

Ausarbeitung der Prüfungsfragen

Version 1.0

Robert Peloschek
aka MacOS X

weitbearbeitet von
Patrick Kastner [huevo]

Kapitel 1 – Sprache

Was beschreibt das Semiotische Dreieck?

Das semiotische Dreieck beschreibt den Bezug zwischen Worten und bezeichneten Objekten.

Es ist deshalb so wichtig, weil wir dazu tendieren, die „drei Eckpunkte“ ineinander verschmelzen zu lassen. Worte, Begriffe und Objekte sind aber nicht das gleiche. Wenn wir über etwas reden, dann verschmelzen der Gegenstand, über den wir reden, das Wort für diesen Gegenstand und der zugehörige Begriff in unserem Denken miteinander. Dies ist das Ziel, denn in einem solchen Moment erscheint uns der Gedanke klar. Die Bedingung dafür ist, dass wir als erfolgreiche Akteure des Denkens die Begriffe genau auf die Gegenstände gerichtet haben und unsere Worte durch vieles Üben zuverlässig mit den Begriffen verknüpft sind. Das ist das Ziel des Denkens und Sprechens. Bis wir jedoch dorthin kommen, ist es ein langer Weg.

Wenn wir den Weg beschreiben wollen, und das müssen wir, wenn wir Denkprozesse verstehen und nachbilden wollen, dann müssen wir das dynamische Zusammenspiel von Begriffen, Gegenständen und Worten verstehen. Dazu ist es äusserst hilfreich, sich die Unterschiede zwischen den drei Eckpunkten des Dreiecks klar zu machen.

Die drei „Eckpunkte“ des semiotischen Dreiecks sind:

Concept = signifier = Signifikans = Symbol:

das Bedeutende (verweist auf anderes); kollektives Lautbild; Zeichen im engeren Sinn; materielles Zeichen

Beispiel: das Lautbild „Tisch“: /tiːs/

Sign = signified = Signifikat = Gedanke/Bezug:

das Bedeutete (auf dieses wird verwiesen); Begriff; Vorstellung, die vom Lautbild ausgelöst wird

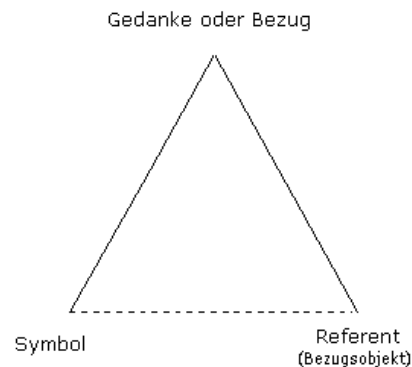
Beispiel: Das Lautbild „Tisch“ löst die Vorstellung von einem Möbelstück aus Holz mit 4 Beinen aus -> Vorstellung ist abhängig von lebensgeschichtlicher Erfahrung (denn bei fast allen Menschen wird ähnliches ausgelöst). Das

Lautbild verweist also auf die Vorstellung im Kopf.

Object = Referent:

das Bezeichnete; das "reale" Objekt in der äußeren Welt

Beispiel: der konkrete wirkliche Tisch



Was bedeuten die Begriffe Konnotation und Denotation?

Die beiden Begriffe dienen der Unterscheidung von Haupt- und Nebenbedeutung eines Zeichens.

Jedes Wort hat eine Grundbedeutung, die man auch *Denotation* nennt. Unter bestimmten Bedingungen können Wörter neben ihrer Grundbedeutung noch zusätzliche Bedeutungen - auch Nebenbedeutungen - haben, die *Konnotationen* heißen. Konnotationen entstehen durch Assoziation (z. B. bei "Mond" > "Nacht", "kühl", "romantisch") beim Leser. Man möchte meinen, die Konnotation hätte somit einen völlig subjektiven Charakter, der es nur noch erschwert, einen Text zu interpretieren. Ganz so ist jedoch nicht. Manche Wörter sind schon im Voraus mit Konnotationen durch die Gesellschaft behaftet. Außerdem kommt hinzu, dass bei der Deutung einzelner Wörter der Text als ganze Einheit berücksichtigt werden muss.

Die Konnotationen ergeben sich also folglich auch aus den Strukturen und Zusammenhängen des Textes, was noch einmal deutlich macht, dass Konnotationen nicht nur durch das Gefühl eines einzelnen Lesers entstehen.

Zusammenfassend heißt das:

Denotation =

- Die wörtliche Bedeutung, wie sie in einem Lexikon vermerkt ist.
- Eine Beziehung erster Ordnung zwischen „signifier“ und „signified“.

Konnotation =

- Zusätzliche (emotionale, stilistische, expressive) Vorstellungen.
- Eine Beziehung zweiter Ordnung zwischen dem Paar „signifier“ – „signified“ einerseits und einem neuen „signified“.

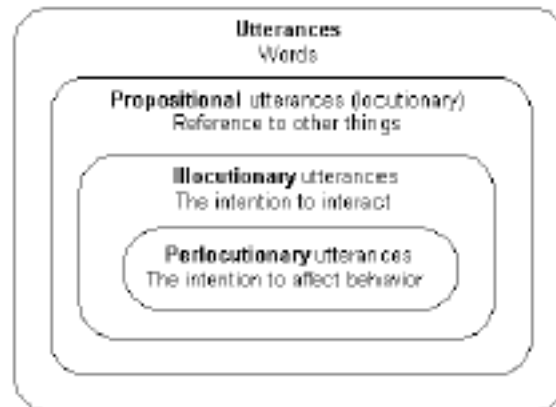
Ist Sprache redundant? Wenn ja, wodurch?

Auf der Ebene von Wörtern ist Sprache redundant.

- Bei > 1.000.000 Wörtern 20 bit = 4 Buchstaben / Wort
- Aber: stark ungleich verteilt und voneinander abhängig

Beschreibe die verschiedenen Ebenen eines Sprechakts.

- Utterance Act
Das Äußern von Lauten, Morphen, Wörtern, Phrasen, ...
- Locutionary (propositional) Act
Referenzieren, Zuweisen von Prädikaten, ...
- Illocutionary Act
Feststellen, Fragen, Befehlen, Versprechen, ...
- Perlocutionary Act
Der intendierte Effekt einer Äußerung.



Welche Kategorien von Sprechakten gibt es?

- Representatives
z.B. Feststellungen, Schlüsse
- Directives
z.B. Bitten, Fragen
- Commissives
z.B. Versprechen, Drohungen, Angebote
- Expressives
z.B. Dank, Entschuldigung, Begrüßung, Gratulation
- Declarations:
z.B. Verheiraten, Taufen, Kündigen, Erklärungen

Was ist ein indirekter Sprechakt? Gib ein (eigenes) Beispiel.

Ein weit verbreiteter indirekter Sprechakt ist, Aufforderungen als Fragen zu formulieren.

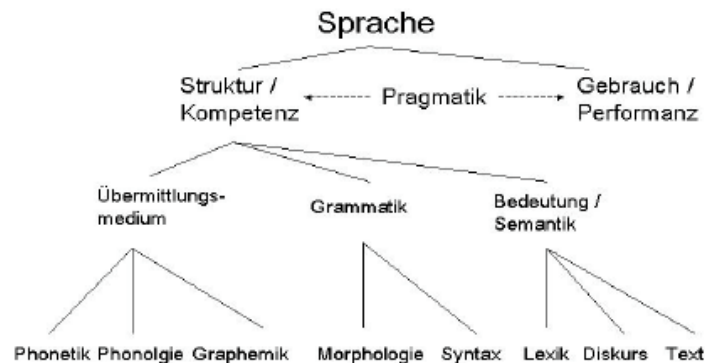
Kontext: Passanten auf der Straße. Wissen sie vielleicht wie spät es ist?
Antwort: Ja, es ist halb zwei.

Die Ebenen der Sprache

- Phonetik
Beschreibung und Klassifikation von Lauten
- Phonologie
Beschreibung des Lautsystems der Sprache
- Graphetik
- Morphologie
Bildung und Struktur von Wörtern

- Syntax
Bildung und Struktur von Sätzen
- Semantik
Bedeutung sprachlicher Äußerungen (kontextfrei)
- Pragmatik
Intention sprachlicher Äußerungen (kontextabhängig)

Am Besten lassen sich die Ebenen der Sprache wohl mit folgendem Baum darstellen:



Was ist Phonetik (bzw. Phonologie, Morphologie, Syntax, Semantik, Pragmatik)?

- Phonetik
Beschäftigt sich mit der Beschreibung und Klassifikation von Lauten; akustisch; analog; segmentale Einheit ist das Phon;
- Phonologie
Beschäftigt sich mit der Beschreibung des Lautsystems von Sprachen; abstract; symbolisch; kleinste bedeutungsunterscheidende Einheit ist das Phonem;
- Morphologie
Beschäftigt sich mit der Struktur und dem Aufbau von Wörtern; kleinste bedeutungstragende Einheit: das Morphem; Morpheme werden durch Morphe realisiert: z.B. sagst = sag + st;
- Syntax
Beschäftigt sich mit der Struktur von Sätzen; "formale Grammatikalitätsbedingungen"; beschrieben durch formale Grammatik;
- Semantik
Beschäftigt sich mit der Bedeutung (dem Sinn) von sprachlichen Äußerungen; unabhängig von Intention und Kontext;
- Pragmatik
Beschäftigt sich mit dem Gebrauch von sprachlichen Äußerungen (kontextabhängig); stellt die Verbindung her zwischen Sprache:
 - als abstraktes Symbolmanipulationssystem und
 - als Kommunikations- und Informationsmedium

Wodurch sind Laute (Phone) definiert?

Phon = die kleinste bedeutungsunterscheidende, aber nicht bedeutungstragende sprachliche Einheit.

Was sind wichtige prosodische Parameter?

- Tonhöhe (Sprachgrundfrequenz)
- Lautheit (Intensität)
- Quantität (Dauer)

Welche Funktion hat Intonation in der Sprache?

Intonation = Sprachmelodie

Funktionen:

- Betonung
- Phrasierung
- Satzmodusmarkierung

Was ist die internationale Lautschrift?

Ist ein Schriftsystem, das für jeden in der menschlichen Sprachen auftretenden Laut ein Zeichen hat. Die Einteilung erfolgt nach Vokalen, Konsonanten, sowie weiteren Lauten (Klicks, Implosives, Ejektive). Daneben umfasst sie auch Zeichen für Töne und für bestimmte suprasegmentale Elemente.

→ Es können mit der internationalen Lautschrift alle Sprachen der Welt (nach ihrem Lautgehalt) geschrieben werden.

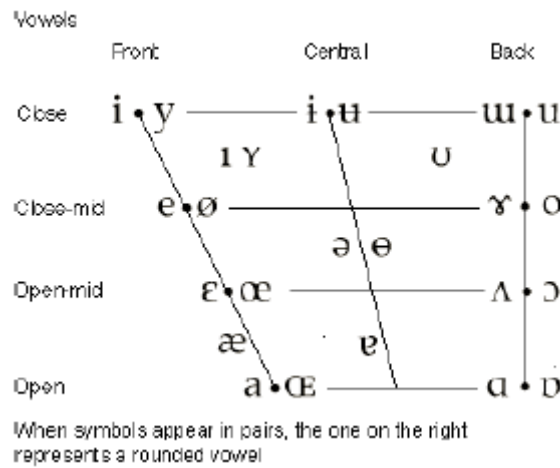
Entwickelt und verwaltet wird die Lautschrift von der International Phonetic Association.

- Die Konsonanten werden in der IPA nach Artikulationsort und Artikulationsart klassifiziert.

Consonants

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			
Ejective stop	pʼ			tʼ		tʼ	cʼ	kʼ	qʼ		
Implosive	ɓ ɗ			ɟ ɗ			ɕ ɟ	ʞ ɟ	ɕ ɟ		

- Die Einzelnen Vokale sind in verschiedenen Sprachen oft ziemlich unterschiedlich realisiert. Konstant ist allerdings ihre relative Position zueinander innerhalb einer Sprache.



Welche Faktoren beeinflussen den Sprechstil?

- Physiologisch
Stimmlage, Stimmqualität
- Emotionen
Freude, Trauer, Zorn, Langeweile, ...
- Sprechergruppen
Dialekt, Soziolekt (kulturell), Ideolekt (personenspezifisch)
- Ausdruck
Sprechstile: flüstern, schreien, singen, ...
Sprachstile: formal, informell, spontan, ...

