

Universität Bielefeld  
Fakultät für Linguistik und Literaturwissenschaft  
Verfahren zur Repräsentation und Verarbeitung  
multimodaler Dokumente  
Prof. Dieter Metzling  
Sommersemester 2006

20.02.2007

---

## Interaktion mit AIBO

---

Unterschiede in der Interaktion von AIBO-erfahrenen und  
-unerfahrenen Teilnehmern

von  
Daniel Jettka  
djettka@uni-bielefeld.de

## INHALT

1	Einleitung .....	1
2	Das Videomaterial .....	2
2.1	Erstellung .....	2
2.2	Sichtung und Hypothesengenerierung .....	3
3	Die Annotation .....	4
3.1	Annotationsschemata .....	5
3.1.1	Sprache: SAMPA und DIDA .....	5
3.1.2	Gestik: Berliner Lexikon der Alltagsgesten .....	7
3.1.3	Mimik: MPEG-4 Facial Animation Parameters .....	8
3.1.4	AIBO: LEDs und akustische Signale .....	9
3.2	Annotation mit ELAN .....	11
3.2.1	Die zu annotierenden Sequenzen .....	11
3.2.2	Durchführung der Annotation .....	13
4	Ergebnisse .....	15
5	Literatur .....	17
	Anhang .....	18
A	Zeicheninventar des SAMPA .....	18
B	Auszug aus dem Berliner Lexikon der Alltagsgesten (BLAG) .....	19
C	Auszug aus MPEG-4 Facial Animation Parameters .....	20
D	AIBOs LED-Anzeigen .....	22
E	AIBO-Sprachbefehle .....	24

## 1 EINLEITUNG

Im Sommersemester 2006 fand an der Universität Bielefeld das Seminar „Verfahren zur Repräsentation und Verarbeitung multimodaler Dokumente“ unter der Leitung von Prof. Dr. Dieter Metzger statt. Als Grundlage für das Seminar entstanden Filmaufnahmen, die die Interaktion<sup>1</sup> von menschlichen Teilnehmern und dem SONY-Hund AIBO<sup>2</sup> zeigen. Weitere Informationen zu den Rahmenbedingungen der Aufnahmen liefert der Abschnitt 2.1.

Unsere Gruppe (gegen Ende des Semesters bestehend aus: Daniel Jettka, Natasa Pohlod und Nico Schneider) entschloss sich nach Sichtung der Filmaufnahmen zur Untersuchung des Unterschieds der Interaktion AIBO-unerfahrener Teilnehmer mit dem AIBO und der Interaktion AIBO-erfahrener Teilnehmer (im Speziellen der AIBO-Besitzerin) mit dem AIBO (vgl. Abschnitt 2.2). Zu diesem Zweck wurden zwei charakteristische Filmausschnitte ausgewählt und als zu untersuchende Daten festgelegt. Im nächsten Schritt mußte nun eine Operationalisierung der Daten erfolgen.

Hierzu bot sich das Annotationstool ELAN<sup>3</sup> an. Das Tool ermöglicht die Annotation von Video- und Audiodaten. Es stellt somit einen guten Ausgangspunkt zur Analyse multimodaler Zusammenhänge dar. Für die Annotation der Daten müssen zunächst Annotationsschemata für die interessierenden Modalitäten aufgestellt werden. Wir entschlossen uns die Modalitäten menschliche Sprache, Gestik und Mimik, sowie die Signale des AIBO in die Analyse aufzunehmen<sup>4</sup>. Der Zusammenhang von Modalitäten, Annotationsschemata und ELAN wird in Kapitel 3 ausführlicher dargestellt.

Auf der Grundlage der Annotation der Videodaten ließen sich die Zusammenhänge der einzelnen Modalitäten analysieren und in Bezug zu dem von uns gewählten Thema stellen. Eine Antwort auf die Frage nach dem unterschiedlichen Charakter der Interaktion AIBO-unerfahrener und -erfahrener Teilnehmer mit dem AIBO konnten wir aus Zeitgründen jedoch nur in Ansätzen liefern. Hierzu verweise ich auf das Kapitel 4, in dem die Ansätze dargestellt werden, aber auch ein Ausblick auf weitere mögliche Analyseschritte gegeben wird.

---

1 Der Begriff Interaktion wird hier in einem vereinfachten, allgemeinen Sinn verwendet.

2 AIBO ist ein Roboterhund, der 1999 von SONY auf den Markt gebracht wurde. Das Ziel bestand darin, einen Haustierersatz für Allergiker zu schaffen. Siehe: <http://www.sonydigital-link.com/AIBO/index.asp?language=de>

3 Eudico Linguistic Annotator; Eudico: European Distributed Corpus

4 Die Aufgabenverteilung sah hierbei wie folgt aus:

Daniel:	menschliche Sprache und AIBO-Signale
Nico:	Gestik
Natasa:	Mimik

## 2 DAS VIDEOMATERIAL

### 2.1 Erstellung

Die Filmaufnahmen unserer Gruppe entstanden am 12. April 2006 im Tonstudio der Fakultät für Linguistik und Literaturwissenschaft der Universität Bielefeld im Raum C01-213. Freundlich unterstützt hat uns hierbei die Leiterin des Studios Ulrike Well. Sie hat uns die später von uns zur Analyse ausgewählten Sequenzen jeweils als .mov-Dateien plus zusätzlicher Tonspur im .wav-Format geliefert.

Der Grundaufbau der Filmaufnahmen sah folgendermaßen aus: Jeweils zwei Personen aus der Projektgruppe sollten sich ohne feste Vorgaben, ausgestattet mit einer Übersicht der wichtigsten Sprachbefehle (vgl. Anhang E), mit dem AIBO beschäftigen. Während der ersten ca. 25-minütigen Aufnahme waren dies Daniel Jettka und Natasa Pohlod. In dieser ersten Aufnahme war außerdem AIBO-Besitzerin Conny Pearce direkt in die Interaktion involviert, um eine kurze Einweisung zu den Funktionen des AIBO zu geben. Die Ausgangssituation der ersten Aufnahme war wie folgt:



Abbildung 1: Ausgangssituation

Die Aufnahmen wurden mit Hilfe zwei fest installierter Kameras und einer Handkamera, geführt durch eine Person der Projektgruppe, erstellt<sup>5</sup>. Auf diese Weise sollte ein möglichst gro-

<sup>5</sup> Der exakte Versuchsaufbau (Abstände der Kameras, Positionierungen im Raum, etc.) wurde leider nicht dokumentiert. Insgesamt befanden sich 4 Personen im Aufnahmezimmer. Die restlichen Anwesenden waren im Schnittraum, der baulich abgetrennt ist.

ßer und genauer Ausschnitt der Handlung erfasst werden. Diese Vorgehensweise ist besonders wichtig, um alle, sich evtl. erst im Nachhinein als interessant herausstellenden, Aspekte zu erfassen. Dieses Problem wurde uns besonders deutlich, als wir während der Analyse der später ausgewählten Sequenzen bemerkten, daß die Signale des AIBO-Displays aus keiner Kameraeinstellung ersichtlich sind. Somit konnte ein wichtiger Ausschnitt der Interaktion nicht in die Analyse aufgenommen werden. Hieraus läßt sich für spätere Aufnahmen ähnlicher Art die Konsequenz ziehen, noch mehr auf die Detailliertheit der Aufnahmen zu achten.

Die weiteren Aufnahmen entstanden in gleicher Weise. Ein Unterschied lag jedoch darin, daß Conny Pearce nicht mehr direkt in die Handlung eingebunden war, sondern sich hinter den Kameras aufhielt. Damit war sie zwar vordergründig nicht an der Aufnahme beteiligt, wurde jedoch des öfteren von den AIBO-Interaktanten um Rat gebeten. Diese Tatsache wird in Abschnitt 3.2 nochmals thematisiert.

