

Als Pfadfinder die Sprachtechnologie erkunden – Das E-Learning-Modul „Gesprochene Sprache“

Zusammenfassung

An der Universität Bonn wird im Rahmen des Verbundprojekts MiLCA das E-Learning-Modul „Gesprochene Sprache“ entwickelt. Das Modul bietet einen Überblick über die Grundlagen der computergestützten Analyse, Synthese und Übertragung von Sprachsignalen. Es umfasst damit phonetisches Grundlagenwissen zur Sprachverarbeitung. Darüber hinaus wird ein wichtiges Aufgabengebiet der akustischen Mensch-Maschine-Kommunikation behandelt: die Sprachsynthese. Inhalte sind u.a. die Vorstellung verschiedener Verfahren zur Sprachsynthese sowie die Erläuterung des Grundaufbaus und der wichtigsten Techniken eines Sprachausgabesystems.

Die Sprachsynthese bildet eine interessante Anwendung, die an der Schnittstelle zwischen Komponenten aus der Computerlinguistik und der Sprachsignalverarbeitung liegt. Auswirkungen von syntaktischer und semantischer Analyse können an der Sprachsynthese unmittelbar experimentell erfahren werden. In diesem Beitrag werden Konzeption und didaktische Struktur des als Web-Based-Training angelegten Moduls „Gesprochene Sprache“ vorgestellt. Weiterer Schwerpunkt des Beitrags ist die Einbindung multimedialer Elemente in das Modul. Es sind insbesondere interaktive Elemente integriert, die das Zusammenspiel von Computerlinguistik und Sprachsignalverarbeitung veranschaulichen.

1. Einleitung

Im Rahmen des vom Bundesministerium für Bildung und Forschung innerhalb der Initiative „Neue Medien in der Bildung“ geförderten Verbundprojekts MiLCA¹ werden Lehrmodule für die Ausbildung in der Computerlinguistik entwickelt. Ziel des Projekts ist die Erstellung verschiedener Lehrmodule mit verstärktem Einsatz multimedialer Elemente. Die Lehrmodule werden innerhalb des Lehrbetriebs unter den am Projekt beteiligten Universitäten² ausgetauscht. Das jeweilige Lehrangebot soll auf diese Weise bereichert und erweitert werden. An der Universität Bonn entsteht das Modul „Gespro-

1 MiLCA – „Medienintensive Lehrmodule in der Computerlinguistik-Ausbildung“
URL: <http://milca.sfs.uni-tuebingen.de>

2 In Zukunft sollen die Lehrmodule, die im Projekt MiLCA entwickelt wurden, auch anderen Universitäten und Bildungseinrichtungen zugänglich gemacht werden.

chene Sprache“, ein Web-Based-Training (WBT) mit Inhalten aus dem Bereich der Sprachtechnologie.

Das Modul besteht aus verschiedenen Teilmodulen. Das Teilmodul „Grundlagen der Sprachsignalverarbeitung“ vermittelt grundlegende Methoden und Techniken, die bei der Darstellung und Verarbeitung von (Sprach-)Signalen durch den Rechner angewendet werden. Darauf aufbauend bietet das Teilmodul „Sprachsynthese“ zunächst einen Überblick über die verschiedenen Ansätze, mittels des Rechners sprachliche Äußerungen zu erzeugen. Anschließend wird die Funktionsweise der einzelnen Komponenten eines Sprachsynthesystems näher erläutert. Dabei wird hier eine Schnittstelle zum Fach der Computerlinguistik geschaffen. Denn fehlerhafte syntaktische oder semantische Analysen, wie sie von computerlinguistischen Komponenten eines sprachverarbeitenden Systems geliefert werden können, wirken sich oft unmittelbar auf die Sprachausgabe eines Sprachsynthesystems aus und können so vom Lernenden direkt an einem Anwendungsbeispiel sinnlich, d.h. hier auditiv, erfahren werden. Diese Schnittstelle zwischen der Phonetik bzw. Sprachsignalverarbeitung und der Computerlinguistik wird vom Teilmodul „Linguistische Aspekte der Verarbeitung gesprochener Sprache“ geleistet.

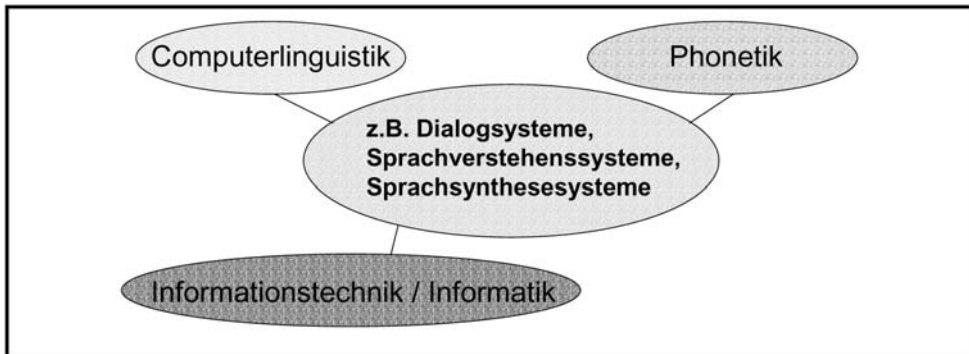


Abbildung 1: Schnittstelle zwischen Phonetik und Computerlinguistik

Dieser Beitrag stellt im Folgenden die Lehrinhalte des Moduls „Gesprochene Sprache“ näher vor. Dabei werden der inhaltliche Aufbau der einzelnen Teilmodule sowie das didaktische Szenario und die verwendeten Kommunikationswerkzeuge beschrieben. Chancen und Probleme, die sich bei Einsatz und Entwicklung eines solchen Moduls als Selbstlernkurs ergeben, sollen hier deutlich werden. Schließlich werden Anmerkungen zur technischen Realisierung des Moduls gemacht: zur Annotation und Codierung der zugrunde liegenden Lernmaterialien, zur Implementierung von animierten Graphiken und Demonstrationsprogrammen sowie zur Einbettung der Kursmaterialien in eine geeignete Lehr-/Lernumgebung.