Welche Faktoren beeinflussen die Stimmlage?

- Geschlecht
- Alter
- Erkrankungen (z.B. der Atemwege)

Welche Stimmqualitäten kennst Du?

Hauchen, Flüstern, Krächzen, Rauheit, Schreien, ...

Beschreibe eine Stimmqualität.

Hauchen: Große Glottisöffnung, hoher Luftdurchfluss

Welche akustischen Faktoren beeinflussen die emotionale Färbung von Sprache?

- Bandbreite der Intonation
- Dynamik der Intonation
- Durchschnittliche Höhe der Sprachgrundfrequenz
- Höhenverlauf der Sprachgrundfrequenz (flach vs. Absinkend)
- Sprechrate, Anzahl und Länge von Pausen
- Artikulation (deutlich vs. Undeutlich)

Welche akustischen Korrelate kennzeichnen Ärger (Furcht, Trauer)?

	Ärger	Furcht	Trauer
Durchschnitt f_0	Hoch	Sehr hoch	Niedrig

<i>Bandbreite f0</i>	Weit	Sehr weit	Eng
<i>Verlauf f0</i>	Fallend		Leicht fallend
<i>Dynamik f0</i>	Groß		
<i>Pausen</i>	Wenige	Wenige	Viele lange
<i>Sprechrate</i>	190 wpm	202 wpm	129 wpm
<i>Anderes</i>			vibrato

Welche Sprachstile kennst Du und wodurch sind sie gekennzeichnet?

- Vorlesen
 - Relativ nahe an geschriebener Sprache
 - Eher neutrale Prosodie
- Präsentation, Vortrag (formal)
 - Im Charakter zwischen Vorlesen und Spontansprache
- Spontansprache (informell)
 - Geprägt durch Grammatikfehler, Hesitationen, Wiederholungen, ...
 - Natürliche, meist lebhaft Prosodie

Welche Ansätze für Schriftsysteme gibt es?

- Buchstabenschrift (typischerweise 20-30 Zeichen)
- Silbenschrift (typischerweise 45-60 Zeichen)
- Zeichenschrift (typischerweise tausende Zeichen)

Was zeichnet Buchstabenschrift (bzw. Silbenschrift, Zeichenschrift) aus?

Buchstabenschrift:

- Zeichen entsprechen (Idealerweise) den Lauten der damit geschriebenen Sprache
- Sprache ändert sich oft schneller als die Orthographie (Extrembeispiel: Englisch)
- Buchstabenschrift wird den unterschiedlichen Lautsystemen der Sprachen angepasst
 - Durch neuen Lautwert, z.B. „c“ (cento (italienisch), cell (englisch))
 - Durch neue Buchstaben bzw. die Adaption vorhandener durch so genannte diakritische Zeichen und Buchstabenkombinationen (z.B. ä, ü, ö, ß, ...)
- In unterschiedlichen Kulturkreisen wurden unterschiedliche Alphabete entwickelt
- Es entwickelten sich zwei unterschiedliche Systeme
 - In manchen Schriften werden alle Laute geschrieben
 - In anderen nur Konstanten und Vokale

Silbenschrift:

- Die Zeichen entsprechen (Idealerweise) den Silben der damit geschriebenen Sprache
 - Werden für Sprachen verwendet, die nur über offene Silben (Konstant – Vokal) verfügen
- Reine Silbenschriften verfügen über ein eigenes Zeichen für jede Silbe

Zeichenschrift:

- Zeichen entsprechen folgenden Einheiten:
 - Logogramme: Bezeichnen Wörter oder Wortteile
 - Manchmal sind solche Zeichen quasi Bilder des dargestellten Wortes: Dann spricht man von Piktogrammen
 - Ideogramme: Bezeichnen abstrakte Begriffe
 - Semantisch-phonetische Zeichen:
 - Verbinden ein semantisches Element mit Hinweisen auf die Aussprache
 - Manche dieser Zeichen stehen nur noch für die entsprechende Aussage

Was bedeuten die Begriffe Referenz und Deixis?

Referenz:

Sprachliche Äußerungen stehen in Bezug zur Außenwelt. Sprachliche Einheiten referieren daher auf Elemente dieser Außenwelt. Jede sprachliche Referenz wird aber erst in ihrem jeweiligen Kontext eindeutig!

Realisierung:

- „Absolute“ Referenz
 - Eigennamen (z.B. George Bush, Berlin)
 - Nominalgruppen (z.B. der Präsident der USA, die deutsche Hauptstadt, die Sonne)
- „Relative“ Referenz = Deixis

Deixis:

Sinn und Zweck: Sprachliche Mittel, Äußerungen und ihre Teilnehmer (Sprecher, Hörer und erzählte Teilnehmer) räumlich und örtlich zu lokalisieren. Im Gegensatz zu absoluten Referenzen ist Deixis immer relativ.

Vgl: Örtliche und zeitliche Deixis.

- Örtliche Deixis lokalisiert die Elemente einer Äußerung im Raum
- Zeitliche Deixis lokalisiert die Elemente einer Äußerung in der Zeit

Welche nichtsprachlichen Mittel werden im Dialog verwendet?

- Blick
- Gestik
 - Zeigegesten, Gestikulation, ikonische Gesten
- Mimik
 - Augenbrauen, Augenlider, Lippen, Nase
- Körperhaltung

Was sind Wortklassen (PoS-Kategorien)? Gib Beispiele.,

Wortklassen (Part of Speech)

Nomina: alle deklinierenden Wörter

Substantiv (Sessel, Kranker)

Adjektiv (groß, rosa, krank)

Artikel (der, ein)
 Numeral (zehn, zehnter)
 Pronomen (sie, mein, wer, der)
 Verben: alle konjugierenden Wörter
 Verb (fragen, rasieren)
 - Modalverb (wollen, mögen)
 - Auxiliar (haben, sein, werden)
 Partikel: alle nicht flektierenden Wörter
 Adverb (unten, heute)
 Präposition (bei, in, entlang)
 Konjunktion (und, weil)
 Partikel (wohl, nicht)
 Interjektion (bitteschön, plumps)

• Welche Funktion hat Wortakzent in der Sprache?

Der Wortakzent liegt auf einer Silbe. Es gibt zwei Arten der Position: Der Wortakzent steht für das gesamte Vokabular fest, so zum Beispiel auf der

- ersten Silbe (Tschechisch, Ungarisch, Finnisch)
- vorletzten Silbe (Polnisch)
- letzten Silbe (Französisch).

Zweite Möglichkeit: Der Wortakzent ist beweglich (lexikalisch bestimmt)

- Deutsch (umfahren - umfahren)
- Englisch (import - import)
- Spanisch (termino - termino)

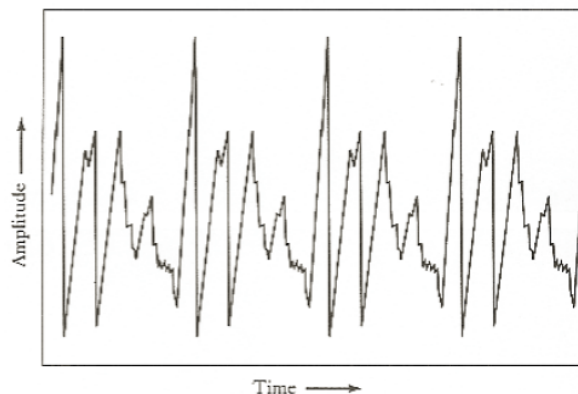
Die Betonung ist (meist) nur auf einem Wortakzent möglich.

Kapitel 2 – Sprache als Signal

Wodurch unterscheiden sich periodische von aperiodischen Schallwellen?

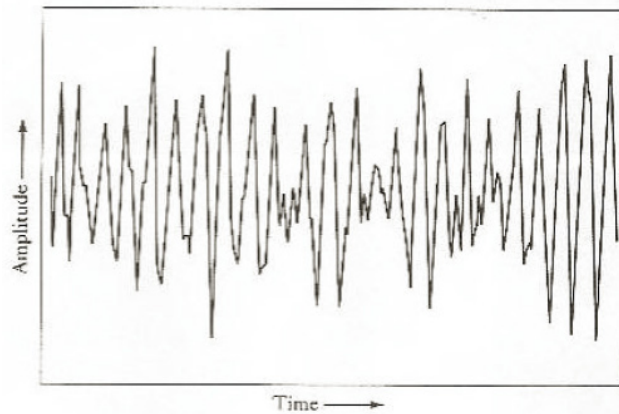
Periodische Schallwellen:

Periodische Schallwellen haben eine bestimmbare Vollschrwingung, die sich ständig (=periodisch) in einer bestimmbaren Frequenz wiederholt. Beispiele für periodische Schallwellen sind Sinus-Schwingungen oder Klänge (Klang = Sinusschwingung + deren Obertöne (Quint + Oktav)). Sie haben eine bestimmbare (durch die Frequenz bestimmte) Tonhöhe. Vokale sind Klänge.



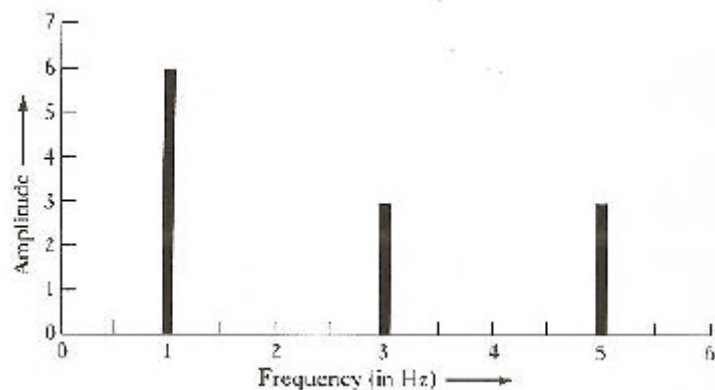
Aperiodische Schallwellen:

Aperiodische Schallwellen haben all diese Eigenschaften nicht und werden daher als Geräusch, Rauschen usw. wahrgenommen. Konsonanten sind zum Beispiel aperiodische Schallwellen.



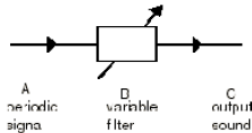
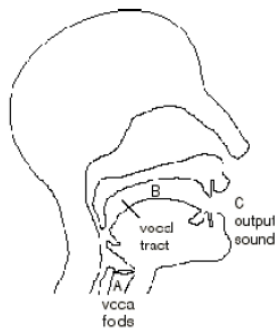
Was ist ein Spektrogramm?

Das Spektrogramm zerlegt ein Sprachsignal in seine (periodischen) Komponenten. Im Gegensatz zum Zeitsignal ist diese Zerlegung bei der Phasenverschiebung konstant.



Beschreibe das Quelle-Filter-Modell der Sprachproduktion.

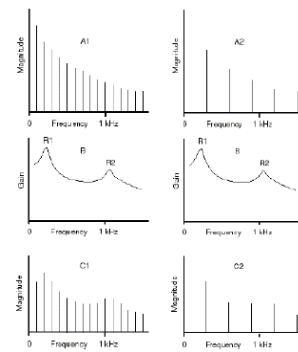
- Quelle
 - Die Glottis produziert Schwingungen (Schall)
 - daraus ergibt sich die Sprachgrundfrequenz (f_0)
 - die Schallwelle setzt sich aus f_0 und Vielfachen von f_0 (Obertönen) zusammen.
- Filter
 - Der Vokaltrakt als Resonanzraum dämpft bzw. verstärkt bestimmte Frequenzen
 - durch seine dynamische Veränderung entstehen die einzelnen Laute
 - durch Engstellen im Vokaltrakt entsteht zusätzlich Rauschen (nichtperiodische Schwingungen)



• Quelle

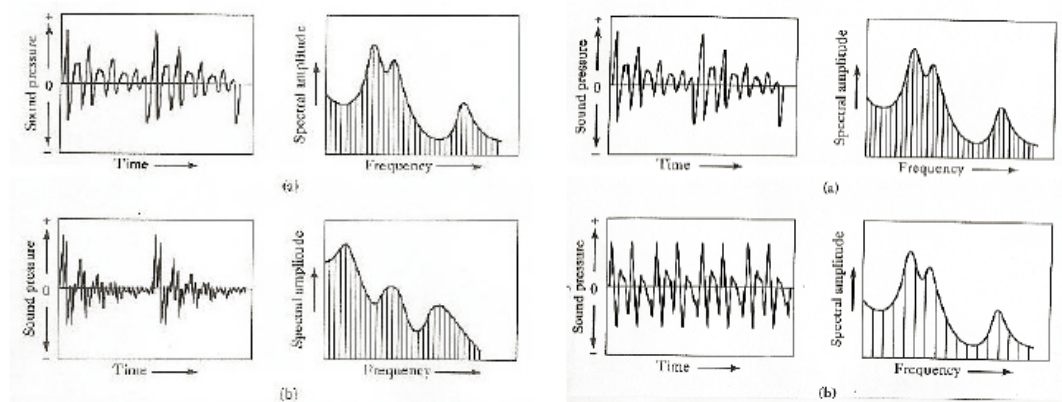
• Filter

• produzierter Schall



Was sind Formanten?

