# Sensoren für smart devices und wearable computers: GPS, Kameras, Mikrophone, Biometrie

Andreas Frei
arfrei@iiic.ethz.ch

Fachseminar SS2000 in Ubiquitous Computing

---

**Abstract**

Sensoren werden immer kleiner und spezifischer in ihrer Funktionalität. Sie werden in Ubiquitous Computing eingesetzt, um den Menschen in seiner Umgebung zu erfassen, seine Bewegungen zu erkennen und biometrische Daten zu erhalten. Anhand der Auswertungen der Sensordaten wird ein Gesamtbild des Menschen erzeugt und in eine digitale Form gebracht.

Dieses Dokument geht in zwei Richtungen, vertieft auf die Mensch zu Maschine Kommunikation ein. Einerseits befindet sich der Mensch dauernd in einem Umfeld (Kontext), das aus Raum, Zeit und Umgebung besteht. Verschiedene Signale aus der Umgebung werden mit Kameras und Mikrophonen aufgezeichnet und zur Kontextinterpretation herangezogen.
Andererseits ist der Mensch begleitet von Emotionen, die ihn in vielen Entscheidungen beeinflussen. Durch Erkennen dieser Emotionen gibt es eine neue Mensch zu Maschine Kommunikation, die den Menschen in verschiedenen Anwendungen unterstützen können.
In verschiedenen Beispielen und Betrachtungsweisen werden auf diese Thematiken eingegangen.

## 1. Kontexterkennung

Der Mensch befindet sich in einem dauernd ändernden Kontext. Dabei wird der Kontext beeinflusst durch den Zusammenhang von Raum, Zeit und Umgebung. In einer Mensch zu Mensch Kommunikation wird ein Hintergrundwissen vorausgesetzt, z.B.: Wissen die Gesprächspartner, wo sie sich im Moment befinden, über welche Themen schon gesprochen wurde oder was die Gründe für dieses Treffen sind. Die Diskussion der Gesprächspartner baut auf diesem Hintergrundwissen auf und erweitert den jeweiligen Wissensstand um die neu gewonnenen Informationen. Für eine produktive Mensch zu Maschine Kommunikation ist dieses Hintergrundwissen die Basis, die zuerst von der Maschine erkannt werden muss.
Das Umfeld eines Menschen ändert sich im Laufe seiner Aktivitäten. Diese Änderungen werden von Sensoren erfasst und nehmen Einfluss auf die digitale Wahrnehmung des Computers vom Menschen.
Aus Informationen, die dem Computer schon zur Verfügung stehen, z.B.: Zugfahrplan oder Vorlesungsverzeichnis, kann der Mensch einem groben, ungenauen Standort zugeordnet werden. Diese Hintergrundinformation, also Standort des Benutzers kann dem Computer helfen, auf Anfragen des Menschen die richtige Information bereit zu halten, z.B.: Wann komme ich in Zürich an oder wo findet meine nächste Vorlesung statt.

Um den sich fortlaufend ändernden Kontext zu erkennen und zu analyse Zwecken weiter zu verarbeiten, braucht es einen Computer oder einen ständigen Begleiter, der die Informationen aufnimmt. Der Begleiter lernt die Gewohnheiten des Benutzers kennen und hat Zugang zu den Informationen, die der Person später nützlich sein können. Ein Beispiel

eines solchen Schattens könnte ein Plüsch-Spielzeug (The Familiar [2]) darstellen, das alle Aktivitäten des Benutzers miterlebt.

Die bis heute eingesetzten Sensoren werden unten beschrieben. In einem weiteren Unterkapitel werden verschiedene Anwendungen und heutige Möglichkeiten zur Kontexterkennung aufgezeigt.