2. Inhaltsübersicht der einzelnen Teilmodule im WBT „Gesprochene Sprache“

Das Modul „Gesprochene Sprache“ besteht zur Zeit aus vier Teilmodulen: „Grundlagen der Sprachsignalverarbeitung“, „Sprachsynthese“, „Linguistische Aspekte der Verarbeitung gesprochener Sprache“ sowie „Mathematische Grundlagen“. Den Einstieg in die einzelnen Teilmodule bildet ein Portal, das eine Übersicht über die Aufgabengebiete der akustischen Mensch-Maschine-Kommunikation bietet. Dies sind die Analyse, Synthese und Übertragung von Signalen sowie Sprachein- und Sprachausgabe mithilfe des Rechners. Vom Portal aus gelangt der Nutzer zu den Lerneinheiten der einzelnen Teilmodule.

2.1. Teilmodul „Grundlagen der Sprachsignalverarbeitung“

Das Teilmodul zur Sprachsignalverarbeitung gibt einen Überblick über die Grundlagen und Grundvoraussetzungen der computergestützten Verarbeitung von Sprachsignalen. Die Analyse, Synthese und Übertragung von Sprachsignalen mit Hilfe des Rechners werden im Einzelnen behandelt.

Zum Einstieg werden einige phonetische Grundlagen zur Sprachsignalverarbeitung vermittelt. Dabei wird die Erzeugung von Sprachsignalen, darunter die verschiedenen Anregungsformen bei der Spracherzeugung, erläutert. Zunächst wird der grundsätzliche Aufbau des menschlichen Sprechtrakts vorgestellt. Daran anschließend erfährt der Lernende, wie Sprachsignale im Rechner gespeichert und dargestellt werden können – eine Grundvoraussetzung, um Signalanalysen durchführen zu können. Dann werden die Grundbegriffe der Systemtheorie beschrieben, um den Prozess der Spracherzeugung modellieren zu können. Einen Ausschnitt aus diesem Themenbereich zeigt Abbildung 2.

Das Prinzip der Kurzzeitanalyse als wichtigste Methode der Behandlung zeitveränderlicher Signale wird erläutert. Des Weiteren werden die wichtigsten elementaren Kenngrößen akustischer Signale, wie z.B. Lautstärke, Frequenz(-anteile), Amplitude oder Pegel, definiert; von diesen Größen abgeleitete Parameter werden beschrieben. Analoge Bezugsgrößen aus der Akustik und Elektrizitätslehre sind zusammengestellt; im Anschluss werden die elementaren Sprachsignalparameter im digitalen Bereich definiert.

Den wichtigsten Teil innerhalb der „Grundlagen der Sprachsignalverarbeitung“ bildet die Erörterung der (Kurzzeit-)Spektralanalyse, die das in der Phonetik mit Abstand wichtigste Messverfahren für Sprachsignale darstellt. Von der Harmonischen Analyse diskreter periodischer Signale (diskrete Fouriertransformation) geht es über die Berechnung von Kurzzeitspektren von Sprachsignalen bis hin zu einem linearen Modell der Spracherzeugung in spektraler Darstellung.

Den Abschluss des Teilmoduls bilden Lerneinheiten zur Glättung von Spektren, insbesondere zum Cepstrum-Verfahren. Der Zusammenhang von Glättung und möglichen Rückschlüssen vom Sprachsignal auf die momentanen Eigenschaften des Vokaltraktes

werden erläutert. Darüber hinaus wird die Problematik der spektralen Darstellung als Visualisierungsmittel der Eigenschaften von Sprachsignalen aufgezeigt. Gewählt wird meist eine Zeit-Frequenz-Darstellung. Das Zeit-Bandbreite-Gesetz macht eine zeitlich wie spektral gleichermaßen feine Auflösung unmöglich. Als Konsequenz existieren zur Darstellung und Analyse von Sprachsignalen Schmalband- und Breitbandspektrogramme nebeneinander. Neben dem Spektrogramm wird eine weitere wichtige Visualisierungsform in der Sprachsignalverarbeitung vorgestellt: das Lautgebirge.

The screenshot shows a web browser window with the following content:

System, Rechenvorschrift, Eingangs- und Ausgangssignal

Anmerkung: Die Begriffe *System* und *Systemtheorie* werden je nach Wissenschaftszweig verschieden definiert und verwendet. Hier werden die in der Nachrichtentechnik gebräuchlichen Definitionen eingeführt. Diese umreißen den Systembegriff sehr eng.

Unter einem *System* (s. folgende Abb.) verstehen wir eine Vorrichtung, die nach einer bestimmten Rechenvorschrift S ein Signal x eindeutig in ein Signal y umwandelt. Hierbei wird x als Eingangssignal, y als Ausgangssignal bezeichnet. Die Signale x und y können Funktionen der gleichen unabhängigen Variablen sein; dies ist jedoch nicht unabdingbar. In verallgemeinerter Form kann ein System auch mehrere Ein- oder Ausgänge haben.

Schüßler (1984) definiert das System wie folgt:

"Ein System ist ein mathematisches Modell eines Gebildes mit Eingängen und Ausgängen, das die Beziehungen zwischen den Eingangs- und Ausgangsgrößen beschreibt." *Schüßler, Hans-Wilhelm, 1984, S. 2*

Blockdiagramm (Blockschaltbild) eines Systems mit Rechenvorschrift S
 $[x(n)]$ Eingangssignal, $[y(n)]$ Ausgangssignal

Das Systemtheorie betrachtet und behandelt die Auswirkung der Rechenvorschrift S auf die Signale unabhängig von der tatsächlichen Realisierung des Systems. S wird auch gelegentlich als **Transformationsvorschrift** oder als **Operator** bezeichnet:

$$y(n) = S\{x(n)\}$$

(Def. 2.1)

Abbildung 2: Inhaltsseite des Teilmoduls „Grundlagen der Sprachsignalverarbeitung“

2.2. Teilmodul „Mathematische Grundlagen“

Das Teilmodul „Mathematische Grundlagen“ vermittelt wichtige Grundkenntnisse, die für das Verständnis der Sprachsignalverarbeitung und sprachtechnologischer Anwendungen unverzichtbar sind. Mit Hilfe komplexer Zahlen bzw. der komplexen Exponentialfunktion lassen sich Schwingungen einheitlich und elegant beschreiben. Das Teilmodul beinhaltet daher Lerneinheiten zur komplexen Zahlenebene, zur Reihenrechnung sowie zum Gebiet der komplexen Exponentialfunktionen. Hier finden sich verstärkt Links zu Internetseiten außerhalb des WBTs, die mathematisches Grundlagenwissen vermitteln.

