Hierarchische kognitive dynamische Systeme zur Sprach– und Signalverarbeitung

Matthias Wolff, Rüdiger Hoffmann, Ronald Römer

***Zusammenfassung*—In diesem Positionspapier stellen wir ausgehend von klassischen Erkenntnissen der Sprach– und Signalverarbeitung das im letzten Jahrzehnt von uns entwickelte „Einheitliche System zur Sprachsynthese und –erkennung“ (UASR) vor. Dieses steht in enger Beziehung zu neueren Ansätzen in der Systemtheorie, namentlich den kognitiven dynamischen Systemen. Diese lassen bislang jedoch weitestgehend die für eine „Erkenntnisfähigkeit“ unerlässliche hierarchische Modellierung und Systemstruktur außer acht (obwohl deren Notwendigkeit als unumstritten gelten kann). Die Konstruktionsprinzipien unseres Systems sind: *hierarchische* Struktur, gemeinsame Daten für Analyse und Synthese sowie einheitliche Algorithmen auf allen Ebenen. Wir argumentieren weiterhin, dass kognitive (Sprach–) Kommunikationsysteme ein inneres Modell ihres Kommunikationspartners haben müssen und zeigen, wie dieses auf naheliegende Weise technisch realisiert werden kann. Sowohl die hierarchische Struktur als auch die „Spiegelung“ des Kommunikationspartners sind durch die Neurobiologie beim natürlichen Vorbild belegt. Als technologische Basis schlagen wir endliche Transduktoren (*finite state transducers*) vor. Wir haben gezeigt, dass diese für unsere Aufgabe universell sind und prinzipiell auf allen Verarbeitungsebenen von der akustischen bis hin zur pragmatischen für Analyse und Synthese eingesetzt werden können. Abschließend geben wir ein Konzept für hierarchische kognitive dynamische Sprach- und Signalverarbeitungssysteme an und benennen die aus unserer Sicht künftig interessanten Forschungsthemen.**

I. UASR – Einheitliches System zur Sprachsynthese und –erkennung

A. Einordnung

Der von uns verfolgte Ansatz befasst sich im weiteren Sinne mit der Weiterentwicklung des Prinzips der „Analyse durch Synthese“ (*analysis by synthesis*, AbS). Historisch betrachtet, taucht diese Bezeichnung bereits in der Frühzeit der elektronischen Sprachsignalverarbeitung auf. Ein Schlüsselbeitrag aus dem Jahre 1961 [1] bietet eine Übersicht über die damaligen Quellen sowie eine strategische Diskussion des Verfahrens, das anhand der praktischen Aufgabenstellung veranschaulicht wird, reale Sprachspektren mit einer Menge von Spektren zu vergleichen, die innerhalb des Analysesystems synthetisiert werden.

Es fällt auf, dass nicht nur die historischen, sondern auch die modernen Beispiele für die Anwendung des AbS-Prinzips im Bereich der signalnahen Verarbeitung angesiedelt sind (vgl. die breite Anwendung in heutigen Sprach-Codecs). Dabei ist schon zeitig klar gewesen, dass die hierarchische Strukturierung der humanen Sprachverarbeitung in den angestrebten technischen Modellen (Spracherkennungs– und –synthesesystemen) berücksichtigt werden muss [2].

UASR (*Unified Approach for Speech Synthesis and Recognition*) stellt eine in dem vergangenen Jahrzehnt an der TU Dresden implementierte Möglichkeit dar, den AbS-Ansatz auf einer hierarchisch organisierten Plattform umzusetzen. Das Projekt wurde hauptsächlich durch die folgenden DFG-Fördervorhaben getragen, die durch eine Anzahl von Anwendungsprojekten flankiert wurden:

- 1997 - 2000: Strukturelles Training hierarchisch organisierter Aussprachewörterbücher (Ho 1674/3)
- 2001 - 2005: Integration von Spracherkennung und –synthese unter Verwendung gemeinsamer Datenbasen (Ho 1674/7)
- 2004 - 2007: Entwicklung von Datenanalyseverfahren für die Qualitätsbewertung technischer Prozesse basierend auf spektralen Repräsentationen akustischer Vorgänge (Ho 1674/8 und He 3656/1)

B. Ein hierarchischer Analyse-Synthese-Ansatz

In einer der klassischen Arbeiten des Fachgebietes [3] werden zwei Grundprinzipien für sprachverarbeitende Systeme wie folgt formuliert:

- „The first fundamental hypothesis is based on consistent evidence [...] indicating that the primary information-bearing attitude of the speech signal is the temporal variation of the short duration amplitude spectrum.“
- „The second foundational rule, which is borne out by the results of [...] psychophysical experiments [...], asserts that speech is a composite signal, hierarchically organized so that simpler patterns at one level are combined in a well-defined manner to form more complex patterns at the succeeding level. [...] The structures at each level of the hierarchy serve to constrain the ways in which the individual patterns associated with that level can be combined.“

