Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Ausarbeitung AW1

Kristoffer A. Witt

Untersuchung der Eignung von Spracherkennung zur Transkription von Radiospots

Fachbereich Elektrotechnik und Informatik
Department of Electrical Engineering and Computer Science

Kristoffer A. Witt

Untersuchung der Eignung von Spracherkennung zur Transkription von Radiospots

Ausarbeitung AW1 eingereicht im Rahmen der Veranstaltung AW1
im Studiengang Master of Science, Verteilte Systeme
Studienrichtung Softwaretechnik
am Fachbereich Elektrotechnik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg
Abgegeben am 15. Dezember 2008

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einleitung

Sobald ein Mensch gelernt hat, Töne und Laute zu artikulieren um daraus Wörter, Wortfolgen und ganze Sätze zu bilden, besitzt er das Handwerkszeug für die nativste Form der menschlichen Informationsübermittlung: die Lautsprache. Seit mehr als einem halben Jahrhundert ((Pfister und Kaufmann, 2008, Seite 283, 10.1)), versuchen Wissenschaftler bereits, dem Computer beizubringen zuzuhören, zu verstehen und sich sprachlich Verständlich zu machen. Diese Bemühungen werden unter dem Themenkomplex Sprachverarbeitung zusammengefasst. Für die resultierenden Techniken und Technologien existieren heute mannigfaltige Anwendungsmöglichkeiten. Besonders der Bestandteil der Spracherkennung, also der Abbildung von phonetischen auf graphemische Daten (Ton in Schrift), eröffnet in Kombination mit Dataminingtechniken neue Perspektiven für die Durchsuchbarkeit von Audioinhalten.

Das Internet enthält heute gigantische Anzahl an Informationen gespeichert in den unterschiedlichsten Medien. Diese Datenflut durchsuchbar zu machen um Sie einer Vielzahl interessierten Anwendern zur Verfügung zu stellen ist der Geschäftsbereich der Suchmaschinen. Dabei muss die andauernde Medientransformation besonders berücksichtigt werden. War es zur Anfangszeit des Internets noch undenkbar, aufgrund der hohen Datenmenge und geringen Übertragungsleistungen, Video- und Audiodaten bereitzustellen ist die heutzutage sehr einfach möglich. Durch die sich stetig verbessernden Komprimierungsalgorithmen in Kombination mit immer schnelleren Leitungen können mehr Daten in weniger Zeit zum Konsumenten übertragen werden. Dadurch steigt die Akzeptanz der multimedialen Formate beim Nutzer, was sich wiederum positiv auf die Popularität der Daten beim Anbieter auswirkt. Die Problematik, die durch diese Transformation der Darstellungsform entsteht ist, dass etwaig vorhandene Nutzdaten bzw. Informationen nun nicht mehr in Textform vorliegen, sondern als Audiodaten. Das bedeutet, um die bisherig entwickelten Such- und Indexierungsalgorithmen weiter verwenden zu können, muss eine Transformation von Audio in Textdaten erfolgen, also Spracherkennung.

In dieser, im Rahmen der Veranstaltung Anwendungen 1 des Master-Studiengangs „Verteilte Systeme“ der Hochschule für angewandte Wissenschaften entstandenen Arbeit, wird der Einsatz von Sprachverarbeitung für die Transkription von Audiodaten hinsichtlich der Eignung für weiterführende Arbeiten analysiert und erörtert.

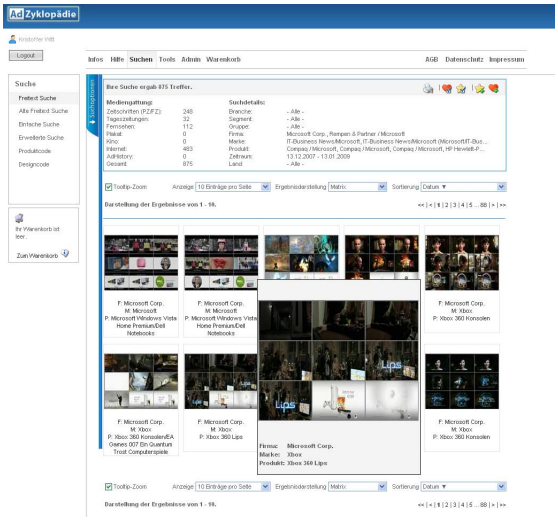

Abbildung 1.1: Screenshot eines Suchergebnis (Suchbegriff „Microsoft“) des GWA-Ädzyklopaedie vom 13.12.2008

1.1 Szenario

Das in Hamburg ansässige Unternehmen AdVision digital GmbH bietet Werbeagenturen, Verlagen und Marktforschungsunternehmen Dienste auf dem Gebiet Medienbeobachtung, speziell im Bereich Werbung, an. Das Hauptprodukt ist die so genannte „GWA-AdZyklopädie“, siehe AdVision Digital GmbH (2008). Ein Internetportal, das Zugriff auf die in den Medien TV, Tageszeitung, Publikums- und Fachzeitschriften und Plakaten geschalteten und erfassten Werbemaßnahmen bietet. Siehe Abbildung 1.1 für eine Beispielansicht. Um den Kunden die Recherche in der Datenbank möglichst einfach zu gestalten, wurden verschiedene Metainformationen festgelegt, die die Durchsuchbarkeit der der aktuell ungefähr 1.5 Millionen Werbemaßnahmen umfassenden Datenbank ermöglichen sollen. Diese Informationen werden aktuell hauptsächlich händisch vergeben. Die Attribute die für alle Medien erfasst werden sind: Konzern, Marke, Produkt sowie ein Produktcode. Zum besseren Verständnis ein Beispiel: Konzern: Daimler, Marke: Mercedes-Benz, Produkt: Mercedes-Benz A-Klasse, Produktcode: PKW Neuwagen. Der Produktcode ermöglicht eine Recherche nach Sparteninformationen, eine Übersicht ist dem Anhang zu entnehmen, siehe Anhang A. Im Zuge der Verbreiterung des Angebots, steht die Aufnahme des Mediums Radiowerbung in die Produktpalette an. Um eine optimale Durchsuchbarkeit dieses Mediums zu erreichen, bietet es sich an, eine automatische Spracherkennung auf den Werbemaßnahmen durchzuführen und somit eine Transkription des Spots zu erreichen, anhand deren eine Bindung an Suchparameter erfolgen könnte. Inwiefern die Qualität einer automatischen Erfassung der Informationen ausreicht und ob eine manuelle Nachbearbeitung notwendig ist, bleibt zu klären.