Das Teilmodul behandelt im Einzelnen folgende Themengebiete:

- Räume und Koordinatensysteme

- Komplexe Zahlen: schrittweise Erweiterung des Systems der natürlichen Zahlen um rationale, reelle, imaginäre bis hin zu komplexen Zahlen in zweidimensionaler Zahlenebene; Erläuterung der Grundrechenarten im Komplexen (u.a. Potenzierung, Radizierung, Logarithmierung)
- Reihenrechnung: endliche und unendliche Reihen, Konvergenz, Majorantenprinzip, Fourierreihe
- Grundkenntnisse über die komplexe Exponentialfunktion: Eulersche Formel, Rechnen mit komplexen Funktionen

Das Teilmodul ist von allen anderen Teilmodulen des Web-Based-Trainings aus erreichbar. Es ist inhaltlich besonders eng mit dem Teilmodul „Grundlagen der Sprachsignalverarbeitung“ verzahnt.

2.3. Teilmodul „Sprachsynthese“

Das Teilmodul „Sprachsynthese“ beschäftigt sich mit einem wichtigen Anwendungsgebiet innerhalb der Sprachsignalverarbeitung, der Sprachausgabe durch den Rechner.

Das Teilmodul bietet zunächst zur Orientierung einen Überblick über verschiedene Sprachausgabesysteme, welcher die historische Entwicklung der Technik veranschaulichen soll. Die Funktionsweise sowohl von einfacher digitaler Sprachwiedergabe als auch von vollständigen Sprachsynthesystemen wird erklärt. Danach wird ein wichtiges Problem bei der Entwicklung eines Synthesesystems verdeutlicht: die Verkettung. Die Verkettung wirkt als Bindeglied zwischen niedrigster symbolischer Ebene (d.h. diskrete phonetische Symbole) und der akustischen Ebene (kontinuierlicher Datenstrom von Sprachsignalparametern). Daran anschließend werden die verschiedenen Prinzipien akustischer Synthese erläutert: parametrische und artikulatorische Verfahren sowie die Synthese durch Signalmanipulation im Zeitbereich. Insbesondere wird hier das PSOLA-Verfahren beschrieben. Integriert ist eine Demonstration verschiedener Synthesysteme.

Die aufeinander aufbauenden Entwicklungsschritte der akustischen Synthese, von der Inventarerstellung bis hin zur Verkettung der natürlichsprachlichen Bausteine, werden aufgezeigt und erläutert. Detailliert behandelt wird eine heute weit verbreitet angewandte Lösung, die Korpussyntese. Das dabei angewandte Prinzip der Unit Selection, welches mit Selektionsalgorithmen und Kostenfunktionen arbeitet, ist hier zentrales Thema.

Weitere Lerneinheiten behandeln Fragen zur Prosodie im Bereich der Sprachsynthese. Die Prosodiesteuerung ist für die Qualität eines Synthesesystems entscheidend. Im Einzelnen werden hier verschiedene Modelle zur Dauer- und Intonationssteuerung erörtert.

Ein weiteres Thema ist die Evaluierung der Qualität von Sprachsynthesystemen. Die Beurteilung der Performanz verschiedener Systeme geschieht durch die Bewertung von synthetischen Stimuli. Dabei kann die Einschätzung der Qualität eines Systems global nach mehreren Kriterien erfolgen: Verständlichkeit, Natürlichkeit, Deutlichkeit der Aussprache etc. Diese Kriterien und die verschiedenen Methoden der Evaluation

werden hier beschrieben und erläutert. Die Vorstellung eines speziellen Verfahrens, um Prosodie und segmentale Qualität eines Sprachsignals unabhängig voneinander beurteilen zu können, zeigt der folgende Ausschnitt aus dem Teilmodul „Sprachsynthese“ (s. Abb. 3). Insbesondere die Lerneinheit über die Evaluierung von Sprachsynthesystemen ist mit zahlreichen Hörbeispielen angereichert und bietet dem Lernenden selbst die Möglichkeit, einen qualitativen Vergleich zwischen verschiedenen Systemen zu ziehen.

Modul A3
"Gesprochene Sprache":
Sprachsynthese

Inhaltsverzeichnis:

0. Einleitung
1. Synthese auf segmentaler Ebene
2. Akustische Synthese
3. Direkte Sprachsynthese vom Sprachkorpus
4. Zur Frage der Prosodie
5. Evaluierung der Qualität
 - Einleitung
 - Evaluierung der Verständlichkeit und Verstehbarkeit
 - Bewertung der Natürlichkeit und zugehöriger Attribute
 - Wie "gut" ist die Sprachsynthese heute?
6. Einige ausgewählte Anwendungen

Übungen

Grundlagen Sprachsignalverarbeitung

Literatur

Glossar

Delexikalisierung zur Prosodiebeurteilung: Das PURR-Verfahren

In einem gewöhnlichen Sprachsignal können Prosodie und segmentale Qualität nicht unabhängig voneinander beurteilt werden. Genau dies ist aber erforderlich, um ein Prosodiemodul gezielt und unabhängig von der segmentalen Qualität zu verbessern. Aus diesem Grund ist es notwendig, ein synthetisches Sprachsignal **delexikalieren**, also die lexikalische Information und ggf. auch die segmentale Qualität aus dem Signal entfernen zu können.

"PURR" = Prosody Unveiling Restricted Representation [siehe dazu auch: Forschungsprojekt am IKP in Bonn]

Abbildung 5: PURR-Verfahren
Jede Grundperiode des Signals wird ersetzt durch eine Summe von drei Sinusschwingungen. Das Verfahren entfernt die segmentale Information vollständig.