## 2.2 Sichtung und Hypothesengenerierung

Der nächste Schritt des Projektes bestand darin, die entstandenen Filmaufnahmen anzusehen und interessante Aspekte herauszuarbeiten. Zu diesem Zweck haben sich die Gruppenmitglieder zunächst einzeln mit den Aufnahmen befasst. In einem späteren Arbeitstreffen der Gruppe wurden dann verschiedene Vorschläge zu möglichen Themen und den dazu passenden Sequenzen diskutiert. Dieses Vorgehen stand natürlich unter der Prämisse besonders auf multimodale Zusammenhänge der Interaktion einzugehen.

Geeinigt haben wir uns auf zwei etwa einminütige Ausschnitte, die in besonderem Maße die Verschiedenheit der Interaktion AIBO-erfahrener und -unerfahrener Interaktanten aufzeigt. Die Unterschiede lagen u.a. in der verschiedenen Verwendung von Modalitäten in der Interaktion mit dem AIBO. Hieraus haben sich für die Gruppe folgende Hypothesen ergeben:

1. AIBO bringt kommunikative Ressourcen in die Interaktion ein, die AIBO-erfahrenen Teilnehmern einen anderen Grad von Interaktion mit AIBO ermöglichen.
2. In der Interaktion von Novizen mit dem AIBO kommt es zu kommunikativen Schwierigkeiten wie turn-taking-Problemen (weil z.B. Signale nicht verstanden

werden).

3. Allgemein: Die Unerfahrenheit der Novizen spiegelt sich in ihrem Verhalten wider.

Diese Hypothesen ergaben sich direkt aus den ausgewählten Filmausschnitten und es ist damit sehr wahrscheinlich, daß sie sich bestätigen lassen. Die Hauptaufgabe bestand nun im Wesentlichen darin, die Beobachtungen durch klar definierte Kategorien festzuhalten und die Hypothesen anhand dieser Kategorien zu belegen.

Die Gruppe hat beschlossen, daß zur adäquaten Beschreibung des Interaktionsgeschehens, die Sprache, die Gestik und Mimik der menschlichen Interaktanten, sowie die von AIBO ausgehenden Signale erfasst werden müssen. Um die Mittel der Erfassung der verschiedenen Modalitäten geht es im folgenden Kapitel.

### 3 DIE ANNOTATION

Die Annotation der gesammelten Daten, d.h. der Filmaufnahmen, stellt einen entscheidenden Schritt bei der Analyse der multimodalen Aspekte der Interaktion dar. Erste Einblicke kann zwar schon die Sichtung des Materials liefern, durch die Annotation können jedoch Beobachtungen manifestiert werden und es lassen sich empirisch Zusammenhänge der verschiedenen Modalitäten (z.B. durch statistische Methoden) erfassen.

Hierbei ist allerdings zu beachten, daß die Annotation systematisch und gründlich durchgeführt werden muss, denn sie ist die Basis für die spätere Analyse. Zur Annotation der Filmsequenzen haben wir das Annotationswerkzeug ELAN verwendet. Auf ELAN und die von uns durchgeführte Annotation werde ich im Abschnitt 3.2 eingehen. Doch zunächst wird es darum gehen, die Annotationsschemata vorzustellen, die wir der Annotation zu Grunde gelegt haben. Sie dienen der Festlegung eines Beschreibungsrahmens mit festen Kategorien. Hierdurch wird ein hoher Grad an Reliabilität der Annotation ermöglicht. Außerdem sind die Daten längerfristig nutzbar, da festgelegte Annotationsschemata in andere Formate überführbar sind.

## 3.1 Annotationsschemata

### 3.1.1 Sprache: SAMPA und DIDA

Bei der Annotation der menschlichen Sprache haben wir uns auf drei Ebenen beschränkt. Die erste Ebene ist die phonetische Umschrift der Äußerungen. Zu diesem Zweck haben wir das ASCII-basierte phonetische Alphabet SAMPA (Speech Assessment Methods Phonetic Alphabet) ausgewählt<sup>6</sup>. Die von uns verwendeten Zeichen sind in einer Tabelle im Anhang auf S. 18 aufgeführt. Für die Zwecke unseres eher kleinen Projekts wäre eine pseudo-phonetische Umschrift evtl. schon ausreichend gewesen; trotzdem haben wir uns bewußt für die formalere Alternative entschieden, um z.B. die Möglichkeit zu bewahren, die Transkription maschinell in andere Formate zu überführen, was eine langfristige Nutzung der Daten ermöglichte. Ein Beispiel für die phonetische Transkription nach SAMPA ist das folgende:

di: fo:ne:tIS@ transkriptslon na:C s@mpa zi:t zo: aUs

Der phonetischen Transkription der Äußerungen haben wir weitere Merkmale hinzugefügt, nämlich die segmentalen Besonderheiten von Äußerungen. Hierbei haben wir uns am Format der DIDA orientiert. „DIDA ist die Bezeichnung für die Diskursdatenbank, die im IDS für die Abteilung 'Pragmatik' für die Eingabe, Verwaltung und weiterer Bearbeitung von Gesprächsdaten entwickelt wurde“ (Institut für Deutsche Sprache, 2001). Das Inventar der Merkmale sieht folgendermaßen aus:

Symbol	Bedeutung
/	Wortabbruch
*	kurze Sprechpause
**	längere Sprechpause
*1,50*	Pause von eineinhalb Sekunden
*3:30*	Pause von drei Minuten dreißig Sekunden
+	direkter Anschluss einer Äußerung an eine eines anderen Sprechers
(kann doch mal)	vermuteter Wortlaut
(mag?sag)	Alternativlautung
(...)	unverständliche Äußerung; wenn mögl.: 3 Punkte pro Silbe
(...2,5)	Länge einer unverständlichen Äußerung
[...]	Auslassung im Transkript
[10min]	Auslassung im Transkript mit absoluter Zeitangabe

Tabelle 1: Suprasegmentale Besonderheiten nach DIDA

6 SAMPA erfasst eine Teilmenge des International Phonetic Alphabet (IPA). Weitere Informationen unter <http://www.phon.ucl.ac.uk/home/sampa/index.html>

Anhand der Annotation dieser Merkmale können im Verlauf der Analyse beispielsweise Bezüge zwischen segmentalen Besonderheiten von Äußerungen eines Sprechers zum Verhalten eines Anderen hergestellt werden. So könnte man z.B. den Grund für einen Wortabbruch eines Sprechers aus der Interaktion mit einem anderen Interaktanten rekonstruieren.

Als weitere Ebene der Annotation von Sprache hatten wir geplant, die Intonation von Äußerungen festzuhalten. Hierfür schien das System GToBI<sup>7</sup> hilfreich zu sein. Im Nachhinein hat sich jedoch herausgestellt, daß das System zu komplex für unsere Zwecke ist. Daher haben wir diese Ebene der Annotation ausgelassen. Zweifelsohne ist die Intonation ein interessanter Aspekt in der Interaktion mit AIBO und ließe sich zu anderen Modalitäten in Beziehung setzen. Aber wie gesagt, sie hätte den Rahmen des Projekts gesprengt.