### 1.1 Sensoren

**Kameras** und **Mikrophone** ermöglichen dem Computer das Sehen und Hören, elektronisch zu erfassen. Vorwiegend werden CCD Kameras dafür eingesetzt.
Die menschliche Wahrnehmung des Sehens und Hörens wird durch physikalische Sensoren erfasst. Wahrnehmung beinhaltet noch vieles mehr, persönliche Empfindungen nehmen beim Sehen und Hören Einfluss. So werden Farben sehr unterschiedlich wahr genommen und lösen je nach Gefühl andere Empfindungen aus. Auch die Sprache hängt stark ab von der Art und Weise wie Sätze betont werden und kann vom Hörer unterschiedlich interpretiert werden. Der Mensch sieht und hört mit seinen Gefühlen!
Das **Global Positioning System** (GPS) wird zur Ortslokalisierung eingesetzt. Damit kann die Position eines Menschen oder Gegenstandes auf 5 cm genau bestimmt werden. Mit Hilfe des differentiellen GPS, mit einer ortsfesten Referenzstation, kann eine Genauigkeit von bis zu 1 cm erreicht werden.
**Wearable Computers** halten im Gebiet des Ubiquitous Computing Einzug. Ein Computer-System, mit dem die verschiedensten Sensordaten gesammelt und direkt „on-body“ ausgewertet werden können. Es besteht also keine Abhängigkeit zu einem Gross-Computer ,der die Daten auswertet. Diese Systeme werden immer leistungsfähiger und kleiner. Sie erlauben dem Benutzer mehr Daten und komplexere Algorithmen anzuwenden.

Neben den Sensoren zur Erkennung eines Personenumfeldes gibt es noch die Geräte zum Visualisieren der gewonnenen Daten. Dazu gehört ein Brillen-Display, das dem Benutzer erlaubt, ein Bild oder Videosequenz auf das Auge zu projizieren.

### 1.2 Anwendungen

Der Kontext kann aus visuellen Betrachtungen, einem Videosystem erkannt werden. Dabei werden im System von Starner, Schiele und Pentland [1] Aktivitäten erkannt. Dieses System hat eine Anwendung gefunden im Spiel „The Patrol“. Als Assistent steht dieses Videosystem den MIT-Studenten zur Verfügung, um in den Gängen des MIT nach Gegnern zu jagen. Es ermöglicht die Ortslokalisierung des Spielers. Aktivitäten oder Aufgaben, die der Spieler nun zu erledigen hat, z.B.: Wechseln des Standortes nach einem Treffer, wird von diesem System erkannt und kann dem Spieler eine Unterstützung sein.

Ein weiteres System zur Erkennung von Kontext bezogenen Situationen stammt von Schiele, Oliver, Jebara und Pentland [3]. Das **DyPERS**, Dynamic Personal Enhanced Reality System, erlaubt es dem Benutzer ein reales Objekt mit einer multimedialen Sicht zu assoziieren. Dabei wird eine multimedia Sequenz abgespeichert. Es stellt ein erweitertes multimediales Gedächtnis dem Menschen zur Verfügung. Die Sequenzen können sobald das Objekt erkannt wurde, abgespielt werden und den Benutzer an diese Information erinnern. In einem Beispiel einer Museumstour werden Erläuterungen des Ausstellers einem Bild zugewiesen und abgespeichert. Diese Daten können nun zu weiteren Auswertungen herangezogen werden, um z.B. Bilder der gleichen Epoche anzuzeigen.

Beim **Baseline System** [4] handelt es sich um ein „supervised“ System, es wird eine Eingabe des Benutzers erwartet. Bei jedem Kontextwechsel spezifiziert der Benutzer den Kontext, z.B. in welchem Raum er sich befindet. Das System stellt nun den Zusammenhang

zwischen den Sensordaten und dem Kontext her, dem Raum. Die Sensoren dieses Systems beschränken sich auf eine CCD Kamera und ein Mikrophon. Zur Eingabe steht dem Benutzer ein Touch-Pad zur Verfügung.

Neben den reinen **Audio/Video Signalen** treten in unserem Umfeld in der Regel vermischte Signale auf. Um diese Signale zu clustern, stellt Clarkson und Pentland [5] ein „unsupervised" System vor. Die Sensoren beinhalten eine CCD-Kamera und ein Mikrophon. Die Signale werden auf Ähnlichkeiten geprüft und eingeordnet. Ähnliche Signale erzeugen die gleichen Rückmeldungen, z.B. können die Signale einem Kaufhaus zugeordnet werden, so erscheint eine Einkaufsliste.

