005
speeds	**0,827**	0,295	-0,090	0,064	0,457	-0,080	-0,041
distance	**0,972**	-0,016	-0,055	0,163	0,061	0,113	0,097
walks	**0,750**	0,258	-0,476	0,065	-0,324	-0,186	-0,007

Tabelle 16 Faktorenladung für die HKA von dnc, gibt die Korrelation zwischen Variablen und Hauptkomponenten an, wobei r=1 vollständig positiv; r=-1 vollständig negativ und r=0 gar nicht korreliert sind.

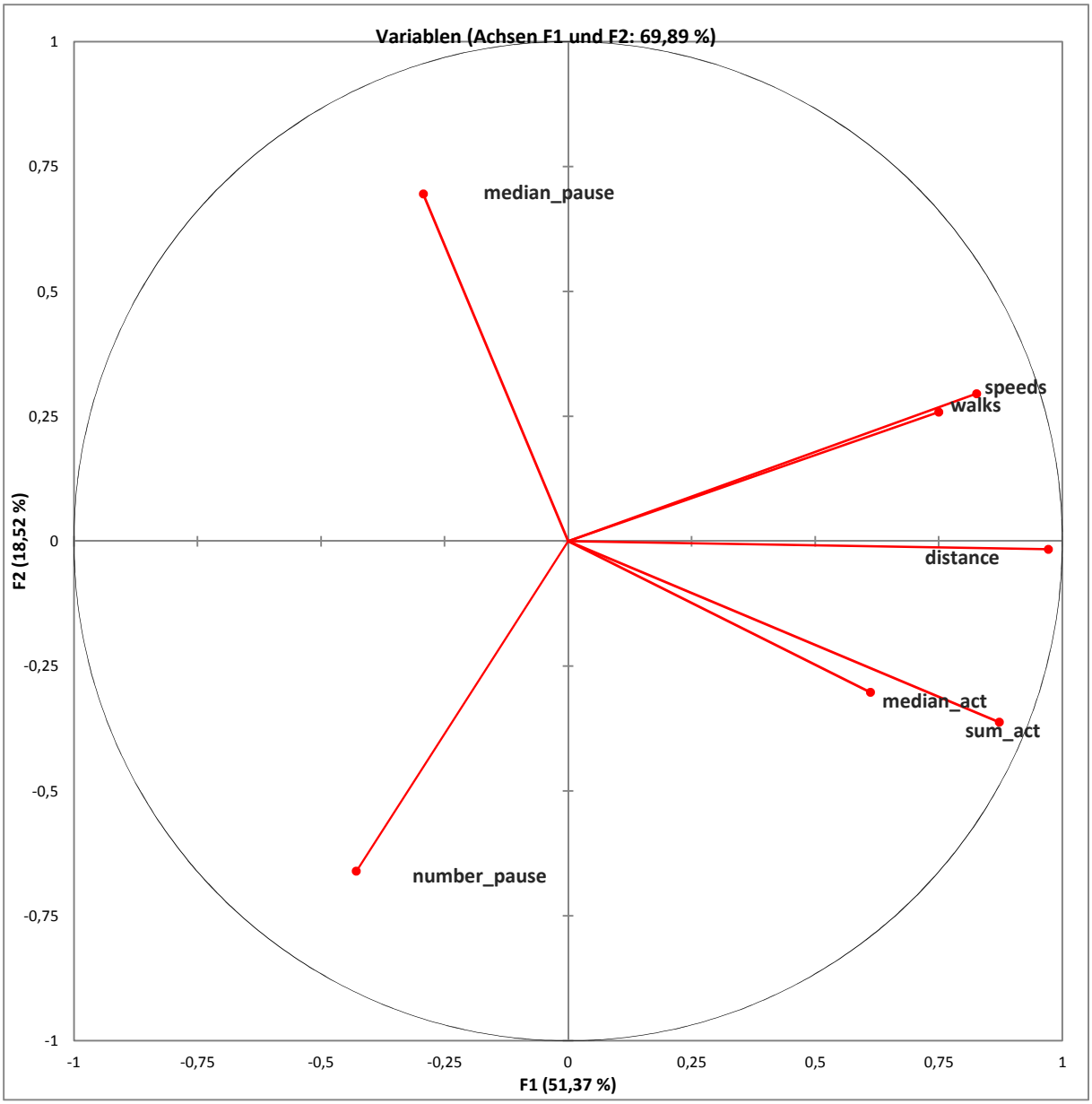


Abbildung 39 Korrelationskreis der HKA von dnc, mit den Achsen F1 (erste Hauptkomponente) und F2 (zweite Hauptkomponente), in Klammern hinter den Achsennamen steht die Variablität, welche durch diese Achse abgedeckt wird. Eingezeichnet sind die Faktorladungen der einzelnen Variablen, je näher die Variablen der Achse oder anderen Variablen sind, desto höher ist die Korrelation.

Wie sich bei den Faktorladungen bereits gezeigt hat, hat die Variable median_act einen größeren Einfluss auf die dritte Hauptkomponente, als auf die übrigen – im Gegensatz zur Gesamthauptkomponentenanalyse, in welcher sie ebenfalls zur ersten beiträgt. Ansonsten zeigen die quadrierten Cosinuswerte (Tabelle 17) ein bekanntest Bild, sum_act; speeds und deistance bestimmten die erste, median_pause und number_pause die zweite Hauptkomponente.

Dieses Ergebnis kann man auch in Prozentwerte umgerechnet in Tabelle18 erkennen, hierbei fällt wieder stark der Einfluss von median_act mit 54,7% auf die dritte Hauptkomponente auf. Auf die ersten beiden verteilen sich die zuvor genannten Variablen recht gleichmäßig.

Quadrierte Cosinuswerte der Variablen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,762**	0,131	0,004	0,035	0,024	0,040	0,005
median_act	0,374	0,092	**0,491**	0,001	0,005	0,037	0,000
median_pause	0,086	**0,483**	0,110	0,313	0,007	0,001	0,000
number_pause	0,184	**0,436**	0,053	0,305	0,010	0,011	0,000
speeds	**0,683**	0,087	0,008	0,004	0,209	0,006	0,002
distance	**0,944**	0,000	0,003	0,027	0,004	0,013	0,009
walks	**0,563**	0,067	0,227	0,004	0,105	0,035	0,000

Tabelle 17 quadrierte Cosinuswerte der HKA von dnc, die fett gedruckten Werte für jede Variable entsprechen dem Faktor, für den der quadrierte Kosinus am größten ist und somit den meisten Einfluss einer Variablen darstellt.

Beiträgeder Variablen (%):

	F1	F2	F3	F4	F5	F6	F7
sum_act	21,178	10,135	0,440	5,018	6,486	28,193	28,550
median_act	10,414	7,065	54,744	0,176	1,285	25,977	0,339
median_pause	2,387	37,254	12,316	45,453	2,022	0,425	0,144
number_pause	5,104	33,656	5,961	44,297	2,839	7,994	0,148
speeds	19,005	6,720	0,910	0,587	57,528	4,419	10,831
distance	26,261	0,020	0,336	3,859	1,015	8,841	59,667
walks	15,650	5,151	25,292	0,609	28,825	24,151	0,321

Tabelle 18 Beiträge der einzelnen Variablen in Prozenten an den Hauptkomponenten von dnc.

Im Folgenden die grafische Darstellung der HKA von dnc, in zwei (Abbildung 40) und drei (Abbildung 41) Dimensionen. Da beide Grafiken dieselben Daten zeigen, bei Abbildung 41, jedoch das Koordinatensystem um eine Achse erweitert wurde, sind die folgenden Feststellungen auf beide Grafiken zu beziehen mit Ausnahme der Aussagen für die Achse F3 ,welche die dritte Hauptkomponente beschreibt. Die Achse F1, welche die erste Hauptkomponente darstellt zeigt eine Streuung im Wertebereich von -3 bis +5, hierbei sind besonders der breite und schmale Streifenmodus vertreten, der streifenlose Modus erreicht den Höchstwert von +3. Die Achse F2, welche die zweite Hauptkomponente beschreibt zeigt ein ähnliches Bild, die Werte streuen von -2,5 bis +2,5. Hierbei ist eine relativ gleichmäßige Verteilung aller Streifentypen gegeben, der streifenlose Modus erreicht jedoch nicht die Maximalwerte und bleibt in einem Bereich von -1,5 bis +1,5 vertreten. Die dritte Achse F3, welche die dritte Hauptkomponente vertritt, zeigt eine unterschiedliche Verteilung der Datensätze. Der breite Streifenmodus deckt einen Bereich von -2,5 bis +2,5 ab, der Streifenlose Einen von -0,5 bis +2,5 und die Werte des schmalstreifigen Modus finden sich im Bereich von -2,5 bis +1,5. Bei allen drei Modi ist dabei eine Konzentration der Punkte um den Achsenschnittpunkt zu erkennen.

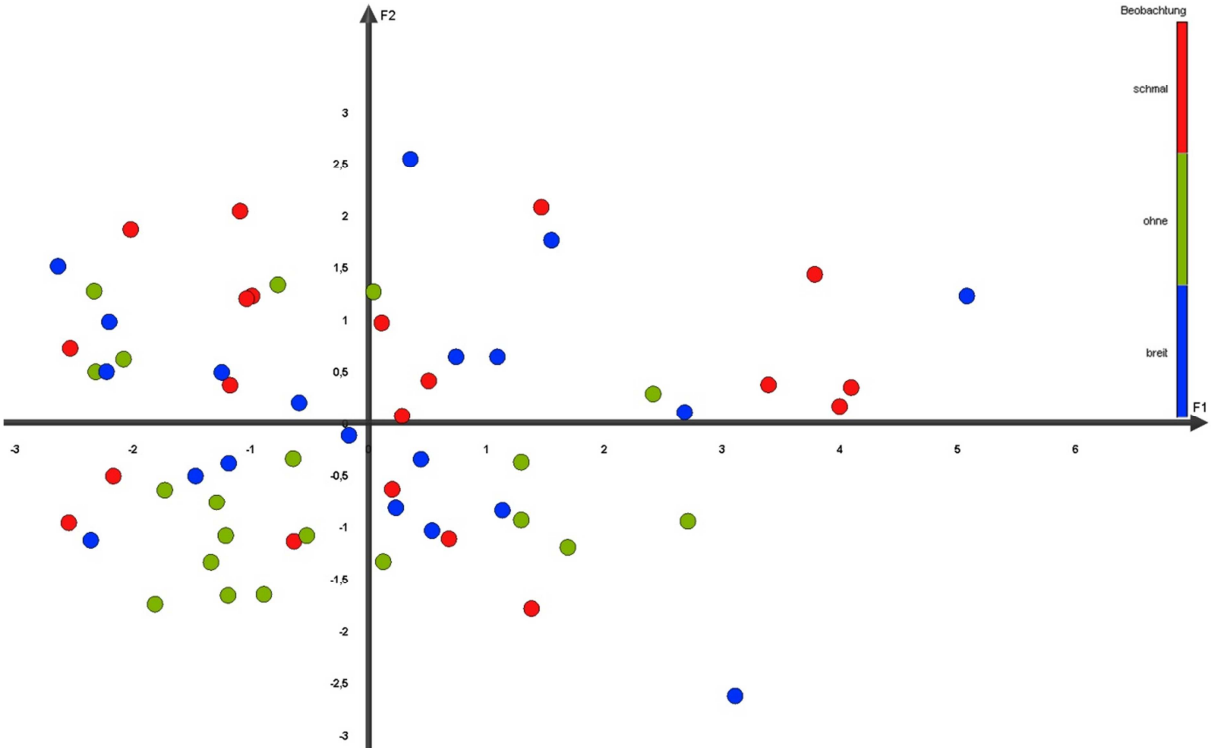


Abbildung 40 Zweidimensionale Darstellung der HKA von dnc, Einteilung nach Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

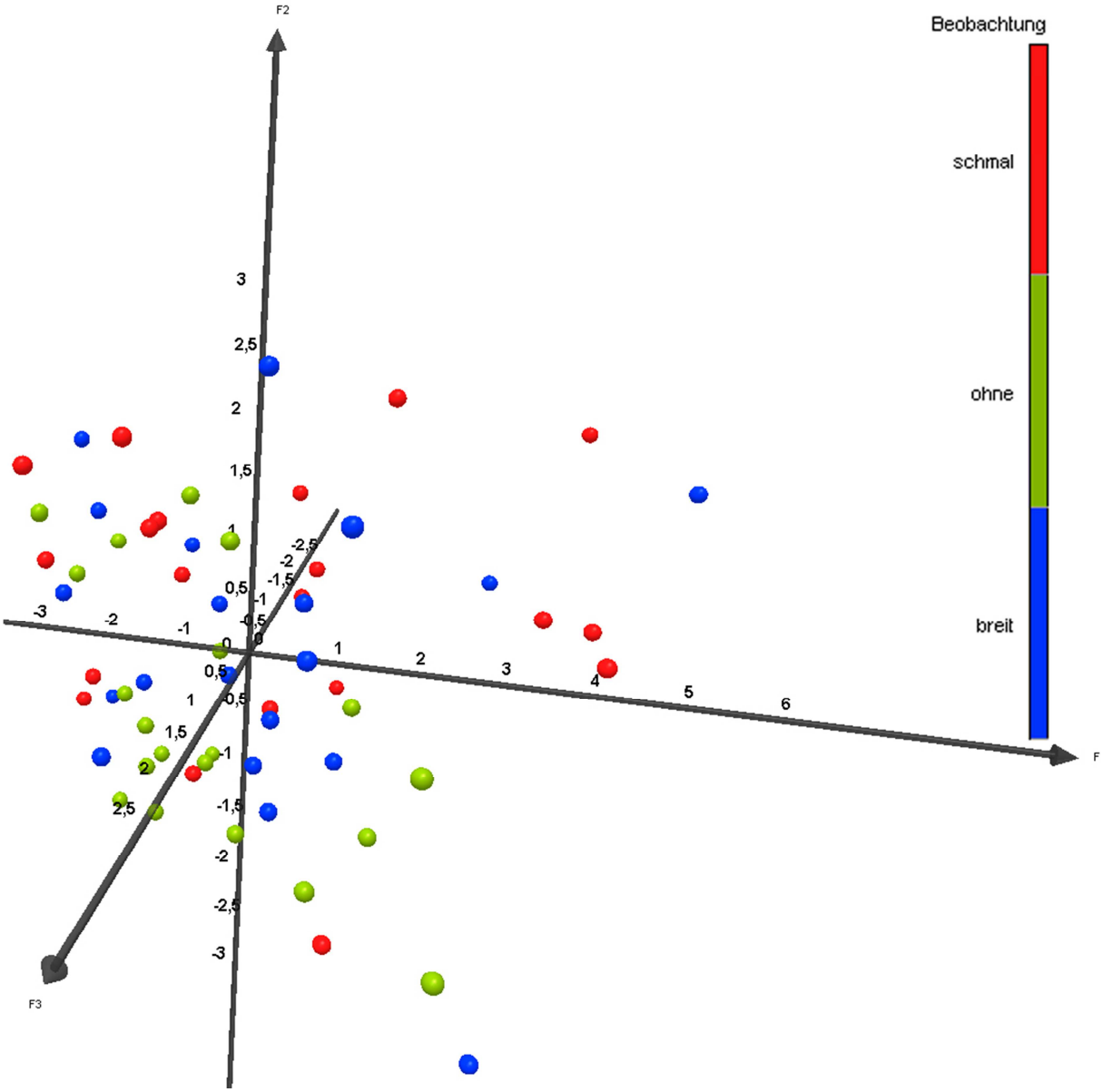


Abbildung 41 Dreidimensionale Darstellung der HKA von dnc, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Zuletzt eine gruppierte Ansicht der Streifentypen (Abb. 42 und 43), hierbei wurden alle Punkte des jeweiligen Streifentyps zu jeweils einem zusammen-gefasst.