Einfacher dagegen ließen sich die Wortarten zu Äußerungen annotieren. Dabei haben wir folgendes Annotationsschema verwendet:

Wortarten	Annotation	Wortarten	Annotation
Abkürzung	<b>Abk</b>	Nomen	<b>N</b>
Adjektiv	<b>A</b>	Partizip1	<b>P1</b>
Adverb	<b>Adv</b>	Partizip2	<b>P2</b>
Artikel	<b>Art</b>	Personalpronomen	<b>PerPro</b>
Auxiliar	<b>Aux</b>	Possessivpronomen	<b>PosPro</b>
Demonstrativpronomen	<b>DemPro</b>	Präposition	<b>Präp</b>
Eigennamen	<b>Eig</b>	Reflexivpronomen	<b>RefPro</b>
Indefinitpronomen	<b>IndPro</b>	Relativpronomen	<b>RelPro</b>
Interjektion	<b>Int</b>	Verb	<b>V</b>
Interrogativpronomen	<b>IntPro</b>	Verbpartikel	<b>VPar</b>
Konjunktion	<b>Konj</b>	keine Wortart gefunden	<b>Y</b>
Modalverb	<b>Mod</b>	Zahlwort	<b>Z</b>

<sup>7</sup> German Tones and Break Indices; nähere Informationen in Grice/Baumann (2000)



### 3.1.2 Gestik: Berliner Lexikon der Alltagsgesten

Die Grundlage der Annotation der Gestik in unserem Projekt bildet BLAG, das Berliner Lexikon der Alltagsgesten (vgl. Serenari et al., 2002). BLAG umfasst 150 Gestenfamilien, die in digitaler Form (Videoaufnahmen) vorliegen. Die Beschreibungen der einzelnen Lexikoneinträge, also prototypischen Gesten, aus BLAG dienen uns als Orientierung zur Erstellung eines Annotationsschemas für Gestik. „The textual part of each entry describes, from an expression perspective, the form of each structural variant, that is the respective parts of the body carrying out the action, as well its place, orientation and movements“ (Serenari et al., 2002: 35). Da BLAG für unsere Annotation zu umfangreich ist, haben wir im Vorhinein nur für uns interessant erscheinende Einträge berücksichtigt. Hieraus hat sich für unser Projekt ein Annotationschema ergeben, das sechs Typen von Alltagsgesten, sowie deren Subtypen enthält (vgl. Anhang B auf Seite 19). Folgende Szene kann als ein Beispiel für die Typisierung einer Geste angesehen werden.



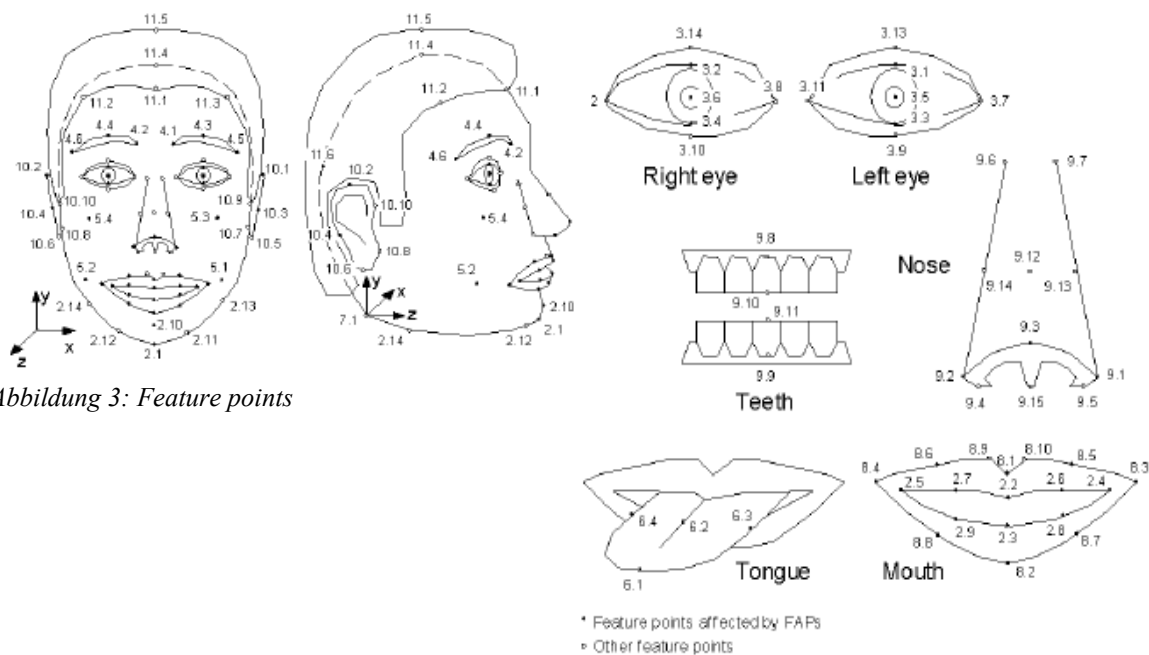
Abbildung 2: Beispiel für eine Geste

Diese Geste wurde von uns auf der Grundlage von BLAG mit „A10 01 - Holding index finger in a certain direction“ charakterisiert. Das Schema ist durchaus erweiterbar, sodaß unter Umständen weitere Einträge aus BLAG aufgenommen werden können. Es hat sich allerdings für uns als ausreichend erwiesen. Im Abschnitt 3.2 wird es um die genaue Verwendung des Schemas bei der Annotation von Gesten gehen. Dort werden auch weitere Beispiele aufgeführt.

### 3.1.3 Mimik: MPEG-4 Facial Animation Parameters und

Das System MPEG-4 SNHC (Moving Pictures Expert Group, Synthetic/Natural Hybrid Coding) schien uns eine gute Basis für die Annotation von Mimik zu sein. Die MPEG ist eine Arbeitsgruppe der ISO (International Standardization Organization) und befasst sich generell mit der Verarbeitung sich bewegender Bilder in Kombination mit Audiodaten (vgl. Kipp et al., 2002: 23). Im Kontext der Arbeitsgruppe ist das Kodierungsmodul MPEG-4 entstanden, welches folgenden Nutzen hat: „MPEG-4 enables integration of face animation with multimedia communications and representations and allows face animation over low bit rate communication channels, for point-to-point as well as multi-point connections with low-delay“ (Kipp et al., 2002: 24). Auf den ersten Blick handelt es sich hierbei um eine ganz andere Anwendung als wir sie anstreben, nämlich um die Animation von Mimik bei künstlichen Agenten. Jedoch stellt das Grundgerüst von MPEG-4 eine nützliche Ausgangssituation für die Annotation menschlicher Mimik dar.

Die Grundlage von MPEG-4 ist ein Modell eines prototypischen neutralen Gesichtsausdrucks, der dem entspannten menschlichen Gesicht nachempfunden ist. Anhand des Modells werden 84 Merkmalsknoten, sogenannte feature points, festgelegt (vgl. Kipp et al., 2002: 26).



Durch die Definition von FAPs, Facial Animation Parameters, und FAPUs, Facial Animation Parameter Units (Maßeinheiten für die relative Entfernung von Merkmalsknoten zueinander),

kann die Bewegung von Merkmalsknoten bestimmt werden und auf diese Weise ein Gesichtsausdruck bzw. dessen Ablauf kodiert werden. Die ursprüngliche Anwendung von MPEG-4 ist die Produktion animierter Mimik. Umgekehrt kann das System jedoch auch verwendet werden, um beobachtete menschliche Gesichtsausdrücke zu kodieren und zu annotieren. Hierzu können die in MPEG-4 zur Verfügung stehenden 68 FAPs (vgl. Tabelle im Anhang C) verwendet werden. Ein Beispiel für die Kategorisierung vom Mimik ist das folgende:



Abbildung 4: Beispiel für Mimik

Die von einem entspannten Gesichtsausdruck abweichende Haltung der Augenbrauen würde mit dem FAP „35 – raise\_l\_o\_eyebrow“ (vertikale Anhebung des äußeren Teils der linken Augenbraue) und „36 – raise\_r\_o\_eyebrow“ (analog mit rechter Seite) kodiert werden.

