Multimodal Interfaces

Seminarbericht

Seminar “Context Awareness”

Institut für Informatik der Universität Zürich

Eingereicht von

Jonas Tappolet, Domenic Benz

Abstract

Dieser Bericht behandelt die Grundlagen von multimodalen Interfaces. Dabei werden zuerst einige Grundlagen über die Mensch-Maschine-Kommunikation behandelt. Im zweiten Teil dieses Berichts wird auf verschiedene Typen von multimodalen Interfaces eingegangen sowie einige Prinzipien für das Design von multimodalen Systemen erläutert.

Inhaltsverzeichnis

Interfaces – die Grundlagen

Interface: Zweck

Seit es die ersten Computersysteme gibt, hat sich schon immer die Frage gestellt, wie die Informationen die in einem solchen System errechnet wurden an den Menschen weitergeleitet werden können. Weiter ist aber auch die Form der Übermittlung von Steuerbefehlen und Informationen vom Menschen an ein Computersystem von zentraler Bedeutung. Da Mensch und Maschine sehr unterschiedliche Arten der (nativen) Kommunikation haben, ist es notwendig, eine gemeinsame Basis zu finden, auf der Mensch und Maschine miteinander kommunizieren können. Solche Kommunikationsschnittstellen werden auch Man-Machine-Interfaces (MMI) genannt. In letzter Zeit hat sich die etwas neutralere Bezeichnung Human-Machine-Interface (HMI) durchgesetzt.

Human-Machine-Interface (HMI)

Es gibt verschieden Szenarien in denen ein HMI zum Einsatz kommt:

- Mensch und Maschine lösen Aufgabe gemeinsam (Interaktion)

  Hierbei handelt es sich um die klassische Arbeit mit dem Computer. Der Mensch ist im Dialog mit dem Computersystem und gibt Befehle welche der Computer ausführt und das Resultat präsentiert, worauf der Mensch erneut Befehle übermittelt.
- Maschine löst Aufgabe selbständig, Mensch überwacht

  Bei besonders rechenintensiven Arbeiten bei denen der Benutzer keinen Einfluss nehmen muss/darf. Zum Beispiel eine Visualisierung bei der

Berechnung eines Wettermodells wo der Mensch kontrollieren kann, ob noch gerechnet wird, und ganz grob ob auch das Richtige berechnet wird.

- Mensch löst Aufgabe selbständig, Maschine überwacht

  Diese Form der Interaktion ist zum Beispiel interessant für Menschen mit Behinderung oder ältere Menschen. Das Computersystem überwacht die Tätigkeiten des Menschen, und hat die Möglichkeit, eine Notsituation zu erkennen und rettende Massnahmen einzuleiten. Mehr dazu im Themengebiet „Assisted Living“.

Grundsätzlich kann die Kommunikation zwischen Mensch und Maschine anhand folgenden Diagramms betrachtet werden:

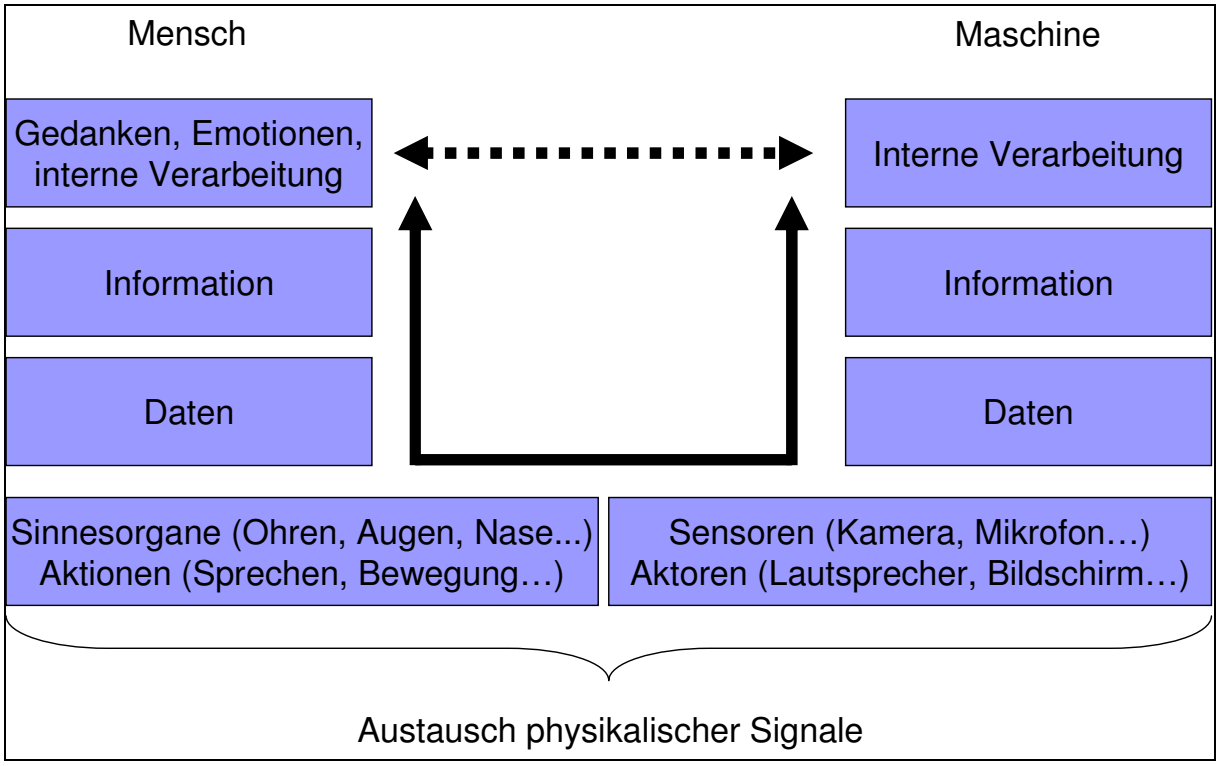

Abbildung 1: Kommunikation zwischen Mensch und Maschine Quelle: Eigene Darstellung

Wichtig bei obiger Darstellung ist, dass die Kommunikation Idealerweise auf der obersten Ebene stattfinden würde. Da dies aber aus technischen Gründen (noch) nicht möglich ist, muss für die Kommunikation ein Umweg gemacht werden. Wenn

ein Mensch dem Computer einen Befehl gebe will, so muss er diese Absicht zuerst in eine Information fassen. Diese Information wird dann in Daten codiert, welche über ein physikalisches Medium weitergegeben werden, und beim Computersystem empfangen werden. Hier werden die Daten wieder zu Informationen decodiert, und als Befehl an die interne Verarbeitung weitergereicht.

Dieser Umweg der gemacht werden muss ist sehr verlustbehaftet. Der Mensch hat im laufe seiner Entwicklung eine gewaltige Fähigkeit entwickelt, bruchstückhafte Kommunikationsfetzen sinnvoll zu vervollständigen um die ursprüngliche Information wiederherzustellen. Computersysteme zeigen hier im Vergleich zum Menschen sehr schlechte Fähigkeiten unvollständige Informationsteile zu ergänzen.

Warum neue Arten der Interaktion?

Die heute typische Form der Interaktion mit einem Computersystem ist Maus und Tastatur. Die Tastatur wurde von der Schreibmaschine übernommen, wobei die Maus eigens für den Computer entwickelt wurde. Dies geschah zu einer Zeit als die Computerindustrie noch in den Kinderschuhen steckte, und finanzielle und technische Einschränkungen die Entwicklung beeinflussten. Zwar sind Maus und Tastatur für einen versierten Computerbenutzer kein grosses Hindernis bei der Bedienung eines Computersystems, es ist aber nicht die natürliche Form der Kommunikation die ein Mensch gewohnt ist. Ein PC Anwender muss den Umgang mit diesen Eingabegeräten zuerst (mühsam) erlernen. Viele Menschen besuchen Wochenlang einen Kurs um das 10-Finger System zu erlernen, oder Sitzen einige Stunden mit der Maus in der Hand vor dem PC und üben die Hand-Augen-Koordination um den Mauszeiger auf dem Bildschirm zu positionieren. Dies zeigt, dass diese klassischen Eingabemethoden für den Menschen nicht intuitiv sind, und er die Bedienung eines Computersystems erst erlernen muss. Ziel sollte es aber

sein, dass ein Computersystem ohne vorgängige, langwierige Schulung bedienbar sein sollte. Dazu muss aber bei den Ein- und Ausgabemethoden auf Verfahren zurückgegriffen werden, die auch in der zwischenmenschlichen Kommunikation vorkommen, der Mensch also natürlich beherrscht, oder in seiner Entwicklung bereits erlernt hat.

