# SCHLÜSSELINNOVATIONEN , PILOTPRODUKTE UND THESEN FÜR SPRACHVERARBEITUNG BEI WEARABLE COMPUTING

*Astrid Ullsperger*

*Juniorprofessur für Tragbare Elektronik und Rechentechnik, BTU Cottbus*

*astrid.ullsperger@tu-cottbus.de*

**Abstract:** Tragbare, d.h. wirklich anziehbare Elektronik und Computertechnologie wird deren Nutzungskonzepte sowie die Mensch-Computer-Schnittstellen (HCI) revolutionieren. Eine Schlüsselinnovation für dieses zweifellos ambitionierte Ziel „Wearable Computing“ stellt nach meiner Überzeugung das Fibercomputing dar, die Integration von High-tech in Fasern bzw. Garne. Unter Mitwirkung der Verfasserin wurde hierzu ein EU-Forschungsprojekt durchgeführt, das die grundsätzliche Machbarkeit bestätigt hat.
Für die künftige kabellose und handfreie Interaktion mit dem Computer stellt neben neuen Inputdevices, die natürliche Spracheingabe die nutzerfreundlichste Eingabeform dar. Entwicklungsleistungen auf DSP (digital signal processing) Plattformen werden völlig neue kundenspezifische Anwendungen ermöglichen. Dazu gehören als einfache Anwendungsfälle erstmals im Rahmen der Konferenz vorgestellte Pilotprodukte für Speech Memory Systeme in Bekleidung. Künftig werden darüber hinaus Verknüpfungen von Sprache, Gestik, Mimik, Gesichts- und Kopfhaltung an Bedeutung gewinnen. Hieraus ergeben sich vielfältige Anforderungen an die Sprachverarbeitung im Kontextbezug.

## 1 Grundsätzliche Anforderungen an Wearable Computer mit Sprachverarbeitung

### 1.1 Bedeutung der Sprachverarbeitung für Wearable Computing

Besonders im Zusammenhang mit internationalen Projekten zu „unsichtbaren“ Computertechnologien und ubiquitären Anwendungen, die im Alltag Informationen und Wissen austauschen, steigt die Zahl der wissenschaftlichen Referenzen und Mediennachrichten zu diesem Thema. Hosen rufen „Ich passe dir!“ (Nottingham-Projekt) [1], Jackets mit intelligentem touristischem Navigationssystem (Carnegie Mellon), beantworten Fragen unter Auswertung von GPS-Daten, Sensorkombinationen und entsprechender Software und machen ein elektronisches Touristen-Tagebuch möglich [2]. Künftig werden Mikrophone im Regenmantel plus Ohrhörer Sprachein- und –ausgabe z.B. Nachfragen via Internet in Datenbanken ermöglichen. Verbunden mit Videokamera und HMD (Head Mounted Display) als multimodale Interfaces wird auch die Einbeziehung von Gesten ermöglicht. Die Software unterstützt die natürliche Sprachverarbeitung, Spracherkennung, Maschinenübersetzung, Handschriftenerkennung und multimodale Mensch-Computer Interaktionen.

### 1.2 In einer Welt zunehmender Informationsfluten brauchen Menschen Unterstützung zur Informationsverarbeitung aber auch zur Informationsbereitstellung und zum - transfer zur richtigen Zeit am richtigen Ort und an den richtigen Empfänger. Sprache stellt dabei eine der wichtigsten und natürlichsten Informationsquellen für den Menschen dar. Für tragbare Geräte

**erfordert die Integration nur minimalen Platzbedarf.Anforderungen an Spracherkennungssysteme im Rahmen von Wearable Computing**

Spracherkennung ist die aufwendigste Eingabemethode bezüglich der Rechenleistung, was den Einsatz leistungsfähiger Systeme erfordert. Gegenüber störenden Umwelteinflüssen ist sie sehr empfindlich. Hauptproblem stellen hierbei vor allem die Umgebungsgeräusche dar. Durch die vorrangige Nutzung im Freien, auf öffentlichen Plätzen, in Autos, Zügen, Industrieanlagen wird die Erkennungsrate erheblich abgesenkt. Bei gezieltem Einsatz können Spracherkennungssysteme für eine entsprechende Umgebung mittels Adaption trainiert werden oder Spezialfilter genutzt werden. Dies wäre allerdings nur für den Einsatz von Wearables im beruflichen Umfeld von Bedeutung, da perspektivisch ein strukturübergreifender Einsatz erst den Vorteil des individuellen Systems ausmacht. Erste Tests, durchgeführt am KSI, mit Xybernaut MA V und ViaVoice [3] haben die gegenwärtigen Grenzen gezeigt. Eine lange individuelle Trainingsphase ist notwendig, um die Technik sprachlich zu steuern. Spracheigenheiten, wie undeutliche Wortabgrenzungen, Dialekt, Denkpausen, Füllwörter, falsche Grammatik, müssen Mustern zuordenbar sein. Persönlich individuelle Fachtermini sowie Anwendungsbereiche (task and language constraints) erweitern Suchraum und Zahl von Zweideutigkeiten und benötigen somit eine längere Lernphase, um die Erkennungsgenauigkeit und –geschwindigkeit zu vergrößern. Dies unterstützt die Theorie der individuellen Sprachinterfacegestaltung für Wearables, auf die nachfolgend noch näher eingegangen werden soll. Gegenwärtig stehen einfache und Einzelworterkennung (isolated word recognition) sowie beschränktes Vokabular im Vordergrund. Obwohl die Erkennung von kontinuierlich gesprochenem Text (Spontaneous Continuous Speech) für Computer teilweise möglich ist, ist es für mobile Geräte bisher nur begrenzt kommerziell nutzbar. Hierfür sind ein wesentlich größeres Vokabular und die Fähigkeit zur Bewältigung der Komplexität und Mehrdeutigkeit der menschlichen Sprache mittels der Natural Understanding – Komponente notwendig.

Von Bedeutung für die Spracherkennung ist auch der Einsatz hoch qualitativer Mikrophone für die entsprechende Qualität des Sprachkanals. Wenn zusätzlich zum Sprachkanal noch eine digitale Leitung, z.B. für GSM oder UMTS genutzt wird, beeinflusst die Digitalisierung des Mikrofonsignals auch die Erkennungsrate. Gerade für Wearbles gehört die Kombination mit einem Navigationssystem aber zu einer Standardapplikation. Neben dem Digitalisierungsrauschen kann es bei analogen Leitungen auch zu Leitungsrauschen kommen, und damit zu systematischen Fehlern. Ein großes Problem stellt in allen Sprachen die Zweideutigkeit des Wörterbuchs dar, die erst mit der verbesserten Einbeziehung von Kontextinformationen gelöst werden kann. Hieraus ergibt sich wiederum eine ideale Verbindung zum tragbaren individuell auf die speziellen Bedürfnisse des Nutzers ausgerichteten Computer.