Natürlich sind Spracherkenner und –synthetisatoren hierarchisch gegliedert. Spracherkenner profitieren von der Redundanzminderung, die beim Übergang von einer Ebene zur nächsthöheren erfolgt. Zum Beispiel ermöglicht die Existenz eines Aussprachewörterbuches auf der Wortebene, dass aus den zahlreichen auf der Ebene der Einheiten (Laute) hypothetisierten Phonemfolgen diejenigen verworfen werden können, die lexikalisch keinen Sinn ergeben. Auf der Gegenseite, der Sprachsynthese, muss in einem „Text-to-Speech-System“

M. Wolff: Brandenburgische Technische Universität Cottbus, vorher Technische Universität Dresden
R. Römer: Brandenburgische Technische Universität Cottbus, vorher Nuance Communications GmbH
R. Hoffmann: Technische Universität Dresden

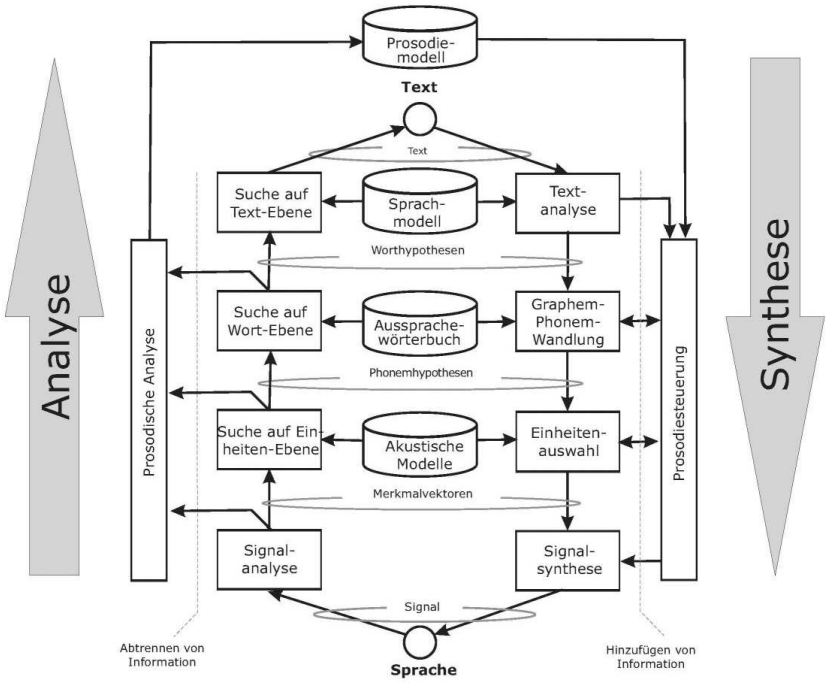

Bild 1: Grundkonzept des UASR (*Unified Approach for speech Synthesis and Recognition*), das seit 2000 an der TU Dresden implementiert wird [6], in der damals aktuellen Terminologie.

(TTS-System) ein gegebener Text linguistisch analysiert, aus einer Schrift– in eine Lautsymbolfolge umgesetzt und schließlich in ein akustisches Signal umgesetzt werden. Für den Übergang von der Schrift– zur Lautebene braucht man entweder ein Regelwerk oder ein Aussprachewörterbuch, wie es auch ein Erkenner benutzt.

Spracherkennung und Sprachsynthese haben sich ursprünglich weitgehend unabhängig voneinander entwickelt. Wie das kleine Beispiel aus dem vorangehenden Abschnitt stellvertretend zeigt, hat es sich nach und nach herausgestellt, dass beide Arbeitsgebiete auf ihren verschiedenen Hierarchieebenen die gleichen Datenbasen wie Lexika, Sprachmodelle usw. benötigen und damit über diese verbunden sind. Obwohl sich diese Erkenntnis jetzt erst in der Breite durchsetzt, ist ihre Bedeutung bereits Ende der 1980er Jahre von J. N. Holmes erkannt worden, wie die beiden folgenden Zitate aus [4] belegen:

- „Advanced systems both for synthesis and for recognition need the same speech knowledge, and there is considerable advantage for the two applications to be studied together."
- „I predict that the most significant progress in the more advanced forms of speech synthesis and recognition will in future come from research teams with a strong interest in both problems."

An der TU Dresden sind die seit den 1960er Jahren betriebenen Arbeiten zur Spracherkennung und –synthese unter den veränderten Bedingungen der 1990er Jahre durch die Beteiligung an dem Verbundprojekt VERBMOBIL [5] in beiden Teildisziplinen stark forciert worden. Die damit verbundene stärkere Hinwendung zu den sprachtechnologischen Wissensbasen, die für Erkenner und TTS-Systeme immer ähnlicher wurden, führte schließlich zu der Idee einer Verschmelzung beider Systeme über gemeinsam genutzte Datenbasen. Dabei entsteht zwangsläufig ein hierarchisch strukturiertes Analyse-Synthese-Schema, das in Bild 1 dargestellt ist. Die Notwendigkeit, ein solches Forschungssystem zu entwickeln, stellte sich Ende der 1990er Jahre heraus. Das Konzept wurde erstmals auf einer Konferenz im Jahre 2000 vorgestellt [6] und in den Folgejahren konsequent ausgebaut [7]. Es erhielt die Bezeichnung „Unified Approach for Speech Synthesis and Recognition" (UASR) und sollte den folgenden Forschungszielen dienen:

- Verbesserung des Verständnisses der Vorgänge bei der Spracherkennung und Sprachsynthese durch Ausbau einer Plattform, die eine Anwendung des Prinzips der Analyse durch Synthese auf verschiedenen Hierarchieebenen ermöglicht.
- Verbesserung des bisher nur gering ausgebauten Verständnisses der Gründe, warum Spracherkenner Fehler machen, durch Aufbau eines Erkennerzweiges mit transparenter Struktur und der begleitenden Möglichkeit, (auch akustisch) zu bewerten, was ein Erkenner eigentlich „hört".
- Schaffung einer Plattform für die Sprachsynthese auf der Basis statistisch trainierter akustischer Modelle (meist als HMM-Synthese bezeichnet), der ein hohes Qualitätspotential zugeordnet wird,
- Schaffung von verbesserten Demonstratoren für die Lehre im Bereich der Signalerkennung und Sprachtechnologie,
- Bereitstellung einer Toolbox für die Gewinnung von Ableitversionen für praktische Anwendungen in der Spracherkennung und –synthese.

UASR war zunächst als reines Sprachverarbeitungssystem konzipiert. Entsprechend der Zielstellung bestand die allgemeine Forderung, dass die Analyse– und Synthesemodule auf einer Hierarchieebene zueinander funktionell invers sein sollen. Die dazu verwendeten Algorithmen entsprachen denen, die in separaten Erkennungs– und TTS-Systemen üblich waren.

Durch die Implementierung von UASR wurde die Notwendigkeit einer vereinheitlichten Algorithmik deutlich. Betrachtet man die verschiedenen benötigten Funktionen, unterscheiden sie sich zunächst stark, da sie auf unterschiedlichste Daten wie HMMs (*Hidden Markov Models*), Wortlisten, reguläre oder stochastische Grammatiken oder Lexika zurückgreifen müssen. Die Systemtheorie zeigt zwar schon lange, dass sich diese unterschiedlichen Funktionen einheitlich in der Sprache endlicher Automaten (FSM, *finite state machines*, oder auch FST, *finite state transducers*) beschreiben lassen [8], jedoch sind erst in den letzten Jahren Bibliotheken entstanden, deren Umfang so groß ist, dass sie in der Sprachverarbeitung erfolgreich eingesetzt werden können. Für UASR wurde eine solche Bibliothek geschaffen und erfolgreich angewendet; ein Überblick kann [7] entnommen werden.

In [21] konnte gezeigt werden, dass die FST-Technologie die HMM-Modellierung einschließt und daher als das allgemeinere Modellierungskonzept verstanden werden kann.

C. Ergebnisse

In [16] findet sich eine Übersicht über einige mit UASR gewonnene Ergebnisse. Erwähnenswert sind Erkennerrealisierungen, Anwendungen in der parametrischen Sprachsynthese (HMM-Synthese) und bei der flexiblen Anwendung von Sprachmodellen und Aussprachewörterbüchern [9]. Die in der Sprachsynthese-Gemeinde rapide zunehmende Orientierung auf die HMM-Synthese hat übrigens dazu geführt, dass die enge Verbindung von Analyse und Synthese als Forschungsgegenstand jetzt breit akzeptiert wird [10].

Zwei aktuelle Anwendungskomplexe von UASR sollen noch erwähnt werden:

- Obwohl UASR als Forschungsplattform konzipiert ist, lassen sich attraktive Anwendungen ableiten. Derzeit entsteht ein kombiniertes Spracherkennungs–/–synthesesystem in Form eines USB-Sticks, das als Kommandointerface für beliebige intelligente Messgeräte verwendet werden kann. In Kooperation mit dem Fraunhofer-Institut IZFP in Dresden wurde ein Prototyp auf FPGA-Basis geschaffen [11], [12].
- Obwohl UASR als Sprachverarbeitungsplattform konzipiert ist, lassen sich die darin implementierten Erkennungsverfahren sehr effektiv auf die Erkennung nonverbaler akustischer Signale anwenden (z. B. [13], [14]). Das bestätigt, dass die strukturelle Erkennung nachdrücklich in Bereiche vordringt, in denen bislang heuristische Ansätze und numerische Erkennungsverfahren dominierten [15]. Von seinen zahlreichen Anwendungen in der akustischen Signalerkennung hat UASR auch im Hinblick auf seine theoretischen Grundlagen deutlich profitiert [16].

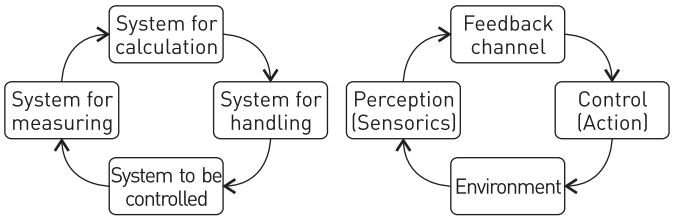

Bild 2: Analogie zwischen dem klassischen Ansatz des Regelkreises (links, nach [25] in der Terminologie der 1970er Jahre) und dem vereinfachten Schema der humanen Kognition, das von S. Haykin als Ansatz für die Klasse der kognitiven dynamischen Systeme verwendet wird (rechts, nach [26]).