Ein weiterer Punkt ist, dass es viele Menschen gibt, die eine Behinderung haben und daher den Computer nicht mit der Maus und Tastatur bedienen können, oder die Ausgabe am Bildschirm nicht lesen können. Daher muss ein Computersystem eine Schnittstelle bieten, die auch für behinderte Menschen bedienbar ist.

Komponenten der natürlichen menschlichen Kommunikation

Die zwischenmenschliche Kommunikation beschränkt sich nicht alleine auf die Sprache. Es ist eine Vielzahl von verschiedenen Kommunikationskanälen mit unterschiedlicher Bandbreite.

- Sprache

  Über die Sprache ist der Mensch in der Lage sehr viele Informationen auszutauschen. Es ist der Kommunikationskanal mit der höchsten Bandbreite, und daher der wichtigste in der heutigen Gesellschaft die einen sehr hohen Bedarf an Informationsaustausch hat.

- Gestik

  Die Gestik wird meistens zur Untermalung der Sprache verwendet. Sie dient der Verdeutlichung des gesagten. Alternativ kann sie die gesprochene Sprache auch ganz ersetzen (Gebärdensprache für Gehörlose).

- Mimik

  Die Mimik ist eine der Hauptausdrucksformen für Emotionen. Es ist dem Menschen kaum möglich, die Mimik bei der Kommunikation zu unterdrücken.

- Blick / Blickrichtung

  Durch den Blick eines Menschen werden Emotionen transportiert. Die Blickrichtung wird zur Kontrolle verwendet, um beispielsweise zu verifizieren mit wem eine Person gerade spricht.

- Lautstärke, Tonfall

  Die Lautstärke der Stimme kann einerseits das Resultat einer Anpassung an die Umgebung sein (Bibliothek, Fabrikhalle), aber auch ein Ausdruck von Emotionen (anschreien, leises Sprechen bei Unsicherheit).

- Lippenbewegung

  Dient zur Kontrolle, wer gerade am Sprechen ist.

- Gerüche

  Ein relativ unerforschtes Gebiet. Die Erkenntnis aus der Tierwelt, wo Gerüche wichtige Mittel der Kommunikation sind, haben dazu veranlasst, diese Reaktionen auch beim Menschen zu testen. Bisher mit widersprüchlichen Resultaten.

- Haptik

  Berührungen können auch Emotionen übermitteln. Beispielsweise bedeutet ein Schulterklopfen Anerkennung und Stolz.

Ziele eines HMI

Eine Mensch-Maschine-Schnittstelle sollte immer auf die entsprechende Anwendung ausgelegt sein. Zum Beispiel ist eine Spracherkennung wenig Sinnvoll für ein System das vorwiegend in sehr lauten Umgebungen eingesetzt wird.
Des Weiteren sollte die Schnittstelle für den Menschen eine möglichst intuitive Bedienung erlauben. Sie sollte also in einer Art und Weise die Möglichkeit zur Kommunikation bieten, wie sie sich der Mensch bereits gewohnt ist, und wie er es bereits gelernt hat aus anderen Bereichen.

Verschiedene Typen

Folgend sollen kurz einige Typen von Interfaces erläutert werden.

Gestenerkennung

Die Gestenerkennung basiert auf den Aufnahmen einer Kamera. Es wird zuerst versucht, die Bewegung aus dem Bildmaterial zu isolieren, um danach die Muster in der Bewegung zu analysieren. Ein grosser Vorteil davon ist, dass ein solches System nicht die volle Aufmerksamkeit des Bedieners benötigt, und daher nebenläufig bedient werden kann. Ein weiterer Vorteil ist die Einfachheit auf der Hardwareseite, und dadurch tiefe Kosten. Es wird lediglich eine Kamera benötigt, welche schon bei niedriger Qualität (Webcam) passable Resultate liefert. Der Nachteil liegt zur Zeit noch darin, dass das System mithilfe einer Initialgeste aktiviert werden muss. Dies dient zur Einsparung von Rechenleistung, da nur auf das Auftreten einer einzigen Geste geprüft werden muss, aber auch zur Verhinderung von Fehleingaben. Da das System meistens den Kontext nicht kennt, kann es nicht unterscheiden zwischen einer Geste die ein Befehl darstellt und einer, die einer Drittperson gilt. An der TU

München wurde ein System entwickelt zur Bedienung eines InfoTainment Systems wie sie in Automobilen zum Einsatz kommen. Der Grund für die Verwendung eines Interfaces mit Gestenerkennung war die Tatsache, dass beim Autofahren die Aufmerksamkeit und damit die Blickrichtung nicht von der Strasse genommen werden sollte.