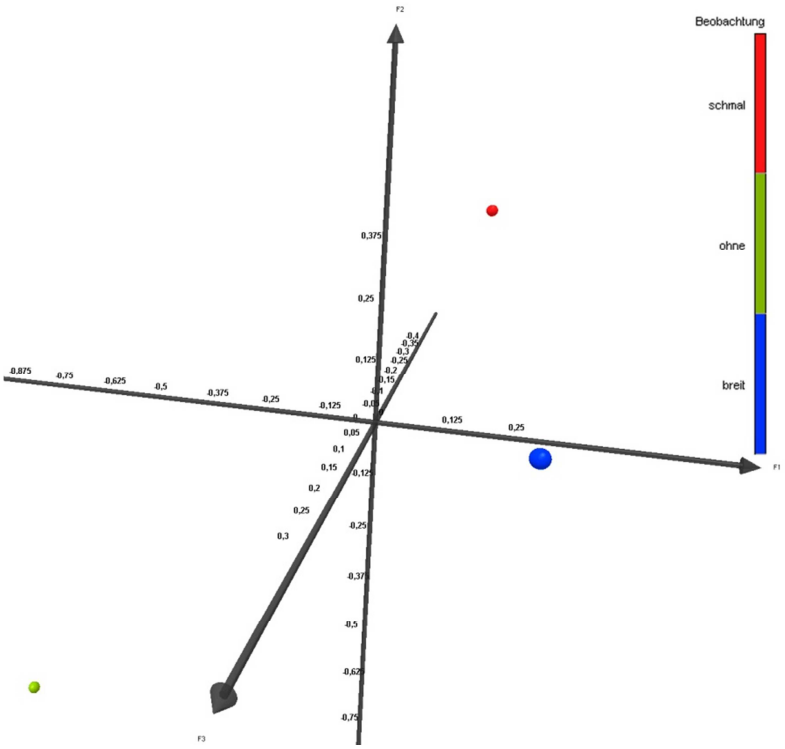


Abbildung 42 Dreidimensionale Darstellung der HKA von dnc, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

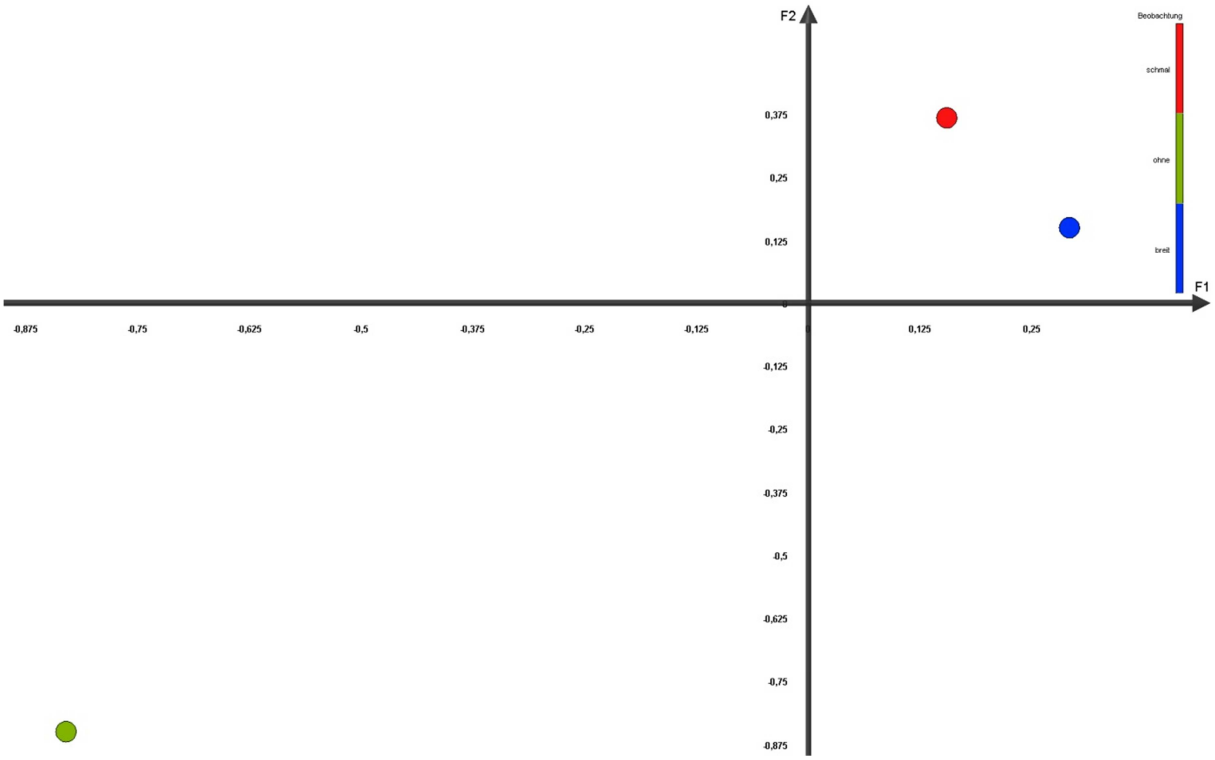


Abbildung 43 Zweidimensionale Darstellung der HKA von dnc, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Hauptkomponentenanalyse RSH

Auch bei rsh sind wie bei der Gesamtanalyse die positiven Korrelationen (Tabelle 19) zwischen Aktivitäts- und Laufvariablen, sowie die Negativen zu den Pausenvariablen erkennbar. Bis auf die Tatsache, dass die Pausenzahl (number_pause) höher positiv mit den restlichen Variablen korreliert, lässt sich kein großer Unterschied zur GesamtHKA feststellen.

Korrelationsmatrix (Pearson (n)):

Variablen	sum_act	median_act	median_pause	number_pause	speeds	distance	walks
sum_act	**1**	**0,779**	**-0,548**	-0,033	**0,791**	**0,933**	**0,624**
median_act	**0,779**	**1**	**-0,263**	**-0,428**	**0,634**	**0,765**	**0,342**
median_pause	**-0,548**	**-0,263**	**1**	**-0,317**	**-0,288**	**-0,414**	**-0,299**
number_pause	-0,033	**-0,428**	**-0,317**	**1**	**-0,268**	-0,245	-0,121
speeds	**0,791**	**0,634**	**-0,288**	**-0,268**	**1**	**0,894**	**0,639**
distance	**0,933**	**0,765**	**-0,414**	-0,245	**0,894**	**1**	**0,695**
walks	**0,624**	**0,342**	**-0,299**	-0,121	**0,639**	**0,695**	**1**

Werte in Fettdruck sind von Null verschieden mit einen Signifikanzniveau von alpha=0,05

Tabelle 19 Korrelationsmatrix nach Pearson für die HKA von rsh. R=1 ist vollständig positiv, r=-1 vollständig negativ und r=0 gar nicht korreliert.

Auch die Eigenwerte (Tabelle 20) zeigen ein Ergebnis, welches dem Gesamtergebnis ähnelt. Fast 90% der Gesamtvariabilität ist auf die ersten drei Hauptkomponenten verteilt (GesamtHKA 86,63%), die erste HKA entspricht mit 59,18% fast genau der ersten HKA der Gesamtanalyse mit 59,19%. Die Eigenwerte sind in der Abbildung 44 eingezeichnet.

Eigenwerte:

	F1	F2	F3	F4	F5	F6	F7
Eigenwert	4,143	1,410	0,686	0,414	0,250	0,073	0,024
Variabilität (%)	59,184	20,146	9,801	5,916	3,572	1,042	0,339
Kummulierter %	59,184	79,330	89,131	95,047	98,619	99,661	100,000

Tabelle 20 Eigenwerte für die berechneten Hauptkomponenten von rsh, Spalten zwei und drei enthalten die prozentuale Umrechnung im Bezug auf Gesamtvarianz und den kumulierten Wert dieser Prozentwerte.

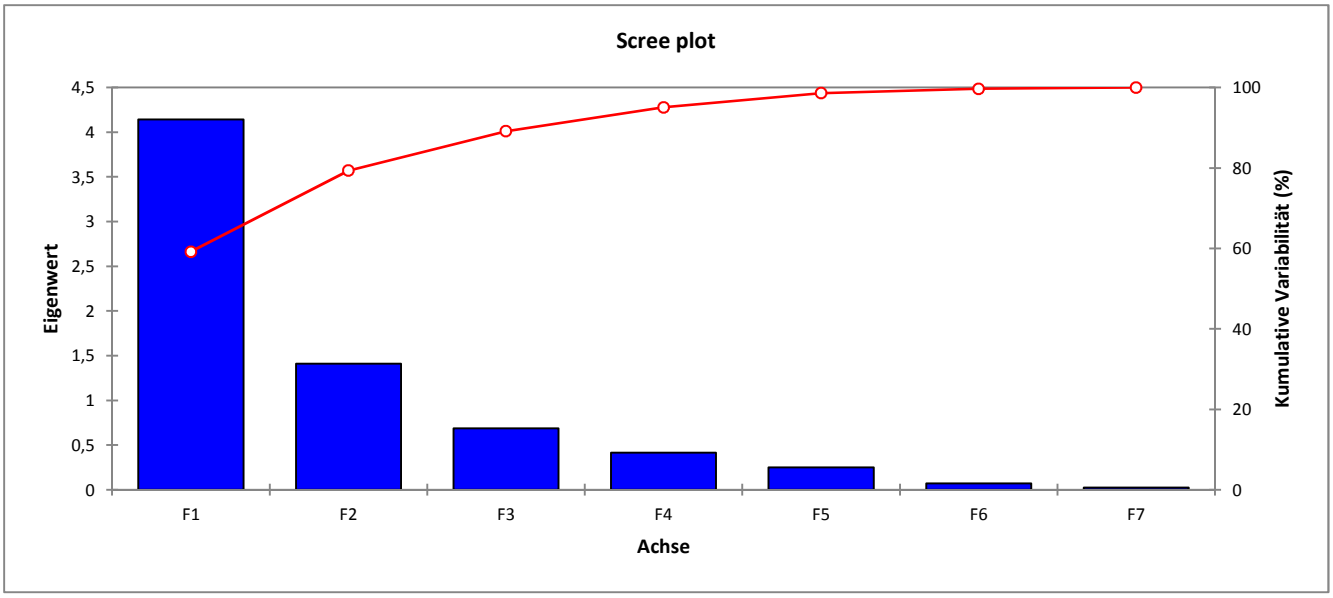


Abbildung 44 Screeplot der HKA von rsh, eingezeichnet sind die Eigenwerte der Hauptkomponenten und der Anteil an der Gesamtvarianz in Prozent

Die Faktorladungen (Tabelle 21) zeigt ebenfalls ein ähnliches Bild, im Vergleich zur GesamtHKA ist die positive Korrelation von sum_act; median_act; speeds; distance; und walks zur ersten Hauptkomponente jedoch stärker. Bei der zweiten Hauptkomponente zeigt sich ein umgekehrtes Bild und die Pausenzeit korreliert negativ, die Pausenzahl hingegen positiv. Außerdem ist eine starke Korrelation von walks zur dritten Hauptkomponente zu erkennen.

Die Darstellung der Faktorenladung ist im Korrelationskreis (Abbildung 45) erkennbar, es fällt auf, dass einige Variablen ihre Position gewechselt haben, die Entfernungen zu den Hauptkomponenten ist jedoch ähnlich. Einzig der Wechsel von number_pause und median_pause ist deutlicher.

Faktorladungen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,948**	0,183	-0,111	0,147	-0,093	0,116	0,106
median_act	**0,808**	-0,281	-0,435	0,091	-0,228	-0,136	-0,027
median_pause	-0,488	**-0,703**	0,286	0,418	-0,102	0,022	0,006
number_pause	-0,251	**0,886**	0,078	0,370	-0,070	-0,053	-0,027
speeds	**0,894**	-0,107	0,156	0,160	0,354	-0,115	0,021
distance	**0,978**	-0,029	0,031	0,087	0,047	0,146	-0,103
walks	**0,730**	0,079	**0,610**	-0,197	-0,216	-0,057	0,000

Tabelle 21 Faktorenladung für die HKA von rsh, gibt die Korrelation zwischen Variablen und Hauptkomponenten an, wobei r=1 vollständig positiv; r=-1 vollständig negativ und r=0 gar nicht korreliert sind.

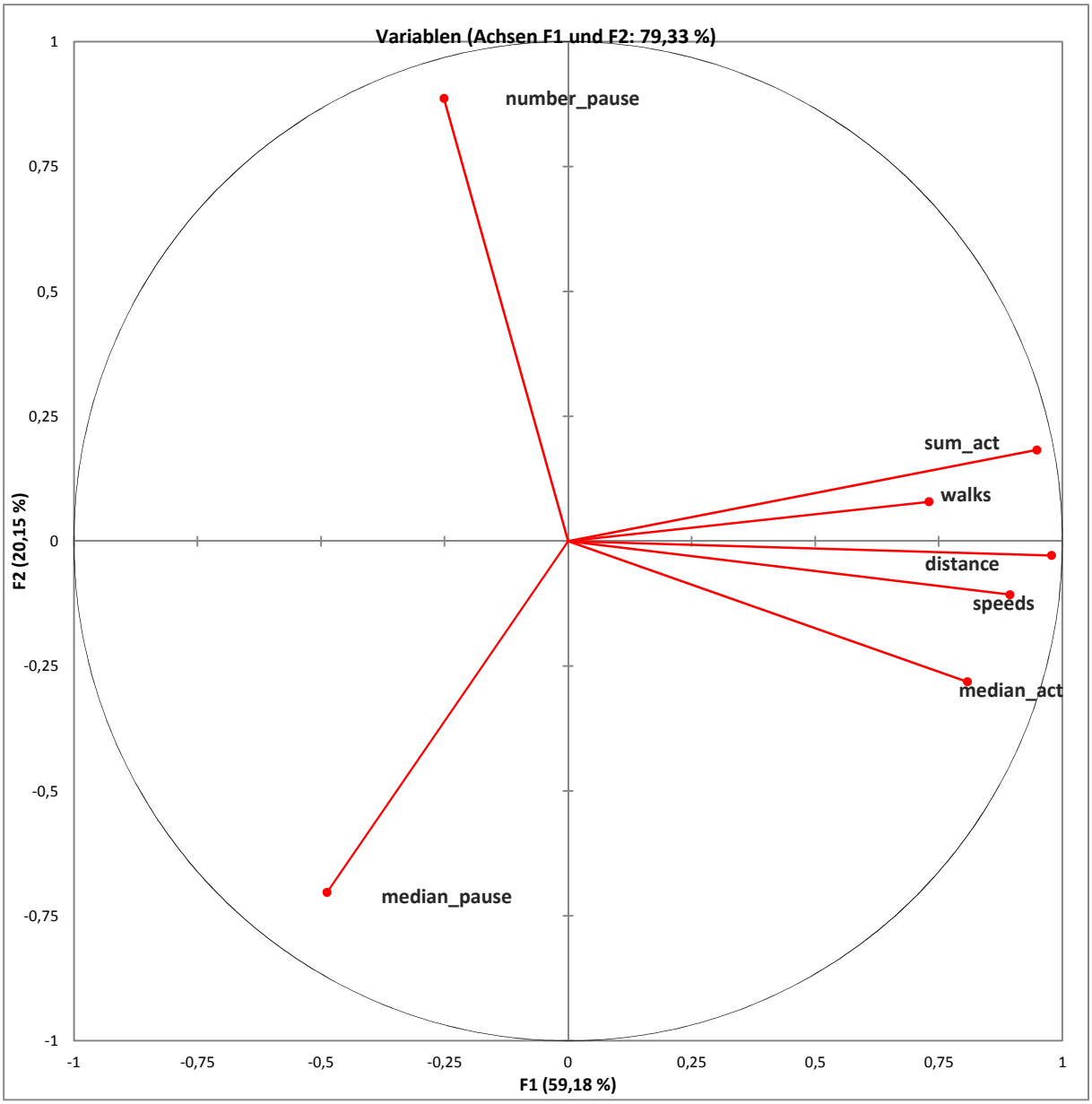


Abbildung 45 Korrelationskreis der HKA von rsh, mit den Achsen F1 (erste Hauptkomponente) und F2 (zweite Hauptkomponente), in Klammern hinter den Achsennamen steht die Variabilität, welche durch diese Achse abgedeckt wird. Eingezeichnet sind die Faktorladungen der einzelnen Variablen, je näher die Variablen der Achse oder anderen Variablen sind, desto höher ist die Korrelation.

