Konversationales Verhalten beim Menschen

Ausarbeitung zum Seminar
“Embodied Conversational Agents”
bei Prof. Ipke Wachsmuth / Stefan Kopp / Timo Sowa
SS 2001

Marco Balke
mbalke@techfak.uni-bielefeld.de

23. Juni 2001

Inhaltsverzeichnis

1 Einleitung & Motivation **3**

2 Funktionen der nonverbalen Kommunikation **4**

3 Regeln der Dialogsteuerung **4**

4 Verschiedene Modalitäten menschlicher Konversation **5**
4.1 Sprache 5
4.1.1 Sprechakte 5
4.1.2 Intonation 6
4.1.3 Paraverbale Äußerungen 6
4.2 Gesten 7
4.2.1 Verschiedene Gestenkategorien 7
4.2.2 Handstellungen 9
4.3 Sonstige Modalitäten 9
4.3.1 Mimik 9
4.3.2 Blickrichtung 10
4.3.3 Kopfbewegungen 10
4.3.4 Atmung 10

5 Problematisierung **11**
5.1 Zeitliche Abhängigkeiten 11

6 Zusammenfassung **12**

1 Einleitung & Motivation

Um die Kommunikation mit dem Computer intuitiver und einfacher zu gestalten, werden sogenannte *Embodied Conversational Agents* entwickelt, Programme, die mit dem Benutzer über eine humanoide Figur kommunizieren. Um eine reibungslose Kommunikation zu ermöglichen, müssen diese ECAs den Regeln menschlicher Kommunikation folgen. Diese Arbeit soll einen Überblick über diejenigen Aspekte menschlichen konversationalen Verhaltens geben, die für die Implementierung eines ECAs von Bedeutung sind.

In dieser Stelle sollen einige elementare Begriffe menschlicher Kommunikation eingeführt werden. Dabei handelt es sich nicht um formale Definitionen, sondern um Beschreibungen der Begriffe.
Unter dem Begriff *Turn-taking* werden alle Begriffe, die das Halten und Weitergeben der Sprecherrolle in einer Konversation steuern, zusammengefaßt. *Turn-giving* bezeichnet die Weitergabe der Sprecherrolle an einen bestimmten Nachfolger durch den aktuellen Sprecher. *Turn-yielding* bezeichnet die Freigabe der Sprecherrolle, jeder Interessierte kann diese nun für sich einfordern. Dieses geschieht durch das sogenannte *Turn-wanting*, das signalisiert, das jemand gerne den Part des Redners übernehmen möchte. Diese Steuerungsfunktionen werden in Kapitel 3 und Kapitel 4 näher betrachtet und analysiert [Duncan, 1972] .
Als *Feedback* bezeichnet man alle die Verhaltensweise der Zuhörer, die dem Sprecher eine Rückmeldung zum Gesagten liefert, wie z.B. nicken oder paraverbale Äußerungen [Yngve, 1970].
Der *Stroke* ist der bedeutungstragende Teil einer Geste, die im allgemeinen in die Phasen *Preparation, Stroke* und *Retraction* eingeteilt werden kann. (Dabei ist zu beachten, daß es weitere Kategorien gibt und auch zweiphasige Gesten, die nur die Phasen Preparation und Retraction haben, vorkommen) [McNeill, 1992].
Jede verbale Äußerung eines Menschen kann als Sprechakt aufgefaßt werden. Sehr plakativ läßt sich sagen, ein Sprechakt ist eine Äußerung, in der zwischen Semantik (was wird gesagt) und Pragmatik (was soll dadurch erreicht werden) unterschieden wird. Eine sehr viel genauere Definition wird in Kapitel 4.1.1 gegeben, in dem auch die weiterführende Betrachtung des Sprechaktes im Rahmen der Funktionen in einer Kommunikation vorgenommen wird [Austin, 1962].
Eine verbale Äußerung läßt sich außerdem meistens in *Thema* und *Rhema* einteilen. Dabei ist das *Thema* der Inhalt des Gesagten, als *Rhema* bezeichnet man die neue Information in einer Äußerung [Halliday, 1967]. Je nach Situation hängt diese Einteilung vom Verlauf des Gespräches ab.
Beispiel: Der Satz:"Joe gibt Tom 50 Euro" kann je nach vorangegangener Fragestellung unterschiedlich eingeteilt werden:
Auf die Frage: "Wer gibt Tom 50 Euro?" wäre *Joe* das Rhema, auf die Frage: "Was gibt Joe Tom?" wäre hingegen *50 Euro* das Rhema.
In Abbildung 1 ist der wohl am weitesten fortgeschrittene *Embodied Conversational Agent* gezeigt: *REA*, der *Real Estate Agent*, der zur Zeit am Massachusets Institute of Technology entickelt wird. Wie der Name bereits ausdrückt, soll REA als Maklerin eingesetzt werden, die Kunden durch virtuelle Gebäude führt.