Abbildung 2: Bedienung eines Bordcomputers mithilfe von Gesten. Quelle: TU München, Lehrstuhl für Mensch-Maschine-Kommunikation

Spracherkennung

Die Spracherkennung hat seitens der Hardware noch geringere Anforderungen als die Gestenerkennung. Es ist lediglich ein Mikrofon und ein A-D Wandler (Soundkarte) notwendig. Wenn die Gesprochene Sprache aufgezeichnet ist, versucht die Spracherkennung die Daten wieder zu decodieren und daraus Informationen zu generieren. Ein Problempunkt ist das Herausfiltern von Umgebungsgeräuschen. Eine computergestützte Spracherkennung wird sehr rasch unbrauchbar wenn die Umgebungsgeräusche einen gewissen Pegel überschreiten. Eine weitere Herausforderung ist die Erkennung der Semantik. Beispielsweise sind die beiden Sätze „Er isst einen Fisch“ und „Er ist ein Fisch“ von der Aussprache her sehr ähnlich. Das Spracherkennungssystem ist hier gefordert um zu Entscheiden

welcher der beiden Sätze gesagt wurde. Hierzu ist es Notwendig, dass das System den Kontext kennt und somit entscheiden kann, welcher Satz im Zusammenhang mehr Sinn macht. Die Erkennung des Kontextes ist gerade bei der Sprache ein Problem, da sie sehr komplex aufgebaut ist. Des Weiteren ist es Möglich, dass Zweideutigkeiten erst durch einen weiteren Kommunikationskanal wie zum Beispiel Gestik geklärt werden.

BCI: Brain-Computer-Interface

Die Idee beim Brain-Computer-Interface ist, dass die elektrische Hirnaktivität des Menschen interpretiert und in Steuersignale umgewandelt wird. Es ist also nur noch Notwendig, die Befehle zu „denken“. Dies basiert auf der Erkenntnis, dass bereits der Gedanke an eine Bewegung eine messbare Hirnaktivität auslöst. So kann ein Benutzer an die Bewegung „rechtes Bein heben“ denken, was vom Computer zum Beispiel für die Auswahl eines nächsten Elements dient. Diese Art von Kommunikation ist zurzeit noch experimentell und findet noch kaum nützliche Anwendung. Es ist eine Vielzahl von Elektroden am Kopf notwendig um die Signale zu empfangen, was das ganze etwas Umständlich in der Handhabung macht. Je weniger Elektroden angebracht werden, umso ungenauer sind die Resultate die erzielt werden können. Eventuell wäre ein längeres Training notwendig was wiederum nicht sehr praktikabel ist für ein Computerinterface. Das grösste Problem ist, dass der Gedanke an eine Bewegung eine Aktion auslöst. Das heisst, dass sich der Anwender nicht bewegen darf, ausser er will bewusst ein Signal an den Computer senden.

Abbildung 3: Brain-Computer-Interface. Quelle: http://mp.tech.org.sg/ITR5/DSC_0172.jpg

Probleme von einzelnen Interfaces

Es gibt kein einzelnes Interface das nur Vorteile und keine Nachteile hat. Jedes Interface hat seine spezifischen Nachteile die andere nicht haben.

Ein weiterer Punkt ist, dass die Bandbreite eines Interfaces mit der es Informationen aufnehmen kann jeweils sehr hoch ist. Der Flaschenhals liegt hier beim Menschen. Er kann zum Beispiel schnell sprechen, aber ab einer gewissen Geschwindigkeit geht es nicht mehr schneller. Der Mensch nutzt dafür aber mehrere Kanäle für die Kommunikation. Wenn sich ein Interface nur auf einen konzentriert, so sind die anderen Kanäle verlorene Bandbreite.

Ein letzter Punkt sind die verschiedenen Anwendungsgebiete. Ein Interface kann in einer Situation die optimale Kommunikationsmethode sein, aber im nächsten Augenblick schon unbrauchbar werden, weil sich die äusseren Umstände geändert haben. Hier muss ein Kommunikationssystem flexibel reagieren können.