Audio-Beispiele:

"Die Frau, die ..." (weibliche Stimme): [Original](#) und [PURR-Signal](#)

"Der Zug fährt ..." (männliche Stimme): [Original](#) und [PURR-Signal](#)

- [Aids-fach-arzt \(PURR\)](#)
- [Chef-volks-wirt \(PURR\)](#)
- [Dampf-schiff-fahrt \(PURR\)](#)
- [Dienst-fahr-rad \(PURR\)](#)

Abbildung 3: Inhaltsseite des Teilmoduls „Sprachsynthese“

Den Abschluss des Teilmoduls bilden ausgewählte Anwendungen, die den Einsatz von Synthesystemen im Behindertenbereich dokumentieren, einige multilinguale Systeme vorstellen sowie die Integration von inhaltsgesteuerter Sprachsynthese in Sprachdialogsysteme beschreiben.

2.4. Teilmodul „Linguistische Aspekte der Verarbeitung gesprochener Sprache“

Computerlinguistik befasst sich traditionell primär mit Sprache in ihrer geschriebenen Form und die meisten Lehrinhalte computerlinguistischer Lehrveranstaltungen sind auf das schriftliche Medium als Ein- oder Ausgabeformat computerlinguistischer Anwendungen ausgelegt. Der zunehmende Einsatz computerlinguistischer Komponenten in

Speech-Anwendungen lenkt die Aufmerksamkeit jedoch auf die spezifischen Erfordernisse der Verarbeitung gesprochenen Sprache.

Auch wenn man davon ausgeht, dass die Schnittstelle zwischen computerlinguistischen und Sprachsignalverarbeitungs-komponenten weitgehend durch phonetisch-phonologische oder orthographische Symbolketten gekennzeichnet ist, so hat man es an dieser Schnittstelle eines Sprachsynthesystems bei der Erkennung meist mit Mengen alternativer Hypothesen zu tun. Die computerlinguistische Generierung muss bei einer orthographischen Ausgabe an das Sprachsynthese-Modul hinreichend viel Zusatzinformation zur richtigen Disambiguierung nicht homophoner Homographen liefern.

Auch suprasegmentale Merkmale wie Prosodie können für die Analyse wichtige Anhaltspunkte für allen linguistischen Analyse-Ebenen liefern. Umgekehrt ist eine adäquate Synthese auf prosodiesteuernde Information des Generierungsmoduls angewiesen.

Die Bedeutung der für Speech-Anwendungen spezifischen Informationen kann durch Experimente mit Sprachsynthese unmittelbar auditiv erfahren werden. Dabei ist insbesondere an Experimente mit variierender morphologischer Zerlegung, variierender Wortbetonung und Wortklassenzuordnung, unterschiedlichen Phrasierungen und Fokus-Hintergrundstrukturen gedacht. Diese Experimente können zur frühzeitigen Sensibilisierung von Computerlinguisten für Phänomene, die für die gesprochene Sprache charakteristisch sind, beitragen.

Derzeit werden webbasierte Experimente entwickelt, die Benutzern die Manipulation morphologischer Analysen und den Eingriff in die Fokus-Hintergrundstruktur von Äußerungen erlauben.

3. Didaktisches Szenario

Das Internet und die Neuen Medien eröffnen die Möglichkeit, Hochschullehre oder Weiterbildungsseminare rechnergestützt oder sogar rein virtuell über das World Wide Web durchzuführen. Formen virtueller Lehre sind z.B. Web-Based-Trainings oder Online-Seminare. Back et al. (2001) nennen einige Stärken von E-Learning-Konzepten gegenüber herkömmlichen Lehrmethoden. Diese lassen sich – wie in diesem Beitrag deutlich werden wird – als wichtige Merkmale des hier vorgestellten Moduls definieren:

- Vernetzungsmöglichkeiten – Standardisierung und Modulstruktur erlauben die Einbindung unterschiedlicher Zielgruppen; damit wird es möglich, eine Lehrveranstaltung für verschiedene Studiengänge mehrerer Universitäten gemeinsam anzubieten.
- Zugriffsmöglichkeiten – die E-Learning-Infrastruktur (World Wide Web, Lernplattform ILIAS) ist ortsungebunden und rund um die Uhr verfügbar; lern- und informationsbezogene Dienstleistungen können ständig aktualisiert werden.
- Personalisierung – die Lernangebote orientieren sich an individuellen Bedürfnissen der Lernenden, da Lerngeschwindigkeiten und Lernziele nicht als konstant angenommen werden können.

- Aktualisierung – die digitale Datenhaltung ermöglicht es, neue Informationen oder Änderungen just-in-time anzubieten.
- Modularisierung – durch Standardisierung von Lerninhalten wird es möglich, diese in unterschiedlichen Modulen zu kombinieren und sie dem spezifischen Lernbedürfnis der Zielgruppe anzupassen.

3.1. Konzeption

Die einzelnen Teilmodule des Moduls „Gesprochene Sprache“³ sind als Web-Based-Training konzipiert (vgl. auch Haas, 2003). Damit haben die Teilmodule den Charakter eines Selbstlernkurses, erfordern also eigenständiges und exploratives Lernen, das durch intensive tutorielle Betreuung unterstützt wird. Im Vergleich zu einem virtuellen Seminar findet in einem WBT keine strenge synchrone Taktung von Lerneinheiten statt. Stattdessen wird auf individuelle Lerngeschwindigkeiten – die speziell im Bezug auf mathematische Inhalte zu erwarten sind – und auf unterschiedliche Lernziele Rücksicht genommen.

Der größte Vorteil eines Selbstlernkurses ist die Zeit- und Ortsunabhängigkeit beim Bearbeiten der Lernmaterialien. Der Lernende kann frei wählen, zu welchen Zeiten und wo er den Kurs „besucht“. Damit ist ein Web-Based-Training anders als herkömmliche Lehrveranstaltungen nicht an einen speziellen Ort gebunden und kann bequem auch in den eigenen vier Wänden belegt werden (vgl. Lemnitzer / Naumann, in diesem Band). Wie zeitintensiv der WBT-Kurs ist, ist ebenfalls individuell unterschiedlich. Der Lernende sucht sich selbst seinen individuellen „Pfad“ durch die Lerneinheiten, kann demnach beliebig lang bei einem Themenkomplex verbleiben. Erreicht werden soll eine verstärkte Interaktion der Lernenden mit den Lernmaterialien.