Die Erkennung der **Zeichensprache** stellt hohe Anforderungen an die Algorithmen zur Auswertung. In [6] werden zwei Methoden zur Erkennung erläutert. Den Systemen steht die Anforderung zugrunde, eine Zeichensprache mit einem Lexikon von 40 Wörtern zu erkennen und in eine gesprochene, Computer generierte, Ausgabe zu formulieren. Das Bild wird auf Handfläche, Orientierung und Bahn gescannt. Diese Eigenschaften werden einem Hidden Markov Model (HMM) als Input zur Auswertung übergeben. Das HMM besteht aus vier Zustandsübergängen, das die drei Eigenschaften der Handbewegung berücksichtigt. HMM's [12] werden unter anderem in der Spracherkennung und Handschriftenerkennung eingesetzt. Beim ersten System ist die Kamera auf dem Kopf montiert, dabei reicht ein Bild von 24x16 Pixel aus, um die Eigenschaften einer Handbewegung zu analysieren.
Das zweite System funktioniert mit einer Kamera, die den Benutzer aus einer Distanz aufnimmt. Sie benötigt eine Auflösung von 320x243 Pixel.
In Zukunft werden in der Zeichensprachenerkennung noch weitere Eigenschaften einfliessen. Körper-, Kopfbewegungen und Gesichtsgesten spielen eine wichtige Rolle in der Zeichensprache. Das System soll durch vier Kameras, Stereo Front, Seiten und wearable Sicht erweitert werden.

## 2. Affective Computing

„... computing that relates to, arises from, or deliberately influences emotions." (R. W. Picard)

Affective Computing beinhaltet eine Mensch zu Maschine Beziehung. Die Maschine soll dabei Emotionen, Gefühle, Empfindungen des Menschen erkennen können. Affective Computing geht auf den Menschen selber ein, das in der Kontexterkennung nicht berücksichtigt wird. Emotionen, die unser Leben stark beeinflussen, werden versucht, zu ergründen und zu erkennen.
In einer Mensch zu Mensch Beziehung spricht man auch von dem emotionalen Quotient. Emotionen interpretieren und die Fähigkeit haben darauf eingehen zu können, wird als wichtiger für Erfolg im Leben gewertet als ein grosses Fachwissen zu besitzen, das sich in einem hohen IQ widerspiegelt.
Emotionen sind körperliche wie auch mentale Ereignisse. Dazu gehören wahrnehmbare Formen von Tonfall, Gesichtsausdruck, Körperhaltung, Verhalten, Haut-Farbe und vieles mehr. Die klassischen Emotionen werden aufgeteilt in Angst, Zorn, Freude, Leid, Interesse und Verwirrung. Picard [8] beschreibt bis zu 50 verschiedene Beispiele, wo Affective Computing Einfluss nehmen könnte. Hier möchte ich nur auf wenige eingehen, die mich besonders angesprochen haben.

In **Computer unterstütztem Lernen** wird anhand des Wissensstandes des Schülers das Lernprogramm angepasst. Ein guter Lehrer erkennt wichtige Eigenschaften des Schülers und geht unterschiedlich darauf ein. Er lässt zum Beispiel Tipps und Hinweise vom Schüler entdecken, um das Selbst-Lernen anzutreiben. Der Lehrer kann auch auf Frustrationen eingehen und Hinweise geben. Die elektronischen Lehrer versuchen nun auf diese