Der Einfluss von Variablen auf die Hauptkomponenten (Tabelle 22) entspricht der Gesamtanalyse. Sum_act; median_act; speeds; distance und walks bestimmen die erste und median_pause, sowie number_pause die zweite Hauptkomponente. Wobei die bei der Gesamtanalyse negative Korrelation von number_pause (-0,604) zu einer starken positiven Korrelation (0,785) wurde.

Quadrierte Cosinuswerte der Variablen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,900**	0,033	0,012	0,022	0,009	0,013	0,011
median_act	**0,653**	0,079	0,189	0,008	0,052	0,018	0,001
median_pause	0,238	**0,494**	0,082	0,175	0,010	0,001	0,000
number_pause	0,063	**0,785**	0,006	0,137	0,005	0,003	0,001
speeds	**0,800**	0,011	0,024	0,025	0,125	0,013	0,000
distance	**0,956**	0,001	0,001	0,008	0,002	0,021	0,011
walks	**0,533**	0,006	0,372	0,039	0,047	0,003	0,000

Tabelle 22 quadrierte Kosinuswerte der HKA von rsh, die fett gedruckten Werte für jede Variable entsprechen dem Faktor, für den der quadrierte Kosinus am größten ist und somit den meisten Einfluss einer Variable darstellt.

Die Variablenbeiträge in Prozent (Tabelle 23) sind bei der ersten Hauptkomponente wieder gleichmäßig verteilt, bei der zweiten ist die stärkere Korrelation von number_pause und somit ein Beitrag von 55,69% vorhanden. Ebenfalls auffällig ist der recht hohe Beitrag von walks zur dritten Hauptkomponente.

Beiträge der Variablen (%):

	F1	F2	F3	F4	F5	F6	F7
sum_act	21,712	2,362	1,791	5,225	3,484	18,340	47,086
median_act	15,754	5,607	27,524	2,007	20,785	25,201	3,122
median_pause	5,745	35,031	11,950	42,282	4,149	0,693	0,150
number_pause	1,524	55,689	0,879	33,088	1,985	3,872	2,964
speeds	19,303	0,812	3,551	6,155	50,088	18,264	1,827
distance	23,087	0,058	0,143	1,839	0,899	29,124	44,850
walks	12,875	0,441	54,162	9,404	18,610	4,507	0,001

Tabelle 23 Beiträge der einzelnen Variablen in Prozenten an den Hauptkomponenten von rsh.

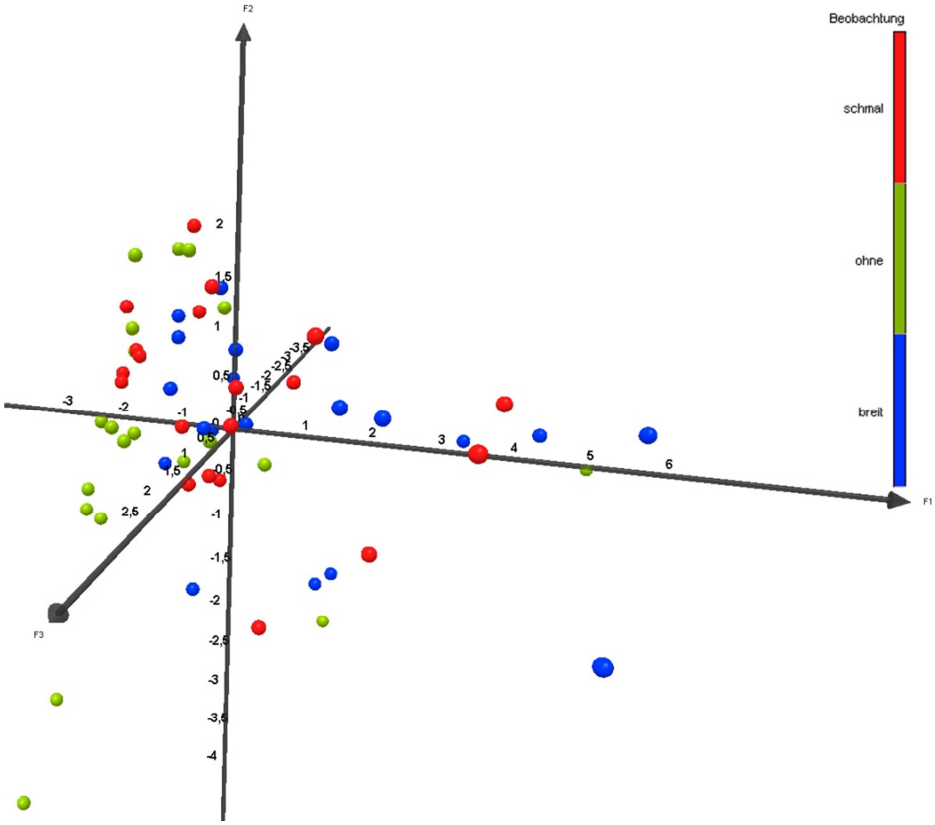


Abbildung 46 Dreidimensionale Darstellung der HKA von rsh, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Im Folgenden die grafische Darstellung der HKA von rsh, in zwei (Abbildung 47) und drei (Abbildung 48) Dimensionen. Da beide Grafiken dieselben Daten zeigen, jedoch das Koordinatensystem um eine Achse erweitert wurde, sind die folgenden Feststellungen auf beide Grafiken zu beziehen mit Ausnahme der Aussagen für die Achse, welche die dritte Hauptkomponente beschreibt. Auf der Achse F1, welche die erste Hautpkomponente beschreibt befinden sich die Werte der drei Streifentypen in unterschiedlichen Abschnitten. Im breitstreifigen Modus kommt es zu einer Streuung im Wertebereich von -1 bis +6, wobei der große Teil der Werte um den Achsenursprung zu finden ist. Der schmalstreifige Modus zeigt Werte im Bereich von -2 bis +4, auch hier sind die meisten Punkte in der Region des Achsenursprungs zu finden. Im streifenlosen Modus befinden sich bis auf drei Ausnahmen alle Datenpunkte im Bereich von -3 bis 0. Auf der Achse F2, welche die zweite Hauptkomponente darstellt befinden sich die Werte des breiten Modus im Bereich von -2 bis +1,5, wobei über 50% im positiven Bereich zu finden sind. Die Datensätze des schmalstreifigen Modus befinden sich im Bereich zwischen -2 und +2, hier sind bis auf 5 alle Datensätze im positiven Bereich. Die Daten des streifenlosen Modus streuen sehr stark und finden sich im Bereich von -4 bis +1,5 und sind dabei über im großen Teil im Bereich von -1,5 bis +1,5 verteilt. Auch bei der Achse F3 ist ein ähnliches Bild erkennbar, der breite Modus ist im Wertebereich von -2 bis 3 zu finden, wobei die meisten Werte um den Achsenursprung zu finden sind. Der schmalstreifige Modus ist auch zum großen Teil um den Achsenursprung verteilt, die Werte finden sich zwischen -0,5 und +2,5. Auch beim streifenlosen Modus sind fast alle Werte um den Achsenursprung gesammelt, die übrigen befinden sich im Bereich von -3,5 bis +1,5.

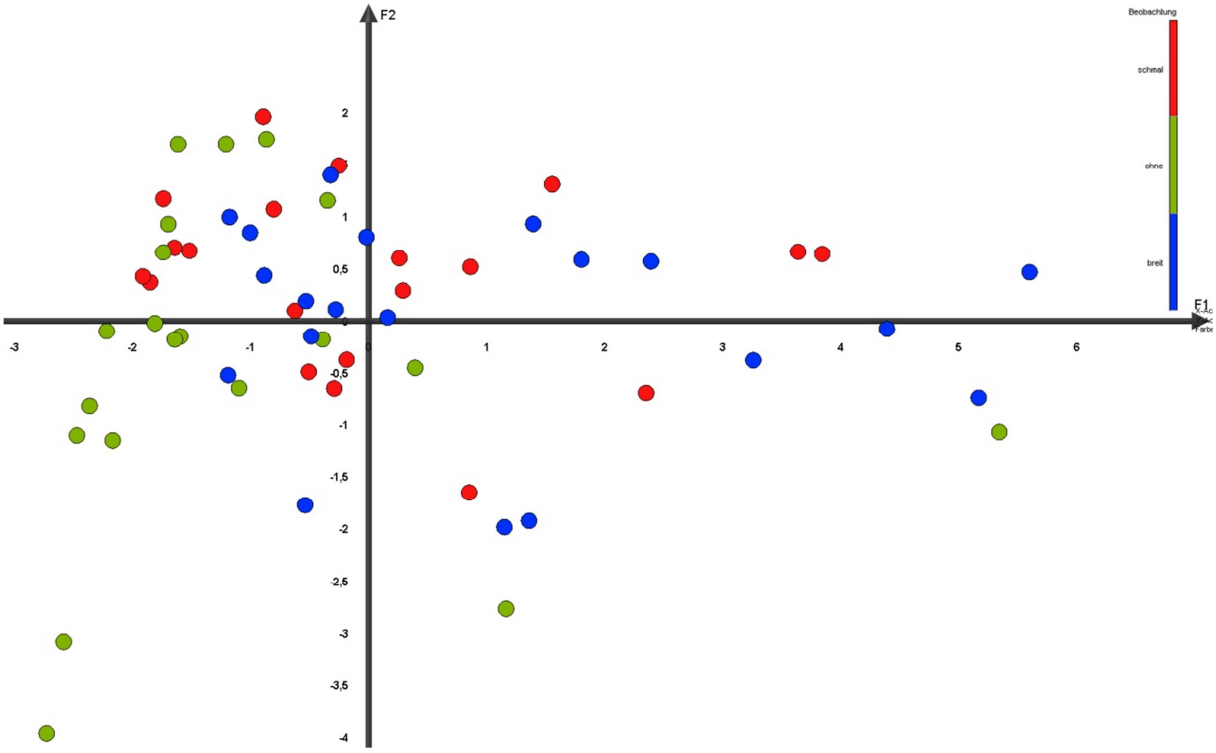


Abbildung 47 Zweidimensionale Darstellung der HKA von rsh, Einteilung nach Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

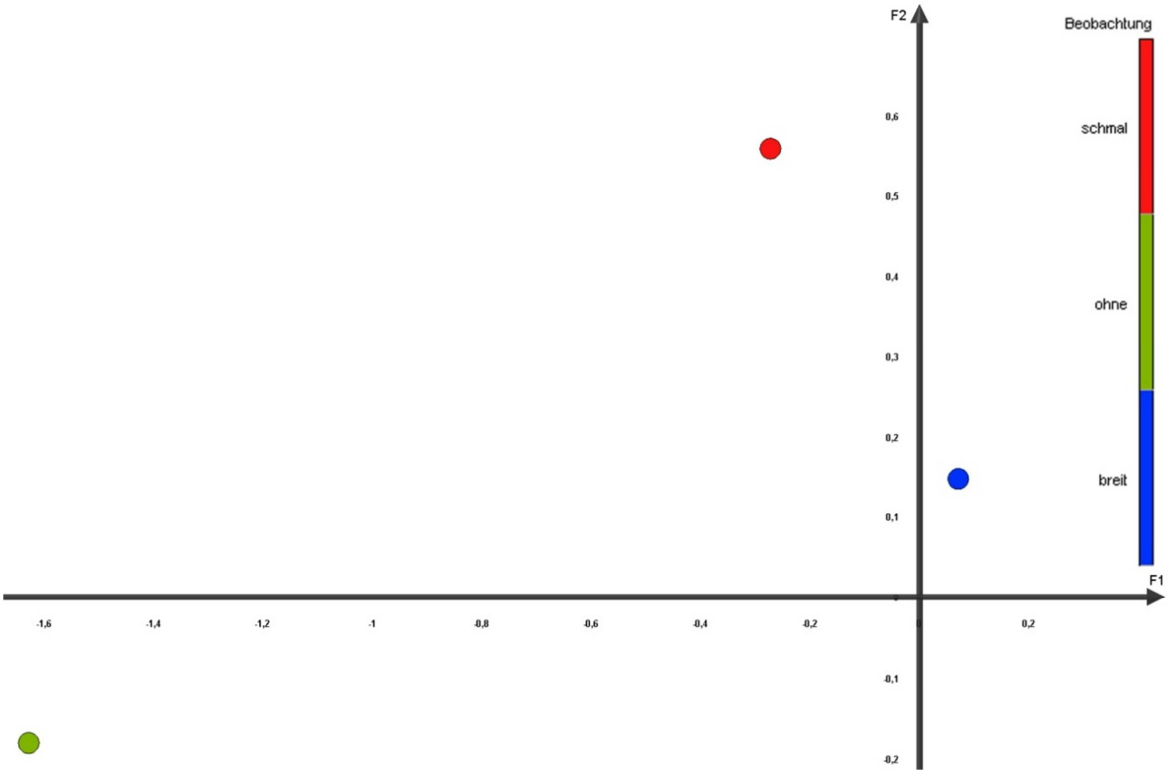


Abbildung 48 Zweidimensionale Darstellung der HKA von rsh, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