1.2 Ideen

Im Folgenden wird kurz erläutert, welche Möglichkeiten sich durch eine automatische Transkription von Audiodaten ergeben und wie diese Daten bezüglich des Szenarios Einsatz finden können.

Indizierung Der offensichtlichste Nutzen der aus den Daten gewonnen Texten ist die Verwendung für Suchindizes. Die erkannten Begriff können über eine einfache Ganzwortsuche gefunden werden und ermöglichen somit einen schnellen Zugriff.

Unterstützung von Gehörlosen Die textuelle Darstellung der gesprochenen Sprache ermöglicht es, gehörlosen Menschen den Inhalt der Werbemaßnahme zu erfassen und zu verarbeiten.

Fingerprinting Bei der Verarbeitung von Werbemaßnahmen bleibt es nicht aus, einzelne Spots mehrfach zu erfassen. möglichst eindeutiger Fingerabdruck generiert werden, um Doppelungen zu erkennen, zu annotieren und zu entfernen. Dies dient auch der Erstellung von sogenannten Streuplänen, die erfassen wann und wo eine Werbemaßnahme gelaufen ist.

Navigation Verknüpft man die Textdaten mit den Zeitpunkten im Audiosignal zu denen sie erkannt wurden, erhält man ein Inhaltsverzeichnis für den Spot. Anhand dessen sich im Spot dann navigieren lässt.

Unterstützung der manuellen Verarbeitung Wie bereits beschrieben, werden aktuell manuell bestimmte Informationen vergeben, wie Produkt- oder Markenname. Ein Transskript des Inhalts kann den Prozess dahingehend beschleunigen, dass ein Mitarbeiter nun nicht mehr den kompletten Spot anhören muss, sondern anhand des Transskripts entscheiden kann welches Produkt beworben wurde.

2 Hauptteil

2.1 Relevante Technologien

Das Kapitel beschreibt kurz die für die Durchführung des angestrebten Projektes benötigten Technologien und deren möglichen Einsatz.

2.1.1 Klassifizierung von Audiodaten

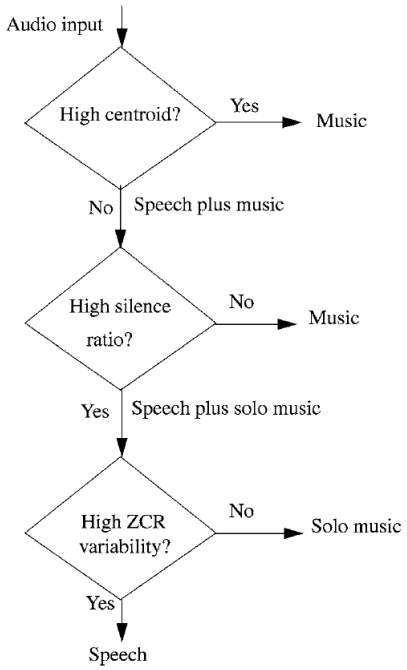

Abbildung 2.1: Beispielhafter Ablauf von Audiodatenklassifizierung, aus (Lu, 2001, Seite 278)

Um Informationen aus Audiodaten gewinnen zu können, ist es nötig, festzustellen um

welche Art von Daten es sich handelt (Lu, 2001, Seite 270). In Abbildung 2.1[1] ist ein möglicher Klassifizierungsprozess dargestellt. Die Klassifizierung teilt das Signal anhand verschiedener Merkmale in Kategorien ein. Anhand der zugeteilten Kategorie kann dann entschieden werden wie und ob die Daten weiteren Informationsextraktionsbemühungen unterzogen werden sollen. Als Beispiel seien Musik und Sprachdaten genannt. Macht es für Sprachdaten Sinn, sie einer automatischen Spracherkennung zu unterziehen, ist für Musik eher eine Einordnung in Genre, Stimmung oder Geschwindigkeit angebracht. Nicht unerwähnt sollte hierbei bleiben, dass in Radiowerbespots auch häufig mit kurzen Gesangspassagen gearbeitet wird, um einen höheren Wiedererkennungsfaktor zu generieren. Dementsprechend sollten diese Gesangspassagen für eine vollständige Indizierung des Inhalts ebenfalls erfasst und verarbeitet werden. Nach erfolgter Einteilung kann auf die geeigneten Signalbestandteile eine Spracherkennung angewendet werden.

Features	Speech	Music
Bandwidth	0–7 kHz	0–20 kHz
Spectral centroid	low	high
Silence ratio	high	low
Zero-crossing rate	more variable	less variable
Regular beat	no existing	often existing

Abbildung 2.2: Beispielhafter Merkmale von Audiodaten, aus (Lu, 2001, Seite 277)

2.1.2 Spracherkennung

Spracherkennung versucht mittels eines Computerprogramms Sprachsignale, zum Beispiel von einem Mikrophon aufgezeichnet, in Textform zu überführen (Transskription). Der grundsätzliche Ablauf lässt sich aus Abbildung 2.3 entnehmen. Für Detailliertere Informationen siehe Pfister und Kaufmann (2008). Dabei gibt es verschiedene Arten von Spracherkennern mit unterschiedlichen Aufgabengebieten, siehe Abbildung 2.4. In dem

[1] Das Audiosignal wird folgendermaßen klassifiziert, enthält es ein hohes Zentroid (>10Khz), siehe 2.2 , also eine hohe Bandbreite an Frequenzen kann davon ausgegangen werden das es sich um Musik handelt, da Sprachfrequenzen meist im Bereich 0-8 Khz anzusetzen sind. Weiter wird anhand des Anteils der Stille, also der Energiearmen Passagen, auf Sprache oder Musik geschlossen. In Musik gibt es weniger stille Passagen als bei Sprache, zum Beispiel Sprechpausen. Die letzte Einteilung ist durch die Varianz der Nulldurchgänge (Zero Crossings) gekennzeichnet. Da Musik meist aus Harmonien besteht, also aus Periodischen Signalen, ist die Anzahl der Nulldurchgänge im Zeitsignal geringer als bei Sprache. Hohe Raten an Nulldurchgängen werden bei Sprache durch sogenannte Frikative verursacht, für weitere Informationen sei auf (Pfister und Kaufmann, 2008, Seite 48) verwiesen.