Vom Grundton unabhängige Bereiche im Klangspektrum, in denen die Teiltöne eine besonders starke Amplitude aufweisen. Jeder Vokal hat charakteristische Formanten. So gelten bestimmte Richtwerte für die Formanten der unterschiedlichen Vokale (a, e, i, o, u, ä, ö, ü).



Kurzum:

Formanten sind unabhängig von der Sprachgrundfrequenz → daher sind sie das wesentliche akustische Merkmal.

Wodurch entstehen Formanten?

Formanten entstehen infolge selektiver Verstärkung bestimmter Frequenzbereiche, entsprechend der Resonanzcharakteristika des menschlichen Sprachapparates.

Was ist die Abtastfrequenz?

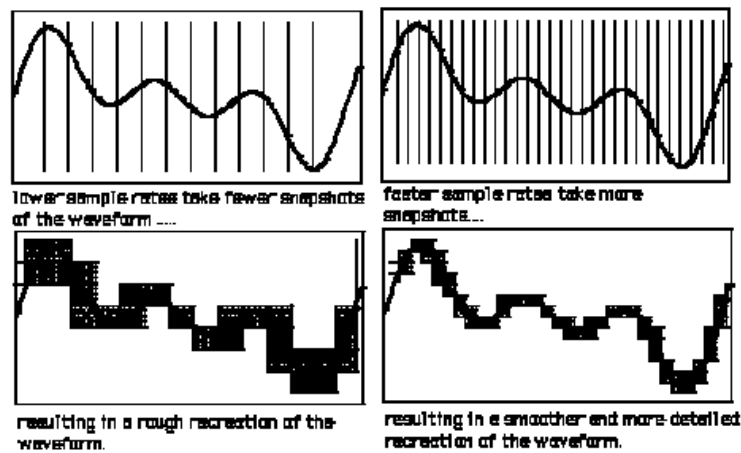
Mit Abtastrate bezeichnet man in der Physik die Rate, mit der Signalwerte aus einem kontinuierlichen Signal entnommen werden.

Wenn man ein analoges Signal in ein digitales Signal umwandeln möchte, kann man die dazu notwendige Abtastrate durch das Abtasttheorem von Shannon/Nyquist bestimmen.

Dabei gilt: Die Abtastrate muss mehr als doppelt so hoch wie die höchste im Signal vorhandene Frequenz sein.

$$f_{\text{abtast}} > 2 \cdot f_{\text{max}}$$

Je höher die Abtastrate (Abtastfrequenz / Samplingrate) ist, desto geringer ist der Fehler, der bei der Digitalisierung entsteht.



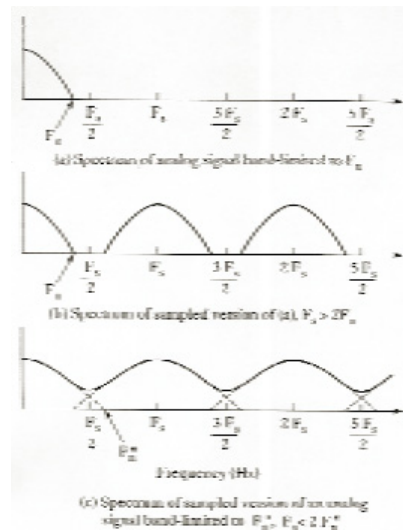
Was ist die Nyquist-Rate?

$$f_{\text{abstast}} > 2 \cdot f_{\text{max}} = \text{die Nyquist-Rate}$$

Es gilt: Die Abtastrate muss mehr als doppelt so hoch wie die höchste im Signal vorhandene Frequenz sein.

Wird die oben genannte Formel nicht eingehalten und f_{abstast} zu klein gewählt, kommt es zu sogenanntem „Aliasing“ (Verzerrungen).

Bei begrenztem Frequenzbereich kann ein Analogsignal daher verlustfrei digitalisiert werden.



Was bedeutet Bandbegrenzung?

Die Bandbegrenzung beschreibt den wahrnehmbaren Frequenzbereich eines Sprach/Ton Signals.

Beispiele:

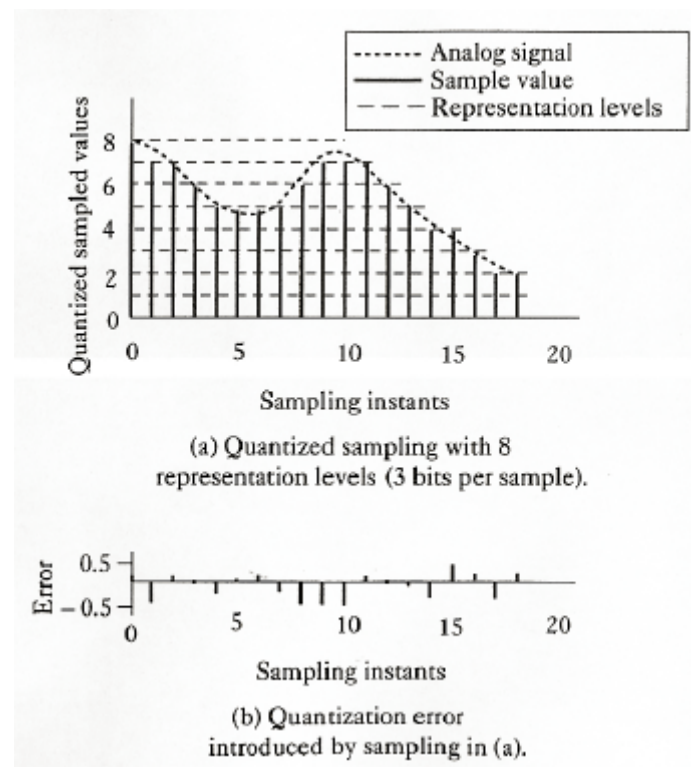
- menschliches Ohr < 16-20 kHz
- Radio (UKW) < 15 kHz
- Telefon: < 4 kHz
- Für Verstehen von Sprach nötig: < 3 kHz

Was bedeutet Quantisierung?

Werte werden mit fixer Genauigkeit abgespeichert.

z.B.: 8bit für Telefonie
16bit für hochqualitative Anwendungen

Zu beachten ist dabei der so genannte Quantisierungsfehler der durch das Mapping von analogen auf diskrete digitale Werte entsteht. Je „feiner“ quantisiert wird, desto geringer ist der Quantisierungsfehler. (siehe Abbildung)



Was ist ein Character Code?

Trivial ausgedrückt bildet ein Character Code Zeichen auf Zahlenkombinationen ab.

Geschichtliches:

Ursprünglich wurden im Computer nur Zahlen verwendet. Mit größeren Speicherkapazitäten und besserer Rechnerleistung wurde auch Text ein Thema. Zunächst wurden nur (englische) Großbuchstaben und einige wenige Sonderzeichen codiert, d.h. auf Zahlenkombinationen abgebildet.

Sehr bald wurde es notwendig:

- die Menge der verfügbaren Buchstaben zu erweitern
- die Codierung zu standardisieren

-> die Entwicklung und Standardisierung von unterschiedlichsten Charakter Codes

Warum sollte Schrift am Computer nicht als Bild dargestellt werden?

Weil der Text als solcher dann nicht weiterbearbeitet werden kann. So wäre zum Beispiel folgendes nicht möglich:

- Suche und Indizierung
- Reformatieren
- Editieren
- Vorlesen
- Etc.

Was ist der ASCII-Code?

Der American Standard Code for Information Interchange wurde 1967 als ANSI X3.4 veröffentlicht. Er ist ein 7-bit Code. Die Bereiche 0-1F und F7 sind für Steuerzeichen reserviert (non printable).

Sprachen die sich damit gänzlich darstellen lassen:
Englisch, Latein, Hawaiianisch, Suaheli

Was ist ISO-8859?

Ist ein 8-bit Charakter Code. Da er eine Erweiterung des ISO-646 ist, ist er zum ISO-646US voll kompatibel. Der weitere Platz wird verwendet, um weitere Zeichen für Gruppen von Sprachen zu kodieren.

Derzeit gibt es mehr als 10 Varianten vom ISO-8859.

z.B.

ISO-8859-1 (Latin 1) für Französisch, Spanisch, Katalanisch, ...

ISO-8859-5 (Cyrillic) für Bulgarisch, Makedonisch, Russisch, ...

Kann man mit einer Version von ISO-8859 auch mehrsprachige Dokumente verfassen?

Ja, aber nur solange man innerhalb einer „Variante“ bleibt. Es gibt in ISO-8859 übrigens über 10 verschiedene Varianten.

z.B.

ISO-8859-1 (Latin 1) für Französisch, Spanisch, Katalanisch, ...

ISO-8859-5 (Cyrillic) für Bulgarisch, Makedonisch, Russisch, ...

Was ist UNICODE?

Wurde entwickelt als 16-bit Code. Das heißt es sind mit UNICODE theoretisch 65.000 Zeichen darstellbar. Derzeit werden aber nur ca. 58.000 genutzt.

Stellt die „basic multilingual plane“ (BMP) dar.

Unicode hat folgende Eigenschaften:

- Jedes Zeichen hat
 - eine eindeutig zugeordnete Zahl
 - einen eigenen (eindeutigen) Namen
 - weitere Informationen
 - normative Eigenschaften (Typ, groß/klein, Wert, etc.)

- zusätzliche Informationen
- Zuordnung der Zeichen nach Schrifttyp (Latein, Kyrillisch, Griechisch, Han, etc.) *nicht* nach der Sprache
- Unterscheidung von Sonderzeichen nach Funktion z.B. “-” als Minus, Bindestrich, Trennzeichen

Es gibt Unicode in mehreren Codierungsformen:

- UTF-8 (8 bit)
- UTF-16 (16 bit)
- UTF-32 (32 bit)

Welche Codierungsformen für UNICODE gibt es?

- UTF-8
Darstellung der Zeichen in Variabler Länge (siehe nächste Frage)
- UTF-16 (16 bit)
Verwendet 16 bit Wörter, Zeichen außerhalb der BMP durch Doppelwort dargestellt
- UTF-32 (32 bit)
Darstellung als 32 bit Wort

Wie wird UNICODE in UTF-8 codiert?

Bei dieser Codierungsform werden die einzelnen Zeichen in variabler Länge dargestellt.

- ASCII Zeichen werden in einem Byte dargestellt wobei das erste Bit 0 ist → 0xxxxxxx
- Alle weiteren Zeichen werden in mehreren Bytes dargestellt. Die ersten Bits sind als Kennzeichnung für die Byte-Folge reserviert.
 - 2 Bytes: 110xxxxx 10xxxxxx
 - 3 Bytes: 1110xxxx 110xxxxx 10xxxxxx
 - 4 Bytes: 11110xxx 1110xxxx 110xxxxx 10xxxxxx
 - Usw.

Wie werden Zeichen in HTML intern dargestellt?

- Character Code
 - Derzeit verwendet HTML ISO-8859-1 (Latin1)
 - Die Umstellung auf UNICODE (UTF-8) ist geplant
- Interne Darstellung von Zeichen
 - ASCII-Zeichen werden als solche dargestellt
 - Jedem Nicht-ASCII Zeichen ist eine dreistellige Zahl zugewiesen
 - z.B. £ für £
 - daneben hat zumindest ein Teil dieser Zeichen auch einen Namen
 - z.B. ö für ö

Was ist Transliteration (bzw. Transkription)?

Dienen der Übertragung einer Sprache von einem Lautsystem in ein anderes.