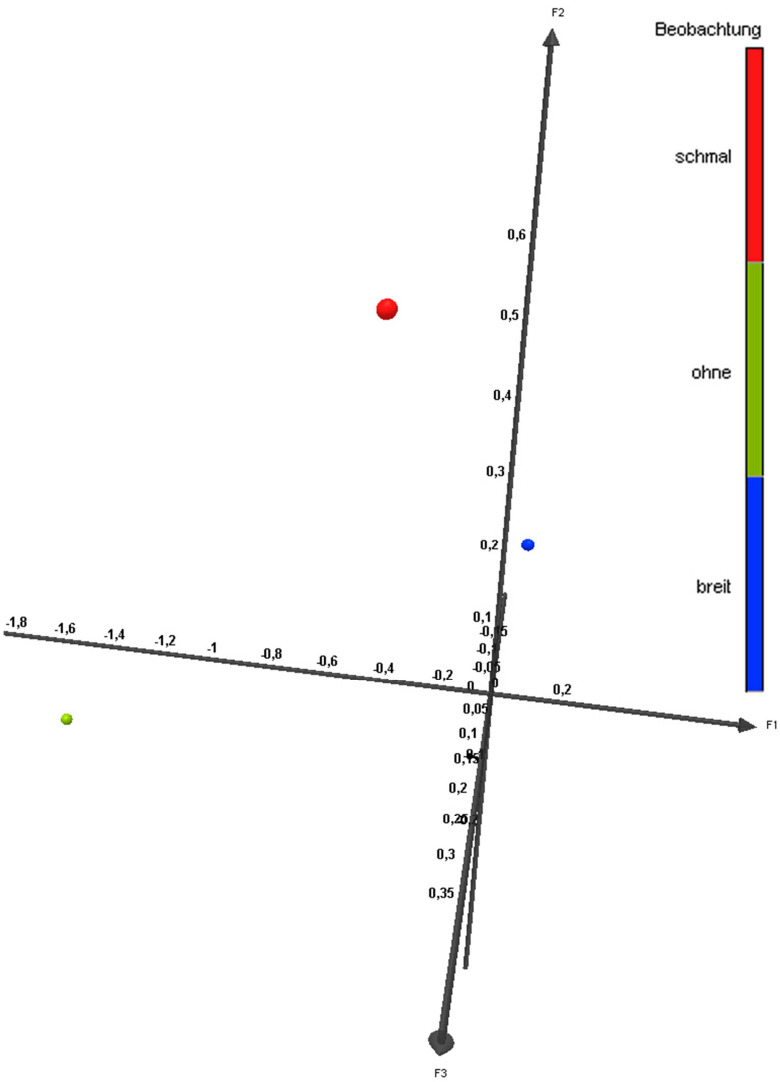


Abbildung 49 Dreidimensionale Darstellung der HKA von rsh, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Hauptkomponentenanalyse RUT1

Die Korrelationsmatrix von rut1 (Tabelle 24) zeigt im Zusammenhang der Pausenvariablen zum Rest ein anderes Bild als in der Gesamtanalyse zu sehen ist. So ist zwar wieder eine recht hohe positive Korrelation zwischen den Aktivitäts- und Laufvariablen erkennbar, jedoch erweitert sich diese positive Korrelation auch auf die Pausenzahl, im Gegensatz zur Gesamtanalyse, bei welcher diese negativ korreliert ist. Median_pause hingegen zeigt ein ähnliches Ergebnis wie in der Gesamtanalyse und ist zu allen andern Werten negativ korreliert.

Korrelationsmatrix (Pearson (n)):

Variablen	sum_act	median_act	median_pause	number_pause	speeds	distance	walks
sum_act	**1**	**0,783**	**-0,537**	**0,315**	**0,633**	**0,880**	**0,390**
median_act	**0,783**	**1**	**-0,493**	0,132	**0,447**	**0,659**	0,253
median_pause	**-0,537**	**-0,493**	**1**	**-0,383**	-0,122	**-0,371**	-0,060
number_pause	**0,315**	0,132	**-0,383**	**1**	-0,107	0,054	-0,109
speeds	**0,633**	**0,447**	-0,122	-0,107	**1**	**0,875**	**0,673**
distance	**0,880**	**0,659**	**-0,371**	0,054	**0,875**	**1**	**0,643**
walks	**0,390**	0,253	-0,060	-0,109	**0,673**	**0,643**	**1**

Werte in Fettdruck sind von Null verschieden mit einen Signifikanzniveau von alpha=0,05

Tabelle 24 Korrelationsmatrix nach Pearson für die HKA von rut1. R=1 ist vollständig positiv, r=-1 vollständig negativ und r=0 gar nicht korreliert.

Die Gesamtvariabiltät von rut1 (Tabelle 25) der ersten drei Hauptkomponenten entspricht mit 86,4% fast exakt derer der Gesamtanalyse mit 86,63%, auffällig ist, dass die erste Hauptkomponente bei rut1 nur 53,72% im Vergleich zu 59,19% hat. Die Eigenwerte sind grafisch im Scree plot (Abbildung 50) aufgetragen.

Eigenwerte:

	F1	F2	F3	F4	F5	F6	F7
Eigenwert	3,760	1,586	0,703	0,483	0,308	0,135	0,026
Variabilität (%)	53,716	22,652	10,039	6,899	4,396	1,933	0,366
Kummulierter %	53,716	76,367	86,407	93,305	97,701	99,634	100,000

Tabelle 25 Eigenwerte für die berechneten Hauptkomponenten von rut1, Spalten zwei und drei enthalten die prozentuale Umrechnung im Bezug auf Gesamtvarianz und den kummulierten Wert dieser Prozentwerte.

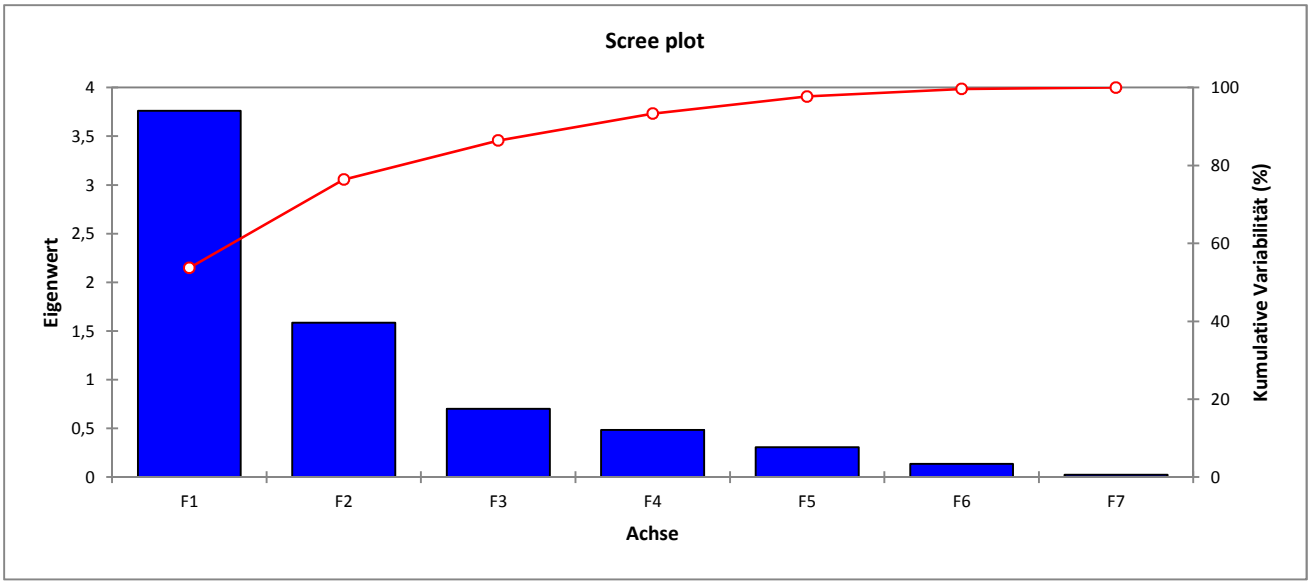


Abbildung 50 Screeplot der HKA von rut1, eingezeichnet sind die Eigenwerte der Hauptkomponenten und der Anteil an der Gesamtvarianz in Prozent

Die Faktorladung (Tabelle 26) zeigt ein völlig anderes Bild als bei der Gesamtanalyse. So reihen sich zwar die Aktivitäts- und Laufvariablen in das bisherige Muster ein, die Pausenvariablen jedoch zeigen stark abweichende Ergebnisse. So ist die negative Korrelation von median_pause viel größer und number_pause ist sogar leicht positiv korreliert mit der ersten Hauptkomponente. Die Korrelation der Variablen mit der zweiten Hauptkomponente unterscheidet sich noch viel stärker, so ändern sich bei beiden Pausenvariablen die Korrelation jeweils von positiv zu negativ und umgekehrt und walks ist im Gegensatz zu einer leicht positiven Korrelation im Gesamten bei rut1 mit -0,527 recht stark negativ korreliert. Dieses Bild wird im Korrelationskreis (Abbildung 51) auch sehr gut dargestellt.

Faktorladungen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,926**	0,219	-0,044	0,176	-0,048	0,227	-0,086
median_act	**0,784**	0,240	-0,411	0,193	0,320	-0,140	0,011
median_pause	**-0,524**	**-0,639**	0,158	0,525	0,125	0,041	0,004
number_pause	0,179	**0,753**	0,602	0,177	0,048	-0,072	0,014
speeds	**0,814**	-0,446	0,111	0,083	-0,264	-0,217	-0,048
distance	**0,961**	-0,172	0,021	0,061	-0,136	0,095	0,124
walks	**0,629**	**-0,527**	0,362	-0,313	0,311	0,031	-0,015

Tabelle 26 Faktorenladung für die HKA von rut1, gibt die Korrelation zwischen Variablen und Hauptkomponenten an, wobei r=1 vollständig positiv; r=-1 vollständig negativ und r=0 gar nicht korreliert sind.

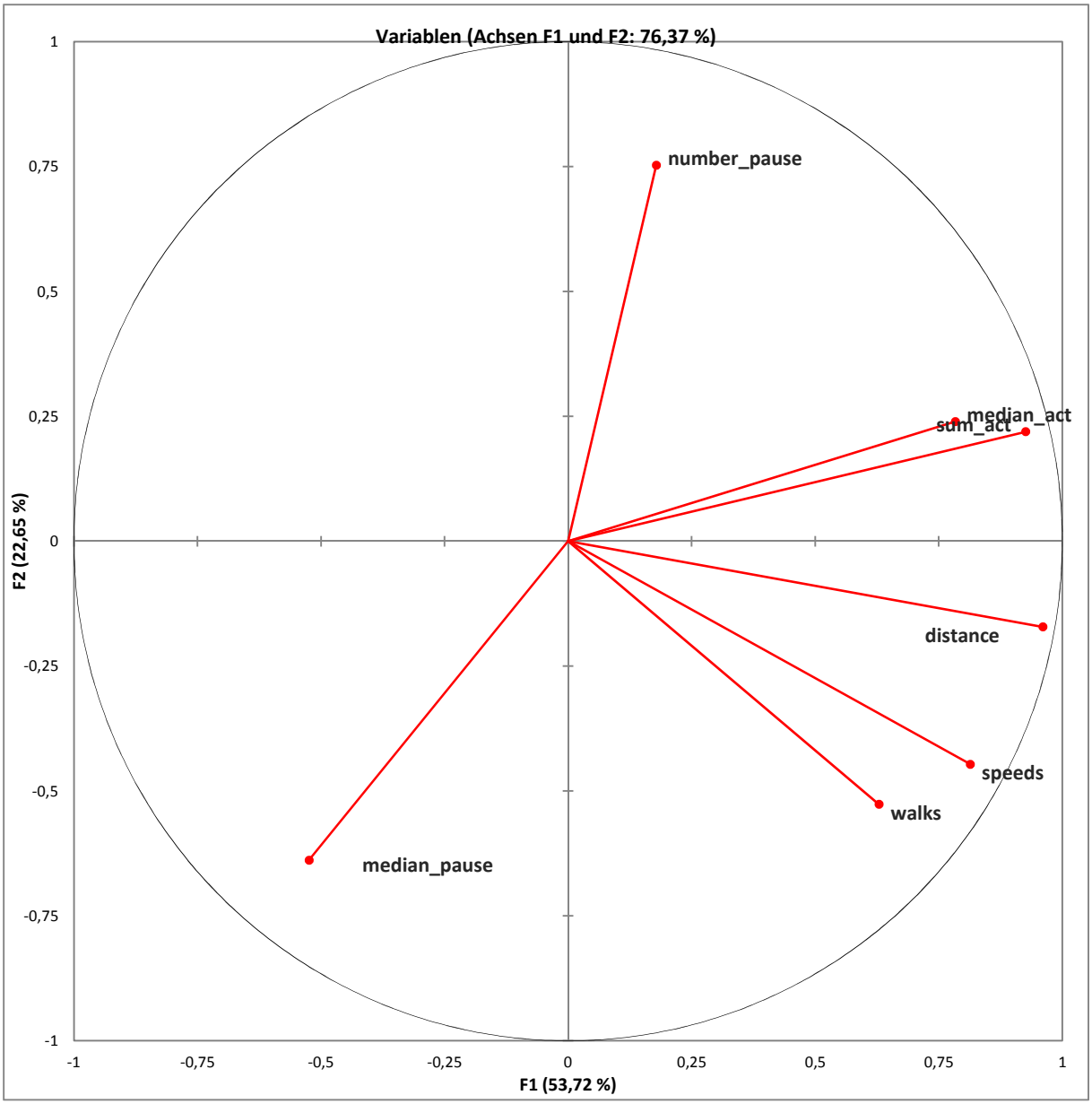


Abbildung 51 Korrelationskreis der HKA von rut1, mit den Achsen F1 (erste Hauptkomponente) und F2 (zweite Hauptkomponente), in Klammern hinter den Achsennamen steht die Variabilität, welche durch diese Achse abgedeckt wird. Eingezeichnet sind die Faktorladungen der einzelnen Variablen, je näher die Variablen der Achse oder anderen Variablen sind, desto höher ist die Korrelation.

Beim Feststellen des Einflusses der Einzelnen Variablen auf die Hauptkomponenten von rut1 (Tabelle 27) zeigt sich wieder ein ähnliches Bild wie bei der Gesamtanalyse, so sind auch hier die Ergebnisse der Pausenvariablen verdreht (die Rohdaten wurden überprüft und sind nicht verdreht). Sonst sind die Ergebnisse nahe der Gesamtanalyse die erste Hauptkomponente wird von sum_act; median_act; speeds; distance und walks bestimmt. Die zweite von median_pause und number_pause. Die prozentuale Verteilung ist ausgewogen und Verteilt sich über die genannten Variablen.

Quadrierte Cosinuswerte der Variablen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,858**	0,048	0,002	0,031	0,002	0,052	0,007
median_act	**0,614**	0,057	0,169	0,037	0,102	0,020	0,000
median_pause	0,275	**0,408**	0,025	0,275	0,016	0,002	0,000
number_pause	0,032	**0,566**	0,363	0,031	0,002	0,005	0,000
speeds	**0,662**	0,199	0,012	0,007	0,070	0,047	0,002
distance	**0,923**	0,029	0,000	0,004	0,019	0,009	0,015
walks	**0,396**	0,277	0,131	0,098	0,097	0,001	0,000

Die fett gedruckten Werte für jede Variable entsprechen dem Faktor, für den der quadrierte Kosinus am größten ist

Tabelle 27 quadrierte Kosinuswerte der HKA von rut1, die fett gedruckten Werte für jede Variable entsprechen dem Faktor, für den der quadrierte Kosinus am größten ist und somit den meisten Einfluss einer Variable darstellt.