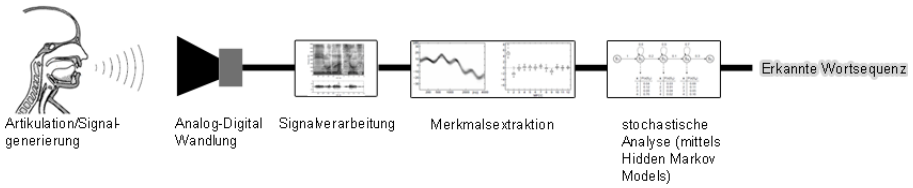

Abbildung 2.3: High Level Modell des Spracherkennungsablaufs

Kontext dieser Arbeit macht dabei nur der kontinuierliche Spracherkenner Sinn, da vollständige Sätze erfasst werden sollen.

Systemklasse	Verarbeitbare Äusserungen
Einzelworterkenner	Einzelne Wörter oder kurze Kommandos isoliert gesprochen, d.h. mit Pausen.
Keyword-Spotter	Einzelne Wörter oder kurze Kommandos in einer sonst beliebigen Äusserung.
Verbundworterkenner	Sequenz von fliessend gesprochenen Wörtern aus einem kleinen Vokabular (z.B. Telefonnummern).
Kontinuierlicher Spracherkenner	Ganze, fliessend gesprochene Sätze.

Abbildung 2.4: Einteilung von Spracherkennungssystemen, aus (Pfister und Kaufmann, 2008, Seite 291)

Das Problem der Spracherkennung bzw. Verarbeitung ist ein interdisziplinäres (Vgl. Abbildung 2.5). Im Folgenden werden kurz die einzelnen Bereiche und ihr Beitrag zum Erkennungsprozess erläutert.

Signalverarbeitung

Sprache ist auf physikalischer Ebene gesehen nichts anderes als Schallwellen, die sich durch Luftdruckunterschiede manifestieren. Durch den Einsatz eines Mikrophons werden aus diesen Luftdruckunterschieden elektrische Signale generiert. Dies ermöglicht die weitere Verarbeitung zwecks Aufbereitung für die Informationsgewinnung. Da die Signalverarbeitung der erste Schritt auf dem Weg zum erkannten Text ist, hat sie Auswirkungen auf alle Folgeoperationen. Sie bildet somit das Fundament für den Erkennungsprozess.

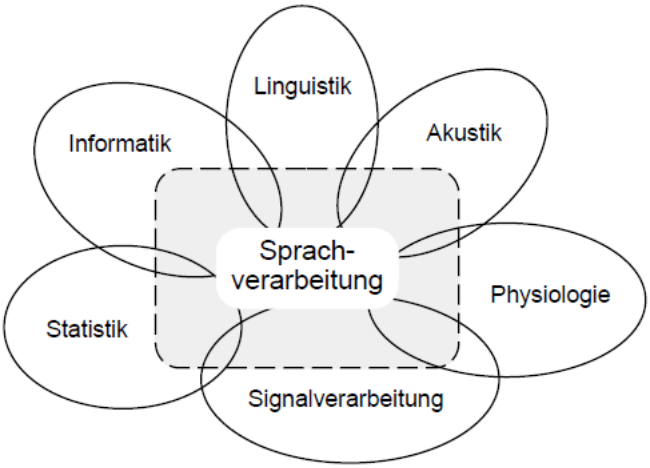

Abbildung 2.5: Disziplinen der Spracherkennung, aus (Pfister und Kaufmann, 2008, Seite 22)

Physiologie

Das Fachgebiet der Physiologie, genauer der Humanphysiologie, beschäftigt sich mit der Funktionsweise der menschlichen Körpers (Vgl. (Guyton, 1991, Seite 3)). Der Aufbau des menschlichen Vokaltrakts beeinflusst die Ausbreitung der Schallwellen von den Stimmbändern. In Abbildung 2.6 sind die Bestandteile des Vokaltrakts zu sehen. Je nach Art ihres Zusammenspiels werden die unterschiedlichen Töne erzeugt (Vgl. Phonetics Flash Animation Project, University of Iowa,USA (2008)). Der Vokaltrakt steht somit in direktem Zusammenhang mit den Frequenzanteilen aus denen Sprachsignale aufgebaut sind. Je detaillierter das Verständnis seines Aufbaus ist, desto höher ist auch die Leistung beziehungsweise die Genauigkeit der Spracherkennung.

Akustik

Wie in den vorhergehenden Punkten erläutert bilden Schallwellen und deren Ausbreitung die Grundlagen der Sprache, die Akustik ist definiert als die Lehre dessen. Durch immer genauere Modelle für die Schallausbreitung und die Interferenz lassen sich die Einflüsse von Lärm auf das Sprachsignal besser approximieren und somit schon vor der Analyse beseitigen oder minimieren.

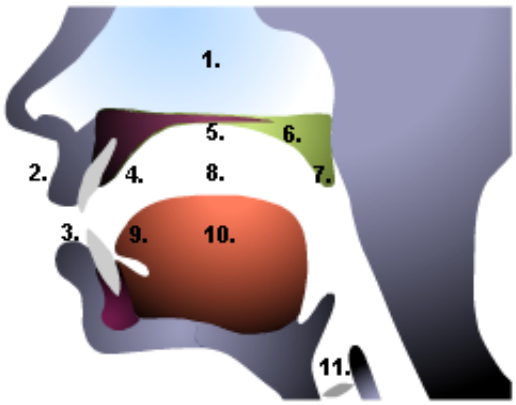

Abbildung 2.6: Aufbau des Vokaltrakts, aus Phonetics Flash Animation Project, University of Iowa,USA (2008)

Linguistik

Die Sprachwissenschaft beschäftigt sich mit der Erforschung und Beschreibung von Sprache. Neben der akustischen Grundlage bildet die Linguistik das semantische Modell der Spracherkennung. Sie beschreibt die Grundzüge der Sprache und wie sie sich definiert. Ohne diese Beschreibung ist eine Erkennung von Sprache unmöglich. Neben der Basisfunktion werden Ergebnisse der sprachwissenschaftlichen Forschung auch für die Verbesserung der Erkennung eingesetzt. Als Beispiel sei hier die Grammatik genannt. Anhand von grammatikalischen Regeln kann entschieden werden ob ein Wort Sinnvoll ist, oder ob besser eine ähnlich klingende Alternative erkannt werden sollte.