Es gibt zwei verschiedene Ansätze:

- *Transliteration* bedeutet, dass es für jedes Zeichen der Quellschrift ein(e) Zeichen(kombination) in der Zielschrift gibt.
 - Nachteil: Nur für Buchstabenschriften geeignet
 - Vorteil: Rückübertragbarkeit
- *Transkription* überträgt die Wörter nach ihrem Lautwert. Es gibt praktisch immer mehrere parallele, unterschiedliche Konventionen der Transkription.
 - Sie wird den Lautwerten der Zielsprache angepasst
 - Dialektale Unterschiede in der Aussprache der Quellsprache bedingen unterschiedliche Transkriptionen
 - Es existieren unterschiedliche Systeme mit unterschiedlichem Normierungsgrad

-> Eine Vielfalt von Schreibweisen!

Welche Schriftarten gibt es?

- Serifenschriftarten (Sprache und Multimedia rulez!)
- Serifenlose Schriftarten (Sprache und Multimedia rulez!)
- Schreibschriften (Sprache und Multimedia rulez!)
- Zierschriften (Sprache und Multimedia rulez!)
- Symbolschriften (□□□□□□□□□□□□□□□□□□□□□□□□□□□□)

Was ist eine Serifenschrift?

Serifen = frz. Füßchen

Als Serifen bezeichnet man die geschwungenen und rechteckigen Enden der Striche. Bei Schriftarten findet man sie bei den Kopf- und Fußansätzen eines Buchstabens.



Der Vorteil der Serifenschriften ist, dass sie besser lesbar sind als Serifenlose Schriftarten. Die Serifen helfen somit bei der Worterkennung. Serifenschriften sind daher die beste Wahl für größere Textmengen und die übliche Grundschrift für Publikationen.

Welche Schriftart ist am besten lesbar?

Ganz allgemein gesagt Serifenschriften. Wieso? -> siehe letzte Frage.
Nicht umsonst ist in Word, OpenOffice & Co. als „Default-Schrift“ eine Serifenschrift eingestellt.

Beispiel: Times New Roman

Was ist die ideale Lesebreite?

Die ideale Lesebreite liegt bei 55 Zeichen.

Ganz allgemein sollte die Länge einer Zeile zwischen 45 und 65 Zeichen liegen. Alles was darunter liegt oder darüber hinausgeht führt zu schlechterer Lesbarkeit. Insbesondere bei Blocksatz führen zu kurze Zeilen zu unregelmäßigen Wortabständen, Worttrennungen und zu einer Hemmung des Leseflusses.

Wann und wie soll man Auszeichnungen in Texten verwenden?

Auszeichnungen dienen der *Hervorhebung* von Text. Man sollte sie daher sehr sparsam einsetzen und den Textinhalt- und Zielsetzung berücksichtigen.

Arten von Auszeichnungen (und wann sie eingesetzt werden sollten):

- Kursive Auszeichnung
 - Engl. Italics
 - Wird *langsamer* gelesen
 - Nur für einzelne Worte geeignet
 - Für Hervorhebungen in ruhigen, fließenden Texten
- Fette Auszeichnung
 - Engl. Bold
 - Greift **stärker** in das Schriftbild ein
 - Unruhiges Schriftbild
 - Hemmt das lineare Lesen
 - Gut für Überschriften geeignet
- Wechsel der Schriftart
 - z.B. von serifenlos auf Serifenschrift
 - Auf deutliche Unterschiede der Schriftarten achten
 - Sehr sparsam einsetzen
 - In einem größeren Schriftgrad wechseln: z.B. für Überschriften
- Weitere Formen
 - GROSSBUCHSTABEN (Versalien)
 - Wirkt im laufenden Text unruhig und unausgeglichen
 - KAPITÄLCHEN
 - Ausgeglichenere Wirkung als Großbuchstaben
 - „Anführungszeichen“
 - Sollte zu Auszeichnung wenn möglich nicht verwendet werden.
 - S p e r r e n
 - Sollte zu Auszeichnung wenn möglich nicht verwendet werden.
 - Unterschreiben
 - Sollte zu Auszeichnung wenn möglich nicht verwendet werden.

• Was sind Glyphen?

Das Zeichen (engl. character) ist die abstrakte Idee eines Buchstabens, die Glyphe ist deren konkrete grafische Darstellung. Elektronische Texte wie dieser hier werden als abstrakte Zeichen gespeichert, und ihre Erscheinungsform hängt von der jeweils gewählten Schriftart ab.

Typischerweise kann ein Zeichen durch unterschiedliche Glyphen dargestellt werden.

• Wodurch wird eine Glyph definiert?

Eine Glyph wird durch unterschiedliche

- Schriftarten (am, am, am, **am**, am)
- Schriftschnitte (am, am, **am**, AM)
- Schriftgrößen (am, am, am, am)

definiert.

Kapitel 3 – Basismethoden der Sprachverarbeitung

Was ist eine POS-Kategorie und wie gross etwa ist ein Set für eine bestimmte Sprache?

Die Wörter einer Sprache sind in POS (Part-of-Speech) Kategorien eingeteilt (z.B. Nomen, Verb, Adjektiv, Präposition).

Ein „Tagset“ für eine Sprache besteht dabei meist aus 30 - 40 Tags.

Wortformen können dabei in mehrere Kategorien fallen (z.B. sieben, meinen, ihr, Leiter). Im Deutschen sind fast 30% der Wörter im Text „ambig“.

Stuttgart-Tübingen Tag Set (STTS)

ADJA	attr. Adj.	NE	Name	PTKNEG	Negation
ADJD	präd. A	PDS	Dem.pron.	PTKVZ	Verbzusatz
ADJA	attr. Adj.	PDAT	Dem.pron.	PTKANT	Antwortpart.
ADJD	präd. Adj.	PIS	Indef.pron.	PTKA	Part. bei Adv
ADV	Adverb	PIAT	Indef.pron.	TRUNC	Komp.
APPR	Präp.	PIDAT	Indef.pron.	VVFIN	Verb finit
APPRART	Präp.+ Art.	PPER	Pers.pron.	VVIMP	Verb Imp.
APPO	Postpos.	PPOSS	Poss.pron.	VVIN	Verb Infin.
ART	Artikel	PPOSAT	Poss.pron.	VVIZU	Verb Infin.
CARD	Zahl	PRELS	Rel.pron.	VVPP	Part. Perf.
FM	Fremdspr.	PRELAT	Rel.pron.	VAFIN	Aux. finit
ITJ	Interj.	PRF	Pers.pron.	VAIMP	Aux. Imp.
KOUI	Konj. uo.	PWS	Fragepron.	VAINF	Aux. Infin.
KOUS	Konj. uo.	PWAT	Fragepron.	VAPP	Aux. Part.
KON	Konj. no.	PWAV	Fragepron.	VMFIN	Mod. finit
KOKOM	Konj. vergl.	PAV	Pron.adv.	VMINF	Mod. Imp.
NN	Nomen	PTKZU	zu	VMPP	Mod. Part.

Was macht POS-Tagging?

Part-of-Speech Tagging bezeichnet das Auszeichnen eines Textes in Wortklassen.

Es gibt mehrere Arten von POS Tagging: Statisches Tagging, Word Sense Tagging, etc.

Die Zuordnung der korrekten Kategorie zu jedem Wort einer Äußerung ist oft ein wichtiger Vorverarbeitungsschritt.

- Unterschiedliche Technologien (N-gramme, Regeln,...)
- Fehlerrate ca. 2-4%

Was bedeutet der Begriff Parsing in der Sprachverarbeitung?

Unter Parsing versteht man die strukturelle (syntaktische) Analyse eines Textes.

Eine vollständige, korrekte Analyse ist oft nicht machbar:

- keine vollständige Grammatik natürlicher Sprache
- fehlerhafte Daten
- Effizienzprobleme
- Ambiguität!

Aus diesen Gründen beschränkt man sich oft auf eine flache und/oder partielle Analyse

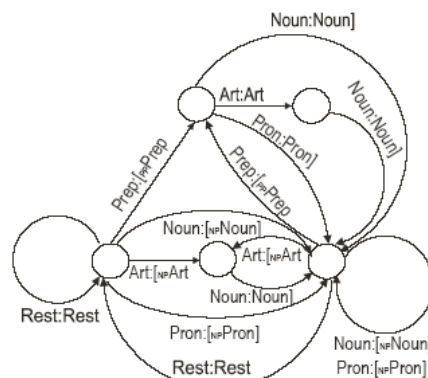
Was versteht man unter flachem, bzw. partiellem Parsing?

Flaches Parsing:

Dabei werden nur eindeutig interpretierbare Konstituenten analysiert.

- flache Struktur (keine tiefen Anbindungen)
- die Analyse bleibt lückenhaft

Zum flachen Parsen werden übrigens oft endliche Automaten (Transducer) eingesetzt. Diese sind zwar schnell und effizient, für das ableiten von ambigen Strukturen jedoch nicht ausreichend.



Partielles Parsing:

Erkennt (wie das Wort „partiell“ ja schon ausdrückt) nur bestimmte Teilstrukturen eines Satzes. Diese Teilstrukturen besitzen einen einfacheren Aufbau und sind daher leicht, schnell (z.B. mit Hilfe endlicher Automaten) und mit großer Zuverlässigkeit erkennbar.

Partielles Parsing kann entweder als der erste Teilschritt bei der vollständigen Analyse oder als Analysewerkzeug für bestimmte Anwendungen betrachtet werden.

Einsatzgebiete: Informationsextraktion, Information Retrieval, Textzusammenfassung

Welche Konstituenten werden bei flachem Parsing identifiziert?

Es werden Konstituenten identifiziert die eindeutig erkennbar sind:

- Kernnominalphrasen (z.B. „dieses sehr teure Auto“)
- Präpositionalphrasen (z.B. „mit 2 Sitzbänken“ oder „um teures Geld“)
- Verbkomplexe (z.B. „kaufen hat müssen“)

Was versteht man unter Namenserkennung und welche Methoden werden verwendet?

Ziel ist die Erkennung von Personennamen, Firmennamen/Organisationen, Geographischen Namen, Datumsangaben, Prozentangaben, Beträgen, etc.

Die Erkennung solcher Bezeichnungen ist ein wesentlicher Vorverarbeitungsschritt.

Die Methoden der Namenserkennung basieren auf:

- manuell erstellten Regeln
- automatischen Lernverfahren

Was versteht man unter Informationsextraktion?

Informationsextraktion ist das gezielte Suchen nach spezifischer Information aus großen Mengen von frei formulierten Texten.

- Eingabe:
 - Die gesuchte Information wird in Form einer Schablone (template) dargestellt.
- Ausgabe:
 - Zu jedem gefundenen relevanten Text wird eine instantiierte Schablone geliefert.
- Anwendungsbereich:
 - Texte mit faktischer Information
 - Keine Spezialkenntnisse für Verstehen nötig

Beispiel für Schablone

Dr. Herrmann Wirth, bisheriger Leiter der Musikhochschule München, verabschiedete sich heute aus dem Amt. Der 65jährige tritt seinen wohlverdienten Ruhestand an. Als seine Nachfolgerin wurde Sabine Klinger benannt. Ebenfalls neu besetzt wurde die Stelle des Musikdirektors. Anneliese Häfner folgt Christian Meindl nach.

- Schablone:
[PersonOut PersonIn Position
Organisation Time]
- Ergebnis1:
[PersonOut Dr.Herrmann Wirth
PersonIn Sabine Klinger
Position Leiter
Organisation MHS München
Time heute]
- Ergebnis2:
[PersonOut Christian Meindl
PersonIn Anneliese Häfner
Position Musikdirektor
Organisation MHS München
Time]

Welche Methoden zur Extraktion domänenspezifischer Information gibt es?

- “Molekularer” Ansatz
 - basiert auf Regeln (Mustern), die den (syntaktischen und semantischen) Kontext berücksichtigen;
 - beginne mit sehr verlässlichen Mustern, erweitere nach Bedarf;
 - Muster können hierarchisch definiert werden.
- “Atomarer” Ansatz
 - basiert auf Interpretation der Typinformation von Nominal- und Verbphrasen ohne Berücksichtigung des Kontexts;
 - jedes Auftreten einer Nomen/Verb-Kombination mit den entsprechenden Eigenschaften bewirkt die Merkmalsinstantiierung in der Schablone.

Wie kann man Informationsextraktionsmodelle evaluieren?