Das UASR-Projekt bildete die Grundlage für mehrere Dissertationen [17]–[20]. Die Zusammenfassung der etwa ein Jahrzehnt umfassenden Arbeiten erfolgte unlängst in der Habilitationsschrift von M. Wolff [21], so dass eine umfassende Dokumentation vorliegt.

II. Kognitive Dynamische Systeme

A. Die Weiterentwicklung der (informationstechnischen) Systemtheorie

Die Theorie der Signale und Systeme in der Form, in der sie in der Informationstechnik gepflegt wird, wurde weitgehend durch die Nachrichtentechnik initiiert [22]. Dank ihres Potentials zur Integration unterschiedlichster Teildisziplinen aus Naturwissenschaft und Technik hat sie sich zu einem Grundlagenfach entwickelt, das an den elektrotechnischen Fakultäten unter unterschiedlichen Überschriften gelehrt wird. Aus der Sicht der Elektrotechnik und Informationstechnik bestanden die wichtigsten Leistungen der Systemtheorie im Ausbau der Theorie der linearen zeitinvarianten Systeme und in deren Erweiterung auf zeitdiskrete Systeme, die zur theoretischen Grundlegung der digitalen Signalverarbeitung geführt hat. Für die Weiterentwicklung der Systemtheorie haben sich die folgenden Gesichtspunkte als besondere Triebkräfte erwiesen:

- Erweiterung der theoretischen Basis zu einer allgemeinen Theorie dynamischer Prozesse [23], [24]
- Erweiterung auf nichtlineare Systeme
- Erweiterung auf zeitvariante Systeme
- Erweiterung auf Systeme mit mehreren Ein- und Ausgängen (MIMO-Systeme)

Weiterhin haben solche Systeme an Bedeutung gewonnen, deren Funktion auf etwas unscharfe Weise mit dem Begriff „intelligentes Verhalten“ umschrieben wird. Dazu gehören Systeme zur Mensch-Technik-Interaktion mit Nutzung unterschiedlicher Medien (Sprache, Bilder, Haptik etc.) oder Systeme, die sich in unbekannten Umgebungen orientieren können (Szenenanalyse, Robotik). Wichtig für diese Systeme ist, dass sie Entscheidungen treffen und ggf. ihr Verhalten aufgrund dieser Entscheidungen zielgerichtet verändern können. Im Gegensatz zu den klassischen Objekten der Systemtheorie bildet bei ihnen also die statistische Lern- und Entscheidungstheorie eine unverzichtbare Grundlage. Offensichtlich hat diese Systemklasse Verbindung zu den regelungstechnischen Wurzeln der Systemtheorie, wie der Vergleich in Bild 2 zeigt.

Beim Menschen stehen naturgemäß die auftretenden Interaktionen zwischen Gehirn und Umwelt im Mittelpunkt der Betrachtungen, sie werden nach Bild 2 als Rückkopplungen verstanden. Diese werden sowohl in der Technik als auch in der Biologie als ein fundamentales Prinzip angesehen; ohne sie wäre kein zielgerichtetes Systemverhalten möglich. Nur wenn die Kommunikationsteilnehmer über Zielvorstellungen verfügen, welche zueinander kompatibel sind, können diese unter Verwendung von Rückkopplungen abgeglichen werden. Kognitive Systeme müssen daher ihre Umgebung wahrnehmen können (Analyse), Entscheidungen unter Unsicherheiten treffen (Steuerung) und zielgerichtet auf ihre Umgebung einwirken können (Synthese). Darüber hinaus ist es notwendig, statistische Modelle für Vorhersagen und Regeln für das Verhalten zu entwickeln.

S. Haykin hat für Systeme, die wie der Mensch ein zielgerichtetes Verhalten aufweisen, die Bezeichnung kognitive dynamische Systeme geprägt [26]. Sie sind durch ihre Sensoren in der Lage, ein internes Modell ihrer Umwelt zu entwickeln und auf dieser Basis gezielt auf die Umwelt einzuwirken.

Überraschenderweise finden sich Anwendungen der kognitiven dynamischen Systeme nicht nur in den Bereichen, in denen „künstliche Intelligenz“ traditionell angesiedelt ist, wozu natürlich auch die Sprachverarbeitung gehört. Haykin sieht ein viel breiteres Anwendungsfeld in einer „kognitiven Signalverarbeitung“. Ausgearbeitete Beispiele sind das kognitive Radar [27] und das kognitive Radio [28], [29].

Unter *cognitive radio* versteht man allgemein ein drahtloses Nachrichtensystem, das seine Parameter dynamisch so einstellt, dass es möglichst effektiv funktioniert. Im engeren Sinne erfolgt eine dynamische Ausnutzung des Spektrums, indem momentan freie Spektralbereiche identifiziert und dynamisch genutzt werden. Das gesamte Gebiet entwickelt sich sehr intensiv; eine Momentaufnahme des damals aktuellen Forschungsstandes lieferte 2008 ein Themenheft des IEEE Signal Processing Magazine [30].