Statistik

Statistische Erkenntnisse über die Häufigkeit von Lautkombinationen ermöglichen erst die stochastische Spracherkennung. Insbesondere der Einsatz von sogenannten Hidden-Markov-Modellen (HMM) hatte einen positiven Einfluss auf die Performanz von Spracherkennern. Wie genau HMMs eingesetzt werden lässt sich bei Rabiner (1989) nachlesen.

Informatik

Die Informatik bildet die Schnittstelle zwischen Audiodaten und Datenverarbeitung. Die empfangenen Signale werden vom Computer ausgewertet und anhand der beschriebenen Merkmale verarbeitet und so in eine Textform überführt. Dabei werden zum Beispiel

Algorithmen der Dynamischen Programmierung verwendet, um möglichst performant große HMMs zu untersuchen.

2.2 Probleme und Risiken

Der nächste Abschnitt beinhaltet eine Analyse der möglichen Probleme und Risiken, die während der Realisierung des Projektes auftreten können.

2.2.1 Heterogenität der Audiodaten

Eines der Hauptprobleme bei der angestrebten Indexierung bzw. Transkription von Audiodaten ist deren Heterogenität. Um ein möglichst effektives und robustes Spracherkennungssystem zu erhalten, wird im Allgemeinen, ein Profil mit den Eigenheiten des jeweiligen Sprechers angelegt. Dieser Schritt ist bei der Erkennung von Werbespots nur sehr selten möglich. Es muss davon ausgegangen werden, dass das System mit einem oder auch mehreren komplett unbekannten Sprechern konfrontiert wird.

2.2.2 Störgeräusche und Umgebung

Weiterhin ist im Vergleich zu einer speziell für die Spracherkennung ausgestatten Umgebung, also personengebundene Mikrophone die Fremdgeräusche auf ein Minimum reduzieren, mit zusätzlichen Stolpersteinen zu rechnen. Als Beispiel seien hier mit hoher Wahrscheinlichkeit auftretende Störgeräusche genannt. Dieser „Lärm“ umfasst zum Beispiel Hintergrundmusik, Nebengeräusche und Soundeffekte. Auch wechselnde Sprecherumgebungen, also zum Beispiel Telefonleitung oder ein Innenraum mit Echo, können den Erkennungsprozess stark negativ beeinflussen.

2.2.3 Segmentierung der Audiodaten

Ebenfalls als Komplikation ist die Segmentierung der Spots zu betrachten. Der Begriff Segmentierung wird in diesem Zusammenhang mehrfach genutzt. Zum Einen ist es wünschenswert aus dem laufenden beziehungsweise aufgezeichnetem Radioprogramm automatisch zu erkennen, wann ein Werbeblock beginnt und wann er endet. Zum Zweiten muss innerhalb dieser Blöcke zwischen den Einzelnen Werbemaßnahmen unterschieden werden.

2.2.4 Uneingeschränkter Wortkontext

In Umgebungen in den Spracherkennung zur steuerung von verschiednen Vorgängen genutzt wird, ist meist eine Grammatik spezifiziert. Diese schreibt vor, welche Wörter in welcher Reihenfolge erkannt werden können, dies hat starken Einfluss auf die Performanz des Erkennungssystems. Aufgrund der Tatsache, dass der Inhalt der Radiospots nicht bekannt ist, lässt sich so eine Grammatik nicht einsetzen. Das führt dazu, dass im Diktiermodus gearbeitet werden muss, in dem jedes Wort erkannt werden muss.

Prosodie und Sprachen

In Werbespots ist es möglich, um ein bestimmtes Bild zu vermitteln, das Sprecher mit stark lokal geprägter Aussprache eingesetzt werden. Dies führt dazu das die Erkennungsrate sinkt. Weiterhin ist nicht bekannt, in welcher Sprache der Inhalt des Spots gehalten ist. Es kann vorkommen, dass ein Slogan oder bestimmte Schlüsselworte auf Englisch, aber der Rest des Spots auf Deutsch geäußert wird.

2.2.5 Unzureichende Lösungsansätze

Aufgrund der Tatsache, dass die Spracherkennung ein relativ altes Fachgebiet ist, gibt es für die oben genannten Problematiken schon verschiedene Lösungsansätze. Trotz dieses Faktes ist es aufgrund der sehr unterschiedlichen Quelldaten möglich, dass diese in der Literatur beschriebenen Ansätze nicht ausreichen. Es muss somit festgestellt werden, in wie weit vorhandene Ergebnisse überhaupt Verwendung finden können. Sollte es nötig werden, neue Ansätze für die Kompensation verschiedener Problematiken finden zu müssen, bedeutet dies einen stark erhöhten Aufwand. Ob dieser in der zur Verfügung stehenden Zeit geleistet werden könnte bleibt abzuwarten.

2.3 Schnittstellen zu anderen AW-Projekten

In dem folgenden Abschnitt werden Schnittstellen zu anderen aktuellen AW-Projekten identifiziert.