Zielgruppe des Bonner MiLCA-Moduls sind Studierende unterschiedlicher Fachrichtungen, insbesondere auch der Computerlinguistik. Es sind Einstiegsmöglichkeiten aus anderen Kontexten als der Sprachsignalverarbeitung zu schaffen, um so den unterschiedlichen Lernzielen gerecht zu werden. Wichtige Voraussetzung dafür ist es, die Kursinhalte in kleine Lerneinheiten einzuteilen. Dies ermöglicht ein den individuellen Bedürfnissen angepasstes gezieltes Ein- und Ausblenden von Lerninhalten. Um das Modul „Gesprochene Sprache“ einem möglichst breitem Spektrum von Studierenden zugänglich zu machen, wird es sowohl in deutscher als auch in englischer Sprache entwickelt.

3.2. Lernpfade

Verschiedene Lernpfade führen durch die Lernmaterialien. Nach einer Kenntnisstandprüfung zu Beginn eines Kapitels sind drei Lernpfade, also unterschiedliche Möglichkeiten

3 Dies sind bisher „Grundlagen der Sprachsignalverarbeitung“, „Sprachsynthese“, „Linguistische Aspekte der Verarbeitung gesprochener Sprache“ sowie „Mathematische Grundlagen“.

des Navigierens durch den Kursinhalt wählbar: Überspringen des Kapitels, das Lesen einer Zusammenfassung oder Einstieg in einen „Intensivkurs“ und damit Zugang zu allen Lerneinheiten des Kapitels. Nach der Bearbeitung eines Kapitels erfolgt eine Lernerfolgskontrolle in Form von Übungsaufgaben, deren nicht korrekte Beantwortung den Studierenden auf einzelne, bereits bearbeitete Lerneinheiten zurückverweist. Alternativ kann der Lernende in der Bearbeitung der Lernmaterialien fortfahren und zum nächsten Kapitel des Teilmoduls wechseln.

3.3. Vernetzung

The screenshot shows a web browser window titled "MILCA - Grundlagen der Sprachsignalverarbeitung - Mozilla". The page content is organized into a table with two columns: a term and its definition. The terms listed are "Anregung", "Artikulator", "Artikulation", and "Bandbreite". Each term is accompanied by a diagram. The "Anregung" diagram shows a complex waveform labeled "Sprechsignal" and a smoother waveform labeled "Anregungssignal (Stimmtonschwingung)" with a period T_0 indicated. The "Bandbreite" diagram shows a waveform $x(t)$ and its corresponding frequency spectrum $|X(f)|$ with a 3 dB bandwidth Δf marked.

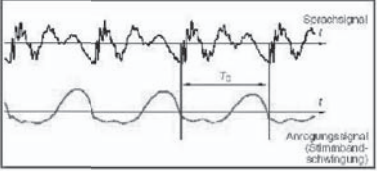
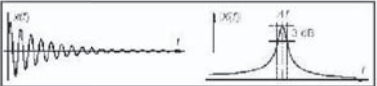
Anregung	<p>Unter Anregung verstehen wir den Vorgang, bei dem im Luftstrom, der den Sprechtrakt durchströmt, Schwingungen oder Geräusche erzeugt werden, die dann als akustische Signale, d.h. als Schall abgestrahlt werden. Beim Sprechen greift der Mensch hauptsächlich auf drei Grundarten der Anregung zurück: die stimmhafte, die stimmlose und die transiente Anregung.</p> 
Artikulator	<p>Sprechorgan, das an der <u>Artikulation</u> beteiligt ist. Ein Artikulator kann aktiv (z.B. Bewegung in Richtung eines anderen Artikulators) oder passiv sein.</p>
Artikulation	<p>Gesamtheit der Bewegungen der Organe des Sprechtrakts und ihre zeitliche Aufeinanderfolge.</p>
Bandbreite	<p>Ist definiert als die Breite des Frequenzbereichs (in Hz), der durch zwei Messpunkte an den beiden Kurvenschenkeln festgelegt ist. Dabei liegen die Messpunkte genau 3 dB (Dezibel) unter dem höchsten Punkt der Kurve (z.B. 3 dB unterhalb der Spitzenamplitude). Die Bandbreite gibt man z.B. bei Filtern oder bei Formanten an.</p> 

Abbildung 4: Ausschnitt aus dem Glossar des Teilmoduls „Grundlagen der Sprachsignalverarbeitung“; das Glossar bildet ein eigenes Element innerhalb des WBTs

Die einzelnen Teilmodule innerhalb des Moduls „Gesprochene Sprache“ sind untereinander vernetzt. Das Teilmodul „Linguistische Aspekte der Verarbeitung gesprochener Sprache“ als Schnittstelle zum Gebiet der Computerlinguistik ist gleichzeitig auch Anknüpfungspunkt zu anderen Lernmodulen, die im Projekt MiLCA entwickelt werden. Auch aufgrund dieser Vernetzung ist die Einteilung in Lerneinheiten zwingend, um einzelne Lerneinheiten von anderen MiLCA-Lernmodulen aus individuell ansteuern zu können.

Die Teilmodule umfassen verschiedene Kursmaterialien, die online verfügbar gemacht werden. Darunter fallen neben konkreten Inhaltselementen wie Texten, Graphiken, Animationen etc. auch jeweils ein Glossar, ein Literaturverzeichnis und eine Formelsammlung. Diese als unabhängig betrachtbaren Elemente sind an inhaltlich relevanten Stellen mit den übrigen Kursinhalten verlinkt. Mathematische Grundlagen sind in einem eigenen Teilmodul zusammengestellt, das ebenfalls innerhalb des WBTs mit den Inhalten der Sprachsignalverarbeitung bzw. der Sprachsynthese verzahnt ist.