Abbildung 1: REA - Der Real Estate Agent.

2 Funktionen der nonverbalen Kommunikation

In der menschlichen Kommunikation ist zwar der Sprachkanal der wichtigste, jedoch leisten nonverbale Signale einen wesentlichen Beitrag zum reibungslosen Ablauf eines Gesprächs. Nach Bente & Krämer[2001] gibt es 4 verschiedene Funktionen nonverbaler Kommunikation:

1. Manipulative Eingriffe in einer realen oder virtuellen Objektwelt, die einen beispielhaften Charakter haben, wie zum Beispiel das Zeigen eines Bewegungsablaufs, werden unter dem Begriff *Modellfunktionen* geführt.

2. *Diskursfunktionen* übernehmen jene nonverbale Verhalten, die in engen Zusammenhang mit der Sprachfunktion steht, wie zum Beispiel sprachbegleitende Gestik.

3. Turn-taking und Feedback-Signale, die den Ablauf der Kommunikation steuern, werden den *Diskursfunktionen* zugeordnet.

4. *Sozio-emotionale Funktionen* übernehmen all jene nonverbalen Signale, die den Bezug von Zuhörer und Sprecher zueinander verändern, also all jene Signale, die zum Beispiel Symphatie oder Antiphatie verursachen.

3 Regeln der Dialogsteuerung

Im allgemeinen folgt jede menschliche Kommunikation einem Satz von Regeln, dabei gibt es jedoch Ausnahmen zu betrachten. Nach Levelt[1989] läßt sich die Steuerung von menschlichem konversationalem Verhalten durch 5 Regeln beschreiben:

1. Der Sprecher kann eine Äußerung machen. Nach Abschluß der Äußerung kann Regel 2, 3 oder 4 in Kraft treten.

2. Der Sprecher kann einen Nachfolger bestimmen.

3. Jeder Gesprächsteilnehmer kann die Sprecherrolle nach dem Abschluß einer Äußerung für sich beanspruchen.

4. Der aktuelle Sprecher kann den Turn behalten und eine weitere Äußerung machen.

5. Der neue Sprecher wird der aktuelle Sprecher und Regel eins tritt in Kraft.

Häufig kommt es vor, das mehrere Personen ein Turn-wanting Signal geben und mit dem Sprechen beginnen, jedoch wird meistens innerhalb einer sehr kurzen Zeit («1s) entschieden, wer die Sprecherrolle übernehmen darf. Dies ist im allgemeinen die Person, die zuerst mit dem Sprechen begonnen hat, allerdings wird hierbei oft *ranghöheren oder qualifizierteren Personen* die Sprecherrolle überlassen. Des weiteren kommt es vor, daß der Sprecher unterbrochen wird, weil jemand anderes sofort zu dem aktuellen Thema etwas beitragen möchte. In diesem Fall werden auch die Regeln verletzt, und eine Konfliktsituation entsteht, die sich im allgemeinen auch sehr schnell löst, aber auch zu einem Streit eskalieren kann, in dem sich keine Seite mehr zu Eingeständnissen in der Weitergabe der Sprecherrolle einläßt.