Sprachverarbeitung, im speziellen Spracherkennung, ist sowohl für die Nutzung mit aufgezeichnetem Inhalt, wie auch für "Live"geäußerten Inhalt geeignet. Daher ergibt sich die implizite Möglichkeit, Sprache als Eingabemodalität zu nutzen. Im Zusammenhang gesehen mit Schnittstellen zu anderen AW-Projekten lässt sich also feststellen, dass überall

dort, wo die Interaktion von Mensch und Computer eine Rolle spielt, der Einsatz von Spracherkennung denkbar wäre. Der Problemraum, der im vorherigen Kapitel erläutert wurde, ist auch bei diesen Projekten ein Ähnlicher. Sprecherunabhängige und robuste, störungsresistente Erkennung sind die Hauptprobleme. So dass Verbesserungen hinsichtlich dieser Schwierigkeiten Auswirkungen auf alle Bereiche haben können. Erwähnenswert bleibt allerdings, dass sich bei den nun folgenden Projekten zum Teil andere Mechanismen zur Lösung beziehungsweise Vereinfachung finden ließen[2].

2.3.1 Einteilung nach Themengebieten

Ein Einsatz von Spracherkennung bietet ist in folgenden Projekten denkbar:

Ambient Assisted Living vertreten durch „Intelligenter Stuhl“

Bei der Metapher Ambiet Assisted Living geht es hauptsächlich um die Unterstützung von möglicherweise eingeschränkten Personen bei alltäglichen Situationen, sowohl in Ihrer Heimstätte als auch außerhalb derer. Der Einsatz von Sprache zur Steuerung eines Systems ermöglicht es, auf andere Periphere Eingabemedien verzichten zu können. Dies ist insbesondere für Personen hilfreich deren Behinderungen eine Extremitäten gesteuerte Eingabe nicht zulassen. Zum Beispiel Personen mit amputierten oder gelähmte Armen bzw. Händen. Ebenfalls wäre der Einsatz eines solchen Sprach-Dialoggesteuerten (beinhaltet zusätzlich Sprachsynthese als Ausgabemodalität) Unterstützungssystems für Sehbehinderte oder blinde Menschen denkbar.

Pervasive Gaming

Eine Einsatzmöglichkeit für Sprachsteuerung in Pervasive Gaming(PG) Umgebungen ist relativ schnell gefunden. Der Begriff PG bedeutet in etwa durchdringendes Spielen. Also grob gesprochen die Nutzung von Informationstechnik um die meist virtuellen Spielumgebungen mehr und mehr in einen realen Kontext zu setzen (Vgl. IPerG Fraunhofer Institut für Angewandte Informationstechnik (2008)). Ansatzpunkte für den Einsatz von Spracherkennung sind:

- Spracheingabe als weitere Modalität für den jeweiligen Spieler um die etwaig benötigte Eingabe zu beschleunigen oder in Situationen wo andere Eingabemittel nicht nutzbar sind zum Beispiel während einer "Flucht".

[2] Beispielsweise Visuelle Unterstützung der Spracherkennung durch Optical Flow Analyse in Multimodalen Umgebungen, siehe Minker u. a. (2005a)

- Spracherkennung für die Kommunikation zwischen Spielern um anhand des Inhalts steuernd auf den Spielverlauf einzuwirken oder auch Teammitgliedern erweiterte Informationen zur Verfügung stellen zu können. Ein einfaches Beispiel wäre die Transkription der Kommunikation zwecks der Dokumentation des Spielverlaufs.
- Die Kommunikation mit einer künstlichen Intelligenz nur über sprachliche Ein- und Ausgaben kann dem Spiel eine enorm gesteigerte Realitätsnähe geben.

Seamless Interaction

Im Zusammenhang mit Seamless Interaction ist Sprache, ähnlich wie in den zuvor genannten möglichen Einsatzgebieten, als eine weitere Modalität für die Interaktion/Kommunikation mit dem Benutzer zu verstehen. Bei Multimodalen Systemen, also Systemen, die über eine Vielzahl unterschiedlicher Ein und Ausgabe Möglichkeiten verfügen, sollte sich die Interaktion so gestalten. Der Benutzer wählt, welche Art der Eingabe für seine aktuelle Situation am besten geeignet ist. Für lange Texte bietet es sich zum Beispiel an diese zu diktieren, während Eingaben die mit einer hohen Frequenz erfolgen müssen eher mit Berührung oder Gesten umgesetzt werden sollten. Weitere Ansätze zur Verwendung von Sprache in Multimodalen Umgebungen können Minker u. a. (2005a) oder Gerd Herzog (2006) entnommen werden.

3 Schluss

3.1 Perspektive

Der Abschnitt Perspektive beschreibt Ansätze für die weitere Vorgehensweise in den folgenden Veranstaltungen AW2 und im Projekt.

Grundlegend wichtig für die Durchführung der in dieser Arbeit erläuterten Ansätze, sind ausgiebige Tests. Daher sollte die Projektzeit dafür verwendet werden, eine detaillierte Testsuite zu erstellen, die es ermöglicht, vergleichbare Resultate verschiedenster Problemlösungen für die im Abschnitt "Probleme und Risikenërörterten Unwegsamkeiten zu erzeugen und auszuwerten. Insbesondere soll diese Testumgebung dazu dienen, verschiedene Spracherkennungs-Systeme mit unterschiedlichen Parametern zu evaluieren um später Ergebnisse mit optimaler Erkennungsrate erzeugen zu können. Ebenfalls soll es die Testsuite ermöglichen etwaig vorhandene, zum Beispiel aus Fachliteratur entnommene, Problemlösungsansätze, speziell in der Signalvorverarbeitung, zu implementieren und auszuwerten. Um überhaupt testen zu können, muss ein Fundus geeigneter Testdaten, also Radiospots angelegt werden. Das bedeutet es müssen entweder Mitschnitte aus dem Radioprogramm angefertigt, oder aber die Spots direkt beim Ersteller angefordert werden. Wahrscheinlich wird, durch die hohe Heterogenität verschiedenster Spots bedingt, zuerst eine Anzahl typischer Vertreter ausgewählt werden müssen. Vorausgesetzt, es existieren ausreichend Gemeinsamkeiten, die eine solche Klassifizierung ermöglichen.