Beiträgeder Variablen (%):

	F1	F2	F3	F4	F5	F6	F7
sum_act	22,817	3,028	0,279	6,421	0,735	38,064	28,657
median_act	16,336	3,618	24,065	7,688	33,245	14,556	0,491
median_pause	7,301	25,717	3,575	56,985	5,094	1,261	0,068
number_pause	0,848	35,717	51,625	6,474	0,763	3,818	0,754
speeds	17,611	12,572	1,758	1,424	22,649	34,878	9,108
distance	24,557	1,859	0,066	0,764	6,016	6,718	60,019
walks	10,529	17,489	18,633	20,242	31,499	0,705	0,904

Tabelle 28 Beiträge der einzelnen Variablen in Prozenten an den Hauptkomponenten von rut1.

Im Folgenden die grafische Darstellung der HKA von rut1, in zwei (Abbildung 53) und drei (Abbildung 52) Dimensionen. Da beide Grafiken dieselben Daten zeigen, jedoch das Koordinatensystem um eine Achse erweitert wurde, sind die folgenden Feststellungen auf beide Grafiken zu beziehen mit Ausnahme der Aussagen für die Achse, welche die dritte Hauptkomponente beschreibt. Auf der Achse F1, welche die erste Hauptkomponente beschreibt unterscheiden sich alle Modi relativ stark in ihrer Verteilung. Der breite Modus streut am meisten und ist gleichmäßig verteilt im Wertebereich von -3 bis +5 zu finden, der schmale Modus streut am schwächsten und ist bis einen Ausreiser im Bereich von -3 bis +1 zu Finden. Im streifenlosen Modus verteilen sich die Datenpunkte im Bereich von -4 bis +3. Bei der Achse F2, welche die zweite Hauptkomponente widerspiegelt finden sich die Werte vom breiten Modus im Bereich von -3 bis +2, die des schmalen Modus im Bereich von -2 bis +3 und der streifenlosen Modus im Bereich von -1,5 bis +2,5. Die dritte Achse F3, welche nur in Grafik X (3d) zu sehen ist spiegelt die Werte der dritten Hauptkomponente wieder. Die Werte des breitstreifigen Modus befinden sich im Bereich von -1,5 bis +0,5, die des

schmalstreifigen im Bereich von -1,5 bis +1,5 und die des steifenlosen Modus im Bereich von -2 bis +0,5 (mit einer Ausnahme bei +2).

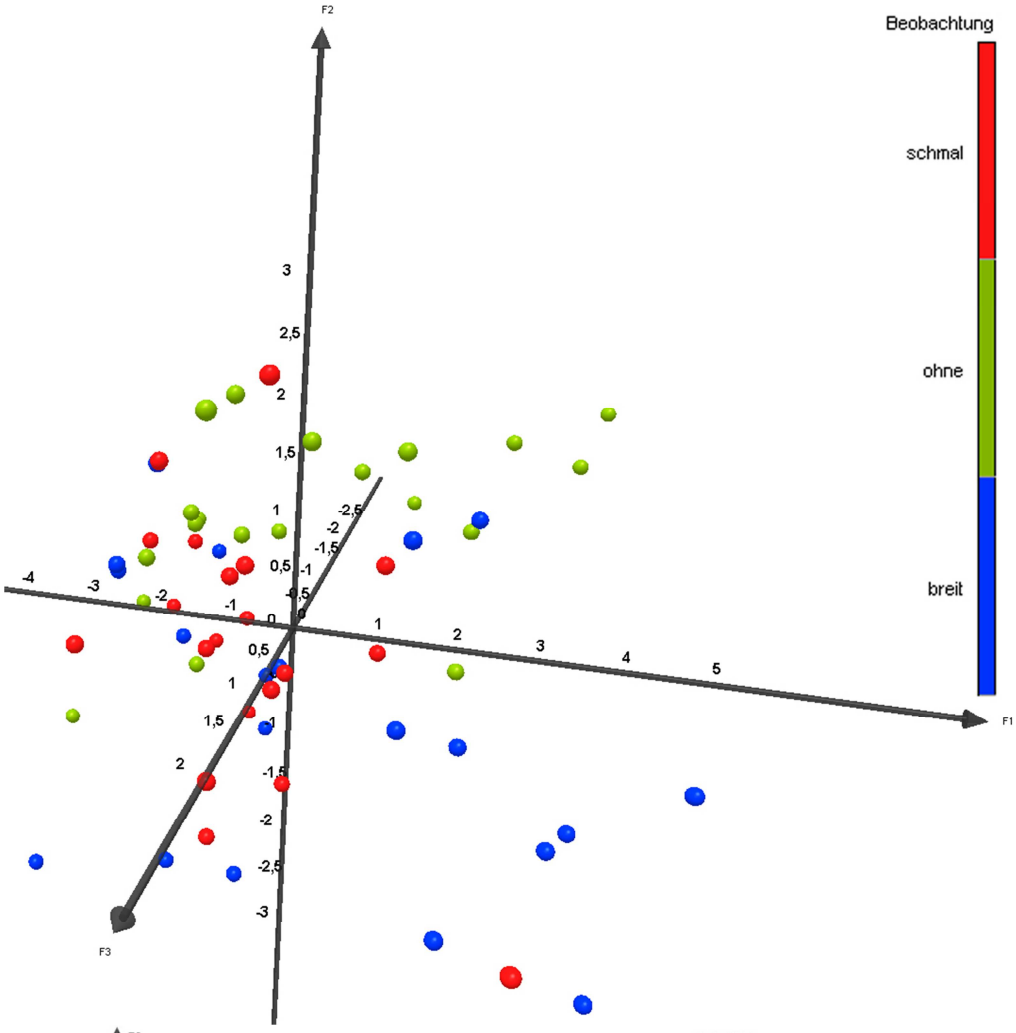


Abbildung 52 Dreidimensionale Darstellung der HKA von rut1, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

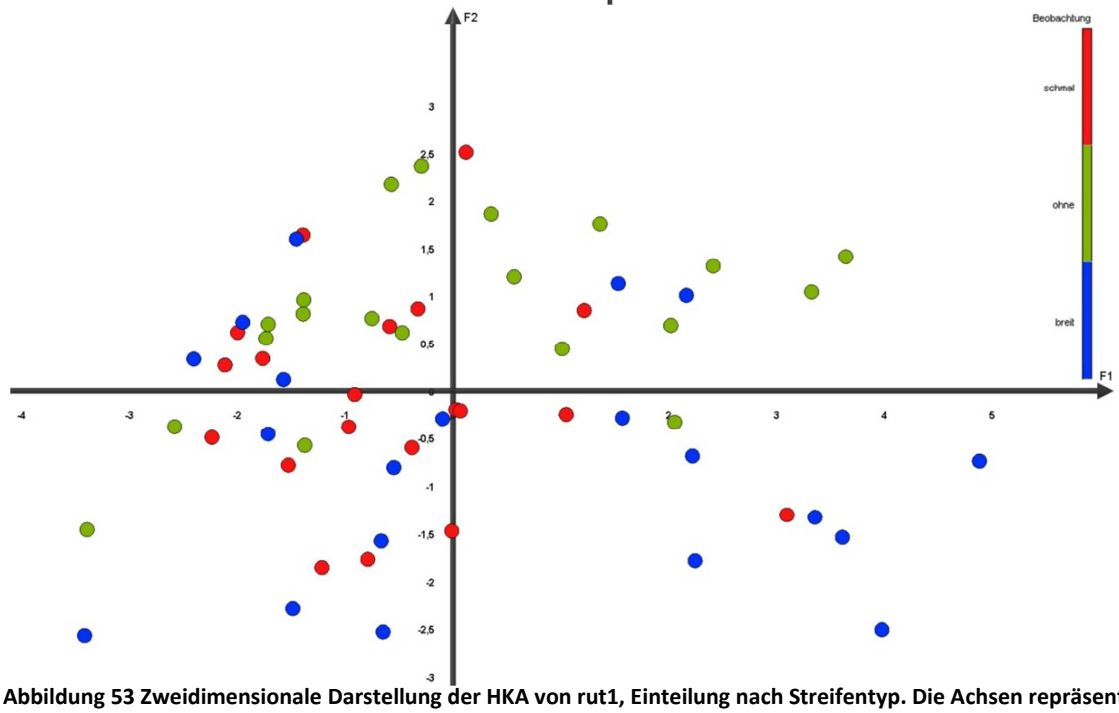


Abbildung 53 Zweidimensionale Darstellung der HKA von rut1, Einteilung nach Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

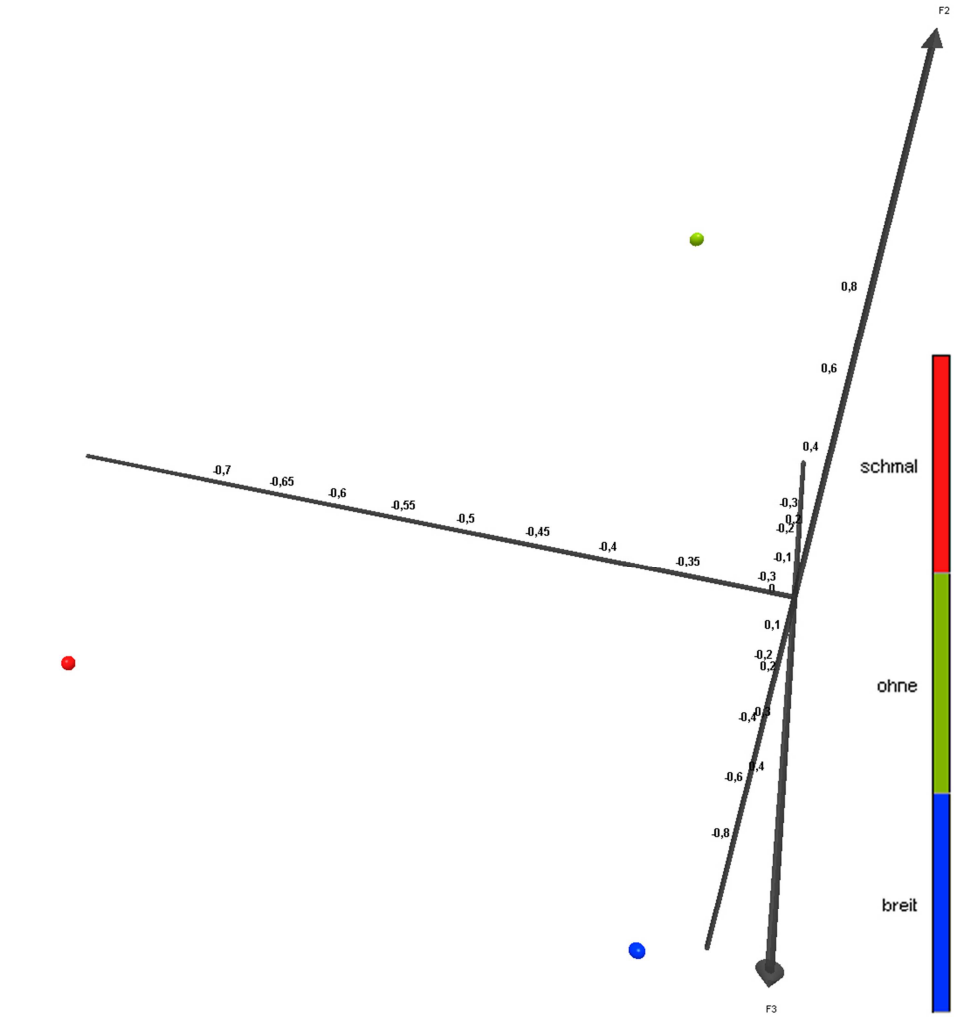


Abbildung 54 Dreidimensionale Darstellung der HKA von rut1, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

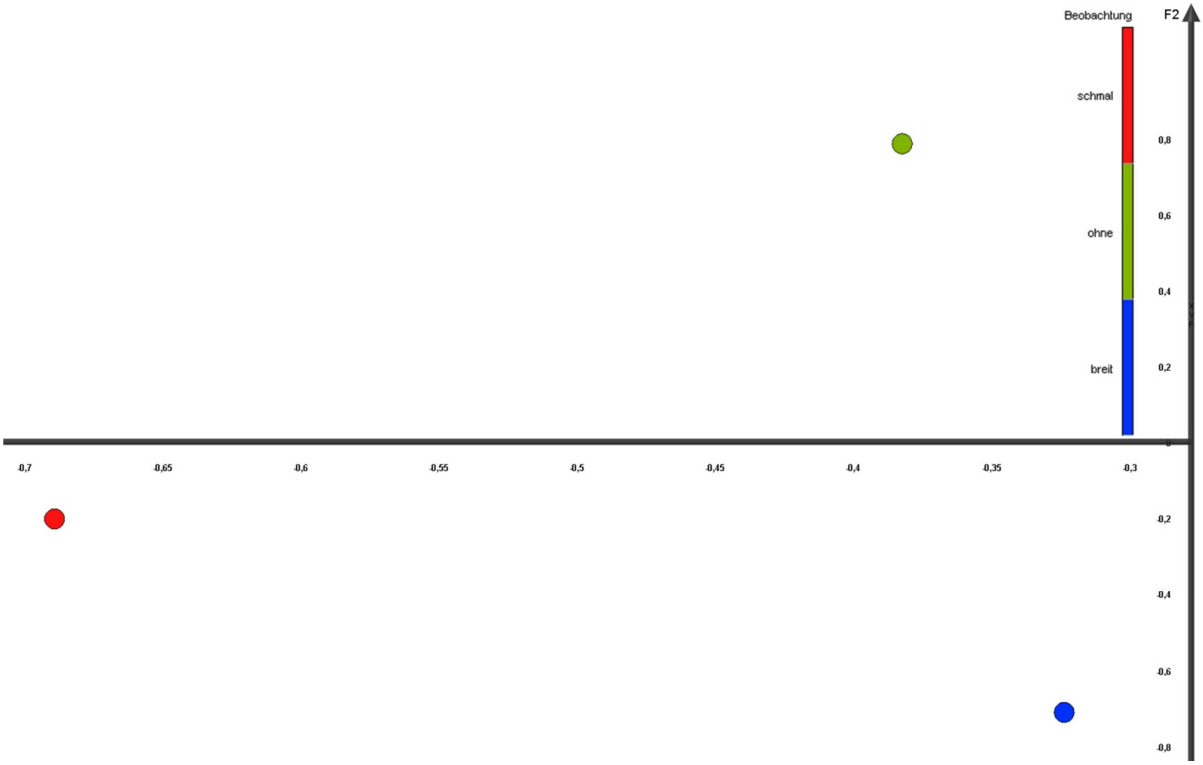


Abbildung 55 Zweidimensionale Darstellung der HKA von rut1, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Hauptkomponentenanalyse RUT2080

Zuletzt die Hauptkomponentenanalyse für rut2080 (Tabelle 29). Bei der Korrelationsmatrix sind kaum Unterschiede zur Gesamtberechnung festzustellen, die Werte schwanken leicht und die positiven Korrelationen innerhalb der Aktivitätsvariablen sind etwas geringer, aber noch immer deutlich im positiv korrelierten Bereich. Die stärksten positiven Korrelationen sind zwischen der Gesamtdistanz und Gesamtaktivität, sowie durchschnittlicher Geschwindigkeit zu erkennen.