4 Verschiedene Modalitäten menschlicher Konversation

In diesem Kapitel sollen alle Modalitäten der menschlichen Kommunikation eingehender Untersucht werden.

4.1 Sprache

4.1.1 Sprechakte

Jede verbale Kommunikation kann als *Handlung* betrachtet werden. Diese Handlung wird als *Sprechakt* bezeichnet. Dabei wird zwischen vier verschiedenen Akten innerhalb eines Sprechaktes unterschieden.

1. Als *lokutionären Akt* bezeichnet man das eigentliche verbale Äußern, wobei völlig von der Bedeutung des Gesagten abstrahiert wird.

2. Der *propositionaler Akt* ist die Semantik der Äußerung, also die Aussage (Information), die durch das Gesagte übermittelt wird.

3. Der *illukotionärer Akt* hingegen ist die Pragmatik der Äußerung, also die eigentliche Intention des Sprechers.

4. Der Begriff perlokutionärer Akt bezeichnet das, was der Sprechakt an Änderungen der Umwelt (inkl. dem Adressaten) bewirkt.

Dabei sind für Entwicklung von *Embodied Conversational Agents* besonders der *propositionale* und der *illokutionäre Akt* entscheidend, da besonders auf den Inhalt des Gesagten und auf die eigentliche Intention des Sprechers eingegangen werden muß, um eine erfolgreiche Mensch-Maschine-Schnittstelle zu implementieren.

Nach Levelt[1989] können einige Hauptklassen von illukotionären Akten unterschieden werden:

- Der Sprecher möchte einen Sachverhalt mitteilen, er trifft eine *Feststellung*.
- Der Sprecher möchte, das etwas geschieht, und gibt einem Gesprächspartner eine *Anweisung*.
- Der Sprecher gibt bekannt, das er selbst einen *Auftrag* ausführen will.
- Der Sprecher möchte seine Gefühle bezüglich eines Sachverhalts mitteilen, dazu verwendet er eine *Äußerung*.
- *Deklarationen* verändern den Zustand der Welt. Dabei ist zu beachten, das jemand, der eine Deklaration macht, die Kompetenzen braucht, diesen Zustand zu ändern. Würde ein Student einem anderen Land den Krieg erklären, so würde dieses nichts am Zustand der Welt ändern, jedoch ein Staatsoberhaupt würde mit einem solchen Sprechakt sehr wohl den Zustand der Welt ändern.

4.1.2 Intonation

Die *Intonation* gibt unter anderem Aufschluß darüber, wie das Gesagte aufzufassen ist (Ernst, Scherz, Ironie, ...). Um eine Veränderung der Intonation zu erkennen, ist es jedoch wichtig die normale Stimmlage eines Menschen zu kennen. Dieses wird für ECAs in der näheren Zukunft nur dann erreichbar sein, wenn zunächst eine Kalibrierung des Systems auf den Benutzer erfolgt, wodurch jedoch die universelle Einsetzbarkeit leiden würde.
Des weiteren ist die Intonation ein wichtiges Turn-taking Signal. Eine gesenkte Stimme zum Ende einer Äußerung ist ein Turn-yielding Signal, während hingegen ein Anheben der Stimme zum Ende einer Äußerung ein Turn-keeping Signal ist. Dabei ist jedoch auch die Blickrichtung zu berücksichtigen (siehe Kapitel 4.3.2), da auch eine Frage mit einer höheren Tonlage abgeschlossen wird, auf die dann der Gesprächspartner antworten soll.
Außerdem zeigt die Betonung auch noch den Focus der Äußerung, d.h. sie markiert das *Rhema* des Gesagten.