Multimodale Interfaces

Als Lösung der oben genannten Problemstellungen kommen Multimodale Interfaces in Frage. Ein Multimodales Interface vereint mehrere Ein-/Ausgabemethoden um so

einerseits die spezifischen Nachteile der einzelnen Interfaces durch die Kombination mit anderen auszumerzen, andererseits kann die Bandbreite erhöht werden durch parallele eingaben und zuletzt kann auch dynamisch auf sich ändernde Umstände reagiert werden. Dadurch ergibt sich ein sehr viel zuverlässigeres System als bei der Verwendung von isolierten Interfaces.

Definition

Multimodale Systeme verarbeiten zwei oder mehrere kombinierte Benutzereingabemethoden wie Sprache, Stift, Berührung (Touchscreen), Gesten, Blickrichtung oder Kopf- und Körperbewegung.

Vorteile

- Fehlertoleranter

  Werden zum Beispiel Spracherkennung und Lippenbewegung kombiniert, so kann bei einem nicht eindeutig erkannten Wort die Lippenbewegung zu Hilfe genommen werden, um zu entscheiden um welche der erkannten Alternativen es sich handelt.

- Schneller

  Durch die parallele Eingabe von Befehlen addieren sich die Bandbreiten der einzelnen Kanäle zu einer gesamten, grösseren Bandbreite. Ein Mensch könnte beispielsweise einen Brief diktieren und gleichzeitig mithilfe von Gesten Anweisungen für die Formatierung geben.

- Natürlicher

  In der zwischenmenschlichen Kommunikation werden fast immer zwei oder mehrere Kommunikationskanäle verwendet. Es für den Menschen daher

selbstverständlich, dass nicht nur einer, sondern mehrere Kanäle beim Gegenüber verarbeitet werden.

Verschiedene Typen von Multimodalen Interfaces

Multimodale Interfaces können je nach ihrer Ausgestaltung verschiedenen Typen zugeordnet werden. Je nach Anwendungsbereich und Anforderungen kommt eine andere Form eines multimodalen Interfaces zum Einsatz. Die wichtigsten verschiedenen Typen [OVI02] oder auch Betriebsmodi sollen nachfolgend erläutert werden.

- **Aktive Interfaces**

  Aktive Interfaces stellen die klassische Art der Mensch-Maschine-Kommunikation dar. Bei Schnittstellen dieser Art will der Benutzer explizit mit dem System kommunizieren und gibt zu diesem Zweck eindeutige, spezifizierte Kommandos an das System weiter. Das System erkennt die erhaltenen Befehle und führt die damit verknüpften Aktionen aus.

- **Passive Interfaces**

  Passive Interfaces werden im Gegensatz zu aktiven nicht entwickelt, um vom Benutzer explizit gegebene Kommandos zu verarbeiten. Sie sollen sich viel mehr dem Benutzer anpassen und auf sein natürliches Verhalten und auf seine jeweilige Umgebung reagieren. Solche Systeme überwachen also einen Benutzer oder eine Umgebung mittels verschiedener Sensoren und führen anhand von erkannten Mustern die jeweils passende Aktion aus.

  Die Schwierigkeiten bei passiven Interfaces liegen darin, zu erkennen, wann das Verhalten des Benutzers analysiert und darauf reagiert werden soll sowie darin, verschiedene, sich allenfalls widersprechende erkannte Muster zu synchronisieren und auszuwerten.

- **Gemischte multimodale Interfaces**

  Aktive und passive Modi lassen sich beliebig miteinander kombinieren, um verschiedene mögliche Ziele zu erreichen. Beispielsweise kann ein System welches mit aktiver Sprachsteuerung arbeitet durch den gleichzeitigen, redundanten Einsatz von passiver Überwachung der Lippenbewegungen weniger fehleranfällig werden.

- **Zeitlich abgestufte Interfaces**

  Zeitlich abgestufte Interfaces sind Schnittstellen, welche verschiedene Modalitäten verarbeiten, die zeitlich aufeinander folgen. Die jeweils später eingesetzten Modalitäten verwenden dabei die Inputs der vorhergegangenen.

  Zeitlich abgestufte Interfaces können dabei sowohl als aktive als auch als passive oder gemischte Interfaces gestaltet sein. Man könnte sich beispielsweise ein Lagerverwaltungssystem vorstellen, welches als zeitlich abgestuftes multimodales Interface gestaltet wird. Bei einem solchen System würde beispielsweise zuerst ein Objekt mittels Blick markiert, anschliessend per Sprachbefehl die mit dem soeben gewählten Objekt auszuführende Aktion bestimmt und schliesslich per Gestik der Ort gewählt, an welchem die erfasste Aktion ausgeführt werden soll.