Gleichzeitig werden auch hohe Anforderungen an Sprachausgabemedien gestellt, die möglichst wenig Einfluss auf die Umwelt haben sollten und im Optimalfall nur vom Nutzer zu hören sind.

## 1.3 Heutige Verbreitung von Systemen der Spracherkennung

Im Moment ist die Spracherkennung in Mobiltelefonen am weitesten verbreitet. Eine Adressbuchauswahl ist einfach durch Aussprache des Namens möglich.

Zwei wichtige Sprachstandards ermöglichen  die Kommunikation, VoiceXML und Java Speech API.  Da eine Trennung von Spracherkennungs- und Sprachausgabemodule von der

Anwendung möglich ist und Module einfach austauschbar sind, kann Weiterentwicklungen schnell Rechnung getragen werden und Spracherkennungsmodule unterschiedlicher Hersteller können zusammenarbeiten. Das ist besonders für die Zunahme von disappearing artefacts ein großer Vorteil. Implementierungsbeispiele von Java Speech API sind IBMs Speech for Java und TTS von Lernout & Hauspie.

Wissenschaftliche Untersuchungen zu Sprachtechnologien sind z.B. in Deutschland vom DFKI-Evaluationzentrum für Sprachtechnologie durchgeführt worden, auch zur Nutzung von Xybernaut-Systemen [4]. Das KSI hat hierzu auf der Anwenderebene einige Tests durchgeführt.

Das Deutsche Forschungszentrum für Künstliche Intelligenz beschäftigt sich u.a. mit unterschiedlichen Aspekten emotionalen Sprachverhaltens und der Sprachsynthese [5]. Eine Übersicht über aktuelle Forschungsprojekte, vorrangig mit Ergebnissen im Bereich multimodaler oder sprachlich emotionaler empirischen Untersuchungen ist im Humain-portal [6] zu finden.

Multimediale Signale setzen sich aus Sprache, Handschrift, Gestik, Zeigen, Buchstabieren, Pupillenbewegung, Mimik und Kopfhaltung etc. zusammen.

Die Effektivität multimedialer Mensch-Computer Interaktion wurde von einer Reihe Wissenschaftler über die letzten Jahre untersucht. [7;8;9]. “Over the last few years, we have focused on developing sensible and useful user interfaces to support multimodal human-computer interaction and accelerated evolution of those”.

Der Wert der Sprachinteraktion ist offensichtlich: Die natürlichste Eigenschaft des Menschen, kann in Bewegung und unterwegs genutzt werden und lässt die Hände frei. Wearables werden vorrangig bei Aktivitäten genutzt, wie umherreisen, rastlose Bewegungen etc.

Viele Firmen arbeiten an Pervasive-Computing-Geräten mit integrierter Spracherkennung. Beim Siemens-Wristphone wurde statt einer sperrigen Tastatur Spracherkennung als Benutzerschnittstelle verwendet. [10]

Für PDA’s werden bereits Sprachanwendungen angeboten (IBM, Lernout & Hauspie). Die Kombination mit einem intelligenten Agenten, der dann im Internet die entsprechenden Fakten zusammenträgt wird die Akzeptanz sowie die Benutzbarkeit wesentlich erhöhen.

Abb. 1 Armbanduhr-Handy (Tri-Band), SiemensTests für einen PDA der Tracking-Input mit einer multimodalen Nutzerschnittstelle verbindet, existieren bereits. [11]

Innerhalb des VRML Standards existiert keine intrinsische Unterstützung für Sprachtechnologie. Die ScanSoft RealSpeak TTS engine und die Siemens ICM Spracherkennungsengine wurden bei unterschiedlichen Test genutzt. Physische Objekte, die für den sprachlichen Austausch bestimmt werden, nutzen Sensoren im Nahbereich. Der Sprach-Interaktions-Manager setzt jeden Sensor in Beziehung zu einer XML-Quelle für die Spezifikation eines Sprachdialogs. Diese zusätzliche XML-Information spezifiziert die Texterkennungsgrammatik und die korrespondierenden gesprochenen Antworten. Ein Beispiel wäre, dass sich ein Ingenieur einem Tank nähert und fragt: „Gegenwärtiger Status?“, worauf der Container antworten würde: „65% mit Wasser gefüllt mit einer Temperatur von 72°C“.

Um Nutzer an Objekte im Sprachlabor von Philips heranzufüren, die aus Unkenntnis des erlaubten Vokabulars die Nutzung vermeiden wollten, erschienen über ein HMD Sprechblasen über einem 3D Objekt um eine Kommunikation zu beginnen. Wenn die Sprachinteraktionsmanagementkomponente eine Info erhält, dass der Sensor aktiviert wurde, wird über einen VRMLtext-Punkt entsprechender Sprachinhalt generiert. Wenn der Sensor wieder deaktiviert wird, verschwindet auch die Sprechblase wieder. Schwierigkeiten bestehen vor allem darin, der Richtung des Nutzers zu folgen oder ggf. die Mimik / Gestik einzubeziehen. Probleme gibt es auch bei der Überlappung von 3D-Objekten im Raum oder mehreren Sensoren. Hier muss eine mögliche Sprachkollision erkannt und eine Nachfrage an den Nutzer gestartet werden.In der mobilen Zusammenarbeit können über ein entsprechendes Rahmenmodell folgende 3 Optionen realisiert werden: ein gemeinsames 3D-Browsen, Unterstützung mit Anmerkungen und Kommentare, Vollduplex Sprachkommunikation über Voice-over-IP.

## 1.4 Grundlagen der Speech to speech Kommunikation beim Wearable Computing

Auf der Welt gibt es etwa 8000 Sprache und noch mehr Dialekte. Dabei ist die Bedeutung unterschiedlich. Mehr als die Hälfte der Weltbevölkerung spricht eine der fünf Sprachen: Chinesisch, Englisch, Spanisch, Russisch und Hindi, im Gegensatz dazu wird die Sprache Elmolo nur noch von 6 Menschen gesprochen und die Sprache Aore nur von einem einzigen Bewohner, womit keine Kommunikation mehr möglich ist. [12].

Die kortikale Informationsverarbeitung beim Menschen erfolgt nicht nacheinander von einfachen zu komplexen Analysen sondern erfolgt auf den verschiedenen Ebenen der Analyse nahezu gleichzeitig. Es erfolgt eine wechselseitige Signalweitergabe über mehrere Stufen, jeweils von einem kortikalen Areal bewältigt. Dabei entstehen Landkarten, der verarbeiteten Information, wie die vom Innenohr kommende, zur Frequenz- oder tonographischen Karte wird.