3.4. Navigation und Kommunikation innerhalb der Lernplattform

Im MiLCA-Projektverbund war zunächst der generelle Einsatz der Lernplattform ILIAS⁴ geplant. Die Lernplattform ILIAS ist jedoch bisher (bis ILIAS 2) nur partiell geeignet, das Modul zu unterstützen. Sie bietet zur Zeit noch nicht die Möglichkeit, in XML archivierte Daten direkt zu importieren, so dass eine Implementierung der Modulinhalt immense Zeitaufwand bedeuten und keine Metadaten beinhalten würde. Zudem wurden die mathematischen Inhalte des Moduls mittels MathML codiert; die dafür notwendige Unicode-basierte Anzeige wird in ILIAS nicht angeboten. In näherer Zukunft ist mit ILIAS 3 aber mit einer Import-Funktion für XML zu rechnen; dies stellt eine vollständige Einbindung der Kursinhalte in die Lernplattform in Aussicht. Die Planungen sehen vor, die Lernplattform nicht nur als Kommunikationswerkzeug, sondern auch als Portal zu den Kursmaterialien, also zu Lerneinheiten, Übungen, Glossar und Literaturverzeichnis einzusetzen.

Bis zur Verfügbarkeit einer ILIAS-Version, die alle benötigten Features unterstützt, wird mit einer Eigenentwicklung als Lernplattform gearbeitet, ILIAS wird zur bidirektionalen Kommunikation zwischen Studierenden und der Dozentin genutzt. Die aus der XML-Form nach XHTML konvertierten Dateien werden in eine Navigationsumgebung eingebunden: Eine Kursseite ist stets in zwei Bereiche (Frames) unterteilt. Die Inhaltsseiten im rechten Frame nehmen den Hauptteil der Bildschirmoberfläche ein. Die Navigationsleiste im linken Frame gibt dem Nutzer durch ein dynamisches Inhaltsverzeichnis Orientierung darüber, wo er sich im Kursinhalt befindet. Daneben sind hier Links zu den Inhaltselementen Glossar, Literaturverzeichnis und Formelsammlung zu finden. Das inhaltlich unterstützend wirkende Teilmodul „Mathematische Grundlagen“ ist ebenso wie die Übungsaufgaben von der Navigationsleiste aus erreichbar. Zusätzlich ist hier ein direkter Zugang zu den verschiedenen, den Kursinhalt und Kursablauf betreffenden Diskussionsforen geschaffen.

Die Kommunikation innerhalb der Kurse – sowohl der Studierenden untereinander als auch zwischen Studierenden und Tutor – findet über die Lernplattform ILIAS statt. Die Kommunikationsmöglichkeiten im WBT umfassen Mail und Diskussionsforen.

4 URL: <http://www.ilias.uni-koeln.de>

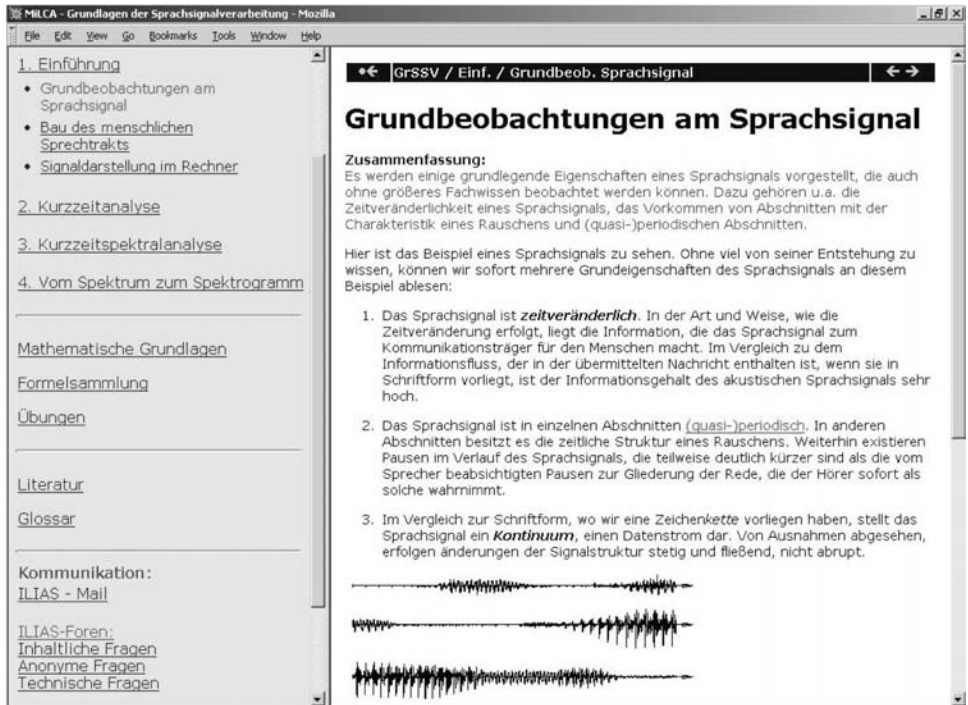


Abbildung 5: Navigationsleiste (linker Frame) und Inhaltsseite (rechter Frame) im WBT „Gesprochene Sprache“

4. Annotierung / Codierung der Lernmaterialien

Die Codierung der Lernmaterialien erfolgt mittels eines MiLCA-einheitlichen Standards, der in einem anderen Beitrag dieses Themenhefts dargestellt wird. Die Orientierung an Standards gewährleistet, dass „Inhalte aus verschiedenen Quellen beliebig kombinierbar sind [und] austauschbare Inhalte entwickelt werden, die leicht wieder verwendet, beliebig ergänzt oder reduziert werden können“.⁵

Die Lernmaterialien werden in XML codiert. Grundlage ist eine DTD, die im Projekt MiLCA entwickelt wurde („GILES“, vgl. Stührenberg, in diesem Band) und sich am LOM-Standard⁶ orientiert. LOM ist ein von einer Arbeitsgruppe des IEEE⁷ entwickelter Standard zur Beschreibung von Lernobjekten mittels Metadaten. Dadurch ist die flexible Handhabung des Lernmoduls in Bezug auf Wiederverwendbarkeit, auch hinsichtlich einer

5 Zitat aus URL: http://www.medien-bildung.net/standards/standards_uebersicht_db.php/spezialthemen/

6 IEEE 1484.12.1-2002, Draft Standard for Learning Object Metadata vom 15.07.2002, URL: http://tsc.ieee.org/doc/wg12/LOM_1484_12_1_v1_Final_Draft.pdf

7 URL: <http://tsc.ieee.org/wg12/>

Nutzung auf anderen Lernplattformen, sichergestellt. Zudem garantiert die innerhalb des gesamten MiLCA-Projekts einheitliche Annotierung der Materialien den problemlosen Austausch einzelner Lerneinheiten.