4.1.3 Paraverbale Äußerungen

Paraverbale Äußerungen sind ein wichtiges Feedback-Signal, die dem Sprecher Interesse, Zustimmung, Ablehnung, etc. signalisiert. Dieses Feedback bestätigt den Sprecher in seiner Rolle, und zeigt zudem an, wie das Auditorium auf seine Aussage reagiert. Studien haben ergeben, daß, wenn dieser Feedback-Kanal wegfällt, der Sprecher nervöser wird, seine Äußerung öfter unterbricht und dann erneut beginnt [Levelt, 1989].

4.2 Gesten

4.2.1 Verschiedene Gestenkategorien

Gestik ist der wichtigste nonverbale Kanal einer Kommunikation, da durch Gesten Sachverhalte näher bestimmt, Emotionen gezeigt, Beispiele gegeben und verschiedenste Turntaking und Feedback-Signale ausgedrückt werden können. Nach Cassell[2000] können dabei folgende verschiedene Gestenkategorien unterschieden werden:

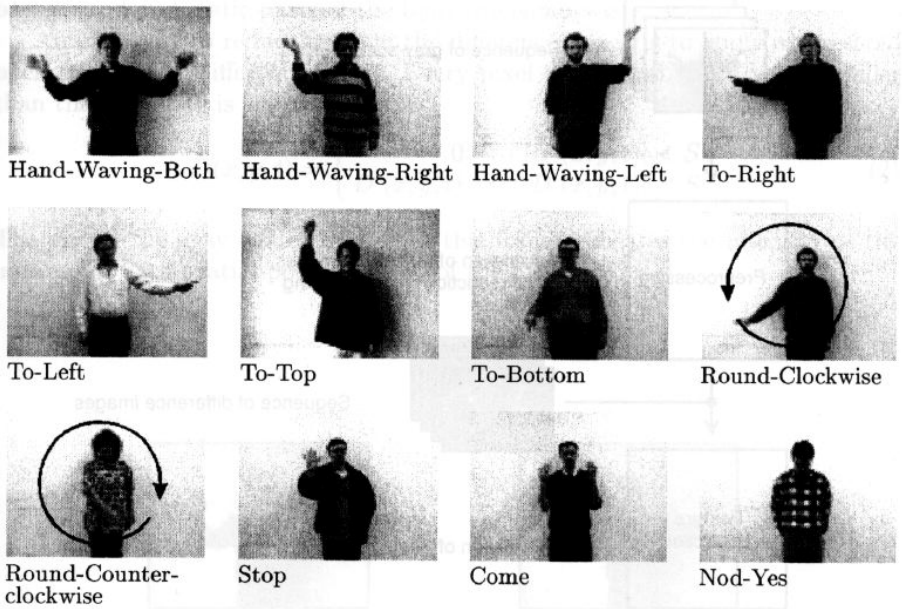

Abbildung 2: Beispiele emblematischer Gesten.

- *Emblematische Gesten* tragen einen festen Bedeutungsgehalt, unabhängig vom Kontext der Sprache. Dabei kann dieser Bedeutungsgehalt zwischen verschiedenen Kulturkreisen durchaus stark variieren. Zum Beispiel wird das *Victory-Zeichen* (gestreckter und gespreizter Mittel- und Zeigefinger) in den USA sowohl mit den Handrücken als auch der Handfläche zum Adressaten als Zeichen für einen Sieg aufgefaßt, in Großbritannien hingegen ist nur die Geste mit der Handfläche zum Beobachter eine Siegesgeste, wird dabei der Handrücken zum Beobachter gedreht, ist dieses eine grobe Beleidigung. In Abbildung 2 sind einige Beispiele für typische emblematische Gesten aufgeführt.

- *Propositionale Gesten* sind Gesten, die sich auf eine Proposition in der Sprache beziehen. Dabei müssen diese insofern von den spontanen Gesten differenziert werden, als das bei *propositionalen Gesten* die Zuordnung von der Geste zur Sprache bewußt erfolgt, während spontane Gesten unterbewußt ausgeführt werden (Beispiel: ”Das [Zeigegeste] Ding da muß weg”).