Korrelationsmatrix (Pearson (n)):

Variablen	sum_act	median_act	median_pause	number_pause	speeds	distance	walks
sum_act	**1**	**0,715**	**-0,384**	**-0,621**	**0,527**	**0,885**	**0,562**
median_act	**0,715**	**1**	**-0,303**	**-0,517**	**0,402**	**0,653**	**0,407**
median_pause	**-0,384**	**-0,303**	**1**	0,004	0,088	-0,173	0,055
number_pause	**-0,621**	**-0,517**	0,004	**1**	**-0,578**	**-0,716**	**-0,538**
speeds	**0,527**	**0,402**	0,088	**-0,578**	**1**	**0,825**	**0,737**
distance	**0,885**	**0,653**	-0,173	**-0,716**	**0,825**	**1**	**0,787**
walks	**0,562**	**0,407**	0,055	**-0,538**	**0,737**	**0,787**	**1**

Werte in Fettdruck sind von Null verschieden mit einen Signifikanzniveau von alpha=0,05

Tabelle 29 Korrelationsmatrix nach Pearson für die HKA von rut2080. R=1 ist vollständig positiv, r=-1 vollständig negativ und r=0 gar nicht korreliert.

Bei den Eigenwerten (Tabelle 30) ist das Ergebnis fast identisch mit der Gesamtrechnung. Die ersten drei Hauptkomponenten decken 87% (GesamtHKA 86,6%) und die Erste 60,3% (GesamtHKA 59,2%) ab. Visualisiert im Scree plot (Abbildung 56).

Eigenwerte:

	F1	F2	F3	F4	F5	F6	F7
Eigenwert	4,218	1,311	0,558	0,399	0,266	0,231	0,017
Variabilität (%)	60,257	18,730	7,965	5,707	3,802	3,303	0,237
Kummulierter %	60,257	78,987	86,951	92,658	96,460	99,763	100,000

Tabelle 30 Eigenwerte für die berechneten Hauptkomponenten von rut2080, Spalten zwei und drei enthalten die prozentuale Umrechnung im Bezug auf Gesamtvarianz und den kummulierten Wert dieser Prozentwerte.

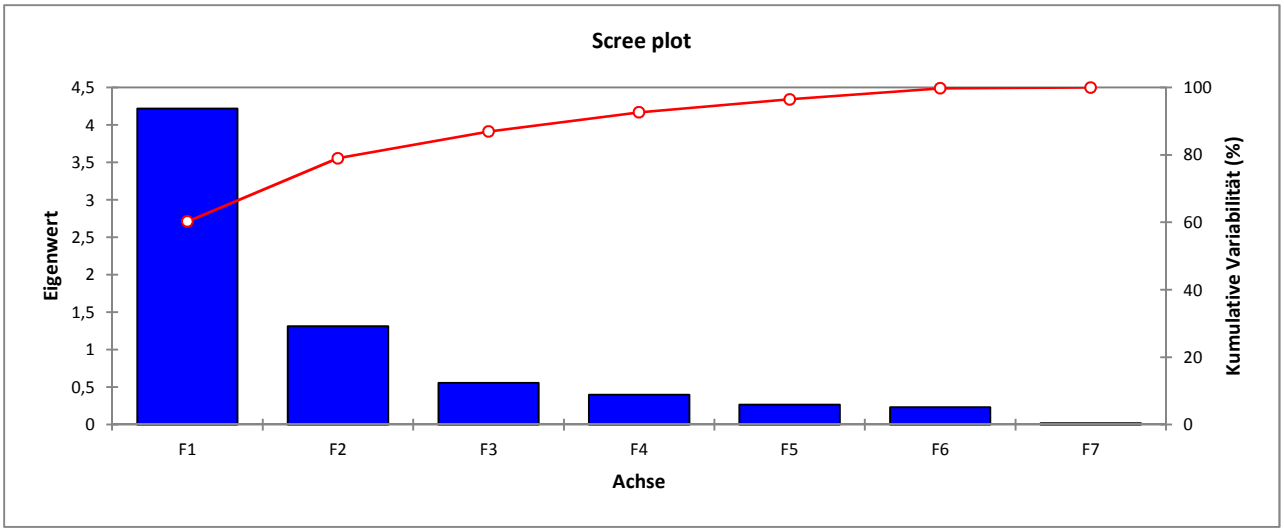


Abbildung 56 Screeplot der HKA von rut2080, eingezeichnet sind die Eigenwerte der Hauptkomponenten und der Anteil an der Gesamtvarianz in Prozent

Die Faktorladung von rut2080 (Tabelle 31) ähnelt dem Gesamtergebnis, so sind wieder die Aktivitäts- und Laufvariablen stark positiv mit der ersten Hauptkomponente korreliert. Den Unterschied zur Gesamtfaktorladung findet man bei number_pause, diese weicht in verschiedenen Punkten ab. Number_pause korreliert stärker negativ mit der ersten Hauptkomponente (-0,784 statt -0,533) und vor allem deutlich schwächer negativ mit der Zweiten (-0,134 statt -0,604). Die restlichen Werte entsprechen weitgehend der Gesamtrechnung und können visualisiert im Korrelationskreis (Abbildung 57) betrachtet werden.

Faktorladungen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,879**	-0,315	-0,034	0,002	-0,129	0,326	-0,060
median_act	**0,740**	-0,364	-0,353	0,405	0,076	-0,162	0,000
median_pause	-0,193	**0,893**	-0,324	0,191	-0,054	0,146	0,000
number_pause	**-0,784**	-0,134	0,400	0,434	0,026	0,134	0,006
speeds	**0,806**	0,386	0,244	0,002	0,375	-0,017	-0,040
distance	**0,979**	0,032	0,099	-0,007	0,018	0,142	0,105
walks	**0,799**	0,339	0,313	0,102	-0,315	-0,198	-0,016

Tabelle 31 Faktorenladung für die HKA von rut2080, gibt die Korrelation zwischen Variablen und Hauptkomponenten an, wobei r=1 vollständig positiv; r=-1 vollständig negativ und r=0 gar nicht korreliert sind.

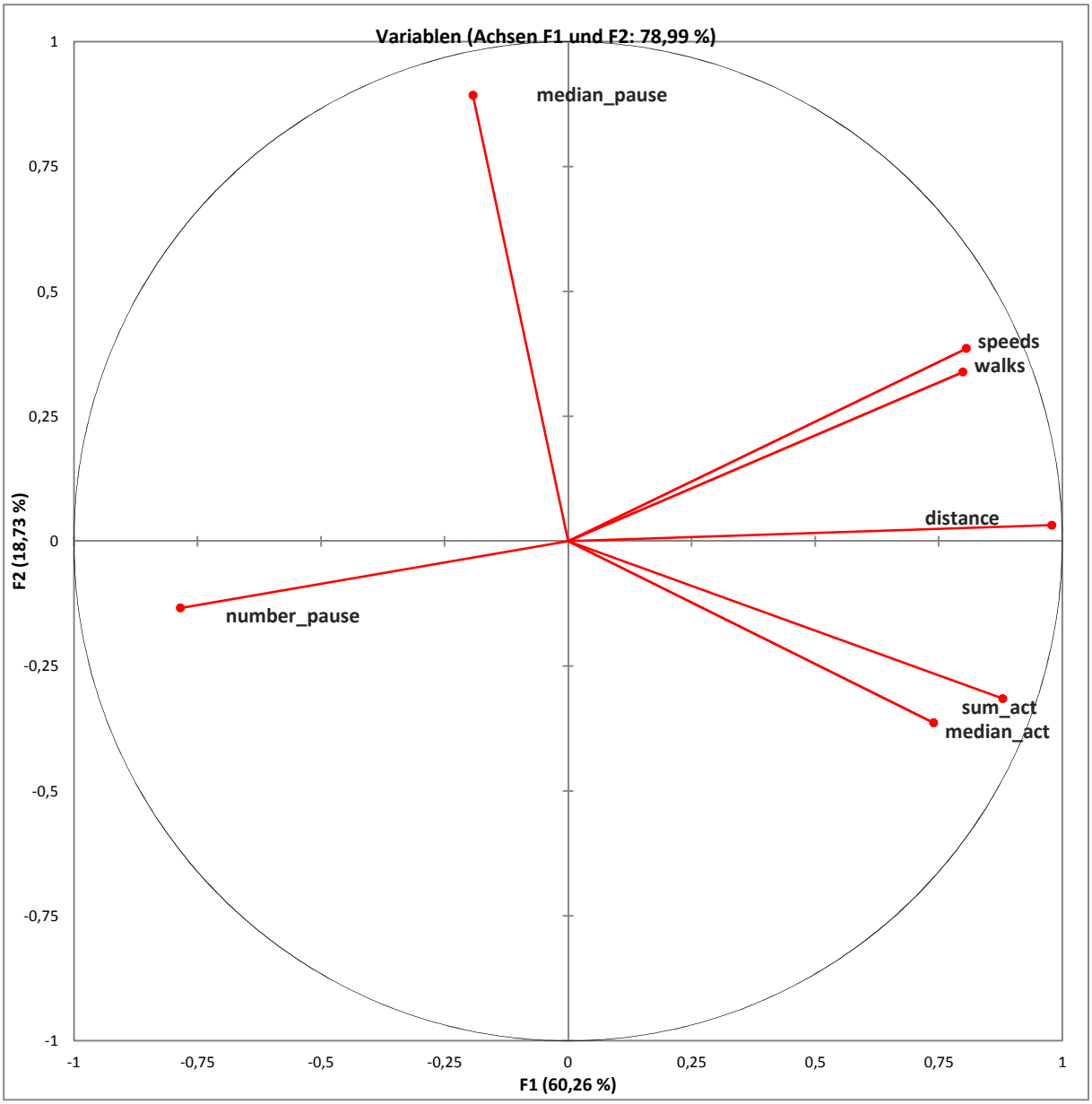


Abbildung 57 Korrelationskreis der HKA von rut2080, mit den Achsen F1 (erste Hauptkomponente) und F2 (zweite Hauptkomponente), in Klammern hinter den Achsennamen steht die Variablität, welche durch diese Achse abgedeckt wird. Eingezeichnet sind die Faktorladungen der einzelnen Variablen, je näher die Variablen der Achse oder anderen Variablen sind, desto höher ist die Korrelation.

Auch bei den quadrierten Cosinuswerten (Tabelle 32) weicht number_pause von der Gesamtanalyse ab, so ist die Korrelation mit der ersten Hauptkomponente am höchsten, dieses hat zur Folge, dass die Erste Hauptkomponente bis auf median_pause von allen Variablen den meisten Einfluss erfährt. Median_pause ist Variable stark positiv mit der zweiten Hauptkomponente korreliert, wobei die Korrelationsstärke nicht groß von der Gesamtanalyse abweicht. Bei der prozentualen Beitrag (Tabelle 33) der Variablen zu den Hauptkomponenten zeigt sich eine gleichmäßige Verteilung unter allen Variablen bis auf median_pause in der ersten und ein hoher Beitrag von 60,8% durch median_pause bei der zweiten Hauptkomponente, wobei dieser Wert sich in gleicher Höhe befindet wie in der

Gesamtanalyse, im Gegensatz zu number_pause, welches bei rut2080 viel geringer ausfällt (1,37% gegenüber 28,34%).

Quadrierte Cosinuswerte der Variablen:

	F1	F2	F3	F4	F5	F6	F7
sum_act	**0,773**	0,099	0,001	0,000	0,017	0,106	0,004
median_act	**0,547**	0,132	0,125	0,164	0,006	0,026	0,000
median_pause	0,037	**0,797**	0,105	0,037	0,003	0,021	0,000
number_pause	**0,615**	0,018	0,160	0,188	0,001	0,018	0,000
speeds	**0,649**	0,149	0,059	0,000	0,140	0,000	0,002
distance	**0,958**	0,001	0,010	0,000	0,000	0,020	0,011
walks	**0,638**	0,115	0,098	0,010	0,099	0,039	0,000

Tabelle 32 quadrierte Kosinuswerte der HKA von rut2080, die fett gedruckten Werte für jede Variable entsprechen dem Faktor, für den der quadrierte Kosinus am größten ist und somit den meisten Einfluss einer Variable darstellt

Beiträge der Variablen (%):

	F1	F2	F3	F4	F5	F6	F7
sum_act	18,334	7,568	0,212	0,001	6,263	45,833	21,788
median_act	12,968	10,082	22,401	41,031	2,175	11,342	0,000
median_pause	0,880	60,800	18,794	9,177	1,083	9,266	0,001
number_pause	14,589	1,365	28,647	47,160	0,262	7,750	0,227
speeds	15,395	11,359	10,656	0,001	52,770	0,124	9,695
distance	22,706	0,079	1,743	0,014	0,119	8,687	66,652
walks	15,129	8,747	17,546	2,616	37,328	16,998	1,636

Tabelle 33 Beiträge der einzelnen Variablen in Prozenten an den Hauptkomponenten von rut2080.

Im Folgenden die grafische Darstellung der HKA von rut2080, in zwei (Abbildung 59) und drei (Abbildung 58) Dimensionen. Da beide Grafiken dieselben Daten zeigen, jedoch das Koordinatensystem um eine Achse erweitert wurde, sind die folgenden Feststellungen auf beide Grafiken zu beziehen mit Ausnahme der Aussagen für die Achse F3, welche die dritte Hauptkomponente beschreibt. Auf der Achse F1, welche die erste Hauptkomponente beschreibt sind die Werte des Breiten Streifenmodus im Bereich von -4 bis +6, des schmalen von -2 bis +4 und des streifenlosen von -4 bis +2. Auf der Achse F2, welche die zweite Hauptkomponente beschreibt befinden sich die Werte des breiten Streifenmodus im Bereich von -2 bis +2,5, des Schmalen von -2 bis +1,5 und die des Streifenlosen von -2,5 bis +0,5. Auf der Achse F3, welche nur in Abbildung 58 erkennbar ist und die dritte Hauptkomponente darstellt befinden sich die Werte des breiten Streifenmodus im Bereich von -1 bis +2,5, die des Schmalen von -1 bis +1 und die des Streifenlosen von -1 bis +0,5 (mit einer Ausnahme bei -3).