Es gibt drei wesentliche Parameter, die aus dem Information Retrieval

stammen, anhand derer man Informationsextraktionsmodelle evaluieren kann:

- Präzision (precision)
 - der Anteil der richtigen an den insgesamt gefundenen Texten
- Vollständigkeit (recall)
 - der Anteil der gefundenen von den insgesamt vorhandenen Texten
- F-Maß
 - üblicherweise $\beta=1$
 - $$F = \frac{(\beta+1) * P * V}{\beta * P + V}$$

Man beachte: Die Güte der Antwort wird nicht berücksichtigt!

Was versteht man unter Spracherkennung (bzw. Sprechererkennung, Sprachenidentifikation)?

- Spracherkennung
 - das automatische Übertragen gesprochener Sprache in Text.
- Sprechererkennung
 - Sprecheridentifikation: identifiziert den jeweiligen Sprecher
 - Sprecherverifikation: überprüft, ob die Stimme einem vorgegebenen Sprecher entspricht.
- Sprachenerkennung (Sprachenidentifikation)
 - die Identifikation der Sprache, in der gesprochen wird.
 - auch für geschriebene Sprache (Texte) relevant!

Welche Parameter beeinflussen die Komplexität bei der Spracherkennung?

- kontinuierlich vs. diskontinuierlich
 - bei einfachen Systemen müssen zwischen den Wörtern deutliche Pausen gemacht werden.
- Einzelsprecher vs. sprecherunabhängig
 - Systeme können an einen Sprecher trainiert werden, um die Qualität zu erhöhen
- beliebiges (großes) vs. spezielles (kleines) Vokabular
 - ein begrenztes Vokabular erhöht die Qualität der Erkennung
- Aufnahmequalität
 - limitierte Bandbreite (Telefon), Nebengeräusche

Welche Faktoren machen Spracherkennung schwierig?

- Phonetische Variabilität
 - Koartikulation, Assimilation, etc.
- akustische Variabilität
 - Kanal, Nebengeräusche, Entfernung, Position, etc.
- Sprechervariabilität intern
 - Stimmqualität, Sprechrate, Emotion, etc.
- Sprechervariabilität extern
 - Geschlecht, Alter, Dialekt, etc.
- Menschen verarbeiten Sprache kontextabhängig
 - "I scream" vs. "Ice cream"

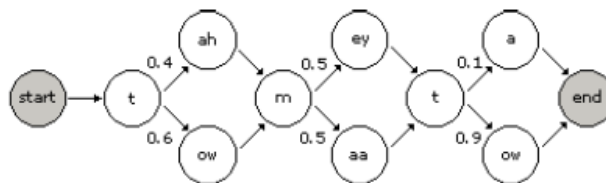
Aus welchen Komponenten besteht ein Spracherkennungssystem?

- Diskretisierung
 - Abtastung, Quantisierung
- Sprachsignalkurzzeitanalyse
 - Fensterfunktion, Merkmalsextraktion
- Vektorquantisierung
 - Abbildung auf digitale Merkmale
- (lautliches) Wortmodell
- (grammatisches) Sprachmodell

Was ist ein Wortmodell in der Spracherkennung?

Die Analyse findet mithilfe von HMMs (Hidden-Markov-Modellen) in zwei Ebenen statt:

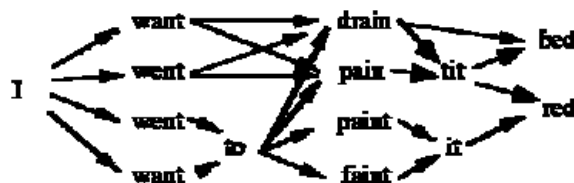
- Phonemanalyse:
 - Phoneme werden als Menge von Allophonen aufgefaßt.
 - Für jedes Allophon wird ein eigenes HMM trainiert.
- Wortanalyse:
 - Einzelne Wörter werden als Netzwerke von Phonemen dargestellt.
 - Jeder gültige Pfad durch das Netzwerk entspricht einer Aussprachevariante.



Auf diese Weise wird die gesamte Äußerung analysiert. Problem: Wortgrenzen werden nicht markiert; Als Ergebnis entstehen dadurch verschiedene Interpretationen unterschiedlicher Plausibilität.

Sie werden dargestellt als

- n-best Wortsequenzen
- Netzwerk mit Übergangswahrscheinlichkeiten



Was ist ein Sprachmodell in der Spracherkennung?

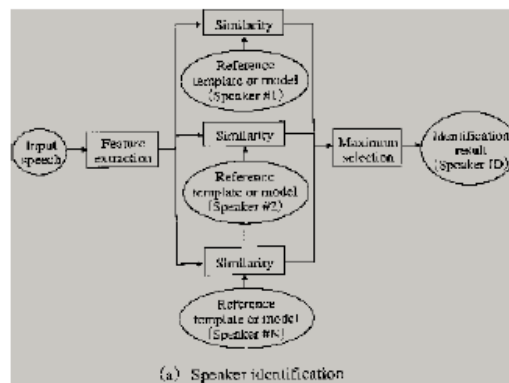
Korrekte Worterkennung basiert nicht alleine auf akustisch/phonetischer Analyse, sondern auch auf dem Kontext. So liefert beispielsweise das Wortmodell unterschiedliche Varianten von Wortfolgen.

Spracherkenner verwenden ein (statistisches) Sprachmodell, das aus einem großen (mehrere Millionen Wörter) Korpus abgeleitet wird.

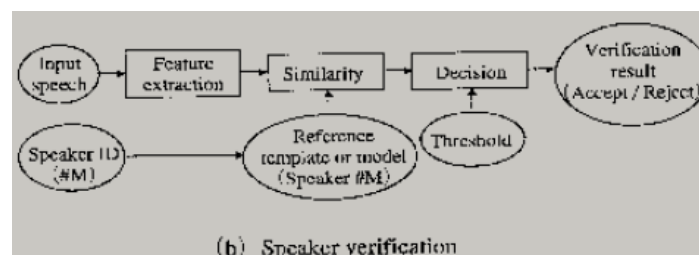
Die Vorgangsweise basiert (ähnlich dem statistischen PoS-Tagging) auf der Identifikation der plausibelsten Wortfolge durch ein HMM.

Was ist der Unterschied zwischen Sprecheridentifikation und Sprecherverifikation?

Sprecheridentifikation identifiziert den Sprecher als bestimmtes Mitglied einer vorgegebenen Gruppe.



Sprecherverifikation überprüft, ob die angegebene Identifikation des Sprechers korrekt ist.



Was versteht man unter reproduktiver Synthese (bzw. Vollsynthese)?

- Reproductive Synthese
 - zusammenhängende Äußerungen bzw. Teile von Äußerungen werden aufgenommen
 - ursprüngliche Stimme erkennbar
 - Vorteil: hohe Sprachqualität
 - Nachteil: wenig flexibel, hoher Aufwand
- Vollsynthese
 - Äußerungen werden ohne direkt auf Aufnahmen zurückzugreifen synthetisiert
 - keine spezielle Stimme erkennbar
 - Vorteil: flexibel, geringer Aufwand
 - Nachteil: schlechtere Sprachqualität

Welche Methoden der Vollsynthese gibt es?

- Konkatentative Synthese
 - aufgenommenes Sprachmaterial wird in kleinsten Einheiten abgespeichert
 - die Einheiten werden nach Bedarf rekonponiert
- Formantsynthese

- verwendet ein akustisches Modell (Quelle-Filter-Modell)
- artikulatorische Synthese
 - verwendet ein (mathematisches) Modell des menschlichen Sprechapparats
 - Äußerungen entstehen durch "Imitation" des Sprechvorgangs

Was versteht man unter konkatenativer Synthese?

Die Synthese greift auf ein Inventar aufgenommener Sprache eines realen Sprechers zurück. Die Länge der aufgenommenen Einheiten ist je nach Methode unterschiedlich. Die aufgenommenen Stücke werden bei der Synthese rekombiniert.

Probleme:

- Die Stücke müssen zusammenpassen (Energie, f_0)
- Die Stücke müssen an die gewünschte Prosodie angepasst werden (f_0 , Dauer)
- Die Stimmqualität ist durch den Sprecher festgelegt.

Welche unterschiedlichen Einheiten werden bei konkatenativer Synthese verwendet?

- Allophone (ca. 50) - ungebräuchlich
- Diphone (einige hundert)
 - von der "Mitte" eines Phons bis zur "Mitte" des nächsten
- Triphone
- Halbsilben (1000 - 2000)
- Gemischte Inventare
 - z.B. Halbsilben, Diphone und Suffixe
- "Unit Selection"
 - dynamische Auswahl zur Laufzeit aus einem großen, entsprechend annotierten Korpus

Was sind die Hauptkomponenten eines Vorlesesystems?

- Vorverarbeitung
 - Tokenization
 - Analyse zur Bestimmung phonologischer Phrasen
 - Abbildung von Schrift auf Lautdarstellung
 - Bestimmung der prosodischen Parameter (f_0 , Dauer)
- Sprachsynthese

Was sind "letter-to-sound" rules?

Sind Regeln für ein Abbildungsverfahren um Ausdrücke die nicht im Lexikon enthalten sind (z.B. Namen) automatisch abzubilden.

Welche Probleme gibt es bei der automatischen Abbildung von Wörtern auf entsprechende Lautfolgen?

- fremdsprachige Ausdrücke
- Akronyme (NATO vs. SPÖ)
- Segmentierung (Häschen)
- Akzent ('durchschauen vs. durch'schauen)

Welche Methoden der Sprachgenerierung gibt es?

- Canned Text
 - verwendet vorgefertigte Phrasen und Texte.
 - Pro: einfach und robust, jeder Stil möglich
 - Con: unflexibel, nur für einfache Anwendungen
- Generierung mit Schablonen
 - verwendet vorgefertigte Bausteine, die aber Stellen beinhalten, an denen spezifisches Material eingefügt werden kann.
 - Pro: ebenfalls einfach und robust, dabei flexibler
 - Con: ebenfalls nur für restringierte Anwendungen
- Volle Generierung
 - produziert weitgehend beliebigen Text, der durch eine entsprechende Eingabestruktur spezifiziert wird.
 - Pro: flexibel, auch für komplexe Anwendungen
 - Con: aufwändig

Was sind Hauptkomponenten voller Generierung?

- Textplanung
 - Festlegen der Inhalte, der zu beschreibenden Objekte, Ereignisse, etc.
 - Sequentialisierung der Beschreibung
- Satzgenerierung
 - Gliederung der Gesamtinformation in Teile, die jeweils als ein Satz realisiert werden sollen.
 - Wortwahl, Pronominalisierung
 - Festlegung stilistischer Parameter (z.B. aktiv - passiv)
- Oberflächenrealisierung bzw. Taktische Generierung
 - die Generierung des Satzes nach den vorgegebenen grammatischen und stilistischen Kriterien.

Was versteht man unter Lokalisierung (bzw. Internationalisierung)?

Lokalisierung ist die Anpassung eines Systems an die sprachlichen und kulturellen Gegebenheiten einer Benutzergruppe. Dies umfasst sprachliche und ikonische Information, aber auch spezielle Formate für Datum, Währung, Zahlen, etc.

Internationalisierung besteht darin, ein System in einer Art und Weise zu gestalten, dass Lokalisierung effizient unterstützt wird. Dies umfasst Trennung textueller Information vom Rest, Berücksichtigen unterschiedlicher Platzanforderungen, Vermeiden nicht lokalisierbarer Elemente (Schrift als Bild), etc.

Welche Elemente der Internationalisierung hat HTML 4?

- Definition des Character-Set
 - Content-type text/html; charset=ISO-8859-1
- Definition der Sprache
 - Top-level: Content-language: de
 - im Text als Attribut: lang =“de“
- Definition der Schreibrichtung

- als Attribut (Text und/oder Dokument): dir=ltr oder dir=rtl
- Language Negotiation
 - wenn der Server (z.B. Apache) das unterstützt, kann eine Sprache präferiert werden, falls ein Dokument mehrsprachig vorhanden ist.

Weitere (nicht so wichtige) Details findet man unter: <http://babel.alis.com>

Was sind Sprachencodes und welche Arten gibt es?

Sprachencodes dienen zur Identifikation der Sprache eines Dokumentes (z.B. im Web).

Es gibt folgende drei Arten:

- 2-Buchstaben Code (ISO-639-1)
 - z.B. de (deutsch), en (englisch), es (spanisch)
- 3-Buchstaben Code (ISO-639-2)
 - z.B. bre (bretonisch), eus (baskisch), mas (Masai)
- RFC-3066 (Sprachencodes für das Internet)
 - ISO-639-1
 - ISO-639-2
 - ISO-639-1 plus ISO-3166, z.B. fr-CA (kanadisches Französisch)
 - ISO-639-1 mit Zusatz registriert bei IANA z.B. enspouse (englischer Dialekt namens "spouse")
 - IANA-registrierter Sprachcode, z.B. i-klingon (beginnt immer mit i-)

Welche Ansätze zur Maschinellen Übersetzung gibt es?