verschiedenen Aspekte des Lernens einzugehen, um sich dem Schüler anzupassen und das Lernverhalten zu fördern.
In **Kunst und Unterhaltung** versucht man, die Aufmerksamkeit des Publikums zu erregen. Volle Aufmerksamkeit erkennt man im Gesicht und der Körperhaltung. Es soll eine dynamische Form der Unterhalter mit dem Publikum geben, wobei die Schauspieler auf die Emotionen des Publikums eingehen können. Um die Reaktionen des Publikums zu erkennen, könnten Kameras, Sensoren an Armstützen und im Boden eingesetzt werden.
**Ausdrucksvolle Emails** worin die Emotionen enthalten sind, können Missverständnisse verhindern. Die Emails sind hauptsächlich auf einen Text beschränkt und führen häufig zu falschen Auffassungen oder werden in einem anderen Ton interpretiert. Es haben sich Symbole „emoticons“ wie ☺ oder ;-) eingebürgert. Diese Symbole sind aber limitiert und Emails führen immer wieder zu Missverständnisse, die wieder geklärt werden müssen.

### 2.1 Sensoren [7]

Durch **Blutdruckmessung** misst man indirekt das sympathische Nerven-System, von da aus wird die Grösse der Blutgefässe kontrolliert.
Die **Haut-Leitfähigkeit** reagiert sehr schnell auf Änderungen im sympathischen Nerven-System. Bei den Emotionen, Überraschungen oder auch bei Angst steigt innerhalb weniger Sekunden die Leitfähigkeit an, erkennbar durch Schweiss.
Unter anderem werden Sensoren zur **Atmungserkennung** und **Muskelaktivitätenerkennung** eingesetzt.

### 2.2 Anwendung

Quantifizieren von **Stress beim Autofahren** [11]
Ziel dieses Versuches ist es, die Emotionen und den Stress, der beim Autofahren auftreten kann, zu analysieren. Es wird grundsätzlich unterschieden, ob es um physischen oder mentalen Stress handelt. Zwei Systeme wurden untersucht. Bei beiden hat man die Haut-Leitfähigkeit, Atmung, Muskelaktivität und Herzschlag gemessen. Das erste System besteht aus einem wearable Computer. Beim Zweiten werden Kameras auf den Fahrer und den Verkehr gerichtet.

## 3. Zusammenfassung

Es wurden auf zwei Aspekte der Mensch zu Maschine Kommunikation eingegangen. Die Kontexterkennung und Affective Computing sind Teilgebiete dieser Kommunikation. Kontexterkennung stellt die Basisinformation über Raum und Zeit des Benutzers her. In Affective Computing geht man auf den Menschen als Individuum mit Gefühlen und Emotionen ein. Durch Erkennen dieser beiden Aspekte soll eine Kommunikation zwischen Mensch und Maschine vereinfacht werden.

## Referenzen


[1] T. Starner, B. Schiele, A. Pentland. *Visual Contextual Awareness in Wearable Computing*. MIT Media Laboratory

[2] B. Clarkson. *The Familiar*. http://vismod.www.media.mit.edu/people/clarkson

[3] B. Schiele, N. Oliver, T. Jebara, A. Pentland. *An Interactive Computer Vision System, DyPERS: Dynamic Personal Enhanced Reality System*. MIT Media Laboratory

[4] B. Clarkson, K. Mase, A. Pentland. *Recognizing User's Context from Wearable Sensors: Baseline System*. MIT Media Laboratory

[5] B. Clarkson, A. Pentland. *Unsupervised Clustering of Ambulatory Audio and Video*. MIT Media Laboratory Perceptual Computing

[6] T. Starner, J. Weaver, A. Pentland. *Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video*. MIT Media Laboratory Perceptual Computing Section Technical Report No. 466

[7] *Affective Computing*. http://www.media.mit.edu/affect/

[8] R. W. Picard. *Affective Computing*. Perceptual Computing, MIT Media Laboratory

[9] R. W. Picard. *Toward Agents that Recognize Emotion*. MIT Media Laboratory

[10] E. Vyzas, R. W. Picard. *Offline and Online Recognition of Emotion Expression from Physiological Data*. MIT Media Laboratory

[11] J. Healey, J. Seger, R. Picard. *Quantifying Driver Stress: Developing a System for Collection and Processing Bio-Metric Signals in Natural Situations*. MIT Media Laboratory Perceptual Computing Section Technical Report No. 483

[12] Lawrence R. Rabiner. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, IEEE