Ein weiteres Beispiel für ein kognitives System wurde von S. Young in [43] vorgestellt. Hierbei handelt es sich um ein Dialogsystem, das ebenfalls auf einem statistischen Modell basiert. Auch dieses System kann als kybernetischer Kreislauf bzw. als Regelkreis verstanden werden. Hier liegt der Schwerpunkt jedoch noch stärker auf der Dialogsteuerung (Regeleinrichtung). In diesem Artikel wurden wichtige Eigenschaften von Dialogmanagementsystemen hervorgehoben; so sollte eine kognitive Steuerung die Fähigkeit zum Schließen und Folgern besitzen und auch Mehrdeutigkeiten aus dem Kontext heraus auflösen können. Weiterhin muss die Steuerung über klar definierte Kommunikationsziele verfügen und sollte auch unter Verwendung von unsicherer Information in der Lage sein, Pläne zur Erreichung des Ziels erstellen zu können. Schließlich wird die Anpassung an die Umgebung sowie Lernfähigkeit und Robustheit gefordert.

Ein weiterer Aspekt, der Auswirkungen auf die Struktur von Kommunikationssystemen haben kann, ist die Richtung der Informationsflüsse. Im Kanalmodell nach C. E. Shannon wird der Informationsfluss nur in einer Richtung verfolgt: vom Sender zum Empfänger. Bei einer Mensch-Maschine-Kommunikation liegt aber ein Informationsfluss in beiden Richtungen vor. Kognitive Kommunikationssysteme müssen daher sowohl über Sende- als auch über Empfangsfunktionen verfügen. Die informationstheoretische Beschreibung eines bidirektionalen Kommunikationsmodells wurde erstmalig in [48] veröffentlicht. Aus den frühen 1970-Jahren stammt ein Modell, mit dem Kommunikationsteilnehmer auf der semiotischen Ebene beschrieben werden können [49]. Im Gegensatz zum einseitig gerichteten Kommunikationsmodell nach Shannon, in welchem ein Sender verbunden durch einen Kanal einem Empfänger gegenübersteht, wird das informationsempfangende, –verarbeitende und –aussendende System (von D. Nauta als I-System bezeichnet) als Ganzes – als Kommunikation*steilnehmer* – betrachtet. Das von Nauta vorgeschlagene Grundmodell basiert auf einer kybernetischen Interpretation der Semiose. Der Prozess der Semiose vollzieht sich im I-System auf folgende Weise [50]: Ein Informationsträger wird vom Empfänger interpretiert und nachfolgend als Zeichen erkannt. Dadurch ändert sich der innere Zustand des I-Systems. Diese Veränderung wiederum regt eine Zeichenartikulation an, welches über den Sender an die Umgebung abgegeben wird. Das Zeichen hat folgende Eigenschaften:

- Es hat einen geordneten inneren Aufbau (syntaktische Komponente), sonst könnte es nicht als Zeichen erkannt werden.
- Es hat eine Bedeutung für das I-System (semantische Komponente), andernfalls würde sich der innere Zustand nicht ändern.
- Es wirkt als Stimulus, der das I-System zu einer Aktion bzw. Reaktion anregt (pragmatische Komponente), sonst könnte es seine Umwelt nicht manipulieren.

Informationstheorien, die diese Erkenntnis nicht berücksichtigen, sind nach [49] unvollständig. Der Bezug des I-Systems zu den Kognitiven Systemen wird daher als komplementäre Beschreibung des Shannonschen Kommunikationsmodells ersichtlich. Allerdings berücksichtigt die reine semiotische Darstellung noch nicht die Ebene der Zielvorstellungen, diese muss das Kommunikationsmodell ebenfalls erfassen. Daher wird die die vollständige Beschreibung von kognitiven Systemen im Sinne von S. Haykin und S. Young erst durch die Integration der Modelle von C. E. Shannon bzw. D. Nauta sowie der Einbindung der Ebene der Zielvorstellungen ermöglicht.

B. Hierarchische Systeme

Im Sinne der vorstehend gegebenen Erläuterung ist der Mensch das universellste kognitive dynamische System. Wenn man sich mit der humanen Informationsverarbeitung auseinandersetzt, kann man nicht außer Acht lassen, dass sie sich auf einer hierarchisch organisierten Struktur vollzieht (Bild 3). Die hierarchische Strukturierung des humanen Systems ist offenbar wesentlich für seine Funktion [31]. Über die Ebenen hinweg erfolgt eine Kombination aus Abstraktion (Bottom-up) und Prädiktion (Top-down). Die bereits erwähnten, technisch realisierten kognitiven dynamischen Systeme benötigen anscheinend diese Strukturierung (noch) nicht, da sie vorwiegend auf der Signalebene arbeiten.