Mathematische Inhalte sind in MathML codiert. Zwar ist die Anzahl der Browser, die dieses Format interpretieren können, heute noch begrenzt⁸ – und auch die Annotierung ist komplexer und aufwändiger als eine graphische Realisierung einer mathematischen Formel. Die Codierung mit MathML bietet aber vielfältige Vorteile: Veränderungen der Zeichengröße mittels der Zoomfunktion des Browsers sind ebenso möglich wie einfache Layoutanpassungen durch Stylesheets, die eine individuelle Ansteuerung und damit eine voneinander unabhängige Gestaltung verschiedener Größen in der Mathematik zulassen, z.B. von Variablen, Indizes, Konstanten, etc. Zudem ist der Code einer mathematischen Formel im Browser anzeigbar und wieder verwendbar – sowohl für den Autor als auch für den Nutzer.

Das häufige Vorkommen mathematischer Variablen und Operatoren, aber auch phonetischer Zeichen stellt Anforderungen an die Codierung, die nur durch Unicode als einheitlichem Zeichenkodierungsstandard erfüllt werden können. Sämtliche Lehrmaterialien des Moduls wurden deshalb Unicode-konform codiert.

```

4530: <Content>
4531:   <Text>□
4532:   <Paragraph>Das lineare Modell der Spracherzeugung wird  $f$  durch die 5
stimmhafte Sprachsignale (auf diese wollen wir uns hier beschreiben) durch die
Periodizität  $t$ , Klangfarbe der Stimme, Amplitude, Vokaltrakt und
Abstrahlung gebildet (ssv\_s241.xhtml: Einfache Modellierung des
Sprechtrakts  $f$  stimmhafte Signale). Weist man diesen Komponenten
entsprechend der formelsammlung.xhtml#Def\_3.30">Formelsammlung Def 3.30a, b
übertragungsfunktionen zu, die wir mit
4533:   <math>□
4534:     <math>□
4535:       <math>P</math>□
4536:       <math>□</math>□
4537:       <math>□
4538:         <math>□</math>□
4539:         <math>m</math>□
4540:         <math>□</math>□
4541:       </math>□
4542:     </math>□
4543:   </math>□
4544: </Formule>□
4545:  $f$  die Periodizität  $t$ ,
4546: <math>□
4547:   <math>□
4548:     <math>□
4549:       <math>G</math>□
4550:       <math>□</math>□
4551:       <math>□
4552:         <math>□</math>□
4553:         <math>m</math>□
4554:         <math>□</math>□
4555:       </math>□
4556:     </math>□

```

Abbildung 6: Beispiel einer Codierung in MathML

8 Das WBT wird zur Zeit unter Verwendung des Browsers „Mozilla“ durchgeführt, der die Anzeige von MathML-Daten ermöglicht.

5. Einsatz multimedialer Elemente

Durch den Einsatz der Neuen Medien sollen Lernprozesse optimiert und persönlichen Bedürfnissen angepasst werden. Die Darstellung komplexer Sachverhalte geschieht daher mit Hilfe verschiedener Medien: Texte, Graphiken, Audiosignale, Animationen, Simulationen. Theoretische Inhalte sollen auf diese Weise konkretisiert und veranschaulicht werden. Speziell in der Sprachverarbeitung empfiehlt sich der verstärkte Einsatz von Audiosignalen. Visuelle und auditive Hilfen sollen gewährleisten, ein besseres Verständnis physikalisch-akustischer Grundgrößen und Zusammenhänge zu erzielen.

So kann beispielsweise ein Grundbegriff wie die Frequenz auf unterschiedliche Weise erfahren werden:

- *textlich* als Begriffsdefinition;
- *formal* als mathematische Gleichung (Frequenz als Kehrwert der Periodendauer);
- *graphisch* mittels der Darstellung eines Signals mit eingezeichneter Periodendauer;
- *auditiv* mithilfe der Hörbarmachung von Signalen unterschiedlicher Frequenz.

Ziel des WBTs ist es, die verschiedenen Medien nicht als voneinander unabhängige Elemente zu betrachten, sondern sie als Einheit zu verstehen, aus der sich der Lernende diejenigen Elemente auswählen kann, die ihn im Lernprozess am besten unterstützen.

5.1. Mathematische Formeln

Wichtige mathematische Formeln bzw. Gleichungen werden durchnummeriert und sind mittels einer vergebenen ID eindeutig referenzierbar. Anhand dieser ID erhält jede im Text vorkommende Definition einen Verweis an die entsprechende Marke in der Formelsammlung, welche bei Aktivierung des gesetzten Links in einem eigenen Browserfenster angezeigt wird.

5.2. Graphiken

Graphiken sind im Fließtext aufgrund der angestrebten Übersichtlichkeit eher klein gehalten. Dem Studierenden bietet sich die Möglichkeit, sich die Abbildung in einem neuen Fenster als Vektorgraphik anzeigen zu lassen. Auf diese Weise ist die Graphik auf individuellen Wunsch in der Größe veränderbar. Ein weiteres Gestaltungselement bilden verlinkte Graphiken. Eine solche Graphik kann direkt auf Tonbeispiele, Glossareinträge, Literaturstellen oder andere Lerneinheiten verweisen.