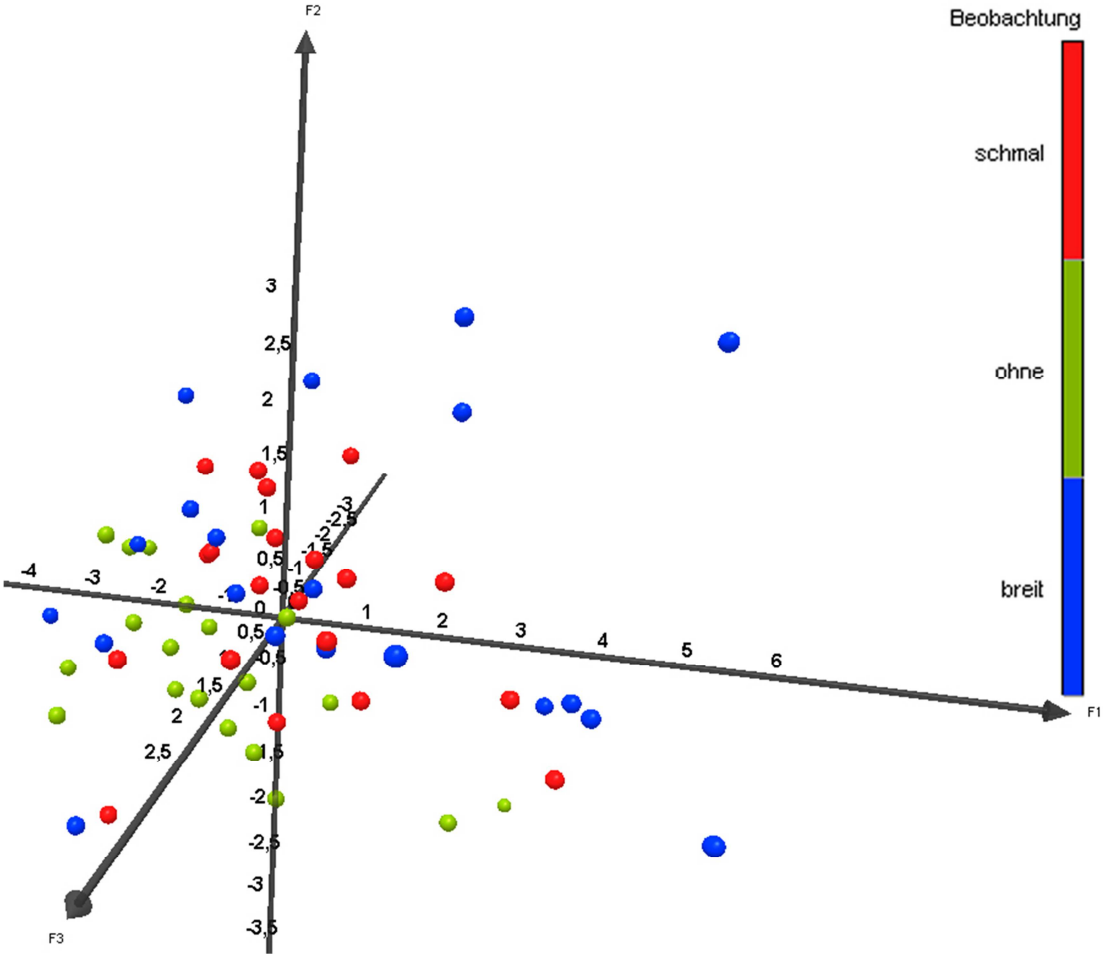
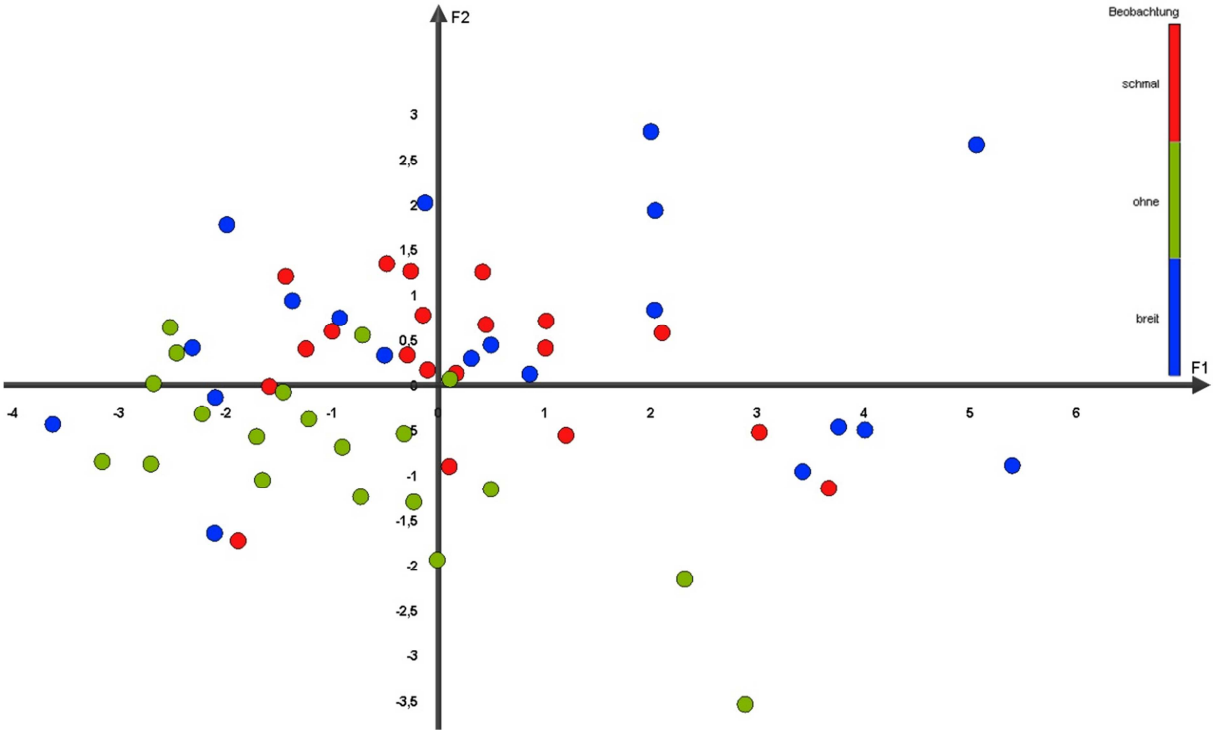


Abbildung 58 Dreidimensionale Darstellung der HKA von rut2080, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.



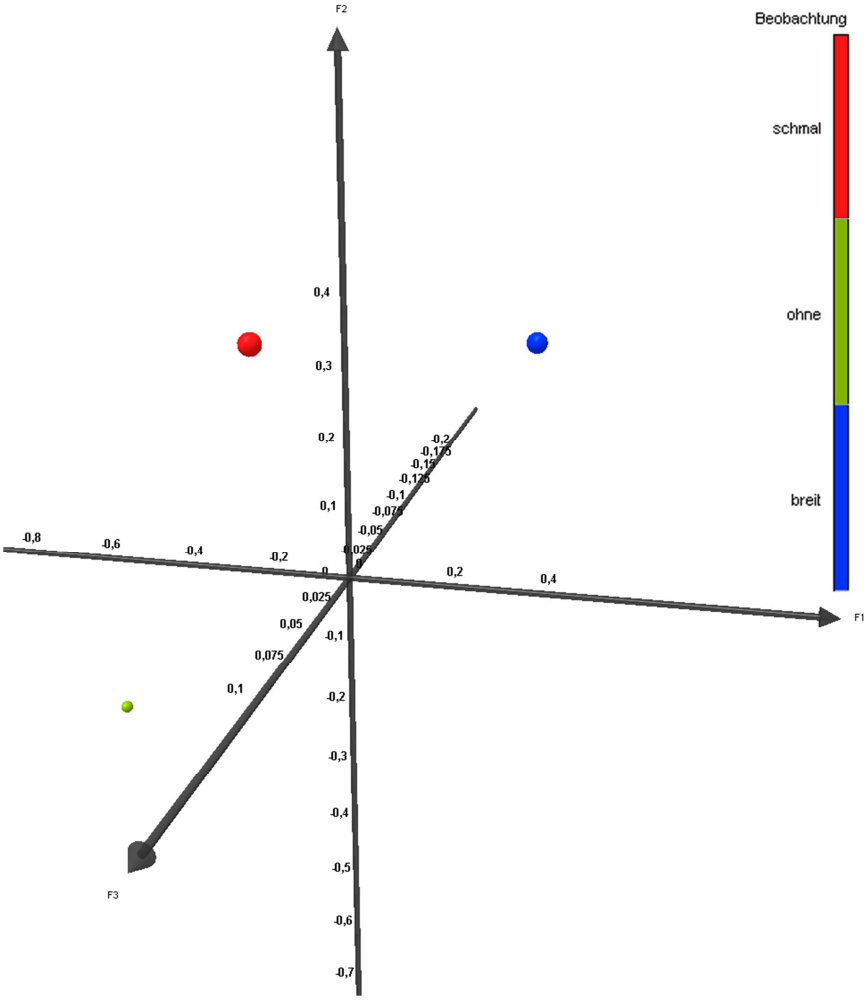


Abbildung 59 Zweidimensionale Darstellung der HKA von rut2080, Einteilung nach Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet sind die Werte der einzelnen Fliegen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Abbildung 60 Dreidimensionale Darstellung der HKA von rut2080, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2; F3=HK3. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

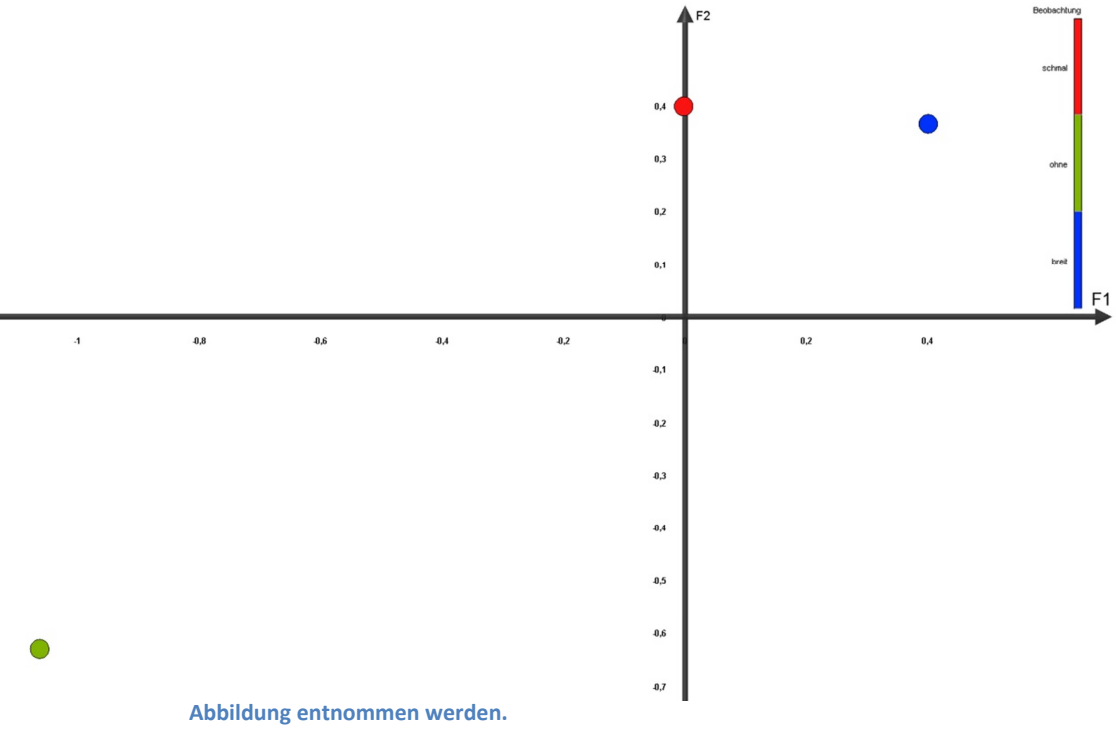


Abbildung 61 Zweidimensionale Darstellung der HKA von rut2080, Einteilung Streifentyp. Die Achsen repräsentieren die jeweiligen Hauptkomponenten F1 = HK1; F2=HK2. Eingezeichnet ist der Median aller Einzelkoordinaten des jeweiligen Streifentypen. Die genaue Zuordnung kann der Legend in der Abbildung entnommen werden.

Diskussion

Wenn man die Resultate im Gesamten betrachtet lassen sich schnell einige Muster erkennen. Schnell fällt auf, dass dnc quasi streifenunabhängig agiert. Einzig bei der Pausenzahl lässt sich ein Unterschied zwischen den Streifenmodi feststellen, so ist diese im streifenlosen Modus erhöht. (Abbildung 16). In den visualisierten Laufplots (Abbildung 22) kann man erkennen, dass dnc sowohl bei schmalen als auch breiten Streifen einen relativ festen Weg zwischen den Streifen einschlägt und sich mittig zwischen den Streifen bewegt, beim streifenlosen Modus ist kein Laufmuster zu erkennen. Was die Aktivitäts- und Pausenwerte angeht, so liegt dnc meist etwas über rut1, welches das Schlusslicht bildet. Umso erstaunlicher ist es, dass die Gesamtvarianz von dnc in der ersten Hauptkomponente nur 51,348% und zusätzlich mit der Zweiten und Dritten 82,691% erreichet (Tabelle 15), da aufgrund der ziemlich homogenen Verteilung der Werte unter den Streifenmodi eine hohe Redundanz zu erwarten wäre, durch diese Varianzverschiebung ist es auch zu erklären, dass median_act mit r=0,7 relativ stark mit der dritten Hauptkomponente korreliert (Tabelle 16). Im Verhältnis zur GesamtHKA (Tabelle 4) ist die Korrelation der Aktivitätsvariablen mit der der ersten Hauptkomponente auch geringer, jedoch nicht stark, so ergibt sich ein ähnliches Bild der Variablenkorrelationen (Abbildung 39). Der Korrelationskreis zeigt, dass die beiden Pausenvariablen negativ zueinander und wenig mit den restlichen Variablen korreliert sind, dieses ist im Bezug auf median_pause nicht weiter verwunderlich, so zeigt diese Variable unabhängig von Streifen- und Genotyp kaum eine Änderung (Abbildung 13) da sich die Daten für number_pause bei dnc kaum unterscheiden ist auch hier keine Korrelation zu erwarten.

Bei dem Einfluss der Variablen auf die Hauptkomponenten fallen im Vergleich zur Gesamtanalyse zwei Sachen auf, zum Einen sind die Korrelationen der Variablen geringer (Tabelle 17), was durch die hohe Varianzverteilung zu erwarten ist zum Anderen hat die Variable median_act wie durch die Faktorladung schon zu ahnen ist einen sehr hohen Einfluss auf die dritte Hauptkomponente und bestimmte diese zu 54,744% (Tabelle 18), was fast 20% höher ist, als bei der Gesamtanalyse.

Grafisch lässt sich das Bild von dnc sehr gut einordnen, auf allen Achsen sind die Werte recht stark am Achsenkreuz gruppiert, auf der Achse F1 (Abbildung 40 bzw 41) eher in niedrigen bis negativen Bereich verteilt, was die geringe Aktivität aus Abbildung 7 wiederspiegelt. Auch im Bezug auf die grafische Auswertung der GesamtHKA (Abbildung 28) zeigt sich die starke Gruppierung um den Achsenursprung, welche mit den geringen Wertunterschieden der Resultate zusammenhängt.

Die gruppierten grafischen Auswertungen (Abbildungen 42 und 43) zeigen auch den großen Einfluss von number_pauses auf. Im Streifenlosen Modus war der Pausenwert höher als in den beiden anderen Modi (Abbildung 16), wodurch die starke Entfernung des Datenwertes für diesen Modus begründen kann.