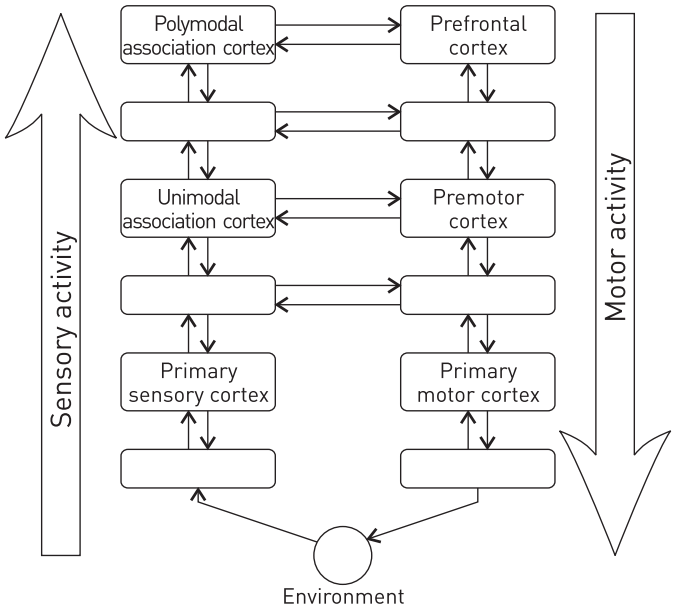

Bild 3: Zusammenwirken des sensorischen Kortex (links) und motorischen Kortex (rechts) im Rahmen des Perzeptions-Aktions-Zyklus. Das Strukturbild wurde aus einer Darstellung von J. Fuster übernommen [31, S. 109], der darauf hinweist, dass die Existenz aller angegebenen Blöcke beim Rhesusaffen nachgewiesen worden ist.

Daher finden wir technische Systeme, die eine ansatzweise vergleichbare hierarchische Struktur haben, bisher nur im Bereich der Mensch-Technik-Interaktion, also der Verarbeitung von Sprache, Bildern, Gesten usw. Daraus ergibt sich auch die formale Ähnlichkeit der UASR-Struktur in Bild 1 mit dem Perzeptions-Aktions-Zyklus in Bild 3. Es fallen zwei formale Unterschiede auf, zu denen kurz Stellung zu nehmen ist:

- Auf eine explizite Darstellung des Arbeitsspeichers, der zwischen dem sensorischen und motorischen Kortex vermittelt, wurde aufgrund seiner Komplexität in Bild 3 verzichtet, während man in einem viel primitiveren System wie in Bild 1 die Datenbasen explizit bezeichnen kann.
- Der in dem Bottom-up-Prozess verlaufende Abstraktionsvorgang ist mit einer schrittweisen Abtrennung irrelevanter Information verbunden. Bei einem Sprachverarbeitungssystem, dessen „Nutzinformation“ der linguistische Inhalt einer Äußerung ist, zählen zu diesen irrelevanten Bestandteilen Signalkomponenten, die den Sprecher als Individuum, seinen emotionalen Zustand oder seine aktuelle Umgebung kennzeichnen. Im UASR werden diese Informationen in einem zusätzlichen Zweig gesammelt, der in Bild 1 in vergröbernder Form als Prosodiemodell bezeichnet ist. Die dort enthaltene Modellinformation muss bei der Synthese im Top-down-Prozess wieder zugesetzt werden, um ein realistisches Sprachsignal zu erhalten. Natürlich ändert sich das Schema sinngemäß, wenn die „Nutzinformation“ umdefiniert wird (z. B. bei einem System, das einen Sprecher und seinen emotionalen Zustand erkennen soll, und bei dem ggf. die linguistische Information als irrelevant einzustufen ist).

C. Folgerung: Hierarchische kognitive dynamische Systeme

An dieser Stelle lässt sich die Feststellungen treffen, dass den aktuellen Herausforderungen der Systemtheorie begegnet werden kann, indem man *sowohl* den kognitiven *als auch* den hierarchischen Aspekt berücksichtigt. Dazu ist anzumerken:

- In der Systemtheorie ist die Berücksichtigung hierarchischer Strukturen seit langer Zeit grundsätzlich erfolgt (z. B. Mesarović et al. [32]), jedoch weitgehend ohne Berücksichtigung des kognitiven Aspekts.
- Ebenfalls bekannt sind hierarchisch angeordnete Sprachnetzwerke, mit denen ein assoziatives Codierungssystem für Texte aufgebaut wurde [33], [34]. Hier liegt ein vielversprechender Ansatz für natürlichsprachliche Systeme vor, mit dem auf die Verwendung von expliziten grammatikalischen Regeln gänzlich verzichtet werden kann.
- Umgekehrt ist die Theorie kognitiver dynamischer Systeme bisher praktisch ausschließlich in Lösungen angewendet worden, für die eine hierarchische Strukturierung nicht wesentlich ist (z. B. *cognitive radio*).
- Eine Theorie hierarchischer kognitiver dynamischer Systeme steht also noch aus. Es muss übrigens darauf hingewiesen werden, dass S. Haykin selbst die Notwendigkeit der Verbindung beider Aspekte herausgestellt hat; so stammt der Hinweis auf die in Bild 3 wiedergegebene hierarchische biologische Struktur aus einem seiner Vorträge [35].
- Als Ingenieurwissenschaftler sind wir der Auffassung, dass die Erforschung der Klasse der hierarchischen kognitiven dynamischen Systeme anhand einer konkreten Implementierung erfolgen sollte.
- UASR ist eine bereits funktionsfähige und in zahlreichen Anwendungen bewährte Implementierung, die sowohl die kognitiven als auch die hierarchischen Aspekte berücksichtigt, und sollte sich deshalb zum weiteren Studium dieser Systemklasse eignen.