- Interlinguaansatz
 - Interlingua erlaubt sprachunabhängige Darstellung der Bedeutung eines Textes
 - Abbildung auf Interlingua und Generierung daraus
- Transferbasierte Übersetzung
 - Analyse des Texts aus Quellsprache
 - sprachpaarspezifischer Transfer
 - Generierung des Texts in Zielsprache
- Beispielbasierte Übersetzung
- Statistische Übersetzung
 - basiert auf der Verwendung großer Korpora

Was ist Beispielbasierte Maschinelle Übersetzung?

Bei dieser Art der maschinellen Übersetzung versucht der Computer aufgrund von Beispielen richtiger Übersetzungen Muster für die Übersetzung unbekannter Texte zu lernen.

Yesterday, 200 delegates met with President Clinton.	Gestern trafen sich 200 Abgeordnete mit Präsident Clinton.
Yesterday, 200 delegates met behind closed doors to discuss the new text code.	Gestern trafen sich 200 Abgeordnete hinter verschlossenen Türen, um über die neuen Steuergesetze zu verhandeln.
Gennifer Flowers is said to have had an affair with President Clinton for many years.	Gennifer Flowers hat angeblich jahrelang eine Affäre mit Präsident Clinton gehabt.
Zielsatz: Gestern trafen sich 200 Abgeordnete mit Präsident Clinton.	

- Quellsatz: Yesterday, 200 delegates met with President Clinton.
- Yesterday, 200 delegates met behind closed doors to discuss the new text code.
- Gennifer Flowers is said to have had an affair with President Clinton for many years.
- Zielsatz: Gestern trafen sich 200 Abgeordnete mit Präsident Clinton.
- Gestern trafen sich 200 Abgeordnete hinter verschlossenen Türen, um über die neuen Steuergesetze zu verhandeln.
- Gennifer Flowers hat angeblich jahrelang eine Affäre mit Präsident Clinton gehabt.

Wozu dient VoiceXML?

Voice XML ermöglicht die einfache Entwicklung von Anwendungen in den Gebieten Telephonie, Gerätesteuerung und Netzanwendungen.

Genauer:

- Sprachausgabe (synthetisierte Sprache)
- Ausgabe von Audio
- Spracherkennung
- DTMF (Tastendruck)
- Sprachaufnahme
- Dialogsteuerung
 - Auswahlmenü
 - Formular
- Telefon-, bzw. Systemsteuerung

Was versteht man unter Word Sense Tagging?

Was versteht man unter artikulatorischer Synthese?

Kapitel 4 – Sprache und andere Medien

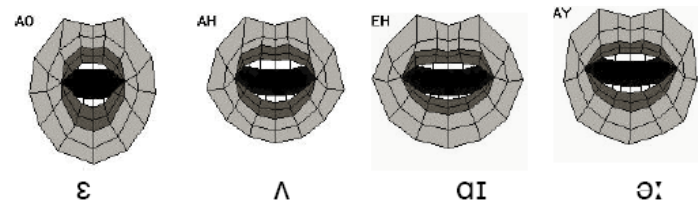
Was sind Viseme?

Ähnlich der Einteilung der Sprache in einzelne Phoneme können auch die Lippenbewegungen in so genannte Viseme eingeteilt werden. Die Auswertung der Aufeinanderfolge der Viseme ermöglicht bereits ohne zusätzliche akustische Informationen die Erkennung von Worten.

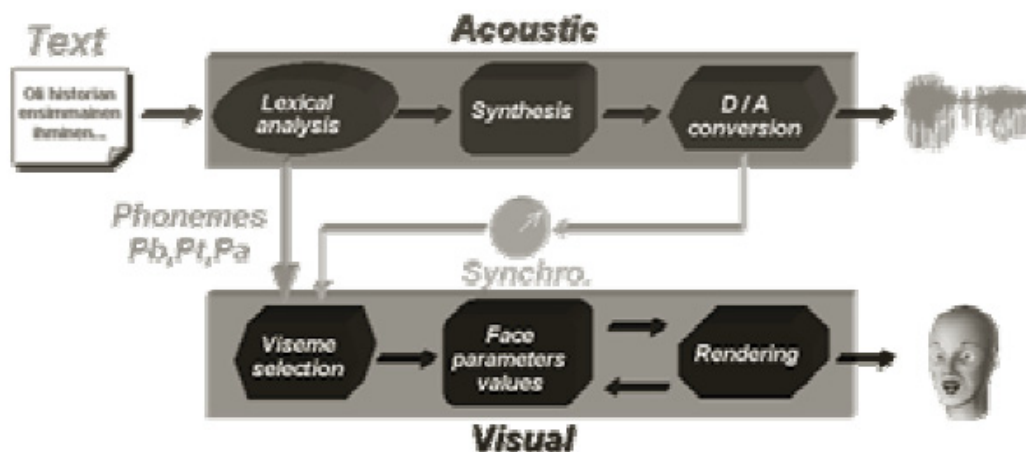
Weiters sind die Unterschiede der Lippenbewegungen zwischen verschiedenen Sprechern weniger stark ausgeprägt, als in der akustischen Artikulation, weshalb eine Adaption an unterschiedliche Sprecher erleichtert wird.

Im Kontext der „Talking Heads“ bedeutet das:

- Ein Visem ist das generische Bild des Gesichtes (insbesondere der Lippenstellung) bei der Erzeugung eines Lauts.
- Viseme sind der graphische Pedant zu Phonemen.
- Ebenso wie bei Phonemen ist das Problem der Koartikulation zu beachten.
- Jedem Phonem ist ein Visem zugeordnet
- Je nach gewünschter Genauigkeit können auch mehrere Phoneme dem selbem Visem zugeordnet sein
- Bessere Qualität durch mehrere Bilder, z.B. zwischen 2 Visemen jeweils ein interpoliertes Bild.



Beschreibe die Architektur eines Talking Head



Welche Ansätze zur Gesichtsmodellierung gibt es?

- Konkatenativer Ansatz (Modell basiert auf real aufgenommenen Videobildern)
- Parametrischer Ansatz (Modell basiert auf einem Gesichtsnetz)
- Anatomischer Ansatz (beschreibt die anatomische Struktur des Gesichtes)

Was sind Vor- und Nachteile des konkatenativen Ansatzes für Talking Heads?

Da das Modell auf real aufgenommenen Videobildern basiert ergeben sich folgende Vor- und Nachteile:

Vorteile:

- hochqualitative Bilder (Videoaufnahmen)

- relativ einfach, da kein explizites Modell erforderlich

Nachteile:

- Kopfhaltung bleibt statisch
- unterschiedlicher Gesichtsausdruck (Lächeln, Erstaunen, etc.) nur möglich, wenn die entsprechenden Bilder vorhanden sind.

Beschreibe den Parametrischen Ansatz für Talking Heads.

Bei dieser Art der Modellierung müssen zwei Probleme gelöst werden:

- Festlegung einer geeigneten Menge von Parametern
 - durch Beobachtung der Oberflächeneigenschaften von Gesichtern
 - durch Untersuchung der Anatomie (Knochengerüst, Muskel, Haut, etc.) -> Anatomisches Modell
 - durch eine Kombination beider Ansätze
- Entwicklung eines Modells auf der Basis dieser Parameter

Parameterwerte beschreiben

- Gestalt und Struktur eines individuellen Gesichts
- den emotionalen Ausdruck

Parameter sind

z.B. Augenfarbe, relative Größe von Nase, Kinn, Stirn, Kieferbreite, ...
aber auch

Ausdruck und Emotionen z.B. Ausdruck des Mundes beim lächeln, ...

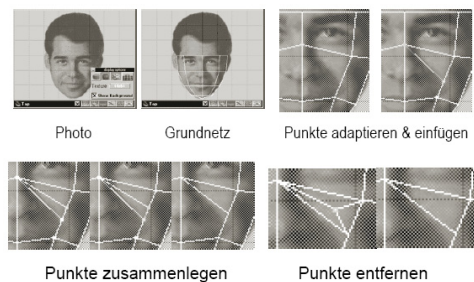
Was ist ein Gesichtsnetz (face mesh) und wie funktioniert es?

Das Gesichtsnetz (face mesh“) ist die Basis für das parametrische Modell.

Das Gesicht wird dabei als eine Menge von Polygonen dargestellt

- Die Grundform des Gesichtsnetzes wird durch die Setzung der Gestaltparameter definiert, bzw. von einer fotografischen Darstellung übertragen
- bestimmte Gesichtsausdrücke sind durch die Setzung der Ausdrucksparameter definiert.

Das Modell muss nun dafür sorgen, dass die Bewegung bestimmter Punkte über das gesamte Netz diffundiert. Das Modell basiert dazu auf einem generischen Gesichtsnetz, das individuell adaptiert wird.

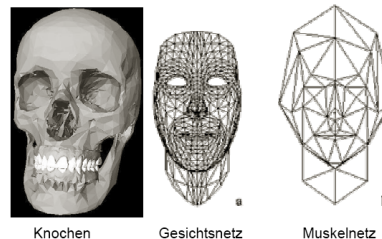


Wodurch unterscheidet sich der anatomische Ansatz beim Talking Head von anderen parametrischen Ansätzen?

Der anatomische Ansatz verwendet ist im Gegensatz zum „normalen“ parametrischen Ansatz (einschichtig) ein dreischichtiges Modell:

- Knochengerüst
- Gesichtsnetz
- Muskeln

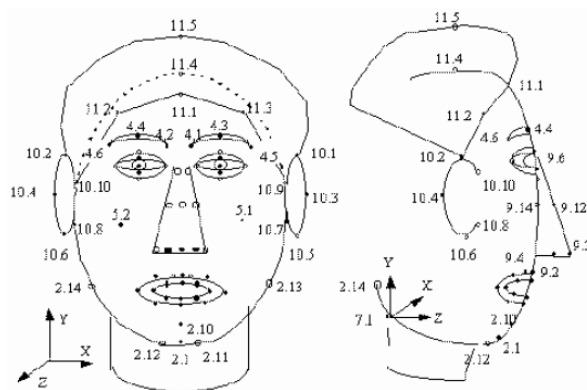
Die Knochen sind dabei wie in der Realität mit der Haut verbunden. Aber auch die Muskeln sind der Menschlichen Anatomie nachempfunden.



Alles in allem wirkt ein auf diese Weise konstruierter „Talking Head“ weit realistischer als ein auf dem einschichtigen Ansatz basierender.

Wie ist Facial Animation in MPEG-4 standardisiert?

- Basiert auf einem Modell eines Gesichts, das durch so genannte Face Feature Points (FFPs) definiert ist, die ein Polygonnetz ergeben.
- Face Definition Parameters (FDPs) dienen dazu, ein individuelles Gesicht festzulegen.
- Face Animation Parameter (FAPs) erlauben es, einen bestimmten Gesichtsausdruck festzulegen.



Was sind FDPs (Face Definition Points) in MPEG-4?

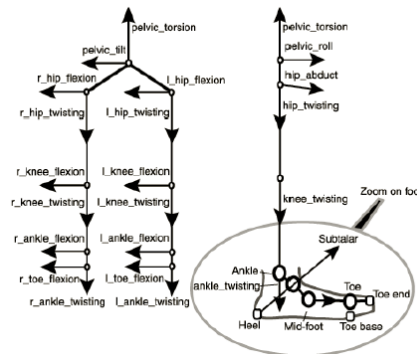
- Definieren ein spezielles Gesicht, indem sie für die FFPs (Face Feature Points) konkrete Werte festlegen.
- Das darunter liegende Modell des Gesichts muss
 - mit der MPEG-4 Definition kompatibel sein,
 - kann aber mehr Information (Punkte, Parameter, etc.) enthalten
- Eine so genannte Face Animation Table (FAT) beschreibt die Abbildung der FAP-Werte auf die FDPs des speziellen Gesichts.

Wie werden in MPEG-4 FAPs auf FDPs eines speziellen Gesichts abgebildet.

Eine so genannte Face Animation Table (FAT) beschreibt die Abbildung der FAP-Werte auf die FDPs des speziellen Gesichts.