Menschen werden mit der Fähigkeit zur Unterscheidung der 70 Phoneme geboren, aus denen die lautlichen Wörter aller Sprachen der Welt gebildet werden.

Dabei ist die Speicherung der Wörter gemäß semantischer Relationen der Ähnlichkeiten oder des Kontrasts verbunden.Eine Hauptfragestellung ist das „problem of constructing „speaker-independent“ speech recognition systems The goal of speaker-independent systems is to recognize what was said regardless to who said it.” [13] Auch wenn Menschen dieselbe Sprache nutzen, ist das Ziel kompliziert, da zwei Personen denselben Satz mit unterschiedlichen Soundsignalen produzieren – unterschiedlichem Akzent, unterschiedlichem Takt und weiteren Unterschieden der Rede. Der Computer hat Probleme den Teil des Sprachsignals vom Signalteil, welches den Redner identifiziert und dessen/deren Ausdruck zu differenzieren. Deshalb ist ein Training auf den individuellen Sprecher in den meisten der heutigen Anwendungsbereiche noch notwendig. Wearable helfen aber dieses universale Spracherkennungsproblem zu lösen. Nicholas Negroponte beschrieb schon vor Jahren eine alternative Lösung in Form eines sprecherabhängigen Verfahrens. Sofern der eigene Computer seinen Träger versteht, kann er das gesprochene Wort dem Rest der Welt weitergeben oder übersetzen. Eine individuelle Sprachkomponente kann so zum universellen Interface für viele unterschiedliche Schichten und Netzwerke werden. Deshalb werden Experimente benötigt, die emotionale gleich bleibende Zustandsmuster verlässlich für das Individuum in gegebenen Kontexten erfassen. In vielen Anwendungen ist es vor allem notwendig, das der individuelle Computer in der Lage ist, den Zustand zu erfassen, um dann gegebenenfalls diese Information bei Bedarf an andere zu kommunizieren. Wearables und

Sprachverarbeitung werden umso effektiver, je mehr Kontext wahrgenommen werden kann, wie z.B. Fühlen, dass die Person Treppen steigt, wenn die Raumtemperatur sich verändert, erschreckende Kriegsmeldungen gelesen werden und andere Situationen. Wahrnehmungscomputer können klarer als andere Systeme autonomen Antworten Ereignissen zuordnen. Für die besten Ergebnisse ist nicht nur die Einbeziehung der allgemeinen Umgebung von Interesse, wie z.B. der Wetterindex sondern auch das familiäre Umfeld, Todesfälle, Erdbeben in Wohnortnähe etc. Die Zustandserkennung kann dabei individuell für ähnliche Umfeldbedingungen sehr unterschiedlich ausfallen, und kann über eine sensorische Kombination aus Blutdruck, Herzfrequenz, Hauttemperatur und –widerstand und Atmung charakterisiert werden. Diese Daten verändern sich für eine Person unter anderen Umgebungseinflüssen. Diese müssen aber zuverlässig erfasst werden. Hierzu werden lernende Systeme benötigt, die eine Klassifizierung für die Kombinationen Mensch-Umfeld erfassen. Sicherlich wird es auch Übereinstimmungen bei Individuen geben, ähnlich, wie sich auch Stimmen ähneln. Individuen werden entsprechend gleicher Eigenschaften in Cluster kategorisiert. Ein „Universalspracherkenner“ würde als erstes analysieren: Wer klingt so in Bezug auf Akzent, Sprachqualität und Rhythmus? Zu jeder Subkategorie existieren Sprachprototypen für eine bestimmte Kategorie. Dadurch können Zustandsausdrücken aufgedeckt werden, die bisher theoretisch noch nicht vorgedacht wurden, worin ein großer Vorteil dieser Methode besteht. Schwierigkeiten werden darin bestehen, dass oftmals Menschen nicht bereit sind, Schlüssel ihrer Emotionen an Computer oder für sie fremde Personen weiterzureichen. Wann ist etwas sarkastisch, ironisch, witzig oder zynisch gemeint? So wie wir Menschen mit der Interpretation Probleme haben, wird es Computern auch nur annähernd gelingen, sprachliche Äußerungen 100%ig zu deuten. Bei Wearable-Systeme besteht allerdings eine große Chance seinen Träger über die Zeit sehr gut kennenzulernen.

## 1.5 Anforderungen an text to speech Systeme

Gegenwärtige „Text zu Sprache-Systeme“ intonieren in der gleichen Weise. Gefühle können so kaum kommuniziert werden: Verärgert zu unterbrechen, Ängste zu äußern, warme Zustimmung oder andere Ausdrucksformen emotionaler Intelligenz widerzuspiegeln.

Sowohl Forscher auf dem Gebiet der Spracherkennung als auch der Sprachsynthese stimmen überein, dass die Einbindung des Sprachrhythmus wesentlich zur Verbesserung der Qualität und damit zur Akzeptanz von Sprachsystemen beitragen könnte [14]. Gegenwärtige kommerzielle Produkte nutzen die Einbindung von Sprachrhythmus kaum. Kommerzielle ASR (Automatic Speech Recognition) sehen erst für hochwertige Anwendungen einen Sinn darin. Die kommerzielle Entwicklung von „Text-to-speech“ (TTS)-Systemen hat durch die dramatische Verbesserung der Qualität der Sprachsynthese eine beschleunigte Entwicklung erfahren. Für die zukünftige Entwicklung ist die Standardisierung solcher Systeme wichtig. Die folgenden Anforderungen sind zwar nicht neu, sollten hier aber noch einmal auch für den Einsatz in Wearables genannt werden:

Anforderungen an ASR output Systeme

- Natürlichkeit der Sprache
- Eindeutigkeit und Klarheit der Sprache
- Ausfiltern der Sachinhalte

Anforderungen an TTS input Systeme

- Schnelligkeit

- Hoher Wert und Training von Laien für Kommentare möglich (sofern nicht, hohe Kosten durch Kommentierung von Sprachrhythmus-corpi durch Experten)
- Hoher Bedarf an realen Daten – für die Analyse von Phänomenen von Unterbrechungen, Stockungen im Redefluss, Durchmischung mit Umfeldgeräuschen

So wurde z.B. ToBe [14] als Standard-System konzipiert und hat unzählige neue Variationen hervorgerufen.

Interessante Anwendungen im Sinne ubiquitous computing beinhalten z.B. Talking goods (Permission marketing) (ident technology AG) [15] hervor, Produkte die zu „kommunizieren“ beginnen, wenn sie berührt werden Die Informationen über das berührte Produkt werden auf das UMTS- Mobiltelefon geleitet und starten dort eine interaktive Kommunikation.