Die Annotation von Mimik und der Bewegung der Merkmalsknoten im Speziellen haben wir in unserem Projekt auf einem relativ niedrigen Level realisiert. Das bedeutet, daß wir zunächst nur die Facial Animation Parameters angewendet haben. Die numerische Kodierung der Bewegung durch FAPUs haben wir, aufgrund des großen Aufwands, hingegen nicht berücksichtigt.

#### 3.1.4 AIBO: LEDs und akustische Signale

Bei der Sichtung der von uns ausgewählten Filmausschnitte hat sich herausgestellt, daß die in AIBO konzipierten Modalitäten eine entscheidende Rolle für die Interaktion der menschlichen Teilnehmer und AIBO einerseits, aber auch für die Interaktion der menschlichen Teilnehmer

untereinander, spielt.

Die offensichtlichsten kommunikativen Ressourcen, die AIBO in die Interaktion einbringt, sind die von ihm ausgehenden akustischen Signale und die LED-Anzeige seines Displays. Es sind zwei Arten von AIBO-eigenen akustischen Signalen zu unterscheiden: Erstens gibt es eine feste Menge von Sätzen bzw. Wortfolgen, die von AIBO ausgehen können und deren Bedeutung durch natürlchsprachliche Kompetenz nachzuvollziehen ist. Diese akustischen Signale haben wir mit Hilfe des in Abschnitt 3.1.1 vorgestellten SAMPA-Zeicheninventars annotiert.

Zweitens kann AIBO aber auch Tonfolgen („Piep“-Töne) produzieren. Die Bedeutung dieser Töne haben wir von der AIBO-Besitzerin interpretieren lassen. Auf der Grundlage dieses Expertenwissens wurden die 8 verschiedenen in den ausgewählten Filmsequenzen enthaltenen Töne mit ihrer vermuteten Bedeutung annotiert.

Sound-ereignis	Interpretation
1	Foto-Countdown (ohne Fotoschießen selber)
2	"ich spiele jetzt einfach mal irgendwas" oder Reaktion auf Streicheln, weil beim Hochnehmen von der Station an den Rückensensor gekommen
3	Sound gehört zu "Don't drop me, please"
4	Sound zu der Ausführung des Befehls "Beg for it"
5	Sound zur Bestätigung des Befehls "Let's dance"
6	Sound zu AIBOs Äußerung "Oh yes"
7	Sound zu AIBOs Äußerung "Here I go"
8	"ich freue mich über Streicheleinheiten"

Die zweite uns interessierende Modalität AIBOs ist die LED-Anzeige an seinem Kopf. Hier kann der Interaktant verschiedene Informationen über den Status von AIBO erfahren.



Abbildung 5: Das AIBO-Display

Leider haben wir die wichtige Rolle des Displays vor allem für die Unterschiede in der Interaktion von erfahrenen und nicht-erfahrenen Teilnehmern mit AIBO erst bei der Sichtung der Filmaufnahmen festgestellt. Daher sind die meisten LED-Anzeigen nicht auf den Aufnahmen zu erkennen. Dieser Umstand stellt eine klare Schwäche unseres Projekts bzw. des Grundaufbaus der Filmaufnahmen dar. Für zukünftige Aufnahmen ergibt sich hieraus, daß auf eine möglichst umfassende Erfassung der gesamten Situation zu achten ist. Trotz des Mangels der Aufnahmen haben wir uns, auch im Hinblick auf weitere Untersuchungen, entschlossen, ein Annotationsschema für AIBOs LED-Anzeigen aufzustellen (vgl. Anhang D). Dieses Schema ist jedoch aus den genannten Gründen bei unserer Annotation noch nicht zur Anwendung gekommen.

### 3.2 Annotation mit ELAN

Nachdem im vorangehenden Abschnitt die zugrunde liegenden Annotationsschemata vorgestellt wurden, geht es nun um die Annotation selbst. Mit Hilfe der Annotation können die durch die Schemata abgedeckten Informationen der Filmaufnahmen auf Dauer fixiert werden.

Im Hinblick auf unsere Fragestellung nach den Unterschieden der Interaktion von erfahrenen und unerfahrenen AIBO-Nutzern mit AIBO, haben wir zwei uns gewinnbringend erscheinende Sequenzen aus den Aufnahmen für die Annotation ausgewählt. Diese werde ich im folgenden Abschnitt kurz darstellen. Danach wird es um die Durchführung der Annotation gehen. Hierbei werde ich allerdings nicht auf die Grundlagen von ELAN, dem von uns verwendeten Annotationstool, eingehen können. Vielmehr wird es darum gehen, den Grundaufbau unserer Annotation zu skizzieren.

#### 3.2.1 Die zu annotierenden Sequenzen

##### Sequenz 1

Die erste Sequenz hat eine Länge von 01:20 min. Die Ausgangssituation ist in der Abbildung 1 auf S. 2 zu sehen. Die Beteiligten der face-to-face Interaktion<sup>8</sup> sind AIBO-Besitzerin Conny, die

---

<sup>8</sup> Den Begriff „face-to-face Interaktion“ habe ich gewählt, um von der Gesamtsituation zu abstrahieren.

Novizen Natasa und Daniel und AIBO. Das Thema der aufgenommenen Interaktion ist die Vorstellung von AIBO durch Conny. Sie erläutert zunächst die Funktion von AIBOs Ladestation. Währenddessen gibt AIBO Laute von sich und zeigt verschiedene LED-Signale auf seinem Display, was für die unerfahrenen Nutzer erst verständlich wird, als Conny einen Kommentar hierzu äußert. Sie meint, daß AIBO gerade ein Foto schießt. Danach streichelt Conny AIBOs Kopf, nimmt AIBO von der Station und setzt ihn auf den Boden. Im weiteren Verlauf der Sequenz erklärt Conny die Sensoren, die AIBO zur Verfügung stehen. Nach AIBOs Aufforderung „give me five“, der Conny durch einen Kontakt zum Sensor in AIBOs Pfote nachkommt, beginnt AIBO „Männchen“ zu machen.

Diese Sequenz ist für unser Projekt besonders interessant, da es die Interaktion von der AIBO-erfahrenen Teilnehmerin Conny zeigt. Das „Fotoschieß“-Signal von AIBO kann nur von Conny interpretiert werden. Das stellt sie selbst durch den von ihr geäußerten Kommentar implizit fest.

## Sequenz 2



Abbildung 6: Ausgangssituation Sequenz 2

Die zweite Sequenz hat eine Länge von 01:06 min. Dieses Mal sind nur Daniel und Natasa, sowie AIBO auf der Aufnahme zu sehen. Conny, die sich hinter den aufgestellten Kameras befindet, wird während der Sequenz jedoch in die face-to-face Interaktion eingebunden.

Die Aufgabe von Natasa und Daniel besteht darin, ein paar Sprachbefehle auszuprobieren. Daniel versucht zunächst die Aufmerksamkeit von AIBO zu erlangen, indem er „AIBO“ äußert. Danach gibt AIBO einen Laut von sich, woraufhin Daniel sich bei Conny erkundigt, ob der Laut bedeute, daß AIBO die Lautkette „AIBO“ wahrgenommen hat. Conny meint jedoch, AIBO habe verstanden, daß er zu tanze beginnen soll. Genau das tut er danach auch. Als nächstes fragt Daniel Conny, ob man AIBO nicht vom Tanzen abbringen könne, indem man ihn streichelt. Conny beginnt ihre Antwort mit „Doch“, was

---

Generell kann man wohl annehmen, daß z.B. auch die Kameralente einen Einfluss auf die Interaktion haben. Das soll hier aber nicht berücksichtigt werden.

für Daniel der Anlass ist, das Streicheln auszuprobieren. Die Aktion wird allerdings abgebrochen, da Conny vom Streicheln abrät, um ein Umfallen AIBOs beim abrupten Abbruch des Tanzes zu verhindern.