Prinzipien für das Design von MM Interfaces

Im Folgenden soll auf einige grundlegende Prinzipien für das Design von multimodalen Schnittstellen gemäss [RAM03] eingegangen werden.

Synchronisation

Wenn verschiedene Modalitäten zum Einsatz kommen, ist es von entscheidender Bedeutung, dass diese sinnvoll synchronisiert werden. Dies ist insofern wichtig, da verschiedene Modalitäten nicht zwingend in derselben Dimension ablaufen. Beispielsweise ist Zeit die wichtigste Dimension bei der Eingabe von Befehlen über Spracherkennung. Bei der Arbeit mit Gestik stellt jedoch der verfügbare Raum eine ebenfalls grundlegend wichtige Dimension dar. Systeme, welche mit verschiedenen Modalitäten und verschiedenen Dimensionen arbeiten müssen darum sicherstellen, dass die verschiedenen Inputs oder Outputs korrekt miteinander synchronisiert und verknüpft werden, damit die richtige Aktion ausgeführt werden kann.
Ungenügende Synchronisation von verschiedenen, redundanten Output-Modalitäten kann beispielsweise zu grosser Verwirrung führen. Als Beispiel könnte ein Routenplanungssystem dienen, welches die vom Benutzer getätigte Auswahl visuell hervorhebt und über eine Sprachausgabe zusätzliche, ergänzende Informationen liefert. Es muss in einem solchen System sichergestellt sein, dass die zusätzlichen Informationen, welche die Sprachausgabe liefert auch zur getätigten Auswahl passen, selbst wenn der Benutzer verschiedene Aktionen kurz hintereinander ausführt. Wird in einem solchen System die Synchronisation vernachlässigt, so wird es schwierig bis unmöglich, dieses sinnvoll zu nutzen.

Abschwächung/Anpassung

Multimodale Systeme sollten sich analog der natürlichen zwischenmenschlichen Kommunikation abschwächen und sich den Gegebenheiten anpassen. Die zwischenmenschliche Kommunikation funktioniert beispielsweise auch dann, wenn bei einem Telefongespräch sämtliche visuellen Kommunikationskomponenten wegfallen. Dies wird möglich durch das hohe Mass an Redundanz in der zwischenmenschlichen Kommunikation.

Bei multimodalen Systemen kann beispielsweise ein Wechsel der Umgebung oder auch des Benutzers zum Wegfall einer oder mehreren Modalitäten führen. So kann beispielsweise die Bedienung eines mobilen Gerätes in einer lauten Umgebung den Einsatz von Gestik, einer Tastatur oder eines Touchscreens erfordern. Wechselt der Benutzer nun beispielsweise in ein Fahrzeug, fallen diese Modalitäten weg und das Gerät sollte per Spracheingabe gesteuert werden können. Bei einem erneuten Wechsel in eine stille Umgebung wie zum Beispiel ein Meeting oder auch eine Vorlesung sollte beispielsweise die Ausgabe-Modalität von Ton (Rufton/Sprachausgabe) auf Haptik (Vibration) und visuelle Signale ändern. Das Gerät sollte in einer solchen Umgebung auch nicht mehr mit Sprachbefehlen gesteuert werden müssen.

Im Folgenden soll kurz auf einige die Anpassungsfähigkeit von multimodalen Systemen betreffende Punkte eingegangen werden.

- **Zusätzliche Modalitäten**

  Ein multimodales System wird umso anpassungsfähiger, je mehr zusätzliche, redundante Modalitäten zum Einsatz kommen.

- **Sich ergänzende Modalitäten**

  Beim Einsatz von komplementären Modalitäten ist Vorsicht geboten. Dies sind typische Punkte, an welchen ein multimodales System versagen kann, da gewisse Funktionen von einzelnen Modalitäten abhängig werden können. In diesem Fall muss sichergestellt werden, dass die notwendigen Modalitäten in der typischen Umgebung und beim typischen Benutzer vorhanden sind, da das System ansonsten nicht bedienbar ist.

- **Sich verändernde Möglichkeiten**

  Bei der Entwicklung von multimodalen Systemen muss analysiert werden, welche äusseren Bedingungen sich verändern können. Es muss berücksichtigt werden, ob das System von verschiedenen Benutzern verwendet und/oder in verschiedenen Umgebungen zum Einsatz kommt.