Ein anderes Beispiel für die Integration von Sprachmodulen könnte ein Wearable Reality Framework“ darstellen.

Ein Ingenieur, der für die Wartung von Flugzeugen zuständig ist und mit dem Cockpit spricht, würde heute wahrscheinlich als verrückt bezeichnet werden. Aber zukünftig könnte eine solche Kommunikation zwischen Techniker und Geräten sowie Wartungsobjekten zur Normalität werden. Ein PocketPC der mit einer hybriden Tracking-Funktion ausgestattet ist, bietet dem Nutzer ein nahtloses, ortsabhängiges, mobiles multi-modales Interface. Die Nutzerschnittstelle legt ein 3dimensionales graphisches Bild mit einem kontextabhängigen Sprachdialog über ein reales Objekt in unmittelbarer Nähe des Nutzers. Zusätzlich wird für die Zusammenarbeit ein shared VRML browsing mit Annotation und einem vollen Duplex Sprachkanal angeboten. Ein einfaches TCP Protocol ermöglicht das gemeinsame 3D-Browsing.

Bereits angedachte Beispiele für anziehbare Sprachanwendungen stellen Memoclips [17] dar, die die Nutzer mit ortsabhängigen und lokal positionierten Nachrichten versorgen, die sowohl auditiv, visuell oder als Kombination aus beiden angeboten werden können. Wenn sich ein Nutzer einem Memoclip in Form eines Artefacts nähert, der über einen Sensor mit der Erinnerungsfunktion verbunden ist, gibt der Memoclip die entsprechende Nachricht wieder.

Ein häufig untersuchtes Forschungsthema stellt dasCyberguide-System [18] dar, welches z.B. eine Führung über einen Universitätscampus mittels Handhelds oder WPC unterstützt. Durch die Bestimmung des Ortes können reale Touren simuliert werden.

# 2 Fibercomputing als Basisinnovation für Wearables

## 2.1 Zielstellung von Fibercomputing (Ficom)

Fibercomputing ist ein Projekt, das im Rahmen des Disappearing Computing Programmes des 5. EU Rahmenprogramms begonnen wurde. Nach der Insolvenz des Initiators *Starlab Brüssel* wurde es von den anderen Partnern *National Microelectronic Research Center (NMRC* in Cork), *Eidgenössische Materialprüfungs- und Forschungsanstalt (EMPA*, Zürich), Integriertes System Design (ISD, Athen) und dem neu hinzugekommenen *Klaus Steilmann Institut für Innovation und Umwelt GmbH (KSI*, Bochum) weiter geführt.

Ziel ist es, große flexible Computersysteme für wearable Anwendungen in Funktionsfasern durch Integration von Silikontransistoren zu entwickeln. Die High-tech - Faserproduktion

realisiert buchstäblich den „unsichtbaren“ Computer und ermöglicht die Integration von Informationstechnologie in tägliche Gegenstände und vorrangig in Hochtechnologie-Gewebeprodukte. Im Rahmen des Ficom Projektes wurden die Chancen und Grenzen solcher „embedded technologies“ erstmals untersucht und praktisch getestet. Sofern weitere Forschungen an die Erfolge dieses ambitionierten Projektes anknüpfen, könnten damit erstmals Alltagsprodukte mit wirklicher Intelligenz ausgestattet werden. Im Folgenden soll zum besseren Verständnis ein erster Überblick zu dieser Zukunftstechnologie gegeben werden.

## 2.2 Überblick über die F&E Stufen von Fibercomputing

Traditionell beginnt auch die Entwicklung von funktionellen Fasern beim Prozess der Faserentwicklung. Bei EMPA in Zürich wurden über einen Extrusions- Prozess $SiO_2$ Fasern mit einer aktiven Beschichtung gefertigt. Die entstandenen Fasern haben eine Zugfestigkeit von 70 GPA, jedoch für den Webprozess noch eine etwas zu geringe Flexibilität. Der Faserdurchmesser beträgt 125 µm. Im Laufe des Projektes wurden noch andere Fasern, später mit größerem Durchmesser, eingesetzt.

Am NMRC in Cork wurde unter Leitung von Alan Mathewson eine neue Technologie entwickelt und patentiert, um flexible funktionelle Fasern zu produzieren.

Die vom Wafer gelösten Fasern haben eine Schlangenform, wie die schematische Darstellung zeigt:

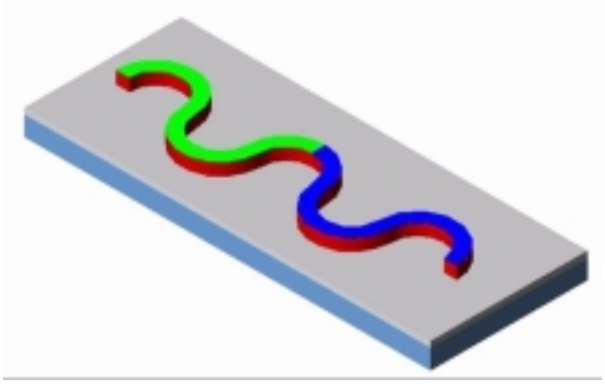

Abb.2: Schematische Darstellung der Dioden - Fertigung auf einer Siliziumfaser, Quelle: NMRC, Beitrag zum Projekt Fibercomputing 2002/2003

Für jede Anwendung ist, ähnlich klassischer Chipherstellung, ein Schaltkreis Design notwendig.

Im Rahmen des Projektes Fibercomputing wurde als erste Demonstrations- und Experimental – Anwendung ein einfacher Ring – Oszillator ausgewählt und das Design hierfür erstellt, wobei der Ring – Oszillator aus einer ungeraden Anzahl von Invertern besteht. Die größte Anforserung besteht hierbei in dem Entwurf eines Designs, dass auf die Größe der Fasern passt, im Falle des Projektes durfte das Design 350µm nicht überschreiten.

Die Produktion der Bausteine geschieht auf single crystal silicon wafers. Der kritischste Schritt ist nicht die Fertigung des Bausteins an sich, sondern nach der Fertigung das Ablösen

des Bausteins von dem Wafer, um eine freistehende, aktive Faser zu erhalten die dann wiederum auf eine Endlos – Unterstützungsfaser aufgebracht wird.

Für diesen Schritt werden auch nach Abschluss des Projektes noch neue, zuverlässig handhabbare Lösungen gesucht.

Erst nachdem die aktiven Fasern gefertigt und auf die Unterstützungsfasern aufgebracht sind, können diese in Form einer textilen Fläche verarbeitet bzw. in eine textile Fläche eingebettet werden.