Im weiteren Verlauf versucht zuerst Natasa mit AIBO zu interagieren: Sie spricht ihn mit seinem Namen „Snowy“ an. AIBO reagiert mit „Oh yes“ und weiteren Lauten, woraufhin Natasa fragend zu Conny schaut. Danach versucht Daniel, AIBO durch „go forward“ zu bewegen. Dieser reagiert jedoch lediglich mit „Here I go“ und weiteren Lauten. Natasa gelingt es schließlich durch „go back“, daß AIBO sich rückwärts bewegt.

Diese Sequenz verdeutlicht die aus Sicht der AIBO-unerfahrenen Teilnehmer nicht gelungene Interaktion mit AIBO. Das Ratsuchen bei der außenstehenden AIBO-Besitzerin Conny verdeutlicht, daß AIBO u.U. keine Rolle als vollwertigem Interaktionspartner zugestanden wird und die Lösung der Interaktionsprobleme nur durch die Kompetenzen von Conny zu lösen sind.

### 3.2.2 Durchführung der Annotation

Die Annotation haben wir mit dem Annotationswerkzeug ELAN durchgeführt. ELAN ermöglicht eine zeitleistenbasierte Annotation von Video- und Audiodaten. Für eine ausführliche Einführung verweise ich auf Van Uytvanck (2006).

Die Annotationsschemata konnten z.T. direkt in ELAN definiert werden. Für die Annotationsebenen Wortarten, Gestik, Mimik und AIBO-Signale konnten wir die Schemata in sogenannte „Benutzerdefinierte Wörterbücher“ überführen. Diese ermöglichen es bei der Annotation eines bestimmten Abschnitts, einfach die passende Kategorie aus dem jeweiligen Annotationsschema (Wörterbuch) auszuwählen.

Als nächstes haben wir für die Beteiligten der Interaktion die zu annotierenden Modalitäten mit Hilfe von Annotationszeilen festgelegt. Im unten aufgeführten Beispiel eines kleinen Annotationsausschnitts sind die Zeilen für die Teilnehmerin Conny zu sehen. Durch die Markierung eines Zeitabschnitts kann ein Annotationsabschnitt festgelegt werden (Beispiel grün markiert). Der Wert für die Annotation kann entweder aus einer Menge von Kategorien ausgewählt werden (falls die Zeile mit einem Wörterbuch verbunden wurde) oder aus einer freien Eingabe bestehen. Freie Eingaben sind im Beispiel nur für die Zeilen „Phonetik – Conny“ und

„Kommentar – Conny“ vorhanden (blau markiert), wobei sich die Phonetik jedoch am Annotationsschema des SAMPA orientiert. Das SAMPA-System wurde aus einem einfachen Grund nicht in ein Wörterbuch überführt: es sollte nicht jedes einzelne Phonem als Zeitabschnitt annotiert werden.

The screenshot shows the ELAN software interface for the file "Elan - 0607101721.AIBO.Video1.kompleaf". The main window is divided into several sections:

- Video Player:** Shows a scene with three people sitting on the floor around a small robot (AIBO). The current time is 00:00:08.950. Below the video are playback controls and a selection range of 00:00:12.720 - 00:00:05.330 (3610).
- Annotation Table:** A table with columns for "Nr", "Annotation", "Anfangszeit", "Endzeit", and "Dauer". The current selection is "Empty".
- Annotation Timeline:** A detailed view of the video's time axis (00:00:09.000 to 00:00:15.000) with various annotations overlaid:
  - Phonetik - Conny:** Shows SAMPA phonetic symbols like (ə)(E)m, dan, vɪt, E, (ə), \*1,2\*, na:, gants, tɔl, fo:to, maCn, \*, ʔm, v, h, k, Stro:, h, vɪ, a, ʔaUto:.
  - Wortarten - Conny:** Shows parts of speech like Y, Ko, Au, Per, Y, Int, Adv, Ad, N, V, Y, K, In, N, A, A, A, P, Adv.
  - Mimik - Conny:** Shows facial expressions like "1 - viseme", "2 - expres", "1 - viseme", "1 - viseme".
  - Expression - Conn:** Shows "glare" and "25 - pitch\_l\_eyeball".
  - eyes - Conny:** Shows "21 - close\_b\_l", "23 - yaw\_l\_ey".
  - head - Conny:** Shows "pitch", "48 - head\_pitch", "49 - head\_yaw", "50 - h".
  - Gestik - Conny:** Shows "A23 01 - T" and "A10 04 - Holding flat hand and und".
  - Kommentar - Conny:** Shows "Connie startet AIBO an" and "Connie beobachtet ruhig AIBO".

Ich werde an dieser Stelle nicht weiter auf die Einzelheiten der Annotation eingehen. Diese sind am besten nachzuvollziehen, wenn man sie sich direkt in ELAN ansieht. Hier kann man die Annotationen der einzelnen Modalitäten zueinander in Beziehung setzen. So ist es zum Beispiel möglich, einzelne Zeilen auszuwählen und sie sich beim Abspielen der Filmsequenz im direkten Vergleich anzeigen zu lassen. Dieses Vorgehen war die Hauptmethode unserer späteren ansatzweisen Analyse. Desweiteren ist aber auch eine direkte Suche von Annotations-



abschnitten anhand der ihnen zugewiesenen Werte und in Abhängigkeit paralleler Annotationen auf anderen Zeilen in ELAN implementiert. Solche Suchen haben wir allerdings nur ansatzweise, eher zu Probezwecken, ausgeführt.

Als weitere wichtige Analysemethode ist zudem die Möglichkeit der statistischen Auswertung der annotierten Daten zu nennen. Da die Annotationen in ELAN mit Hilfe der Auszeichnungssprache XML kodiert werden, gibt es eine Vielzahl von Anwendungen zu ihrer Auswertung. Mit deren Hilfe könnten zum Beispiel Werte für Redeanteile, dem zeitlichen Verhältnis von sprachlichen und gestischen Einheiten und noch vieles mehr ermittelt werden. Die Methode der statistischen Auswertung wäre für ausführliche Analysen sicherlich notwendig.

Die Ergebnisse unserer Analyse der Annotationen im Hinblick auf die Hauptfragestellung nach den Unterschieden der Interaktion erfahrener und unerfahrener Teilnehmer mit dem AIBO werden Thema des folgenden Kapitels sein.

## 4 ERGEBNISSE

Leider ist die Analyse der Annotationen aus Zeitgründen gegen Ende des Semesters in unserem Projekt etwas zu kurz gekommen. Trotzdem möchte ich die Ansätze vorstellen, die sich ergeben haben. Die Ausgangshypothesen, die wir nach der Sichtung des Videomaterials aufgestellt hatten waren:

1. AIBO bringt kommunikative Ressourcen in die Interaktion ein, die AIBO-erfahrenen Teilnehmern einen anderen Grad von Interaktion mit AIBO ermöglichen.
2. In der Interaktion von Novizen mit dem AIBO kommt es zu kommunikativen Schwierigkeiten wie turn-taking-Problemen (weil z.B. Signale nicht verstanden werden).
3. Allgemein: Die Unerfahrenheit der Novizen spiegelt sich in ihrem Verhalten wider.