Wie ist Body Animation in MPEG-4 standardisiert?

- Ist ein Objekt, das virtuelle Körper auf der Basis eines 3D Netzes beschreibt.
- Body Definition Parameters (BDPs) erlauben es, auf der Grundlage eines generischen Körpers eine bestimmte virtuelle Person zu beschreiben (Gestalt, Textur, Farbe).
- Body Animation Parameters (BAPs) erlauben es, bestimmte Bewegungen, Gesten, etc., direkt anzugeben.



Wie wird Text-to-Speech in MPEG-4 unterstützt?

In MPEG-4 kann der zugrunde liegende Text zur Verbesserung der Qualität der Sprachsynthese mit zusätzlichen Parametern annotiert sein.

Mögliche Parameter sind:

- Prosodie
- Phonetische Annotationen (spezielle Aussprache)
- Lippensynchronisation
- Sprache, bzw. Dialekt
- Sprechstil (Alter, Geschlecht, Sprechrate)
- Verweis auch Mimik (über FAPs)

Bezeichnet wird dies als „Speech Markup“.

Wozu dient Speech Markup?

Dient dazu, um die Synthese gesprochener Sprache aus Text besser kontrollieren zu können.

Was ist SSML und wozu dient es?

SSML (Speech Synthesis Markup Language) wird entwickelt durch das W3C (World Wide Web Consortium) und stellt eine Erweiterung von XML dar.

Es dient zur Beschreibung von Sprachsynthese. Zu sprechende Passagen sind so beispielsweise mit <speech> gekennzeichnet.

Übrigens kann SSML auch in andere Markup Sprachen integriert werden (z.B. in SMIL, ACSS und VoiceXML).

Nenne 3 Konstrukte von SSML und beschreibe Sie näher.

< speak >	kennzeichnet zu sprechenden Text
< sub >	ersetzt Text durch einen anderen String
< audio >	erlaubt das Einfügen von Audiofiles

Was ist SAPI und wozu dient es?

SAPI = (Microsoft Speech API) und wurde wie der Name schon sagt von Microsoft entwickelt. Wie bei SSML und SABLE können Texte mit XMLartigem Markup versehen werden (zur Sprachsynthese).

Natürlich gibt es somit auch hier eine Reihe definierter Tags:

<bookmark>, <context>, <emph>, <lang>, ...

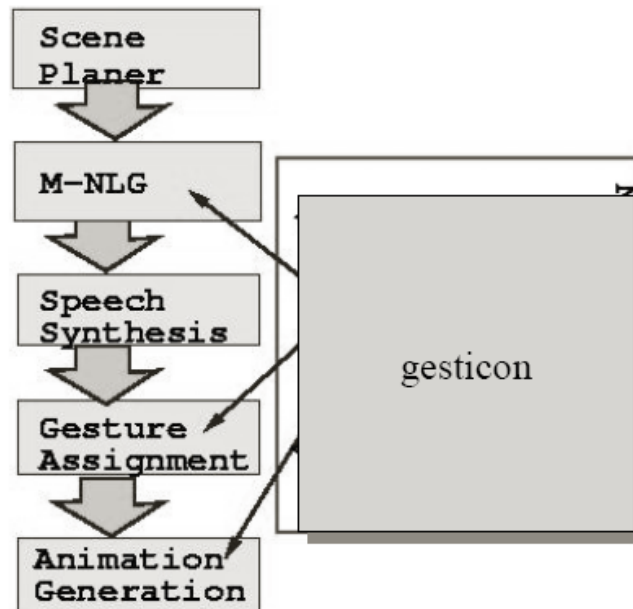
Welche Arten von Gesten unterscheidet man?

- ikonischen Gesten: ihre Bedeutung ist
 - kontextspezifisch und
 - kulturspezifisch
- deiktischen Gesten (Zeigegesten)
- sprachbegleitenden Gesten: sie sind
 - idiosynkratisch (sprecher-spezifisch)
 - Bedeutung ist vage und oft unklar.

Was ist ein Gestenlexikon?

- Idealerweise ein zentrales Inventar des "gesamten" Wissens über Gesten
- Codiert die Abbildung von Bedeutung auf Form
- Inhalt:
 - Anwendungsrestriktionen (z.B. Geschlecht, Sprecher, Emotion)
 - verbale Beschreibung
 - Form
 - beschreibt abstrakt die spatialen und temporalen Eigenschaften der Geste
 - Playercode
 - technische Spezifikationen für das Rendering

Welche Ebenen müssen bei der Integration von Gesten mit Sprache berücksichtigt werden?



M-NLG = Multi-modal Natural Language Generator

Beschreibe das Linguistic Description Scheme in MPEG-7.

- Ein speziell für linguistische Daten entwickeltes Description Scheme ist das Linguistic Description Scheme (LDS)
- ist für beliebige Sprachen anwendbar
- beschreibt die semantische Struktur (propositionalen Gehalt, Dialogstruktur) von Texten, Geschriebenem und Gesprochenem auch in Kombination mit Audio und Video.
- Erweitert den Datentyp Dependency Structure
 - erweitert den Skopus von Sätzen auf größere Einheiten (Absätze, Kapitel, etc.)
 - kann durch neue Klassifikationsschemata erweitert werden.

Worum geht es beim “Semantic Web”?

Die Grundidee des „Semantic Web“ ist, die Inhalte des World Wide Web auch für automatische Prozesse verstehbar zu machen.

Vorschlag: Inhalte mit Metadaten annotieren, die automatische Verarbeitung möglich machen.

Anwendungen:

- “Wissensentdeckung” (Knowledge Discovery)
- Katalogisierung
- Intelligente Software-Agenten
- Bewertung von Inhalten
- Kollektionen verteilter Seiten
- Administrative (IPR, privacy policy, digitale Signatur, etc.)

Was versteht man unter Multimedia-Authoring?

Unter Multimedia Authoring versteht man die automatische Generierung von Dokumenten, die Text mit Graphik, Bild, Ton und Animation verbinden.

- Vorteile:

- effektive Koordination verschiedener Medien in konsistenter Weise
- adaptiv durch kontext-sensitive Generierung
- reflexiv durch explizite Repräsentation der Struktur und Semantik von Multimediadokumenten
- Anwendungen:
 - Informationskioske
 - e-Learning
 - Technische Dokumentationen

Welche Faktoren beeinflussen die Zuordnung von Inhalt und Medium beim Multimedia-Authoring?

- Charakteristik der Information
- Charakteristik des Mediums
- Präsentationsziele
- Benutzercharakteristik
- Aufgaben des Benutzers
- Ressourcenbegrenzung

Welche Art von Verweisen treten in Multimedia- Dokumenten auf?

- Multimedia-Verweise
 - Verweisen auf Objekte unter Verwendung von zumindest zwei Medien
 - sprachliche Ausdrücke (drücke diesen roten Knopf) und Zeigegesten
 - Kombination von Bild und zugeordnetem Text (drücke den roten Knopf an der Vorderseite)
- Cross-Media-Verweise
 - Verweisen auf ein anderes Medium innerhalb des Dokuments
 - im Text (im oberen Teil von Bild 3) auf Bild oder Grafik
- Anaphorische Verweise

Welche Architekturansätze gibt es im Multimedia-Authoring?

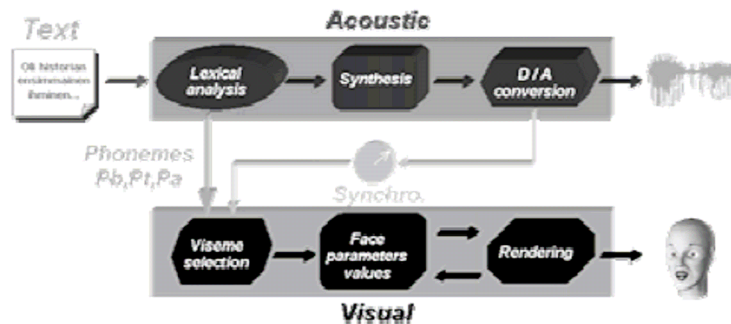
- Schemabasiert
 - Für verschiedene Arten und Formen von Präsentationen werden Schemata definiert, die jeweils ein prototypisches Dokumentendesign darstellen.
- Operatorbasiert
 - Mithilfe von Planern wird aufgrund der Parametersetzung ein Dokumentendesign entwickelt.
- Agentenbasiert
 - Die Aufgabe wird distribuiert gelöst.
 - Hierarchisch: Masteragent kann Serveragenten „anwerben“.
 - Offene Architektur miteinander kooperierender Agenten.

Wozu dient SMIL?

Ist ein Repräsentationsformat für Hypermedia-Präsentationen im WWW. Es dient dem Aufbau komplexer dynamischer Hypermediapräsentationen aus (vorhandenen und standardisierten) Medienobjekten.

- Repräsentationskonstrukte für:
 - Spezifikation komplexer räumlicher und zeitlicher Layoutkonfigurationen
 - Mediensynchronisation
 - Vernetzung von Medienobjekten untereinander

Beschreibe die Architektur eines Talking Head.



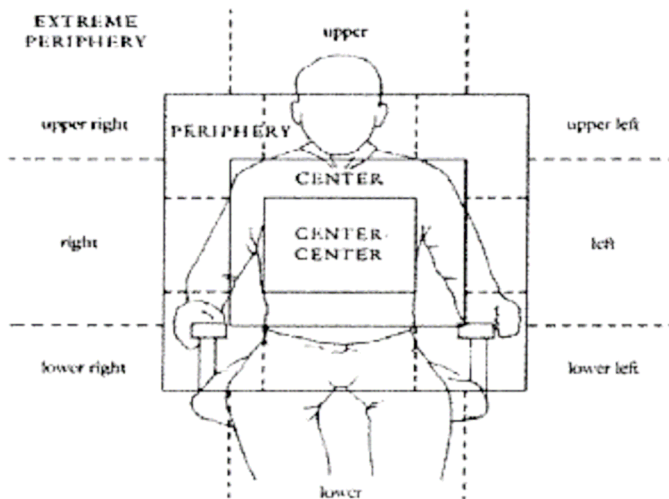
Welche Aspekte sollten in einem Text annotiert werden, der von einem Talking Head vorgetragen wird?

- Intonation:
 - Spezifiziert nach dem ToBI (Tone and Break Indices) Standard
 - Silbenakzente ($L+H^*$, $!H^*$, $L+!H^*$)
 - Grenztöne ($H-$, $L-$, $L-L\%$)
- Augenbrauenbewegungen:
 - Facial Action Unit (AU)
 - 1 – inneres Heben
 - 2 – äußeres Heben
 - 3 – Stirnrunzeln
 - Lächeln
 - Augenzwinkern
- Kopfbewegungen:
 - TL – tilting nod (auf Phrase)
 - D* – downward nod (auf Silbe)
 - U* – upward nod (auf Silbe)

Wie kann man eine Geste beschreiben?

- Bewegung in einzelne Phasen zerlegen:
 - Preparation, stroke, hold, retraction etc.
- N Phasen \square 1 Geste
- Jede Geste hat Attribute
 - Lemma
 - Händigkeit
 - Lexikalische Begleitphrasen
 - Temporale Beziehung Geste-Sprache

Wie können Gesten im Raum lokalisiert werden?



Welche Strategien zur Höflichkeit gibt es?

- Direct
 - „Ich mag kein dunkles Brot!“
- Approval-Oriented
 - „Ich mag dunkles Brot. Aber ich habe schon zuviel davon gegessen!“
- Autonomy-Oriented
 - „Wie wäre es, wenn wir ein paar Semmeln kaufen?“
- Bald record
 - „Haben wir genug Brot?“

Kapitel 5 – Multimediale & multimodale Anwendungen

Wie kann man Videodokumente segmentieren?

Auf zweierlei Arten:

- Erkennen von Wechseln (Schnitt, Überblendung, Bildwechsel, etc.)
 - auf der Basis von Änderungen von Farbe und Intensität
 - Bewertung von Einzelbildunterschieden
 - relativ einfach und schnell
- Erkennen von Bewegungen, Personenwechsel, etc.
 - wesentlich langsamer

Diese Methoden alleine erlauben allerdings noch keine inhaltliche Segmentierung.

Wie funktioniert Topic-Erkennung in Videodokumenten?