Für die Zukunft ist es Ziel, die Funktionsfasern in einer Qualität zu fertigen, die eine Fertigung im industriellen Maßstab ermöglichen, im Projekt erfolgte dies noch per Hand.

### 2.3 Sprachein- und Ausgabe im Rahmen von Fibercomputing

Bevor die textile Technik in Produkte integriert werden können, sollten die Schnittstellenfragen geklärt sein.

Auch hier spielen die Sprachein- und –Ausgabeschnittstellen eine wichtige Rolle, an die gleichzeitig besondere Anforderungen gestellt werden.

Während bei der Einbettung diskreter Bauelemente in Bekleidung die Verbindung mit kleinen Lautsprechern und Mikrophonen kein großes Problem darstellt, sind diese Lösungen für die Verbindung mit Fibercomputing wenig geeignet, zielen diese doch darauf hin, möglichst wenig Hardware am Körper zu tragen.

Für die Sprachausgabe könnten hauchdünne, flexible Biegewellenwandler, wie sie auch schon in tragbaren Lautsprechersystemen im Einsatz sind, zum Einsatz kommen. Die technologische Basis dafür müsste wiederum auf den Fasern realisiert werden. Um einen entsprechenden Resonanzkörper zu erhalten, können Lufttaschen in Kleidung eingebaut werden.

Eine ähnliche Technologie müsste für Mikrophone gefunden werden. Dabei sollten sich neue technische Entwicklungen das Prinzip des Fibercomputings zunutze machen. Durch die Verteilung über eine große Fläche kann eine punktuell kleine Leistungsfähigkeit durch die entsprechende Verbindung aufsummiert werden. Wenn man die laufenden Entwicklungen im Rahmen von Polymerelektronik und Piezofasern verfolgt, sollte es hierfür langfristig Lösungsmöglichkeiten geben.

## 3 Sprachein- und Ausgabeschnittstellen im Rahmen des 7$^{th}$ skin modells

### 3.1 Das 7 – Skin Modell

Durch ubiquitäre und tragbare Anwendungen entstehen vielschichtige Vernetzungen. Für die Untersuchung der dabei entstehenden Kommunikations- und Informationsprobleme wurde ein 7-Schichten-Modell [19] konzipiert. Grundlage bilden hierbei die unterschiedlichen Integrationsstufen von Computertechnologie, beginnend unter der Haut in Form von Implantaten, weitergehend auf die Haut, z.B. Haarfasern, über Bekleidung, hautnah und –fern bis zur infrastrukturellen Umfeldebene. Unter Berücksichtigung der Notwendigkeit zur

Standardisierung und Modalisierung, besteht die Fragestellung, zwischen welchen Ebenen Sprachmodulintegration sinnvoll und gegenwärtig technologisch machbar ist.

Die folgende Abbildung zeigt schematisch das 7 – Schichten Modell.

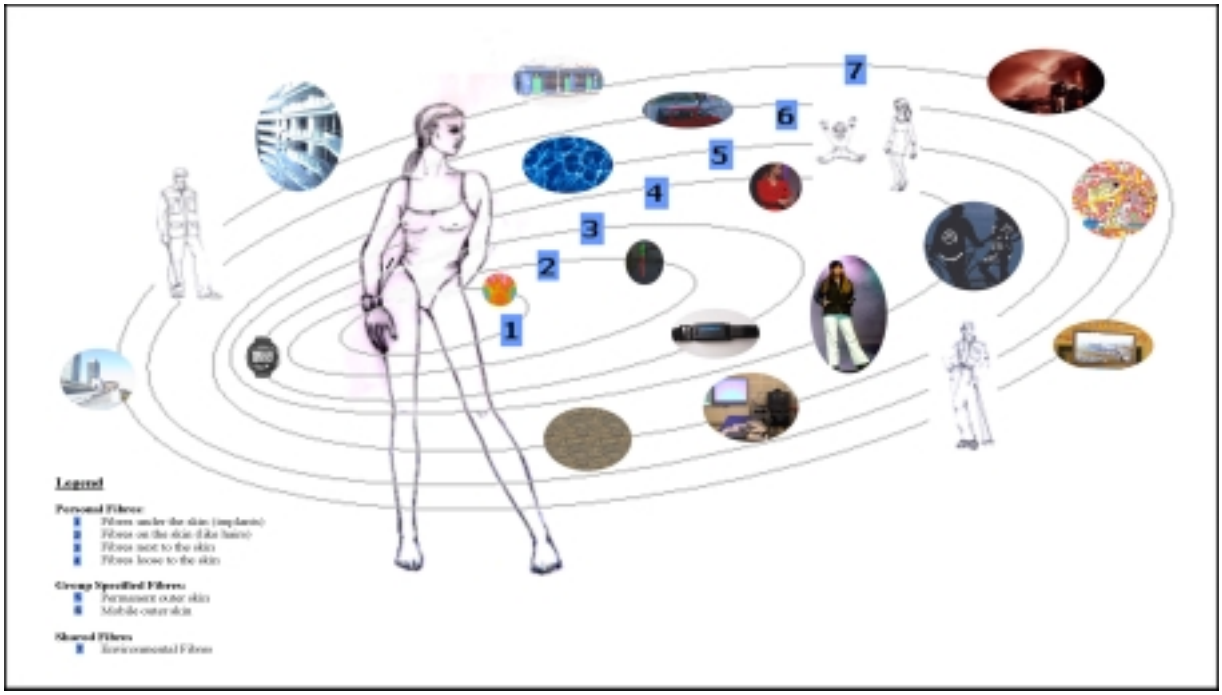


Abb.3: 7 Schichten Modell, Klaus Steilmann Institut

## 3.2 Sprachtechnologien für Metaebenen

Besonders für den Austausch zwischen den Metaebenen – personal specified fibers (PSF) – grouped specified fibers (GSF) und shared fibers (SHF) sind Sprachtechnologien zu implementieren.

Für Sprachanwendungen sind folgende Beispielszenarien denkbar:

**PSF** – Sensoren in Fasern, die die Körperfunktionen des Trägers überwachen. Sie geben z.B. auf die Frage: „In welchem Allgemeinzustand befinde ich mich gerade?“ die Auskunft: „Blutdruck 160 zu 90, etwas zu hoch, Hauttemperatur und –feuchtigkeit grenzwertig, Puls und Herzschlag sehr schnell. Vielleicht sollten Sie eine Pause machen und sich hinsetzen.“ Das Verhältnis: 1 konkrete Person kommuniziert mit unterschiedlichen individuellen Devices.