- *Spontane Gesten* beziehen wie *propositionale Gesten* ihren Bedeutungsgehalt aus dem Kontext der Sprache, werden jedoch unterbewußt ausgeübt. Dabei sind vier Unterkategorien zu unterscheiden:

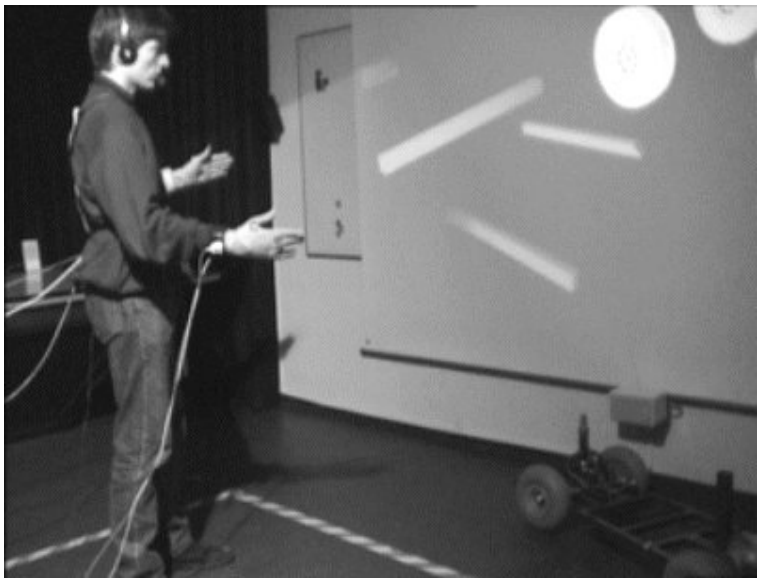

Abbildung 3: Ikonische Geste.

- *Ikonische Gesten* beschreiben die Form oder Lage von Objekten mit den Händen (siehe Abbildung 3).
- *Mimetische Gesten* imitieren die Interaktionen mit einem Objekt (siehe Abbildung 4).

Abbildung 4: Mimetische Geste.

- *Deiktische Gesten* sind Zeigegesten (siehe Abbildung 5).
- Ein *Beat* ist eine sehr kurze Geste, die Bedeutungsgehalt unterstreicht und häufig parallel zu betonten Worten vorkommt. Dabei werden die Hände kurz und ruckartig bewegt.

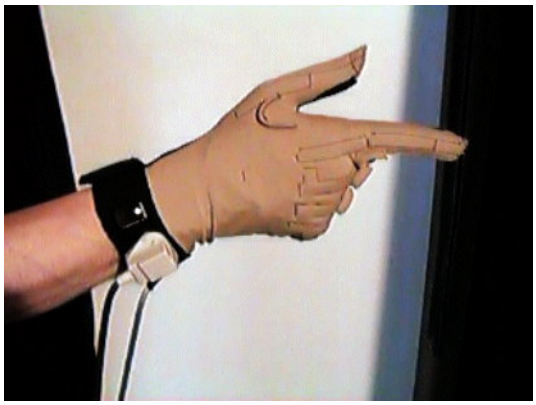

Abbildung 5: Deiktische Geste.

Die Erkennung von *emblematischen Gesten* ist inzwischen sehr gut implementiert, jedoch machen diese nur einen Bruchteil der Gesten in einem normalen Gespräch aus. Wesentlich mächtiger sind die spontanen Gesten, deren Bedeutungsgehalt entscheidend von dem Inhalt des Gesprochenen abhängt. Die Erkennung und Deutung dieser *spontanen koverbalen Gesten* ist noch aktuelles Forschungsgebiet.