Zu Hypothese 1 läßt sich festhalten, daß sie sicherlich den Hauptgrund für die Unterschiedlichkeit der Interaktion von Conny (AIBO-erfahren) mit dem AIBO und der Interaktion von Daniel und Natasa (AIBO-unerfahren) mit dem AIBO darstellt. In der ersten Sequenz (zwischen Sekunde 07:00 und 14:00) gibt AIBO akustische und LED-Signale von sich, die Conny

anzeigen, daß er ein Foto schießt. Diese Signale können von den unerfahrenen Teilnehmern nicht verstanden werden.

Noch deutlicher wird die fehlende Kompetenz der Novizen in Sequenz 2, in der permanent Störungen der Interaktion zwischen den Novizen und AIBO zu beobachten sind. Man könnte vermuten, daß die Novizen aufgrund der nicht verstandenen Signale AIBO gar nicht als Interaktionspartner annehmen können. Um diese Annahme zu untermauern, könnte man Kriterien für eine intakte Interaktion aufstellen oder einen prototypischen Verlauf von Interaktion annehmen. Auf dieser Grundlage sollte es möglich sein, das Verhalten von erfahrenen AIBO-Nutzern und den unerfahrenen gegenüberzustellen. Dieses Vorgehen würde auf die Überprüfung der Hypothese 3 abzielen.

Die zweite Hypothese ist, wenn man die erste Hypothese als bestätigt ansehen wollte, im Grunde ein Teilaspekt der Nicht-Nutzbarkeit der kommunikativen Ressourcen AIBOs durch die Novizen. Die Sequenz 2 stellt, wie oben bereits angedeutet, ein Musterbeispiel hierfür dar. Bei genauerer Betrachtung lassen sich in den Filmaufnahmen eine Vielzahl von kommunikativen Schwierigkeiten im Umgang mit AIBO entdecken. Schon die langsamere Verarbeitung von Signalen durch AIBO stellt für die menschlichen Interaktionspartner Probleme dar. Doch dieses Argument zielt eher auf die generellen Probleme von künstlichen Agenten.

Allgemein läßt sich festhalten, daß den Kommunikationskanälen, die von AIBO eingeführt werden, bei der Analyse der Interaktion, große Beachtung zu schenken ist. Das Verständnis dieser Kanäle stellt einen entscheidenden Unterschied in der Interaktion von erfahrenen und unerfahrenen Teilnehmern mit AIBO dar.

## 5 LITERATUR

- Grice, Martine / Baumann, Stefan (2000): Deutsche Intonation und GToBI.  
In: *Linguistische Berichte*, 181/2000, Hamburg: Helmut Buske Verlag.
- Institut für Deutsche Sprache (2001): *Transkriptionsrichtlinien für die Eingabe in DIDA*.  
[ <http://www.ids-mannheim.de/prag/dida/dida-trl.pdf> (04.02.2007) ]
- Kipp, Michael/Reithinger, Norbert/Bernsen, Nils Ole/Dybkjær, Laila /Wegener Knudsen, Malene/Machuca, Maria/Riera, Montse (2002): *Best practice gesture, facial expression, and cross-modality coding schemes for inclusion in the workbench*.  
[ <http://nite.nis.sdu.dk/deliverables/NITE-D2.3-F.pdf> (15.02.2007) ]
- Serenari, Massimo/Dybkjær, Laila/Heid, Ulrich/Kipp, Michael/Reithinger, Norbert (2002): *Survey of existing gesture, facial expression, and cross-modality coding schemes*.  
[ <http://nite.nis.sdu.dk/deliverables/NITE-D2.1-sept02-F.pdf> (15.02.2007) ]
- Van Uytvanck, Dieter (2006): *EUDICO Linguistic Annotator (ELAN) version 2.6 . Manual*.  
[ [http://www.mpi.nl/tools/elan/ELAN\\_Manual.pdf](http://www.mpi.nl/tools/elan/ELAN_Manual.pdf) (04.02.2007) ]

## ANHANG

## A Zeicheninventar des SAMPA

Symbol	Wort	Annotation
<b>p</b>	<i>Pein</i>	paIn
<b>b</b>	<i>Bein</i>	baIn
<b>t</b>	<i>Teich</i>	taIC
<b>d</b>	<i>Deich</i>	daIC
<b>k</b>	<i>Kunst</i>	kUnst
<b>g</b>	<i>Gunst</i>	gUnst
<b>?</b>	<i>Verein</i>	fE6"?aln
<b>pf</b>	<i>Pfahl</i>	pfa:l
<b>ts</b>	<i>Zahl</i>	tSa:l
<b>tS</b>	<i>deutsch</i>	dOYtS
<b>dZ</b>	<i>Dschungel</i>	"dZUN=l
<b>f</b>	<i>fast</i>	fast
<b>v</b>	<i>was</i>	vas
<b>s</b>	<i>Tasse</i>	"tas@
<b>z</b>	<i>Hase</i>	"ha:z@
<b>S</b>	<i>waschen</i>	"vaS=n
<b>Z</b>	<i>Genie</i>	Ze"ni:
<b>C</b>	<i>sicher</i>	"zIC6
<b>j</b>	<i>Jahr</i>	ja:6
<b>x</b>	<i>Buch</i>	bu:x
<b>h</b>	<i>Hand</i>	hant
<b>m</b>	<i>mein</i>	maIn
<b>n</b>	<i>nein</i>	naIn
<b>N</b>	<i>Ding</i>	dIN
<b>l</b>	<i>Leim</i>	laIm
<b>R</b>	<i>Reim</i>	RaIm
<b>l</b>	<i>Sitz</i>	zItS
<b>E</b>	<i>Gesetz</i>	g@"zEtS
<b>a</b>	<i>Satz</i>	zats
<b>O</b>	<i>Trotz</i>	trOtS
<b>U</b>	<i>Schutz</i>	SUtS

:	<i>auffällige Dehnung</i>
“	<i>auffällige Betonung</i>

Symbol	Wort	Annotation
<b>Y</b>	<i>hübsch</i>	hYpS
<b>9</b>	<i>plötzlich</i>	"pl9tSIIc
<b>i:</b>	<i>Lied</i>	li:t
<b>e:</b>	<i>Beet</i>	be:t
<b>E:</b>	<i>spät</i>	SpE:t
<b>a:</b>	<i>Tat</i>	ta:t
<b>o:</b>	<i>rot</i>	ro:t
<b>u:</b>	<i>Blut</i>	blu:t
<b>y:</b>	<i>süß</i>	zy:s
<b>2:</b>	<i>blöd</i>	bl2:t
<b>al</b>	<i>Eis</i>	als
<b>aU</b>	<i>Haus</i>	haUs
<b>OY</b>	<i>Kreuz</i>	krOYtS
<b>@</b>	<i>bitte</i>	"blt@
<b>6</b>	<i>besser</i>	"bEs6
<b>i:6</b>	<i>Tier</i>	ti:6
<b>l6</b>	<i>Wirt</i>	vl6t
<b>y:6</b>	<i>Tür</i>	ty:6
<b>Y6</b>	<i>Türke</i>	"tY6k@
<b>e:6</b>	<i>schwer</i>	Sve:6
<b>E6</b>	<i>Berg</i>	bE6k
<b>E:6</b>	<i>Bär</i>	bE:6
<b>2:6</b>	<i>Föhr</i>	f2:6
<b>96</b>	<i>Wörter</i>	"v96t6
<b>a:6</b>	<i>Haar</i>	ha:6
<b>a6</b>	<i>hart</i>	ha6t
<b>u:6</b>	<i>Kur</i>	ku:6
<b>U6</b>	<i>kurz</i>	kU6tS
<b>o:6</b>	<i>Ohr</i>	o:6
<b>O6</b>	<i>dort</i>	dO6t