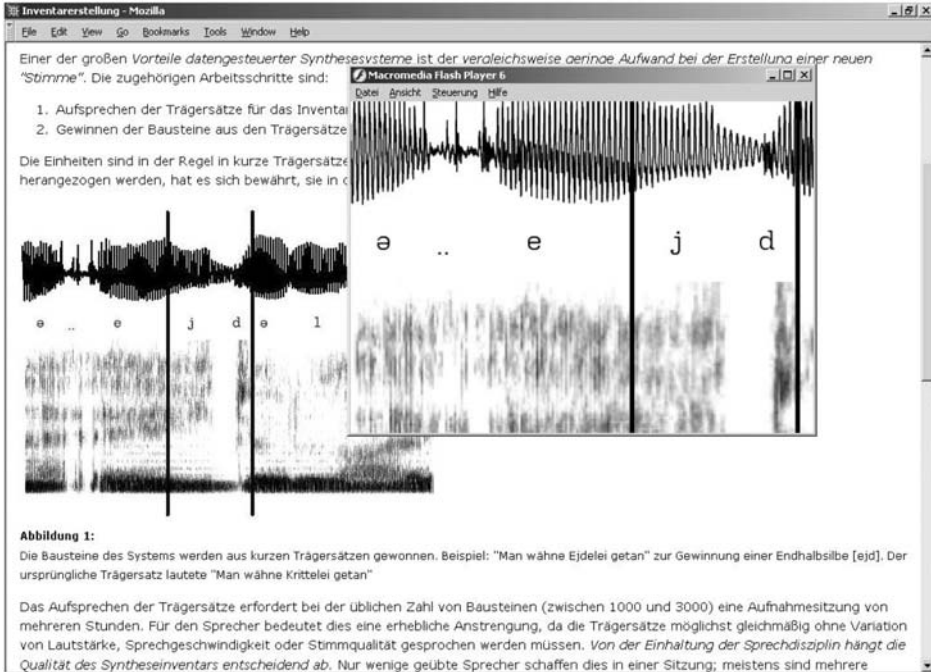


Abbildung 7: Beispiel einer Graphik aus dem Teilmodul „Sprachsynthese“ – die Vektorgraphik ermöglicht eine Ansicht im Detail

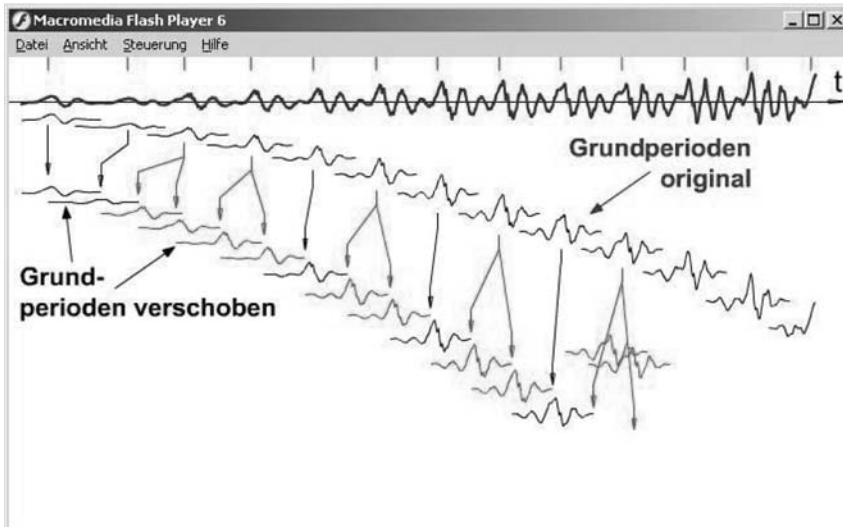


Abbildung 8: Beispiel einer Flash-Animation aus dem Teilmodul „Sprachsynthese“ – einzelne Schritte einer Signalmanipulation (PSOLA-Verfahren) werden dadurch anschaulicher

5.3. Interaktive Elemente

Physikalisch-akustische Abläufe oder Prozesse in der Signalverarbeitung werden im WBT mithilfe von animierten Graphiken, Flash-Animationen, Java-Applets und Demonstrationsprogrammen dargestellt. Dadurch können beispielsweise einzelne Schritte der Manipulation eines Sprachsignals in ihrem zeitlichen Ablauf durch eine Animation erläutert werden, die vom Lernenden selbst gesteuert wird. Visualisierung und Interaktivität sind didaktisch gesehen sehr wichtige Methoden, um logische und zeitliche Zusammenhänge besser verständlich zu machen.

6. Ausblick

Bisher wurden die innerhalb von MiLCA entwickelten Lehrmodule für den Lehrbetrieb unter den Projektpartnern ausgetauscht. Geplant ist jedoch ein weiter gehender Einsatz der Module, der von einer Verwertungsgesellschaft, der mehrere MiLCA-Partner angehören, koordiniert wird. Die im Rahmen des Projekts durchgeführte Evaluation der in MiLCA als Lehrveranstaltungen abgehaltenen Module soll deren Stärken und Schwächen in Konzeption, Aufbau und Inhalt aufdecken und erlaubt so eine gezielte Verbesserung der Lehrmaterialien sowie des Kursablaufs für den Einsatz in der virtuellen Lehre.

6.1. Nutzungsszenarien

Für ein Lehr- und Lernmodul, wie es in diesem Beitrag vorgestellt wurde, sind speziell in der Hochschullehre verschiedene Nutzungsszenarien denkbar. Dazu gehören:

- der Einsatz *als reguläre Lehrveranstaltung* mit Beteiligung Studierender mehrerer Universitäten als echter Selbstlernkurs mit tutorieller Unterstützung
- der parallele Einsatz des WBT *als Ergänzung zu einer zeitgleich stattfindenden Präsenzveranstaltung* als Möglichkeit zur Vor- und Nacharbeit
- der Einsatz *als Repetitorium zu in einer Lehrveranstaltung besprochenen Lerninhalten* als Wiederholung und/oder Prüfungsvorbereitung

Virtuelle, multimediale Lernmodule stellen eine Lehrform der Zukunft dar, die nicht nur an Universitäten, sondern auch im Weiterbildungsbereich verstärkt eingesetzt werden wird. Allerdings lösen sie in dem hier beschriebenen Szenario die Dozentenlehre nicht ab, sondern ergänzen sie in einer zeitunabhängigen anschaulichen, explorativ-multimedialen Form.

Literatur

- Back, Andrea; Bendel, Oliver; Stoller-Schai, Daniel (2001): „E-Learning im Unternehmen. Grundlagen – Strategien – Methoden – Technologien“. Orell Füssli Verlag, Zürich.
- Haas, Felicitas (2003): „Methoden und Einsatz des Web-Based-Trainings „Grundlagen der Sprachsignalverarbeitung“ in der virtuellen Lehre“, in: Fortschritte der Akustik – DAGA 2003, Aachen.