4.2.2 Handstellungen

Die Stellung der Hände ist ein wichtiges Turn-taking Signal. So signalisieren erhobene Hände zum Abschluß einer Äußerung, daß der Sprecher noch nicht bereit ist, diese Rolle abzugeben, während hingegen gesenkte Hände ein Turn-yielding Signal sind. Hebt ein Zuhörer die Hände, so ist dies ein Turn-wanting Signal.

4.3 Sonstige Modalitäten

In diesem Abschnitt werden all die Kommunikationsmittel betrachtet, die sich nicht in die großen Kategorien *Sprache* und *Gestik* einordnen lassen.

4.3.1 Mimik

Auch Mimik ist ein wichtiges Signal in einer menschlichen Konversation. In der Mimik eines Menschen lassen sich leicht Emotionen erkennen, so daß der Gesichtsausdruck auf jeden Fall ein sehr bedeutendes Feedback-Signal ist. Dabei sind insbesondere die Stellung von Lippen und Augenbrauen von Interesse: Der Mund zeigt zum einen sehr offensichtlich die Gemütsverfassung eines Menschen an (zusammengekniffene Lippen => Anspannung, lächeln => Wohlwollen, etc.), zum anderen hilft die Beobachtung der Lippen dem Zuhörer bei dem akustischen Verständnis der Äußerungen des Sprechers. Die Augenbrauen sind sowohl Feedback- als auch Turn-taking Signal: Hochgezogene Augenbrauen signalisieren Interesse, zusammengekniffene Augenbrauen jedoch Skepsis. Erhobene Augenbrauen in Kombination mit einer Fixierung einer bestimmten Person zum Abschluß einer Äußerung ist eine Aufforderung an diese fixierte Person, zu dem Gesagten Stellung zu beziehen. Allgemein gelten erhobene Augenbrauen zum Ende einer Äußerung als Turn-yielding Signal.

4.3.2 Blickrichtung

Die Blickrichtung des Sprechers hat verschiedene Funktionen. Zum einen kann die Blickrichtung als *deiktische Geste* aufgefaßt werden, zum anderen dient die Blickrichtung als Turn-taking Signal. Wendet der Sprecher zum Abschluß einer Äußerung den Blick ab,

Abbildung 6: Beispiel für Turn-keeping.

so ist dies ein Turn-keeping Signal, wird vom Sprecher zum Abschluß seines Gesprächsbeitrag ein Zuhörer fixiert, so wählt der Sprecher diesen als Nachfolger (siehe Regel 2, Kapitel 3).

Der Blickkontakt von Zuhörer zu Sprecher ist ein Feedback-Signal, mit dem Interesse signalisiert wird.

4.3.3 Kopfbewegungen

Kopfbewegungen wie zum Beispiel Nicken oder Kopfschütteln sind ein häufig benutzter Feedback-Kanal, mit dem der Zuhörer seine Stellung zum Gesagten ausdrücken kann. Dabei gibt es jedoch wie bei emblematischen Gesten von Kulturkreis zu Kulturkreis abweichende Bedeutungen der vollführten Kopfbewegung.

4.3.4 Atmung

Die Atmung liefert Informationen über den Grad der Erregung der einzelnen Gesprächspartner (Schnelle, flache Atmung signalisiert Aufregung, tiefe ruhige Atmung strahlt Ruhe aus), ein deutliches Luftholen kann aber auch als Turn-wanting Zeichen aufgefaßt werden.