Beispielsweise durch Indizierung der gesprochenen Inhalte im A/V Dokument durch:

- Erkennen von Wörtern und Phrasen durch klassische Spracherkennung.
(dies ist allerdings aufwendig und relativ fehleranfällig)

- Aufbau von Phonemverbänden, aus denen Schlüsselwörter abgeleitet werden können. (relativ einfach und robust)

Ein weiteres Beispiel für ein solches „Multimedia Retrieval“ ist z.B. das System „ARIA“ (Bildatenbank), das am MIT entwickelt wurde.

Was ist der Unterschied zwischen Medium und Modalität?

Medium:

In der Physik und im allgemeinen Sprachgebrauch ist ein Medium Träger(-material) und (Ver-)Mittler gewisser Eigenschaften oder Informationen. Beispiele sind Text, Graphik, Bild, Geräusch, Musik und Bewegtbild. Diese Medien sind nicht unbedingt an einen bestimmten Wahrnehmungskanal oder ein Sinnesorgan gebunden. Gleichwohl spielen menschliche Sinne eine wichtige Rolle bei der Wahrnehmung der mittels Medien transportierten Information.

Modalität:

Der Modus (oder die Modalität), über den die Wahrnehmung geschieht, wird von den menschlichen Fähigkeiten/Sinnesorganen und dem gewählten Medium bestimmt: Musik wird (fast ausschließlich) akustisch wahrgenommen, Texte können akustisch oder visuell (oder taktil via Braille-Schrift) wahrgenommen werden.

Anders ausgedrückt: Das Medium ist nicht per Definition an einen menschlichen Sinn oder Modus gebunden. Gesprochene Sprache transportiert zwar unter Umständen zusätzliche Information zum Text, das Medium Text bleibt davon jedoch zunächst unberührt. Seine Basisinformation wird übertragen, egal, ob akustisch oder visuell. Ähnliches gilt für Grafiken, die natürlich primär für den visuellen Kanal geeignet sind, nichtsdestotrotz aber auch taktil erfahrbar sein können.

Aus der Unterscheidung zwischen Medium und Modus folgt auch die Unterscheidung zwischen multimedialen und multimodalen Systemen. Letztere sprechen mehrere Sinne an, während erstere sich mehrerer Medien, also Informationsträgern bedienen.

Was ist ein Multimodales Interface?

Eine Systemschnittstelle die den (sequentiellen oder nebenläufigen) Einsatz multipler Modalitäten erlaubt.

Was sind wesentliche Anforderungen an Multimodale Interfaces?

- 1) Auswahl geeigneter Modalitäten
 - Zusätzliche Modalitäten sollten auch zusätzlichen „Nutzen“ bringen.
- 2) Zuordnung zwischen Modalität und Funktionalität
 - Benutzer kann grundsätzlich beliebige Modalität wählen und/oder
 - klare Zuordnung der Modalität zu bestimmten Funktionen.
- 3) Symmetrie zwischen Ausgabe- und Eingabemodalitäten

- Benutzer und System sollten über etwa die gleichen Möglichkeiten verfügen.
- Besonders wichtig, wenn durch Personalisierung auch die emotionale Ebene zum Tragen kommt.

Welche unterschiedlichen Arten von Multimodalen Interfaces gibt es?

Es gibt zwei verschiedene Arten von multimodalen Interfaces:

- „Klassisches“ Interface, das mehrere Modalitäten zur Kommunikation anbietet.
- Verwendung eines Agenten als „Ansprech-partner“.
 - Nicht optisch präsent: etwa nur als Stimme.
 - Verwendung eines animierten Agenten (Präsentationsagent, Embodied Conversational Agent, Virtual Reality mit Avatar)

Was zeichnet Agenten aus?

Sie sind:

- *Reaktiv*: reagiert auf Änderungen der Umgebung.
- *Autonom*: hat Kontrolle über seine Aktionen.
- *Zielorientiert*: proaktiv, nicht nur reagierend.
- *Zeitlich kontinuierlich kommunizierend*: kontinuierliche Kommunikation mit anderen Agenten, bzw. Benutzern
- *Lernend bzw. adaptiv*: ändert sein Verhalten aufgrund von Erfahrungen.
- *Flexibel*: Aktionen sind nicht global definiert.
- *Haben Charakter*: glaubwürdige „Persönlichkeit“ (u.U. mit emotionalen Zuständen)

Wovon hängt die Glaubwürdigkeit von Agenten ab?

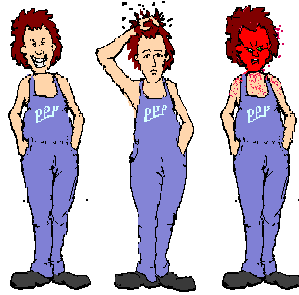
- Aussehen ist wichtig
 - Interessanterweise wirken einfache Zeichentrickfiguren oft glaubwürdiger als Fotorealistische
- Verhalten ist wichtig
 - Bewegungen, Gesten
 - Gesichtsausdruck und Haltung, um Emotionen auszudrücken
- Interaktionsmodus
 - „normale“ natürliche Sprache oder Artefakt

Was ist ein Präsentationsagent?

Ein Präsentationsagent repräsentiert einem Benutzer einer Applikation gegenüber „das gesamte System“. Er führt den Dialog mit dem Benutzer und muss diesem gegenüber die Systemzustände visualisieren. Um intelligent auf Benutzerreaktionen einerseits und auf Systemaktivitäten andererseits reagieren zu können, muss der Präsentationsagent den Dialogverlauf verfolgen und möglichst jeden Dialogzustand adäquat kommentieren.

Es können verschiedene Systemzustände durch die Mimik des Präsentationsagenten visualisiert werden. z.B.:

- Lachen: Freundliche Begrüßung des Benutzers oder das Passwort war gültig
- Am Kopf kratzen: System bearbeitet eine Eingabe
- Ärger, Wut: Mehrere Anmeldeversuche sind fehlgeschlagen



Was ist ein Multimedia Interface Agent?

Ist ein animierter Agent (Cartoon oder 3D-Humanoider) zur Unterstützung von Informationspräsentation.

Vorteile:

- Erweitertes Repertoire an Präsentationstechniken (z.B. Zeigegesten)
- Navigationshilfe beim Erfassen der Information (Vorgeben einer optimalen zeitlichen Reihenfolge, insbesondere bei paralleler Informationspräsentation)
- Instruktion durch Vormachen (erlaubt einfaches Vermitteln komplexer räumlich-zeitlicher Abläufe)

Man beachte: Anthropomorphisierung (Vermenschlichung) des Computers!

Was ist ein Embodied Conversational Agent?

Ist ein animierter Agent (Cartoon oder 3D-Humanoider) der in der Lage ist, eine Konversation zu führen. Bzw. ein Interface, das über die natürlichen Modalitäten eines Menschen verfügt

Embodied Conversational Agents

- können Sprache sowohl verstehen als auch produzieren
- produzieren passende Mimik und Körperhaltung
- produzieren Gesten (unterstützend und Zeigegesten)

Von ihrer Funktionalität her sind Embodied Conversational Agents also Dialogsysteme mit optischer Präsenz.

Was sind wesentliche Eigenschaften menschlicher Konversation?

- Multiple Modalitäten: Sprache, Mimik, Gestik, Körperhaltung
- Modalitäten werden parallel und in Kombination miteinander eingesetzt.
- Die Initiative im Diskurs wechselt:
 - Initiierung des Diskurses
 - Turn taking and turn giving
 - Feedback
- Propositionale und interaktionale Information

Was sind Anforderungen an die Architektur eines Embodied Conversational Agents?

- Multimodale Eingabe und Ausgabe
 - Das System muss einen glaubwürdigen Agenten präsentieren können
 - Der Benutzer muss dem System multimodalen Input liefern können (Mikrofon, Datenhandschuh, etc.)
- Echtzeitverhalten
- Verstehen und Produktion von propositionaler und interaktionaler Information.
 - Sowohl auf sprachlicher als auch auf nicht-verbaler Ebene.
- Modell der Diskursfunktionen
 - Das System muss einen natürlichen Diskurs initiieren und aufrechterhalten können.

Was ist „Microsoft Agent“?

Microsoft Agent ist ein einfaches von Microsoft entwickeltes Werkzeug um Multimedia Agents zu entwickeln.



Komponenten sind:

- Kernarchitektur
- Vordefinierte Charaktere
- Text-to-Speech
- Spracherkennung

Mehr dazu unter: <http://www.microsoft.com/msagent>

Was sind Diskursfunktionen? Gib Beispiele.

- Erkennt Anwesenheit des Benutzers
- Feedback
- Rollenwechsel (turn-taking)
- Funktionen mit zusätzlichem propositionalem Gehalt:
 - Begrüßen und Verabschieden
 - Betonungsfunktion

Die einzelnen Diskursfunktionen sind im „Diskursverhalten“ realisiert.

Wie ist turn-taking in REA realisiert?

Turn-Taking = „Rollenwechsel“

Beim FEMBOT REA wird dieses Turn-Taking unter folgenden Umständen ausgelöst:

<u>State</u>	<u>User Input</u>	<u>Input Function</u>
REA	Gesture	Wanting Turn
speaking	Speech	Taking Turn
User	Pause < 0.5sec	Wanting feedback
speaking	Imperative phrase	Giving turn
	Interrogative phrase	Giving turn
	Decl. Phrase & pause > 0.5sec & no gesture	Giving turn
	Decl. Phrase & long gesture	Holding turn

In REA gibt es drei verschiedene Arten von Turn-Taking:

<i>Conversational Functions:</i>	<i>Communicative Behaviour:</i>
Give Turn	Looking, Raise Eyebrows
Wanting Turn	Raise hands into gesture space
Take Turn	Glance away, Start Talking

Was sind (im Bereich VR) Avatare?

Avatar (Sanskrit) bezeichnet ursprünglich die Inkarnation einer Gottheit.

Im Bereich der Virtual Reality versteht man unter Avataren 3D-Objekte (Körper, Köpfe, abstrakte Objekte, etc.), die den Benutzer im virtuellen Raum repräsentieren.

Über den Avatar kann der Benutzer mit anderen Akteuren (synthetischen Agenten oder Avataren) kommunizieren. Dabei kann (muss aber nicht) der Avatar auch über natürliche Sprache kommunizieren.

Wozu dienen Spezifikationsprachen für virtuelle Charaktere?

- Schnittstelle zwischen Verhaltensplanung und Animationsengine
- Wiederverwendung und Modalisierung von Definitionen
- Trennung von Funktion und Realisierung
- Unabhängigkeit von der Animationsengine
- Zur Beschreibung des motorischen und sprachlichen Verhaltens virtueller Charaktere
 - AML – Avatar Markup Language
 - CML – Character Markup Language
 - VHML – Virtual Human Markup Language
 - MPML – Multimodal Presentation ML
- Zur Beschreibung emotionalen und sozialen Verhaltens virtueller Charaktere
 - APML – Affective Presentation ML
 - SIML – Social Identity Markup Language

Beschreibe eine Spezifikationsprache für virtuelle Charaktere näher. Character Markup Language (CML)

- zur Definition „glaubwürdigen“ Verhaltens von synthetischen Charakteren

- schließt die Lücke zwischen kommunikativen Aspekten (Emotion, Intention) und gezeigtem Verhalten
- Konzentration auf visuelle Aspekte:
 - animierter Gesichts- und Körperausdruck, bzw. – bewegung
 - Persönlichkeit, Rolle, Emotion, Gestik
- (limitierte) Steuerung der Sprache
- MPEG4-kompatibel
- Mediensynchronisation mithilfe von SMIL

Welche kommunikativen Funktionen kennst du?

- Information über die Welt
- Information über die Identität des Sprechers
 - Physiognomie
 - Stimme
- Information über die Psyche des Sprechers
 - Intention
- Topic-comment
- Rhethorische Relationen
- Turn-taking
 - Emotion
 - Überzeugungen
- Metakognitive Information (Sicherheit, Quelle,...)

Was versteht man unter Interaction Process Analysis (IPA)?

- Bekannte Methodik zur Beobachtung der Interaktion in kleinen Gruppen
- Vorteile:
 - Standardmenge von Kategorien, um soziale Interaktionen zu codieren.
 - Signifikante Menge von empirischen Daten, die existierende Gruppen beschreiben
- Kommunikation wird in Beziehung gesetzt zur Gruppenaufgabe und zur sozial-emotionalen Effizienz und Zufriedenheit.
- Die Beobachtung der Kommunikation eines Mitglieds der Gruppe führt zu Eindrücken über Führungsstil und Beliebtheit durch andere Mitglieder.