Bei der technischen Realisierung hierarchischer kognitiver dynamischer Systeme muss den Erkenntnissen aus den Neurowissenschaften Rechnung getragen werden. Die auf diesem Gebiet entwickelten biologischen sensomotorischen Modelle zeigen eine überraschende strukturelle Ähnlichkeit zum UASR-System (vgl. Bilder 1 und 3). Der wesentliche Unterschied zwischen diesen beiden Modellen besteht darin, dass die vertikale Informationsverarbeitung zwischen den Hierarchieebenen bidirektional erfolgt. Erste konzeptuelle Modellierungsvorschläge zur bidirektionalen Informationsverarbeitung in hierarchischen Systemen sind in [51] publiziert worden. Die Berücksichtigung bidirektionaler Signalpfade ist ein wesentliches Merkmal des sogenannten Cortikalen Algorithmus [52], [53]. In Anlehnung an den Modellierungsvorschlag von [51] wurde in [36] ein Konzept zur Realisierung des Cortikalen Algorithmus unter Verwendung von kaskadierten bidirektionalen Hidden Markov Modellen (CBHMM) vorgestellt. Dieses Konzept basiert auf der Erweiterung von kaskadierten Hidden Markov Modellen (CHMM), welche zur statistischen Modellierung der Zielsuche in hierarchischen Systemen eingeführt wurden [54].

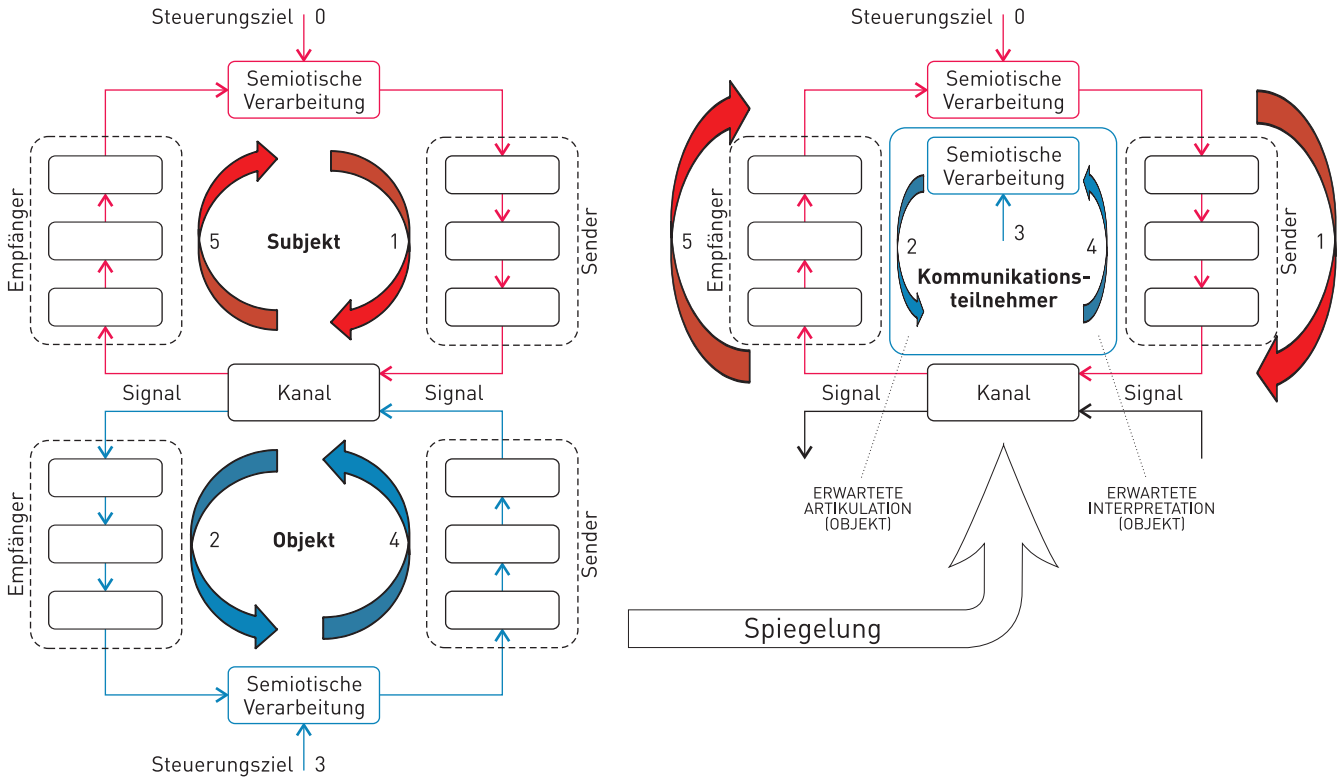

Bild 4: Kommunikationszyklus zwischen zwei Kommunikationsteilnehmern (links). Das Kommunikationssubjekt verfügt über ein inneres Modell des Kommunikationsobjekts, durch Spiegelung des Kommunikationobjekts kann eine Simulation des Kommunikationsobjekts erfolgen (rechts).