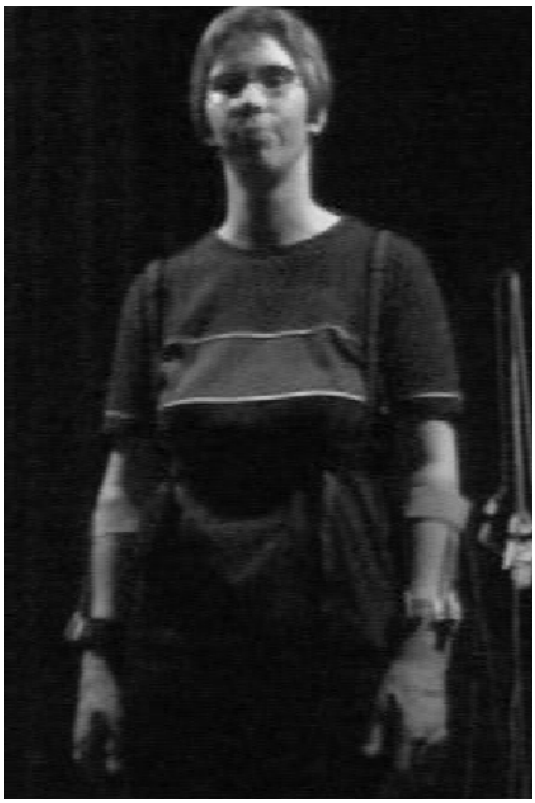

Abbildung 7: Beispiel für Turn-giving.

5 Problematisierung

Aus den vorangegangen Kapiteln läßt sich leicht erkennen, daß es eine extrem schwierige Aufgabe ist, menschliches konversationales Verhalten in einem *Embodied Conversational Agent* nachzubilden. Dabei erweist sich hauptsächlich die Kontextabhängigkeit der einzelnen Kommunikationskanäle im Zusammenhang mit der differenzierten Zeitlichkeit als Hürde. Zur korrekten Auswertung der nonverbalen Kommunikation muß eine semantische und pragmatische Analyse des Gesprochenen erfolgen (Bestimmung des propositionalen und illokutionären Akts des Sprechaktes), um mit diesem Wissen dann Gestik, Mimik, etc. zu interpretieren. Dabei ist eine sehr hohe Geschwindigkeit, mit der eine menschliche Konversation abläuft, zu erreichen.
Allein die Auswertung *spontaner koverbaler Gestik* ist noch immer Forschungsgebiet, so daß ein perfekte Interpretation sämtlicher menschlicher Kommunikationskanäle noch nicht verwirklicht ist. Auch die Generierung der Mimik und Gestik eines ECAs ist problematisch, da die komplexen Zusammenhänge der einzelnen Modalitäten immer abhängig vom Kontext der Sprache ist. Also muß zuerst das Thema erkannt werden, die Äußerung formuliert, um dann die nonverbalen Kanäle berechnen zu können.
Wenn diese Anforderungen nicht erfüllt werden, kommt es zu einer gestörten Kommunikation, die den Benutzer frustriert und davon abhalten wird, den ECA zu verwenden.

5.1 Zeitliche Abhängigkeiten

In diesem Abschnitt sollen einige Beispiele für die komplizierten zeitlichen Zusammenhänge gegeben werden.

- Der Stroke einer Geste kommt immer kurz vor oder simultan mit dem Bezugswort.
- Pausen beim Sprechen können als Stilmittel, Luftholen oder Turn-yielding Signal verstanden werden.
- Die zeitlichen Abhängigkeiten zwischen den einzelnen Modalitäten sind extrem kurz.
- Die Reaktionen der Zuhörer auf den Sprecher liegen ca. im Bereich 100ms - 1s.