Literaturverzeichnis

[AdVision Digital GmbH 2008] AdVision Digital GmbH: GWA Ädzyklopaedie. http://v2.adzyklopaedie.com. 2008. – [Online; accessed 12-December-2008]

[Comerford u. a. 1997] Comerford, Richard ; Makhoul, John ; Schwartz, Richard: The voice of the computer is heard in the land (and it listens too!). In: IEEE Spectr. 34 (1997), Nr. 12, S. 39–47. – ISSN 0018-9235

[Gerd Herzog 2006] Gerd Herzog, Norbert R.: The SmartKom Architecture: A Framework for Multimodal Dialogue Systems. In: Wahlster, Wolfgang (Hrsg.): SmartKom - Foundations of Multimodal Dialogue Systems, Springer, 7 2006 (Cognitive Technologies), S. 55–70

[Guyton 1991] Guyton, Arthur C.: Textbook of Medical Physiology. Philadelphia PA : W. B. Saunders Company, 1991. – ISBN 0-7216-3087-1

[Horchani u. a. 2007] Horchani, Meriam ; Caron, Benjamin ; Nigay, Laurence ; Panaget, Franck: Natural multimodal dialogue systems: a configurable dialogue and presentation strategies component. In: ICMI '07: Proceedings of the 9th international conference on Multimodal interfaces. New York, NY, USA : ACM, 2007, S. 291–298. – ISBN 978-1-59593-817-6

[IPerG Fraunhofer Institut für Angewandte Informationstechnik 2008] IPerG Fraunhofer Institut für Angewandte Informationstechnik: Integriertes Projekt über Pervasive Gaming. http://www.fit.fraunhofer.de/projects/mixed-reality/iperg.html. 2008. – [Online; accessed 12-December-2008]

[Lu 2001] Lu, Goujun: Indexing and Retrieval of Audio: A Survey. In: Multimedia Tools Appl. 15 (2001), Nr. 3, S. 269–290. – ISSN 1380-7501

[Minker u. a. 2005a] Minker, Wolfgang (Hrsg.) ; Bühler, Dirk (Hrsg.) ; Dybkjær, Laila (Hrsg.): Text, Speech and Language Technology. Bd. 28: Spoken Multimodal Human-Computer Dialogue in Mobile Environments. Dordrecht : Springer, 2005. – 37–57 S. – ISBN 978-1-4020-3073-4

[Minker u. a. 2005b] MINKER, Wolfgang (Hrsg.) ; BÜHLER, Dirk (Hrsg.) ; DYBKJÆR, Laila (Hrsg.): Text, Speech and Language Technology. Bd. 28: Spoken Multimodal Human-Computer Dialogue in Mobile Environments. Dordrecht : Springer, 2005. – 37–57 S. – ISBN 978-1-4020-3073-4

[Pfister und Kaufmann 2008] PFISTER, Beat ; KAUFMANN, Tobias: Sprachverarbeitung - Grundlagen und Methoden der Sprachsynthese und Spracherkennung. Springer, 2008. – ISBN 978-3-540-75909-6

[Phonetics Flash Animation Project, University of Iowa,USA 2008] PHONETICS FLASH ANIMATION PROJECT, UNIVERSITY OF IOWA,USA: Phonetics: The sounds of German. http://www.uiowa.edu/~acadtech/phonetics/german/frameset.html. 2008. – [Online; accessed 12-December-2008]

[Rabiner 1989] RABINER, L. R.: A tutorial on hidden Markov models and selected applications in speech recognition. In: Proceedings of the IEEE 77 (1989), Nr. 2, S. 257–286. – URL http://dx.doi.org/10.1109/5.18626

[Scheirer und Slaney 1997] SCHEIRER, E. ; SLANEY, M.: Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator. In: Acoustics, Speech, and Signal Processing, IEEE International Conference on 2 (1997), S. 1331

Glossar

Centroid/Zentroid

Bezeichnet den Mittelwert eines Spektrums.

Datamining

Maschinelles auswerten von Daten zwecks Informationextraktion.

Vokaltrakt

Als Vokaltrakt werden Rachen, Mund und Nasenraum bezeichnet, da Sie die Resonanzräume für die von Stimmlippen erzeugten Schwingungen bilden.

ZCR

Zero Crossings Rate - Anzahl von Null Übergängen in der Zeitdarstellung eines Audiosignals. Die ZCR ist in der Spracherkennung ein Indikator für einteilung von Stimmhaften und Stimmlosen Phonemen.