Gemeinsamer Status für verschiedene Modalitäten

Das erfolgreiche Lösen eines Problems in der zwischenmenschlichen Kommunikation bedingt, dass alle an der Konversation Teilnehmenden einen gemeinsamen mentalen Status besitzen. Dies bedeutet beispielsweise, dass die beteiligten Personen ein Kopfnicken des Vorsitzenden alle im selben Kontext auf dieselbe Art interpretieren.

Für den Einsatz von verschiedenen Modalitäten in der Mensch-Maschine-Kommunikation trifft das oben genannte ebenfalls zu. Das System muss eine Möglichkeit haben, bei einem Wechsel der Modalität herauszufinden worauf sich der neue Input bezieht oder abstützt. Nur durch den Einsatz eines gemeinsamen Status für die beteiligten Modalitäten kann ein multimodales System die über verschiedene

Kanäle erhaltenen Inputs sinnvoll verarbeiten und mit dem Benutzer einen sinnvollen multimodalen Dialog führen.

Nachfolgend soll auf einige den gemeinsamen Interaktionsstatus betreffende Punkte eingegangen werden.

- **Wechsel der Modalität**

  Wechselt der Benutzer inmitten eines Dialogs aufgrund von äusseren Umständen oder Präferenzen die Input-Modalität, so können bereits erfasste Inputs verloren gehen. Damit dies nicht geschieht, muss das System einen zentralen, allen Modalitäten zur Verfügung stehenden Interaktionsstatus verwalten.

- **History**

  Das System kann diesen Interaktionsstatus nutzen, um eine History-Funktion zu verwirklichen. Es kann sodann mittels der vergangenen Interaktionsdaten einen auf den Benutzer abgestimmten Dialog verwenden, der es dem Benutzer erlaubt, schnellstmöglich und gemäss seinen Präferenzen und Möglichkeiten mit dem System zu arbeiten.

Multimodale Interfaces sollten vorhersagbar sein

Multimodale Interfaces geben dem Benutzer eine Vielzahl an Möglichkeiten, mit dem System zu interagieren und zu arbeiten. Damit der Benutzer aber auch effektiv und effizient mit einem solchen System arbeiten kann, muss er wissen, welche Modalitäten er zu welchem Zweck benutzen kann und welche Ergebnisse aus ihrem Einsatz resultieren. Falls der Benutzer gewisse Modalitäten für spezielle Funktionen

nicht verwenden kann, zum Beispiel kann eine Unterschrift nicht mittels Sprachbefehlen gegeben werden, muss das System dies dem Benutzer in geeigneter Weise mitteilen und seine Abfragen entsprechend präsentieren.
Insbesondere für folgende Punkte sind vorhersagbare Schnittstellen von zentraler Bedeutung.

- **Entlockung des korrekten Inputs**

  Zweckdienlich gestaltete Dialoge helfen, den Benutzer durch das System zu führen. Durch ein intelligentes Design der verschiedenen Dialoge mit dem Benutzer kann das System dem Benutzer die richtigen respektive einen Passenden Input entlocken. Dies fördert die Geschwindigkeit, mit welcher die anfallenden Aufgaben gelöst werden. Darüber hinaus steigt die Zufriedenheit und Akzeptanz des Benutzers, da unnötige Wiederholungen der Eingaben wegfallen.

- **„Lost in Space“ Problem**

  Bereits durch herkömmliche Grafische Benutzerschnittstellen sind die Probleme bekannt, welche eine mit zu viel Funktionalität ausgestattete Schnittstelle verursachen kann. Der Benutzer fühlt sich schlicht überfordert von der Vielzahl der sich ihm bietenden Möglichkeiten. Er weiss folglich nicht mehr, was er als nächste Aktion ausführen kann respektive soll.

  Werden dem Benutzer nun noch verschiedene Interaktions-Modalitäten zur Verfügung gestellt, verschärft sich dieses Problem noch zusätzlich.

  Ein auf den individuellen Benutzer und seine Möglichkeiten abgestimmter Dialog ist darum unerlässlich.

Context Awareness

Ein wichtiger Punkt, welcher in diesem Dokument bereits mehrmals erwähnt wurde ist die Kontextsensitivität eines multimodalen Systems.

Multimodale Systeme müssen sich dem Benutzer, seinen Möglichkeiten und seiner Umgebung anpassen, um sinnvoll eingesetzt werden zu können. Es muss sichergestellt werden, dass dem Benutzer die bestmögliche Kombination der vorhandenen Modalitäten zur Bearbeitung einer bestimmten Aufgabe zur Verfügung gestellt werden.