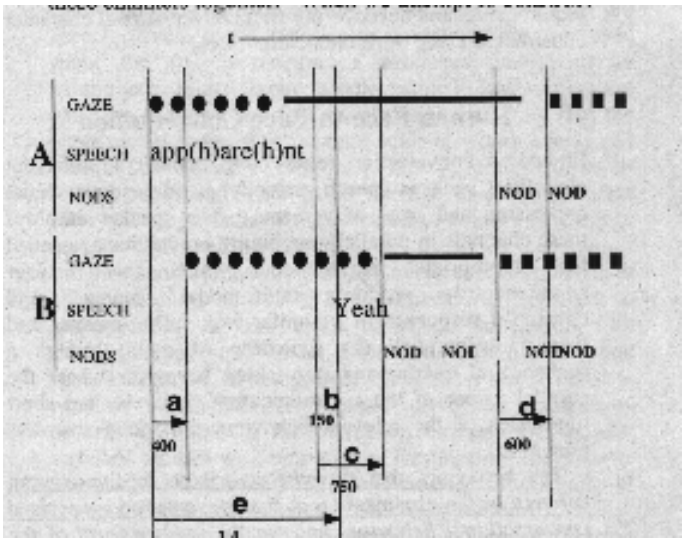

Abbildung 8: Betrachtung einer multimodalen Kommunikation: Kreise bedeuten das sich die Blickrichtung zum Gesprächspartner bewegt, Linien bedeuten Blickkontakt zum Partner, Quadrate stehe für einen abgewandten Blick. (h) steht für eine Anhebung der Stimme (aus [Cassel,1998]).

In Abbildung 8 werden die Modalitäten eines kurzen Gesprächsausschnitte analysiert. Auffällig sind hier die Parallelität der Modalitäten und die sehr schnellen Interaktionen der Gesprächspartner. Parallel zu der sprachlichen Äußerung wird über den Blickkontakt die Weitergabe der Sprecherrolle vereinbart. Das Nicken ist ein Feedback-Signal, mit dem der Zuhörer Zustimmung zu dem Gesagten anzeigt.

6 Zusammenfassung

Abschließend läßt sich sagen, daß menschliches konversationales Verhalten inzwischen recht gut erforscht ist, allerdings gibt es kontroverse Ansätze, wie man dieses modellieren soll. ECAs haben sicherlich das Potential zu einem mächtigen Werkzeug in der multimodalen Mensch-Maschine-Kommunikation, allerdings ist die Übertragung der Ergebnisse der Erforschung menschlicher Kommunikation auf *Embodied Conversational Agents* sehr schwierig:

- Die Vielzahl der Kommunikationskanäle sowie deren wechselseitige Abhängigkeiten sind immer noch Forschungsgebiet.
- Die zeitlichen Abhängigkeiten zwischen Sprache und nonverbaler Kommunikation sind sehr schwierig zu modellieren. Besonders die wechselnden Abhängigkeiten der Modalitäten im Zusammenhang mit dem Kontext der Sprache sind sehr komplex und können noch nicht umgesetzt werden.

Literatur

[1] [Cassel, 2000] Cassell, Sullivan, Prevost, Churchill (ed.),"Embodied Conversational Agents", The MIT Press, 2000.

[2] [Levelt, 1989] Levelt, "Speaking", The MIT Press, 1989.

[3] [Cassel, 1998] Cassel, Bickmore, Billinghurst, Campbell, Chang, Vilhjalmsson, Yan, "An Architecture for Embodied Conversational Characters" , Proceedings of the First on Embodied Conversational Characters, 1998

[4] [Bente & Krämer, 2001] Bente & Krämer, "Psychologische Aspekte bei der Implementierung und Evaluation von nonverbal agierender Interface-Agenten" , Proceedings Mensch & Computer 2001, S. 275-285, Teubner, 2001.

[5] [McNeill, 1992] McNeill, David, "Hand and Mind: What Gestures Reveal about Thought", University of Chicago Press, 1992.

[6] [Halliday, 1967] Halliday, M., "Intonation and Grammar in British English.", The Hauge: Mouton.

[7] [Austin, 1962] Austin, J. L.,' 'How to Do Things with Words", Oxford University Press, London, 1962.

[8] [Duncan, 1972] Duncan, S., "Some signals and rules for taking speaking turns in conversations." Journal of Personality and Social Psychology 23, 1972.

[9] [Yngve, 1970] Yngve, Victor H., "On getting a word in edgewise." Proceedings of Chicago Linguistics Society 6, S. 567-577, 1970.