**GSF** – Kinder im Kindergarten befragen den intelligenten Kochtopf zum Essen: „Was gibt es heute oder morgen zu Mittag? Sind in der Suppe Mohrrüben? Ist Puddingessen gesund?“ Das Verhältnis: n Personen (Kinder) mit ($\Sigma$ $f_1$ …..$f_n$) der Summe charakteristischer Eigenschaften – hohe Stimmlage, einfache Vokabeln, themenspezifische Fragestellungen : 1 Artefact.

SHF: n Personen (ohne charakteristische Beschränkungen) zu 1 Artefact.

GSF und SHF Systeme stellen dabei besonders hohe Anforderungen an die Sprachsysteme.

# 4 Pilotprodukte mit Sprachverarbeitung in Bekleidung im Test

## 4.1 Wearable Reality

Gegenwärtig werden Spracherkennungstechnologien vorrangig in PCs für Diktate, das Ausführen von Kommandos, die Bereitstellung von audiobasierten Anfragen, bei Telefonen für den Zugriff auf oder Transaktionen mit Informationen wie Fahrplänen, Serviceangeboten, Salesfunktionen, Touch-tone-Systeme in Kinos, Hotels, Flughäfen, Sprachsynthese für das Vorlesen von Textdokumenten (e-Mails) angewandt. Erste Pilotprojekte beschäftigen sich mit der Anwendung in embedded devices.

Auf der persönlichen Ebene steht besonders das Training für die Verbesserung der Qualität der Spracherkennung im Vordergrund, wofür technologisch bereits sehr anspruchsvolle Systeme existieren. Bei der Gruppen- und Umfeldebene als technologisch hoch anspruchsvoll einzustufen.

Ein zukunftsweisendes Spracherkennungssystem wurde von der Carnegie Mellon University im Rahmen der ISWC 1998 als fiktives Einstein-Interview präsentiert. Das Auditorium konnte Fragen stellen, die von Einstein entsprechend seinem Know-How und seinen Erfahrungen beantwortet wurden. Fast ausnahmslos wurde auf unterschiedlichste Fragen bis hin zu seiner Haarfrisur auf unterschiedlichste Dialekte, Akzente und Sprachrhythmen korrekt von ihm geantwortet. Im Zusammenspiel mit sprachlichen Äußerungen waren Mimik und Gestik stimmig. Das System ist für zukünftige ubiquitäre Anwendungen im Entertainment- aber auch Edutainmentbereich denkbar und würde auch für shared und grouped interfaces anzuwenden sein.

Für wearable aber auch pervasive computing gibt es zum einen den Ansatz Mobiltelefon und auf der Gegenseite ein Voice Gateway zu errichten, dass die Sprachverarbeitung übernimmt, ein interessanter Ansatz ist allerdings die Integration von Sprachverarbeitung direkt im Gerät (embedded speech recognition).

Die Voice-Gateway-Architektur kann vorteilhaft Milliarden von bereits vorhandenen Mobil- und Festnetztelefonen nutzen, um Informationen aus dem Web oder einem Back-End-System abzufragen. Klassische Anwendungsbereiche sind Flug- und Eisenbahngesellschaften, die Transport- und Unterhaltungsindustrie mit transaktionsbasierten Systemen oder Informationssystemen. Positiv ist dabei auch der individuelle Zugriff auf Informationen, sofern noch kein WPC vorhanden ist und nur ein Mobiltelefon genutzt werden kann.

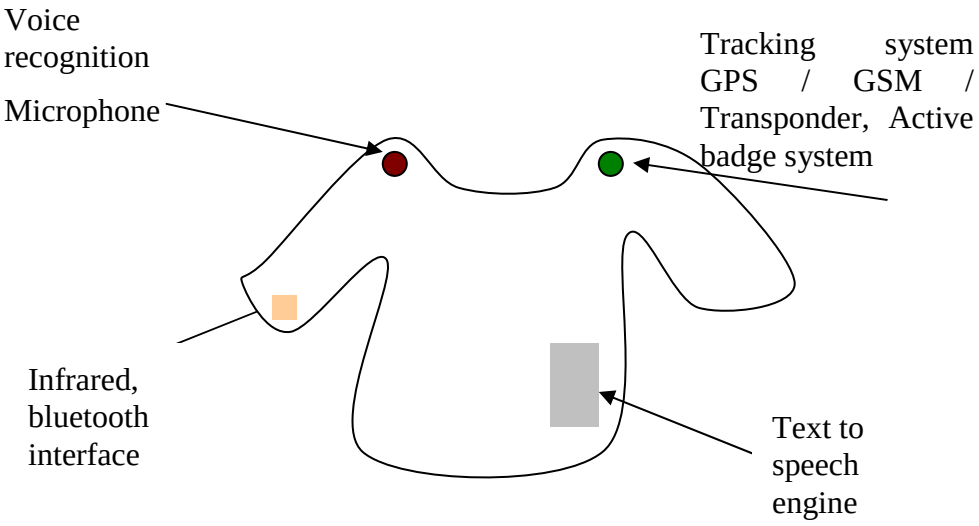

Abb. 4: A functional overview of Wearable Reality

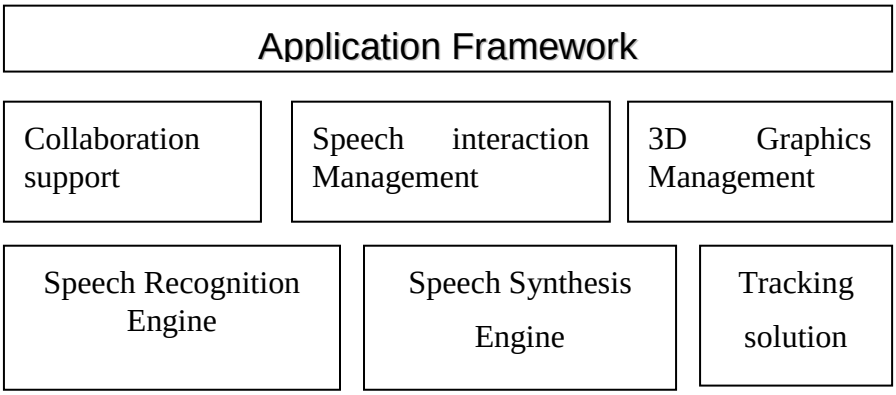


Abb.5: nach: Pervasive Computing. Technology and Architecture of Mobile internet applications

## 4.2 Intelligente Fashion Interfaces als zukünftige Sprachverarbeitungartefakte

Mode ist ein wichtiger Bestandteil unseres Lebens. Zunehmend werden Konzepte realisiert die Technologie mit Mode verbinden. Ein Beispiel ist hierfür ein Modell des KSI – ein Ohrstecker mit integriertem Chip zur Verstärkung von Audiosignalen.