::	<i>sehr lange Dehnung</i>

## B Auszug aus dem Berliner Lexikon der Alltagsgesten (BLAG)

<b>A01</b> ____	<b>Wafting a hand in front of one's eyes</b>
A01 01	Sliding flat hand to the side in front of face
A01 02	Sliding flat hand to the side in front of face and looking to the floor
A01 03	Waving splayed hand to and fro in front of face
A01 04	Waving splayed hand to and fro in front of face with head pushed forwards
<b>A10 &amp; C04-E03</b> ____	<b>Pointing</b>
A10 & C04-E03	<01><02><03><04><05><06><07><08><09> Pointing with hand
A10 & C04-E03 01	Holding index finger in a certain direction
A10 & C04-E03 02	Holding index finger forwards with touch side directed downwards
A10 & C04-E03 03	Holding index finger forwards with touch side directed upwards
A10 & C04-E03 04	Holding flat hand and underarm in a particular direction
A10 & C04-E03 05	Holding flat hand with touch side directed downwards in a particular direction
A10 & C04-E03 06	Holding pointed out thumb in a particular direction
A10 & C04-E03 07	Directing index finger to chest
A10 & C04-E03 08	Holding index finger alternately in two different directions
A10 & C04-E03 09	Pointing index finger to addressee and to chest
A10 & C04-E03	<10><11><12> Pointing with other parts of body
A10 & C04-E03 10	Pushing chin forwards
A10 & C04-E03 11	Tilting head
A10 & C04-E03 12	Rolling eyes in a particular direction
<b>A13</b> ____	<b>Lowering flat hands</b>
A13 01	Lowering hand and arm at stomach height
A13 02	Lowering finger of flat hand to stomach height
A13 03	Pushing vertical flat hand at stomach height forwards and downwards
<b>A19</b> ____	<b>Twiddling one's thumbs</b>
A19 01	Twiddling thumbs around each other
A19 02	Twiddling thumbs around each other quickly
A19 03	Turning thumbs forwards and backwards around each other
<b>A23</b> ____	<b>Twisting one's hand back and forth</b>
A23 01	Twisting horizontally splayed out hand
A23 02	Turning splayed out hand to and fro and displaying the side
A23 03	Repeatedly turning away vertical, frontally displayed, splayed out hand
<b>A26</b> ____	<b>Raising flat hands</b>
A26 01	Raising horizontal flat hand
A26 02	Raising horizontal flat hand quickly and repeatedly
A26 03	Tipping flat hand upwards
A26 04	Raising flat hand horizontally after circling movement
A26 05	Tipping flat hand several times upwards and continuing to raising it

(Quelle: Serenari et al., 2002: 36ff)

## C Auszug aus MPEG-4 Facial Animation Parameters


#	FAP name	FAP description	Units	Uni- or bidirectional	Position motion	Group	FDP subgroupnumber
1	viseme	Set of values determining the mixture of two visemes for this frame (e.g. pbm, fv, th)	na	na	na	1	na
2	expression	A set of values determining the mixture of two facial expression	na	na	na	1	na
3	open_jaw	Vertical jaw displacement (does not affect mouth opening)	MNS	U	down	2	1
4	lower_t_midlip	Vertical top middle inner lip displacement	MNS	B	down	2	2
5	raise_b_midlip	Vertical bottom middle inner lip displacement	MNS	B	up	2	3
6	stretch_l_cornerlip	Horizontal displacement of left inner lip corner	MW	B	left	2	4
7	stretch_r_cornerlip	Horizontal displacement of right inner lip corner	MW	B	right	2	5
8	lower_t_lip_lm	Vertical displacement of midpoint between left corner and middle of top inner lip	MNS	B	down	2	6
9	lower_t_lip_rm	Vertical displacement of midpoint between right corner and middle of top inner lip	MNS	B	down	2	7
10	raise_b_lip_lm	Vertical displacement of midpoint between left corner and middle of bottom inner lip	MNS	B	up	2	8
11	raise_b_lip_rm	Vertical displacement of midpoint between right corner and middle of bottom inner lip	MNS	B	up	2	9
12	raise_l_cornerlip	Vertical displacement of left inner lip corner	MNS	B	up	2	4
13	raise_r_cornerlip	Vertical displacement of right inner lip corner	MNS	B	up	2	5
14	thrust_jaw	Depth displacement of jaw	MNS	U	forward	2	1
15	shift_jaw	Side to side displacement of jaw	MW	B	right	2	1
16	push_b_lip	Depth displacement of bottom middle lip	MNS	B	forward	2	3
17	push_t_lip	Depth displacement of top middle lip	MNS	B	forward	2	2
18	depress_chin	Upward and compressing movement of the chin (like in sadness)	MNS	B	up	2	10
19	close_t_l_eyelid	Vertical displacement of top left eyelid	IRIS D	B	down	3	1
20	close_t_r_eyelid	Vertical displacement of top right eyelid	IRIS D	B	down	3	2
21	close_b_l_eyelid	Vertical displacement of bottom left eyelid	IRIS D	B	up	3	3
22	close_b_r_eyelid	Vertical displacement of bottom right eyelid	IRIS D	B	up	3	4
23	yaw_l_eyeball	Horizontal orientation of left eyeball	AU	B	left	3	5
24	yaw_r_eyeball	Horizontal orientation of right eyeball	AU	B	left	3	6
25	pitch_l_eyeball	Vertical orientation of left eyeball	AU	B	down	3	5
26	pitch_r_eyeball	Vertical orientation of right eyeball	AU	B	down	3	6
27	thrust_l_eyeball	Depth displacement of left eyeball	ES	B	forward	3	5
28	thrust_r_eyeball	Depth displacement of right eyeball	ES	B	forward	3	6
29	dilate_l_pupil	Dilation of left pupil	IRIS D	B	growing	3	5
30	dilate_r_pupil	Dilation of right pupil	IRIS D	B	growing	3	6
31	raise_l_i_eyebrow	Vertical displacement of left inner eyebrow	ENS	B	up	4	1
32	raise_r_i_eyebrow	Vertical displacement of right inner eyebrow	ENS	B	up	4	2