A Auszug GWA Ädzyklopaedie Produktcode

1. Automobile/Kfz
   - 1.1. Auto/Pkw
     - 1.1.1. Pkw - Gebrauchtwagen
     - 1.1.2. Pkw - Geländefahrzeuge/Jeeps
     - 1.1.3. Pkw - Neuwagen
     - 1.1.4. Range-/Programmwerbung
     - 1.1.5. Sponsoring/Image-/Firmenwerbung
   - 1.2. Nutzfahrzeuge
     - 1.2.1. Busse
     - 1.2.2. Industriefahrzeuge/Gabelstapler (nur Fahrzeuge)
     - 1.2.3. Lkw
     - 1.2.4. Pick-Ups/Transporter
     - 1.2.5. Range-/Programmwerbung
     - 1.2.6. Sponsoring/Image-/Firmenwerbung
   - 1.3. Reifen/Felgen
     - 1.3.1. Felgen
     - 1.3.2. Range-/Programmwerbung
     - 1.3.3. Reifen
     - 1.3.4. Sponsoring/Image-/Firmenwerbung
   - 1.4. Roller/Motorräder/Mopeds/Zubehör
     - 1.4.1. Mopeds/Roller
     - 1.4.2. Motorräder
     - 1.4.3. Pflege/Zubehör
     - 1.4.4. Range-/Programmwerbung
     - 1.4.5. Sponsoring/Image-/Firmenwerbung
   - 1.5. Tanken/Kraftstoffe/Öle
     - 1.5.1. Benzin
     - 1.5.2. Diesel
     - 1.5.3. Gas/Wasserstoff
     - 1.5.4. Hochleistungs-Kraftstoffe
     - 1.5.5. Motoröle/Schmierstoffe
     - 1.5.6. Range-/Programmwerbung
     - 1.5.7. Sponsoring/Image-/Firmenwerbung
     - 1.5.8. Tankkarten/Bonuskarten
     - 1.5.9. Tankstelle/Shop/Service
     - 1.5.10. Öko Kraftstoffe
   - 1.6. Wohn-/Reisemobile/Zubehör
     - 1.6.1. Caravan/Wohn-/Reisemobil
     - 1.6.2. Range-/Programmwerbung
     - 1.6.3. Sponsoring/Image-/Firmenwerbung
     - 1.6.4. Zubehör
   - 1.7. Zubehör/Teile/Service
     - 1.7.1. Anhänger
     - 1.7.2. Autobatterien
     - 1.7.3. Autopflege/Autolack
     - 1.7.4. Collection/Mode/Accessoires
     - 1.7.5. div. Teile/Zubehör
     - 1.7.6. Kfz Audio-Systeme/Kommunikation
     - 1.7.7. Kfz DVD-Player/Recorder
     - 1.7.8. Kfz Navigationssysteme
     - 1.7.9. Kfz Systemkombinationen/Multifunktionssysteme
     - 1.7.10. Kindersitze
     - 1.7.11. Service/Werkstatt
2. Bekleidung
   - 2.1. Baby-/Kinderkleidung
     - 2.1.1. Bekleidung
     - 2.1.2. Range-/Programmwerbung
     - 2.1.3. Schuhe
     - 2.1.4. Sponsoring/Image-/Firmenwerbung
   - 2.2. Oberbekleidung
     - 2.2.1. Bademoden
     - 2.2.2. Berufsbekleidung
     - 2.2.3. Damenoberbekleidung
     - 2.2.4. Hemden
     - 2.2.5. Herrenoberbekleidung
     - 2.2.6. Jeans
     - 2.2.7. Krawatten
     - 2.2.8. Lederbekleidung/Pelze
     - 2.2.9. Pullover
     - 2.2.10. Range-/Programmwerbung
     - 2.2.11. Sponsoring/Image-/Firmenwerbung
     - 2.2.12. Sport-/Freizeit-/Outdoorkleidung
     - 2.2.13. T-Shirts
   - 2.3. Schuhe
     - 2.3.1. Arbeitsschuhe/Stiefel/sonstige Schuhe
     - 2.3.2. Freizeitschuhe/Sportschuhe
     - 2.3.3. Pflege/Zubehör
     - 2.3.4. Range-/Programmwerbung
     - 2.3.5. Sponsoring/Image-/Firmenwerbung
   - 2.4. Taschen/Koffer/Accessoires
     - 2.4.1. Aktentaschen
     - 2.4.2. Gürtel/Schirme
     - 2.4.3. Handschuhe/Hüte/Mützen
     - 2.4.4. Handtaschen
     - 2.4.5. Koffer/Rucksäcke
     - 2.4.6. Range-/Programmwerbung
     - 2.4.7. Sponsoring/Image-/Firmenwerbung
     - 2.4.8. Tücher/Schals
   - 2.5. Wäsche/Strümpfe
     - 2.5.1. Damenwäsche/Dessous
     - 2.5.2. Herrenwäsche
     - 2.5.3. Nachtwäsche
     - 2.5.4. Range-/Programmwerbung
     - 2.5.5. Sponsoring/Image-/Firmenwerbung
     - 2.5.6. Strümpfe/Strumpfhosen
3. Büro/Telekommunikation/IT
   - 3.1. Büroausstattung (nicht Möbel)
     - 3.1.1. Büroartikel/Papier
     - 3.1.2. Diktiergeräte/Schreibmaschinen
     - 3.1.3. Klebstoffe
     - 3.1.4. Schreib-/Rechengeräte
   - 3.2. Hardware
     - 3.2.1. Computerzubehör/CD-DVD-Rohlinge/Kabel etc.
     - 3.2.2. Desktop Computer/PC
     - 3.2.3. Groß-/EDV-/Spezialrechner
     - 3.2.4. Handheld-PCs (Palms)
     - 3.2.5. Mikroelektronik/Prozessoren
     - 3.2.6. Monitore/Beamer/Projektoren
     - 3.2.7. Notebooks/Pocket-PCs
     - 3.2.8. Range-/Programmwerbung
     - 3.2.9. Router/Modem
     - 3.2.10. Scanner/Kopierer/Drucker
     - 3.2.11. Server
     - 3.2.12. Sponsoring/Image-/Firmenwerbung
   - 3.3. Kommunikationsgeräte
     - 3.3.1. Faxgeräte
     - 3.3.2. Funkgeräte/-sender/Systeme
     - 3.3.3. Handys/Fotohandys/Zubehör
     - 3.3.4. Range-/Programmwerbung
     - 3.3.5. sonstige Telekommunikationsgeräte/Zubehör
     - 3.3.6. Sponsoring/Image-/Firmenwerbung
     - 3.3.7. Telefon/-Anlagen
   - 3.4. Software
     - 3.4.1. Betriebssysteme
     - 3.4.2. Datenbanken (nicht Online Dienste)
     - 3.4.3. Entwicklungs-/Programmierungssoftware
     - 3.4.4. Grafik-/Video-/Musiksoftware
     - 3.4.5. Lernsoftware
     - 3.4.6. Office-/Finanzen-/Buchhaltungssoftware
     - 3.4.7. Professionelle Unternehmenssoftware
     - 3.4.8. Range-/Programmwerbung
     - 3.4.9. sonstige Anwendungssoftware