Abb.6: IFI Audio, KSI

Zunehmend werden Komponenten für die natürliche Spracheingabe als embedded Technologien entwickelt. Die Systeme basieren auf einer Client-Server-Architektur.

Besonders die Anwendung der digitalen Sprachverarbeitung in der mobilen Kommunikation (GSM, TETRA, UMTS) werden als modische Sprachverarbeitungsartefakte hohe Marktpotentiale erschließen können.

Einen ersten Schritt stellt die patentierte Memory-Weste des KSI dar, die durch folgende Merkmale charakterisiert ist

- Baugruppen Mikrofon, Lautsprecher, Speicher, Batterie, sind aus Gründen der Flexibilität getrennt und in „IN-Line“ in der Vorderseite eines Kleidungsstückes untergebracht.
- Die Unterbringung erfolgt in Durchbrüchen des Obermaterials, die von innen verschlossen werden, nach außen sichtbar den Charakter „Elektronik“ unterstreichen, sowie die Bedienelemente hervorheben.
- Die Anbringung erfolgt nicht sichtbar im Futter des Kleidungsstücks und wird über Knöpfe, Klettbänder oder Reißverschlüsse befestigt..
- Die Bedienelemente sind im Obermaterial durch Aufnäher oder Label gekennzeichnet.
- Das Mikrofon wird vorzugsweise auf der Vorderseite des Kragens oder Revers angebracht.
- Neben der manuellen Bestätigung der START-Taste ist auch eine sprachgesteuerte (VCO) Aufzeichnung möglich.
- Der Lautsprecher ist bei der Benutzung des Ohrhörers abgeschaltet.
- Der Lautsprecher kann sowohl auf der Vorderseite des Kleidungsstückes, als auch auf der Schulter getragen werden (Schulterpolster).

Abb. 7 /Abb. 8: Memory-Spech (SMILE), IPEKDERI & KSI

Eine neue Kollektion wurde als Memory-Innovative-Smile-Lederkleidung entwickelt. [21]

Das System ist bewusst einfach gehalten und unterscheidet sich dadurch von den im Handel erhältlichen Voice Recordern. Es ist speziell dafür gedacht, schnell unterwegs kurze Notizen aufzusprechen.

Zukünftig soll dieses System in Richtung Spracherkennung weiter ausgebaut werden, so dass für die Aufnahme kein Knopf mehr betätigt werden muss, sondern die Aufnahme per Sprachkommando erfolgt.

# 5 Thesen zu künftigen Trends der Sprachverarbeitung bei Wearable Computing

## 5.1. Verbindung mit affective computing

Die Verbindung von Sprachtechnologien mit affective computing bietet, wie weiter oben bereits angedeutet, die größten Chancen für zuverlässige Sprachsysteme, die möglicherweise auch versteckte Botschaften der Sprache mit entschlüsseln können. Besonders die Einbindung von emotionaler Intelligenz ist dabei von zentraler Bedeutung. Nicht alle HCI-Anwendungen profitieren allerdings von emotionalen Erweiterungen. Wenn Fakten oder Daten ohne lokalen Kontext benötigt werden, steht eine zuverlässige, schnelle und exakte Bereitstellung im Vordergrund

## 5.2. Verbindung mit context awarness und context sensing

Für wearable computer – Anwendungen ist neben der Einbeziehung emotionaler Zustände für die sprachliche Kommunikation auch die Verbindung mit Umfeldbedingungen von Vorteil z.B. bei der Erteilung von Ratschlägen (Beruf, Weiterbildung, Familie). Generell wird durch die Zunahme von ubiquitous Anwendungen, wie bei künstlichen Verkaufsassistenten, bei künstlichen Tutoren, künstlichen Freunden als reale Artefakte affektive computing an Bedeutung zunehmen. Auch wenn die Reaktionsfähigkeit von Menschen durch Alter oder Krankheit abnimmt, sind die Integration von AC-und CA/CS-Anwendungen von Vorteil. Sie können auch helfen, verhaltenstechnische Lernprozesse zu unterstützen, wie z.B. klare und eindeutige Kommunikation.

## 5.3. Systemverbund zu Ambient Intelligence Anwendungen

Sprache ist nicht fest sondern passt sich den Erfahrungen an bzw. wird von uns den Erfahrungen angepasst. Sie wird von kreativen und kompetenten Sprechern kontinuierlich weiterentwickelt. Da Sprache eine Art kollektives Gedächtnis einer bestimmten Kultur darstellt, ein einzigartiges System von verbundenen Begriffen und Regeln ist, sind Sprachverarbeitungssysteme in lernende komplexe Beziehungen zu setzen und werden zum Schlüssel für die Akzeptanz von Wearable Computing-Anwendungen.

Als ein Teil von Ambient Intelligence Anwendungen werden sie wesentlich zur intuitiven Nutzung beitragen – so wie atmen, essen, sprechen oder laufen. Nutzerschnittstellen haben multimodal gestaltet zu sein (multi-user, multilingual, multi-channel, multi-purpose) für Sprache, Gestik und Mustererkennung. Sie sollten sich an Nutzerbedürfnisse anpassen und kontext-sensitive Schnittstellen mit Informationsfiltern, -präsentationen und Cross-Media-Inhalten anbieten. Die Mustererkennung (eingeschlossen Sprache und Gestik) ist der Schlüssel für künstliche Intelligenz, die sich rasch weiterentwickelt. Die Spracherkennung wird wesentlichen Einfluss auf die Miniaturisierung von Systemen und die Erweiterung von Objekten haben, die freihändige Nutzung von individuellen Geräten ermöglichen. Zukunftsszenarien gehen von einer synchronisierten Nutzung von Stimme, Gestik, automatischer Identifizierung und Ortung aus, so dass Dienstleistungen sofort verfügbar sind. Übereinstimmend mit einer Reihe von Experten werden nichts desto trotz vorerst Redeanweisungen (speech instructions) wichtiger als die Sprachsynthese (speech synthesis).

[26] für Wearable Computing Anwendungen sowie deren Kommunikation mit intelligenten Artefacten werden.

### 5.4. Steuerung durch Gedanken

Die Zukunft könnte allerdings auch stärker „by reading brainwaves“ bestimmt sein [25], da die sprachliche Konversation im Sinne der wearable-Anwendungen eine Anzahl von Störpotentialen für das Umfeld, wie auch praktische Tests zeigten, aufweisen. Hierdurch könnte stärker die diskrete individuelle Nutzung von Technologien ermöglicht werden.