	FAP name	FAP description	Units	Uni- or bidirectional	Position motion	Group	FDP subgroupnumber
33	raise_l_m_eyebrow	Vertical displacement of left middle eyebrow	ENS	B	up	4	3
34	raise_r_m_eyebrow	Vertical displacement of right middle eyebrow	ENS	B	up	4	4
35	raise_l_o_eyebrow	Vertical displacement of left outer eyebrow	ENS	B	up	4	5
36	raise_r_o_eyebrow	Vertical displacement of right outer eyebrow	ENS	B	up	4	6
37	squeeze_l_eyebrow	Horizontal displacement of left eyebrow	ES	B	right	4	1
38	squeeze_r_eyebrow	Horizontal displacement of right eyebrow	ES	B	left	4	2
39	puff_l_cheek	Horizontal displacement of left cheek	ES	B	left	5	1
40	puff_r_cheek	Horizontal displacement of right cheek	ES	B	right	5	2
41	lift_l_cheek	Vertical displacement of left cheek	ENS	U	up	5	3
42	lift_r_cheek	Vertical displacement of right cheek	ENS	U	up	5	4
43	shift_tongue_tip	Horizontal displacement of tongue tip	MW	B	right	6	1
44	raise_tongue_tip	Vertical displacement of tongue tip	MNS	B	up	6	1
45	thrust_tongue_tip	Depth displacement of tongue tip	MW	B	forward	6	1
46	raise_tongue	Vertical displacement of tongue	MNS	B	up	6	2
47	tongue_roll	Rolling of the tongue into U shape	AU	U	concave upward	6	3, 4
48	head_pitch	Head pitch angle from top of spine	AU	B	down	7	1
49	head_yaw	Head yaw angle from top of spine	AU	B	left	7	1
50	head_roll	Head roll angle from top of spine	AU	B	right	7	1
51	lower_t_midlip_o	Vertical top middle outer lip displacement	MNS	B	down	8	1
52	raise_b_midlip_o	Vertical bottom middle outer lip displacement	MNS	B	up	8	2
53	stretch_l_cornerlip_o	Horizontal displacement of left outer lip corner	MW	B	left	8	3
54	stretch_r_cornerlip_o	Horizontal displacement of right outer lip corner	MW	B	right	8	4
55	lower_t_lip_lm_o	Vertical displacement of midpoint between left corner and middle of top outer lip	MNS	B	down	8	5
56	lower_t_lip_rm_o	Vertical displacement of midpoint between right corner and middle of top outer lip	MNS	B	down	8	6
57	raise_b_lip_lm_o	Vertical displacement of midpoint between left corner and middle of bottom outer lip	MNS	B	up	8	7
58	raise_b_lip_rm_o	Vertical displacement of midpoint between right corner and middle of bottom outer lip	MNS	B	up	8	8
59	raise_l_cornerlip_o	Vertical displacement of left outer lip corner	MNS	B	up	8	3
60	raise_r_cornerlip_o	Vertical displacement of right outer lip corner	MNS	B	up	8	4
61	stretch_l_nose	Horizontal displacement of left side of nose	ENS	B	left	9	1
62	stretch_r_nose	Horizontal displacement of right side of nose	ENS	B	right	9	2
63	raise_nose	Vertical displacement of nose tip	ENS	B	up	9	3
64	bend_nose	Horizontal displacement of nose tip	ENS	B	right	9	3
65	raise_l_ear	Vertical displacement of left ear	ENS	B	up	10	1
66	raise_r_ear	Vertical displacement of right ear	ENS	B	up	10	2
67	pull_l_ear	Horizontal displacement of left ear	ENS	B	left	10	3
68	pull_r_ear	Horizontal displacement of right ear	ENS	B	right	10	4

(Quelle: Kipp et al., 2002: 28ff)

## D AIBOs LED-Anzeigen

AIBOs Hauptbefindlichkeiten		
LED-Anzeige	Erklärung	Annotation
	Glücklich	<b>glüchl.</b>
	Schlecht gelaunt	<b>schl. gel.</b>
	Traurig	<b>traurig</b>
	Böse	<b>böse</b>
	Ängstlich	<b>ängst.</b>
	Überrascht	<b>überr.</b>

Zustandsmodi		
LED-Anzeige	Erklärung	Annotation
	leuchtet blau: AEP eingeschaltet (autonomer Modus)	<b>AEP autonom</b>

AIBOs Grundbotschaften		
LED-Anzeige	Erklärung	Annotation
	AIBONE entdeckt	<b>AIBONE erk.</b>
	<b>rosa Ball entdeckt</b>	<b>Ball erk.</b>
	menschliches Gesicht erkannt	<b>Gesicht erk.</b>
	Herrchen/ Frauchen erkannt	<b>Herrchen erk.</b>
	im Kopf angebrachter Infrarot-Entfernungssensor hat Hindernis erkannt	<b>Kopf Hind.</b>
	in Brust angebrachter Infrarot-Entfernungssensor hat Hindernis erkannt	<b>Brust Hind.</b>
	rechte Beine haben Hindernis erkannt (Zusammenstoß)	<b>re. Beine Hind.</b>
	linke Beine haben Hindernis erkannt (Zusammenstoß)	<b>li. Beine Hind.</b>



	blinkt blau: AEP eingeschaltet (Fernsteuerungs modus)	<b>AEP ferng.</b>
	leuchtet rosa: Spielmodus	<b>Spielmodus</b>
	leuchtet grün: AIBO ruht sich aus	<b>ruht aus</b>
	langsam grün blinkend: AIBO möchte hochgenommen werden	<b>will hoch</b>
	leuchtet gelb oder blinkt gelb: <b>Klinikmodus</b>	<b>Klinik</b>
	leuchtet rot: Alarmmodus (zu hohe <b>Temperatur</b> )	<b>Alarm</b>

	Kopf hat Hindernis erkannt (Zusammenstoß)	<b>Kopf Ges. Hind.</b>
	AIBO hat Hand eines Menschen erkannt	<b>Pfötchen Hand erk.</b>
	Entfernungssens or in Brust hat schräg vor sich fallenden Gegen- stand registriert	<b>Brust fall. Geg.</b>
	rosafarbenen Gegenstand erkannt	<b>rosa Geg.</b>

Lieblingssituationen		
LED- Anzeige	Erklärung	Annotation
	AIBO untersucht seinen Lieblingsplatz	<b>Liebl.platz</b>
	AIBO hat einen Lieblingsgegenst and entdeckt	<b>Liebl.geg.</b>
	AIBO wird angefasst	<b>Anfassen</b>

(Quelle:  
AIBO ERS-7 – die Illume-Face-Funktion  
<http://www.sonydigital-link.com/aibo/downloads/de/ledface.pdf> )

## E AIBO-Sprachbefehle

### Bewegung

"Stand up"  
 "Sit down"  
 "Lay down"  
 "Go forward"  
 "Go back"  
 "Turn left"  
 "Turn right"  
 "Turn around"  
 "Come here"  
 "AIBO come here"  
 "come here AIBO"  
 "Over here"  
 "Walk around"  
 "Go to the station"  
 "Stop"  
 "Wait"  
 "Go"  
 "Move it"

### Fragen

"What's up"  
 "How are you?"  
 "What's going on?"  
 "What are you doing?"  
 "What time is it?"  
 "Are you sleepy?"  
 "Are you tired?"  
 "Are you hungry?"  
 "What's your name?"  
 "What's your owner's name?"  
 "Do you love me?"  
 "Where is your ball"  
 "Where is your pink ball"  
 "Where is your AIBOne"  
 "Where is your room"  
 "Where is your station"  
 "I love you AIBO"  
 "I am your owner"

### Interaktion mit Spielzeug

"Find your AIBONE"  
 "Pick up your AIBONE"  
 "Bring me your AIBONE"  
 "Give it to me"  
 "Open your mouth"  
 "Find your ball"  
 "Kick the ball"

### Einstellungen

"Name registration"  
 "Owner registration"  
 "Set alarm"  
 "Favorite thing registration"

### Unterhaltung

"Snowy"  
 "AIBO"  
 "yes"  
 "no"  
 "Hello"  
 "Good morning"  
 "Good Evening"  
 "I'm here" "I'm starving"  
 "Say hello" "I'm hungry"  
 "Good Bye" "It was fine"  
 "Bye Bye" "It was nice"  
 "Good night" "It was okay"  
 "I'm leaving" "It was good"  
 "I'm leaving now" "it was boring"  
 "See you later" "Be Quiet"  
 "Good boy" "just kidding"  
 "Good AIBO" "I'm sorry AIBO"  
 "Good girl" "I'm sorry"  
 "How cute" "AIBO I'm sorry"  
 "Go for it" "work"  
 "school"  
 "nowhere"

### Einfache Aktionen

"Bark"  
 "Bark Bark"  
 "Let's Dance"  
 "Shake hands"  
 "The other paw"  
 "Beg for it"  
 "Chase your tail"  
 "Take a picture"  
 "Snapshot"  
 "Look around"  
 "Seven Strike!"  
 "Don't do that"  
 "Be careful"

"I have good news"  
 "I had a bad day"  
 "I'm sleepy"  
 "I'm tired"  
 "I'm fine"  
 "I'm okay"  
 "I'm good"