     - 3.4.10. Sponsoring/Image-/Firmenwerbung
     - 3.4.11. Vernetzungssoftware
     - 3.4.12. Viren-/Datenschutzsoftware
   - 3.5. Telekommunikation
     - 3.5.1. Calling Cards/Telefonkarten
     - 3.5.2. Festnetz/Tarife
     - 3.5.3. Internet-Provider Festnetz/DSL/Tarife
     - 3.5.4. Internet-Provider Mobilfunknetz/UMTS/Tarife
     - 3.5.5. IPTV/Triple-Play-Angebote(IP-basiert Telefon/TV/Internet)
     - 3.5.6. Mobilfunk/Tarife
     - 3.5.7. Range-/Programmwerbung
     - 3.5.8. Sponsoring/Image-/Firmenwerbung
     - 3.5.9. VoIP/Internettelefonie
     - 3.5.10. Vorwahlen/spezielle Festnetztarife
4. Finanzen/Versicherungen
   - 4.1. Geldinstitute/Kapitalanlagen
     - 4.1.1. Banken/Sparkassen
     - 4.1.2. Bausparkassen/Baufinanzierung
     - 4.1.3. Börse/Börsengang/Emissionen
     - 4.1.4. Finanz-/Anlage-/Vermögensberatung/Makler/Service
     - 4.1.5. Förder-/Landes-/Bundes-/Zentralbanken
     - 4.1.6. Inkasso Gesellschaften/Pfand-/Leihhäuser
     - 4.1.7. Investment/Sparen/Aktien/Fonds/Zertifikate/Immobilien
     - 4.1.8. Kredit-/Kunden-/Rabattkarten
     - 4.1.9. Online Banking/Online Brokerage
     - 4.1.10. Range-/Programmwerbung
     - 4.1.11. Sponsoring/Image-/Firmenwerbung
   - 4.2. Leasing/Finanzierung
     - 4.2.1. Leasing/Finanzierung Kfz
     - 4.2.2. sonstiges Leasing/Finanzierung(kein Kfz)
   - 4.3. Versicherung
     - 4.3.1. Alters-/Renten-/Rentenzusatzversicherung
     - 4.3.2. Berufsunfähigkeit/Unfallversicherung
     - 4.3.3. Kfz-Versicherung
     - 4.3.4. Krankenkasse/-versicherung gesetzlich
     - 4.3.5. Krankenkasse/-versicherung privat
     - 4.3.6. Krankenkasse/Zusatzversicherungen
     - 4.3.7. Lebensversicherung
     - 4.3.8. Range-/Programmwerbung
     - 4.3.9. Rechtschutzversicherung
     - 4.3.10. Sach-/Haftpflichtversicherung
     - 4.3.11. sonstige Versicherungen
     - 4.3.12. Sponsoring/Image-/Firmenwerbung
     - 4.3.13. Versicherungsagenturen
5. Food
   - 5.1. Baby-/Kleinkindernahrung
     - 5.1.1. Baby-/Kleinkinderkost
     - 5.1.2. Bio-/Ökoprodukte
     - 5.1.3. Range-/Programmwerbung
     - 5.1.4. Sponsoring/Image-/Firmenwerbung
   - 5.2. Brot/Kuchen/Teigwaren
     - 5.2.1. Backmischungen/-zutaten
     - 5.2.2. Brot/Brötchen/Toast/Aufbackbrot
     - 5.2.3. Kekse/Gebäck
     - 5.2.4. Knäckebrot/Zwieback
     - 5.2.5. Kuchen/Torten
     - 5.2.6. Nudeln/Pasta/Teigwaren
     - 5.2.7. Range-/Programmwerbung
     - 5.2.8. Sponsoring/Image-/Firmenwerbung
   - 5.3. Eiscreme/Dessert
     - 5.3.1. Eis am Stiel
     - 5.3.2. Eispackungen/Eiswürfel
     - 5.3.3. Pudding/Puddingpulver
     - 5.3.4. Range-/Programmwerbung
     - 5.3.5. sonstige Desserts
     - 5.3.6. Sponsoring/Image-/Firmenwerbung
   - 5.4. Fertiggerichte
     - 5.4.1. Fertiggerichte (nicht Tiefkühlkost)
     - 5.4.2. Range-/Programmwerbung
     - 5.4.3. Sponsoring/Image-/Firmenwerbung
   - 5.5. Fette/Öle
     - 5.5.1. Bio-/Ökoprodukte
     - 5.5.2. Butter
     - 5.5.3. Margarine
     - 5.5.4. Pflanzenfette/Speiseöle
     - 5.5.5. Range-/Programmwerbung
     - 5.5.6. Sponsoring/Image-/Firmenwerbung
   - 5.6. Frische Nahrungsmittel
     - 5.6.1. Bio-/Ökoprodukte
     - 5.6.2. Fisch
     - 5.6.3. Fleisch/Wurst
     - 5.6.4. Geflügel/Eier
     - 5.6.5. Obst/Gemüse/Kartoffeln
     - 5.6.6. Range-/Programmwerbung
     - 5.6.7. Sponsoring/Image-/Firmenwerbung
   - 5.7. Gewürze/Suppen/Saucen
     - 5.7.1. Gewürze/Flüssiggewürze/Salz
     - 5.7.2. Range-/Programmwerbung
     - 5.7.3. Saucen
     - 5.7.4. Senf/Mayonnaise/Ketchup
     - 5.7.5. Sponsoring/Image-/Firmenwerbung
     - 5.7.6. Suppen
   - 5.8. Konserven
     - 5.8.1. Fleisch-/Wurst-/Fisch-/Geflügelkonserven
     - 5.8.2. Obst-/Gemüsekonserven
     - 5.8.3. Range-/Programmwerbung
     - 5.8.4. Sponsoring/Image-/Firmenwerbung
   - 5.9. Milchprodukte
     - 5.9.1. Bio-/Ökoprodukte
     - 5.9.2. Crème fraiche
     - 5.9.3. Dosenmilch/Sahne
     - 5.9.4. Joghurt
     - 5.9.5. Käse
     - 5.9.6. Quark
     - 5.9.7. Range-/Programmwerbung
     - 5.9.8. Sponsoring/Image-/Firmenwerbung
     - 5.9.9. Milchreis / Grießbrei
   - 5.10. Nährmittel
     - 5.10.1. Bio-/Ökoprodukte
     - 5.10.2. Diät-/Schlankheitskost/Nahrungszusätze
     - 5.10.3. Feinkost/Salate
     - 5.10.4. Getreideerzeugnisse/Mehl/Müsli
     - 5.10.5. Kartoffel-Erzeugnisse
     - 5.10.6. Konfitüre/Honig/Brotaufstrich
     - 5.10.7. Range-/Programmwerbung
     - 5.10.8. Reis/Hülsenfrüchte
     - 5.10.9. Sponsoring/Image-/Firmenwerbung

AdVision digital GmbH ▪ Wandalenweg 30 ▪ 20097 Hamburg
Tel.: +49 (40) 24 42 48 -0 ▪ Fax: +49 (40) 24 42 48 -48 ▪ E-Mail: service@advision-digital.de

Stand: 04.07.2008