### 5.5. Berücksichtigung ethischer Fragestellungen

Ähnlich wie die Untersuchungen von Steve Mann [22], wurden auch im Projekt HUMAINE [Human-Machine-Interaction Network on Emotions] [23] die Bedeutung eines mehr ethisch geprägten Ansatzes zur Nutzung von unsichtbaren Technologien, zu denen auch Spracherkennungsprogramme gehören, hervorgehoben. Durch emotional geprägte Computertechnologien werden viel stärker als je zuvor Befindlichkeiten in Form von „Frankenstein“ oder „Big-Brother-Visionen“ hervorgerufen. Daher werden Entwicklungsarbeiten zur Stärkung der Nutzerkontrolle und für neue Sicherheitsstandards in Zukunft weiter an Bedeutung gewinnen.

Eine Studie von Yacoub et al. [24] beschreibt ein grundlegendes Problem von Technologie sehr schön, die Möglichkeiten werden herausgestellt, Einschränkungen oder Probleme werden aber nicht benannt . Damit disappearing computer-Lösungen zu einer wirklichen Verbesserung der Lebensqualität beitragen, sollten bei jeder Entwicklung Technikfolgen- und Innovationsbewertungen einen wesentlichen Beitrag zur Forschung liefern.

## Literatur


[1] Nottingham-Projekt. http://www.nottingham.ac.uk/research-innovation/press-releases

[2] Smart Sight: A Tourist Assistant System, Jie Yang, Weiyi Yang, Matthias Denecke, Alex Waibel Interactive Systems Laboratories, Carnegie Mellon University Pittsburgh, PA 15213

[3] Hartmann, Wolf D., Klaus Steilmann und Astrid Ullsperger: High-tech-fashion. Mit einem Lexikon von avatar bis wearable computer. – Heimdall-Verlag, 2001. und Xybernaut http://www.infosyncworld.com/news/n/5107.html

[4] LT-EVAL: Multitasking mit dem Xybernaut, www.lt-eval.org

[5] http://emotion-research.net/Members/DFKI

[6] http://emotion-research.net/wiki/Databases

[7] Nakagawa, S. and Zhang, J.X., "An input interface with speech and touch screen," Trans. Inst. Elec.Eng. Jpn. C (Japan), Vol. 114-C, No. 10, pp. 1009-1017, 1994.

[8] Ando, H., Kitahara, Y., and Hataoka, N., "Evaluation of multimodal interface using spoken language and pointing gesture on interior design system," Proc. ICSLP'94, Vol. 2, pp. 567-570, Yokohama, Japan.

[9] Sawhney, N. and Schmandt, C., ‘Speaking and listening on the run: design for wearable audio computing’, Proc. 2nd international symposium on wearable computers, pp.108-115, October 19-20,1998.

[10] http://www.testberichte.de/test/produkt_tests_siemens_wristphone_p16419.html

[11] Goose, Stuart, Heiko Wanning and Georg Schneider: Mobile Reality: A PDA-Based Mulitmodal Framework Synchronizing a Hybrid Tracking Solution with 3D Graphics and Location-Sensitive Speech Interaction. – In: UbiComp 2002: Ubiquitous Computing. 4th International Conference Göteborg, Sweden, September/October 2002. Proceedings. – p. 33 – 47

[12] Spitzer, Manfred: Geist im Netz. Modelle für Lernen, Denken und Handeln. – Heidelberg, Berlin: Spektrum, 2000.

[13] Picard, Rosalind: Affective computing. MIT Press 1997, p. 32.

[14] Wightman, Colin W.: ToBi or not ToBi? In: Speech Prosody 2002. 1st International conference. Aix-en-Provence, France, 11-13 April 2002

[15] Donat, Stefan: Unfallverhütung durch Datenübertragung über die Haut. ident technology AG, Bayrischer Innovationspreis 2004 www.bayern.de/Wirtschaftsstandort/ Innovationspreis/Preistraeger_2004/ident.html

[16] Nishimoto, T., Shida, N., Kobayashi, T., and Shirai, K., "Multimodal drawing tool using speech, mouse and keyboard," Proc. ICSLP'94, Vol. 3, pp. 1287- 1290, Yokohama, Japan.

[17] Beigl, M.: Memoclip: A location-based Rembrance Applicance, Journal of Personal Technologies, 4(4), Springer Press, 2000.- p. 230 – 234.

[18] Abowd, G., Atkeson, S., Dey, A. , Hong, J., Long, S., Kooper R. and Pinkerton M.: Cyberguide: A mobile context-aware tourguide, ACM Wireless Networks Volume 3, p. 421 – 433, November 1997.

[19] Fibercomputing http://www.fibercomputing.net

[20] Smart Innovation Leather Wear (SMILE) ® IPEKDERI and KSI, 2004

[21] Mann, Steve with Hal Niedzviecki: Cyborg. Digital destiny and human possibility in the age of the wearable computer. – Canada: Randomhouse, 2001.

[22] HUMAINE. Report on ethical prospects and pitfalls for emotion-oriented systems. – Deliverable IST-programm, 28 May 2004, Version 1.0

[23] Yacoub, S., Simske, S., Lin, X. & Burns, J. 2003. Recognition of emotions in interactive voice response systems. *Proc. Eurospeech 200*3, Geneva.

[24] Smailagic, A., D.P. Siewiorek, R. Martin, D. Reilly, “CMU Wearable Computers for Real-Time Speech Translation,” Third International Conference on Wearable Computers, 1999.6

[25] Siewiorek, Dan: Power Consumption and Performance Analysis of Real – Time Speech Translator Smart Module, Human Computer Interaction Institute, Institute for Complex Engineered Systems Carnegie Mellon University, Pittsburgh, PA 15213, dpr@andrew.cmu.edu, {dps,asim}@cs.cmu.edu

[26] C.K. Christakos, “Optimizing a Language Translation Application for Mobile Use,” Master’s Project Report, Carnegie Mellon University, Electrical and Computer Engineering Department, 1997.

[27] Oviatt, S.L., Cohen, P.R., and Wang, M., "Toward interface design for human language technology: modality and structure as determinants of linguistic complexity," Speech Communication (Netherlands), Vol. 15, Nos. 3-4, pp. 283-300, 1994.

[28] Ravishankar, M., Efficient Algorithms for Speech Rec-ognition. Ph.D Thesis, Carnegie Mellon University, May 1996, Tech Report. CMU-CS-96-143

[29] Ducatel, K. And M. Bogdanowicz, F. Scapolo, J. Leijten & J-C. Burgelmann: SCENARIOS FOR AMBIENT INTELLIGENCE IN 2010. Final Report. February 2001, IPTS-Seville.

***`*Es gilt das gesprochene Wort.***