Gemäss [RAM03] orientiert sich die bestmögliche Kombination in diesem Kontext an den folgenden Kriterien:

- Bedürfnisse und Möglichkeiten des Benutzers
- Möglichkeiten des Systems / des Geräts
- Zur Verfügung stehende Bandbreite zwischen Gerät und Netzwerk
- Zur Verfügung stehende Bandbreite zwischen Gerät und Benutzer
- Durch die Umgebung des Benutzers diktierte Anforderungen (z.B. hands-free als Lenker eines Fahrzeugs)

Gerüchte über multimodale Interaktion

Nachfolgend sollen zwei interessante Gerüchte über multimodale Interaktion gemäss [OVI99] besprochen werden.

1. Gerücht: If you build a multimodal system, users will interact multimodally.

Die Benutzer haben eine starke Präferenz, multimodal zu interagieren. Dies bezieht sich jedoch in den meisten Fällen auf räumliche Interaktions-Domänen. Beispielsweise interagierten 95% der Benutzer multimodal in einer räumlichen Problem-Domäne, als sie die Wahl hatten zwischen Spracheingabe und Pen-Input [OVI97]. Dies bedeutet jedoch nicht, dass die Benutzer aufgrund solcher Präferenzen automatisch jeden Befehl an das System in multimodaler Form geben. Die Benutzer verwenden in den meisten Fällen eine Mischung aus unimodaler und multimodaler Interaktion. Eine Studie [OVI97b] hat gezeigt, dass die Benutzer typischerweise 20% der Befehle in multimodaler Form äussern, während in den restlichen Fällen die Befehle entweder per Sprache oder über Pen-Input gegeben wurden.

Ob Benutzer multimodal oder unimodal interagieren hängt zu einem grossen Teil von der zu lösenden Aufgabe ab. Bei räumlichen Aufgaben (z.B. Auswahl einer Lokation, einer Grösse oder einer Richtung) interagieren die Benutzer in fast allen Fällen multimodal.

Es lässt sich also sagen, dass die Benutzer die Möglichkeiten schätzen, welche multimodale System bieten, trotzdem interagieren die meisten Benutzer in einer Mischform aus unimodaler und multimodaler Kommunikation.

10. Gerücht: Enhanced efficiency is the main advantage of multimodal systems.

Oft wird davon ausgegangen, dass der wichtigste Vorteil eines multimodalen Systems in erhöhter Schnelligkeit und Effizienz liegt. Beispielsweise betrug der Geschwindigkeitszuwachs in einem multimodalen Pen-Voice System 10% verglichen mit einem System, welches nur durch Sprache bedient wurde [OVI97]. Zu beachten ist jedoch, dass dieser Geschwindigkeitsvorteil nur in einer räumlichen Problemdomäne beobachtet wurde.

Multimodale Interaktion bringt jedoch andere Vorteile, welche nennenswerter sind als der moderate Geschwindigkeitsgewinn. So kann beispielsweise die Fehlerrate bei Spracheingabe durch multimodale Interaktion um 36-50% gesenkt werden [OVI97].

Ein weiterer Vorteil liegt in der Berücksichtigung der Präferenz der Benutzer, multimodal zu interagieren. Im Weiteren erscheint eine erhöhte Flexibilität bei der Interaktion als grosser Vorteil gegenüber unimodalen Systemen. Da der Benutzer zwischen den verschiedenen Modalitäten wählen kann, kann ein multimodales System beinahe von jedem an jedem Ort eingesetzt werden.

Quellen

[OVI02] Oviatt, S.L., Multimodal Interfaces,
http://www.cse.ogi.edu/CHCC/Publications/zHCI%20Handbook%20MMI-%20Oviatt%20SS.pdf

[RAM03] Raman, T.V., User Interface Principles For Multimodal Interaction**,**
http://www.almaden.ibm.com/cs/people/tvraman/chi-2003/mmi-position.html

[OVI97] Oviatt, S.L. Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction 12*, (1997), 93–129.

[OVI97b] Oviatt, S.L., DeAngeli, A. and Kuhn, K. Integration and synchronization of input modes during multimodal human-computer interaction.
In *Proceedings of Conference on Human Factors in Computing Systems CHI'97* (March 22–27, Atlanta, GA). ACM Press, NY, 1997, 415–422.