Rut2080 und cstp kristallisieren sich bereits bei den frühen Resultaten als aktivste Genotypen heraus, so zeigt cstp im schmalen Streifenmodus mit 500s (Abbildung 3) die meiste Gesamtaktivität aller Genotypen, jedoch sind hier die anderen beiden Streifentypen mit 400s deutlich weniger aktiv. Rut2080 hingegen zeigt bei allen Streifentypen eine mit 450s sehr hohe Aktivität (Abbildung 5), dieses durchgehend aktive Verhalten von rut2080 weitet sich auch auf die mittlere Zeit der Einzelaktivitäten (Abbildung 12) aus, wobei rut2080 hier wieder mit Abstand am aktivsten ist. Interessant ist, dass CSTP bei den schmalen Streifen wieder hohe Aktivitätswerte zeigt und nah an rut2080 herankommt, jedoch in den anderen Streifenmodi deutlich weniger aktiv (Abbildung 9) ist. Es hat den Anschein, als würden die rut2080 Fliegen den Unterschied zwischen breiten und schmalen Streifen nicht zu erkennen und dabei zusätzlich eine sehr hohe Gesamtaktivität zu besitzen. Dieses setzt sich ebenfalls in den Geschwindigkeiten (Abbildung 17) und Läufen zwischen den Streifen (Tabelle 3) fort. Auch das Verhalten von cstp setzt sich in dem Sinne fort, dass der schmale Streifenmodus der aktivste ist, so zeigt cstp mit 120 Pausen (Abbilding 14) die wenigsten Pausen aller Genotypen und mit 5800mm die längste Gesamtdistanz (Abbildung 19), auch höher als die von rut2080, jedoch jeweils nur auf den schmalstreifigen Modus angewandt.
Diese Verbindung zwischen den Beiden Genotypen setzt sich auch bei der HKA fort.

Beide Genotypen zeigen eine sehr ähnliche Korrelationematrix (Tabellen 9 für cstp und 29 für rut2080), bei welcher die Aktivitäts- und Laufvariablen positiv untereinander korreliert und negativ mit den Pausenvariablen korreliert sind, bei welcher aber auch die Pausenvariablen untereinander negativ korreliert sind. Dieses entspricht auch der GesamtHKA (Tabelle 4). Auch bei den Eigenwerten gibt es keine großen Unterschiede, so deckt rut2080 genauso wie die GesamtHKA 86% der Gesamtvarianz mit den ersten drei Hauptkomponenten ab (Tabelle 5 und Tabelle 30), cstp (Tabelle 11) ist mit 90,15% etwas höher und hat somit auch die höchste Gesamtvarianz aller Genotypen mit drei Hauptkomponenten erreicht.

Bei der Faktorladung kommt es zu ersten auffälligen Unterschieden, die Faktorladung weicht in verschiedenen Punkten ab. Der Hauptunterscheid zwischen rut2080 und cstp liegt darin, dass median_act von cstp mit r=0,513 (Tabelle 11) im Gegensatz zu r=-0,353 (Tabelle 31) von rut2080 sehr stark positiv mit der dritten Hauptkomponente korreliert, die weiteren Werte ähneln bei beiden

Genotypen, rut2080 weist lediglich eine geringere Korrelation von sum_act zur ersten Hauptkomponente auf und die negative Korrelation von number_pause zur zweiten Hauptkomponente ist geringer. In einem Punkt unterscheiden sich aber beide von der GesamtHKA, denn bei dieser ist die negative Korrelation von number_pause mit -0,533 (Tabelle 6) gegenüber ca. -0,750 bei den beiden Genotypen im Bezug auf die erste Hauptkomponente geringer. Dieses zeigt sich auch prompt im Korrelationskreis, während der Korrelationskreis von cstp (Abbildung 33) dem der GesamtHKA ähnelt (Abbildung 27) fällt bei rut2080(Abbildung 57) sofort die stärkere Korrelation von number_pause zur zweiten Hauptachse und die entgegengesetzte Korrelation zu distance auf.

Auch beim Einfluss der Variablen auf die Hauptkomponenten zeigen rut2080(Tabelle 32) und cstp(Tabelle 12) ein sehr ähnliches Ergebnis und weichen dabei von der GesamtHKA(Tabelle 7) ab. Während bei der GesamtHKA die Zuteilung der Aktivitäts- und Laufvariablen zur Ersten und der Pausevariablen zur zweiten Hauptkomponenten vorgenommen werden kann ist sowohl bei rut2080, als auch bei cstp der Einfluss von number_pause auf die erste Hauptkomponente deutlich höher als auf die zweite.

Auch bei der grafischen Auswertung beider Genotypen sind die genannten Ähnlichkeiten zu finden, dieses lässt sich auf den ersten Blick in den gruppierten Ansichten aller Fliegen (Abbildungen 30 und 31) ausmachen, vor allem auf den Achsen F1 und F2 finden sich rut2080 und cstp in den höheren Werten, was auf eine hohe Aktivität(F1) schließen lässt und durch den auf den hohen Einfluss der Pausenvariable number_pause hindeutet, welche bei rut2080 und cstp besonders viel Einfluss auf die zweite Hauptkomponente hat. Die Streifenverteilungen von cstp (Abbildungen 34 und 35) und von rut2080 (Abbildungen 59 und 60) zeigt ein ähnliches, wenn auch unerwartetes Bild. So finden sich bei beiden Genotypen die breiten Streifenwerte im positiven Bereich von F1 und F2, sowie im negativen von F3. Die Schmalen im positiven von F2 und negativen von F3, bei F1 befindet sich rut2080, in etwa bei 0 und cstp im negativen Bereich. Der streifenlose Modus ist überall im negativen Bereich angesiedelt. Dieses Ergebnis verwundert da zumindest bei cstp durch die stets höhere Aktivität im schmalstreifigen Modus ein höherer Wert für F1 zu erwarten gewesen wäre - die Gruppierung der Daten könnte als Fehlerquelle dienen, denn im Gesamtplot von cstp (Abbildung 34) erkennt man recht viele schmalstreifige Datenpunkte im positiven Bereich von F1.

Die übrigen beiden Genotypen rut1 (Abbildung 7) und rsh (Abbildung 4) stechen vor allem durch ihre niedrige Aktivität hervor. Sowohl in der Gesamtaktivität, als auch in der Einzelaktivität (Abbildungen 9 und 10). Rsh verhält sich hierbei ähnlich wie cstp und hat in den schmalstreifigen Modi die meiste Aktivität, bei rut1 hingegen dominiert der Streifenlose Modus, wobei dieses durch die hohe Standardabweichung relativiert wird. Im Allgemeinen zeigt rut1 die geringste Aktivität. Unterstützend für dieses Verhalten zeigt sich ebenfalls die Zahl der Pausen so zeigen rsh (Abbildung 16) und rut1 (Abbildung 15) durchgehend die meisten Pausen, dieses auf den breiten und streifenlosen Modus verteilt. Auch bei der Laufgeschwindigkeit (Abbildung 17) ist rut1 das Schlusslicht, rsh wird hier einzig durch dnc (Abbildung 18) unterboten, wobei dieses vom Streifenmodus abhängt, dieses setzt sich auch bei der Laufdistanz fort (Abbildung 19/Tabelle 2), wieder gibt rut1 das Schlusslicht und rsh wird zeigt höhere Werte im schmalen Modus als dnc (Abbildung 20) - beide Genotypen zeigen bei den Laufvariablen höhere Werte im schmalstreifigen Modus.

Bei den Laufplots zeigen die Genotypen ein prinzipiell ähnliches Bild im breit- und schmalstreifgen Modus ist bei rut1 (Abbildung 24) in den Occupancy- und Einzelplots die Tendenz einer Bewegung

zwischen den Streifen erkennbar, beim streifenlosen Modus ist erwartungsgemäß kein Muster auszumachen, rsh zeigt ein deutlicheres Bild (Abbildung 23), sowohl im breit- wie auch schmalstreifigen Modus sind im allen Plottypen stark gerichtete Bewegungen zwischen den Polen erkennbar, erstaunlich ist eine ähnliche Bewegungsverteilung im streifenlosen Modus, welcher - da ohne Reiz - keine Laufpräferenz zeigen sollte.

Die Korrelationsmatrix der HKA zeigt bei rsh (Tabelle 19) nur geringe Wertunterscheide zur GesamtHKA (Tabelle 4), rut1 (Tabelle 24) hingegen zeigt Unterschiede zur Gesamten und der Korrelationsmatrix von rsh. Zuerst fällt auf, dass number_pause positiv mit sum_act korreliert ist, auch die Korrelation zu den anderen Aktivitäts- und Laufvariablen ist höher, jedoch nicht signifikant. Außerdem ist walks durchgehend geringer mit den andern Variablen korreliert. Da rut1 bei allen diesen Variablen niedrige Werte gezeigt hatte wäre eine ähnliche Korrelation wie im Gesamten oder bei rsh zu erwarten gewesen.

Bei der Varianz zeigt rsh (Tabelle 20) mit drei Hauptkomponenten sogar eine etwas höhere Gesamtvarianz als im Gesamten (Tabelle X) zu finden ist, entspricht aber in etwas dem Gesamten. Rut1 zeigt hier vor allem in der ersten Hauptkomponente mit 53,716% (Tabelle 25) eine deutlich geringere Varianz als im Gesamten oder bei rsh.

Das abweichende Verhalten von rut1 setzt sich mit der Faktorladung (Tabelle 26) fort, so finden man zwar wie im Gesamten (Tabelle 5) und bei rsh (Tabelle 21), die hohe Korrelation von Lauf- und Aktivitätsvariablen mit der ersten Hauptkomponenten, jedoch zeigt sich bei rut1 ein deutlich höhere negative Korrelation von median_pause mit der ersten und vor allem zweiten Hautkomponenten. Diese Starke negative Korrelation bei der zweiten Hauptkomponenten findet sich auch bei rsh. Außerdem ist eine sehr starke negative Korrelation von walks zur zweiten Hauptkomponente zu finden, was mit der sehr geringen Anzahl der Läufe zwischen den Streifen begründet sein dürfte.

Im Korrelationskreis ergibt sich deshalb auch ungewöhnliches Bild, sowohl bei rsh (Abbildung 45), alsauch bei ru1 (Abbildung 51) kommt es durch die im Gegensatz zur GesamtHKA (Abbildung 27) umgedrehten Korrelationen von number_pause und median_pause zu einem Seitenwechsel dieser Variablen. Außerdem wird die starke positive Korrelation von number_pause erkennbar und die starke Streuung der restlichen Variable im Wertebereich der Achse F2, während die restlichen Werte bei rsh eher der Gesamtanalyse entsprechen, auch wenn es hier zu teilweisen Vorzeichenwechseln kam.

So zeigen auch die quadrierten Kosinuswerte, dass sowohl rsh (Tabelle 22), als auch rut1 (Tabelle 27) zumindest für die erste Hauptkomponente der GesamtHKA entsprechen (Tabelle 7), bei der Zweiten zeigen beide dieselbe Abweichung, so korreliert number_pause stärker mit der F2 als median_pause - dieses ist bei der GesamtHKA umgekehrt.

Die grafische Analyse der beiden Genotypen zeigt wieder Unterschiede. So reicht sich rsh (Abbildung 47) mit der Streuung von breiten und schmalen Streifenwerten über die gesamte, die der streifenlosen nur im negativen Bereich (bis auf Ausbrecher) Achse F1 in das Gesamtbild der anderen Genotypen (Abbildung gesamt) ein. Im Gegensatz hierzu steht rut1 (Abbildung 53), bei welchen die Werte um alle Achsen stark streuen und sogar die streifenlosen Werte die meiste Aktivität bescheinigen. Dieses setzt sich auch in der gruppierten Ansicht fort, hier verliert rut1 (Abbildung 55) sogar alle positiven F1 Werte und bezeugt dem schmalen Modus die geringste Aktivität, durch den sehr niedrigen Wert bei F1. RSH zeigt ein etwas bekannteres, wenngleich abgeschwächtes Bild

(Abbildung 49) gegenüber den anderen Genotypen. So zeugt der Wert für die breiten Streifen von der höchsten Aktivität und einem geringen Einfluss der Pausen, mehr Einfluss durch Pausen und eine etwas niedrigere Aktivität zeigt sich im schmalen Modus. Der Streifenlose Modus ist abgeschlagen mit wenig Aktivität.

Im Gesamten ergibt sich ein klares Bild, rut2080 zeigt wie bereits in den Resultate aus der Projektarbeit gezeigt eine durchgehend hohe Grundaktivität, diese ist fast unabhängig vom Streifentypen (Abbildung 28) und wird nur in manchen Punkten von cstp übertroffen, wobei dieses dann im schmalen Streifenmodus passiert. Interessant, ist die Tatsache, dass rut1 und rut2080, welche einen sehr nahe genetischen Ursprung haben komplett unterschiedliche Verhaltensweisen aufzeigen.
Dnc weist ebenfalls ein streifenunabhängiges Verhalten wie rut2080 auf, jedoch mit einer stark eingeschränkten Aktivität – da in anderen Arbeiten von XYZ bereits festgestellt wurde, dass dnc unter einer Lernschwäche leidet[5], könnte dieses damit zusammenhängen und auch bei rut2080 eine Rolle spielen.

Die Hauptkomponentenanalyse bestätigt auch, dass die Pausendauer offensichtlich keinen Einfluss auf eine der anderen Variablen zu haben scheint (Abbildung 27) und so unter Umständen genetisch bedingt ist. Man müsste untersuchen ob *Drosophila melanogaster* in der Lage ist bei veränderten Experimentierdesigns auch andere Pausenlängen zu zeigen.

Die Korrelation der Aktivitäts- und Laufvariablen untereinander und zu den Hauptkomponenten, zeigt, dass in diesem Bereich die größte Varianz und die meisten Unterscheide (Tabelle 7) zwischen den Fliegen vorliegen, wohingegen die Pausen nur eine untergeordnete Rolle spielen.

Mit Hilfe der HKA kann man auch sehr gut sehen, dass die Streifentypen eine Rolle in der Aktivität spielen und erkennt dieses schnell in den grafischen Plots (Abbildungen 28 und 29). Bis auf die beiden rut-Genotypen, welche entweder so gut wie keine Aktivität (rut1) oder durchgehend sehr viel Aktivität zeigen (rut2080), erkennt man, dass die schmalen und breiten Streifentypen eine höhere Aktivität zeigen. Dabei dominieren die schmalen Streifentypen in der Regel. Die durchgehend schlechte Leistung von rsh könnte auch im starken Lern- und Gedächnitsdefizit verankert sein [6].

Literaturverzeichnis:

[1] Jedrzej Blaszkiewicz. Buridan – Laufexperimente fünf genetisch unterschiedlicher Drosophila melanogaster Stämm. Unveröffentlicht. 2010
[2]Götz K.G. 1980: "Visual guidance in Drosophila." Basic Life Sci. 16: 391-407
[3] Leyer, Wesche. Multivariate Statistik in der Ökologie. 2007
[4]A. Handl, Multivariate Analysemethoden, Statistik und ihre Anwendungen, 2nd ed. 2010
[5] Tully et al. „differential effects of dunce mutations on associative learning and memory in drosophila“ Neurogenetics 1993
[6] Folkers et al. „Radish, a Drosophila mutant deficient in consolidated memory” Neurobiology 1993