Im Folgenden wird kurz darauf eingegangen, wie die Berücksichtigung bidirektionaler Signalpfade durch gespiegelte Kommunikationsprozesse begründet werden kann.

Die Zielstellung eines Kognitiven Systems besteht darin, die Kommunikation zwischen den Teilnehmern zu optimieren und deren Vorstellungen abzugleichen. Um zwischen den Kommunikationsteilnehmern unterscheiden zu können, wird im Folgenden von einem Kommunikations*subjekt* (dominant) und einem Kommunikations*objekt* (subdominant) die Rede sein. Aus der Sicht des Kommunikationssubjekts ist es sinnvoll, bei der Analyse die erwartete Wirkung des Kommunikationsobjekts zu berücksichtigen. Dies kann beispielsweise in Form einer Prädiktion entgegen der Analyserichtung erfolgen. Andererseits soll auch sichergestellt werden, dass die vom Kommunikationssubjekt erzeugte Wirkung möglichst gut vom Kommunikationsobjekt verstanden wird. Die dafür notwendigen zusätzlichen Signalwege können auf natürliche Weise begründet werden, wenn das Kommunikationssubjekt über ein inneres Modell des Kommunikationspartners verfügt, welches simuliert und zu Vorhersagen genutzt werden kann [55].

Die Entwicklung dieser Idee für den Kommunikationsprozess soll im Folgenden kurz skizziert werden (vgl. [58]). Dabei wird davon ausgegangen, dass sich beide Kommunikationsteilnehmer als kognitives System (vgl. Bild 2) modellieren lassen, allerdings werden Sensorik und Motorik nun hierarchisch ausgeführt.

Zunächst wird im linken Teil von Bild 4 der zeitliche Ablauf des kybernetischen Zyklus für zwei Kommunikationsteilnehmer verdeutlicht. Unter Vorgabe eines Kommunikationsziels (0) wählt das Kommunikationssubjekt eine Aktion aus. Diese Aktion wird entlang des Signalwegs (1) artikuliert und über den Kanal übertragen. Das Kommunikationsobjekt interpretiert das ankommende Signal (2) und wählt seinerseits unter Berücksichtigung der eigenen Kommunikationsziele (3) eine entsprechende Reaktion aus und artikuliert diese entlang des Signalwegs (4).

Schließlich interpretiert das Subjekt das ankommende Signal (5) und kontrolliert die Einhaltung des Kommunikationsziels. Dieser Zyklus kann bis zur Erreichung des Kommunikationsziels mehrfach wiederholt werden. Der Kommunikationsteilnehmer kann die Kommunikation optimieren, wenn er über ein inneres Modell des Kommunikationspartners verfügt, welches er simulieren kann. Dieser Prozess kann als Spiegelung des Kommunikationsobjekts verstanden werden. Die Idee von der Entwicklung eines inneren Modells ist aus der Simulationstheorie [55] bekannt, einige Indizien zur Stütze dieser Theorie konnten durch die Entdeckung der Spiegelneuronen erbracht werden [56], [57].

Unter Verwendung des inneren Modells ist der kybernetische Zyklus des Kommunikationssubjekts durch folgenden zeitlichen Ablauf gekennzeichnet (vgl. Bild 4, rechts): Nach Vorgabe eines Kommunikationsziels (0) wählt das Kommunikationssubjekt eine Aktion aus, diese Aktion wird entlang des Signalwegs (1) artikuliert und über den Kanal übertragen. Nun wird die erwartete Interpretation des Objekts wieder zurückgeführt (2) und die erwartete Reaktion – über das

innere Modell vom Kommunikationsobjekt – vorhergesagt. Diese Reaktion wird entlang des Signalwegs (4) artikuliert und als Vorhersage der erwarteten Artikulation des Objekts verstanden. In (5) erfolgt die Interpretation des ankommenden Signals.

Für den Synthesezweig kann das gleiche Prinzip verwendet werden, allerdings muss dann die Reihenfolge der Operationen (1) und (2) vertauscht werden, ansonsten könnte die artikulierte Aktion (1) die erwartete Interpretation des Objekts (2) nicht berücksichtigen. Dies gelingt, wenn das generierte Signal (1) verzögert wird und erst im nächsten Syntheseschritt für eine Bottom-Up Prädiktion verwendet wird. Weitere Vereinfachungen ergeben sich, wenn die Kommunikationsziele von Objekt und Subjekt identisch sind. Dann kann der semiotische Block des Subjekts – ausgehend von der Bottom-Up Prädiktion – die erwartete Reaktion des Objekts nach einer optimalen Strategie [43] auswählen und als Erwartung (4) artikulieren.

Zur Realisierung der gespiegelten Kommunikationsprozesse sind bidirektionale Signalpfade notwendig, diese wurden in 5 nun auch explizit eingetragen. Neben den bidirektionalen Signalwegen sind in dieser Abbildung die zwei typischen Informationsformen von Analyse-Synthese-Systemen berücksichtigt (vgl. mit Bild 6). Dabei finden sich die redundanten (vorhersagbaren) Bestandteile in den Modellen der einzelnen Hierarchieebenen wieder. Diese werden sowohl von der Analyse als auch von der Synthese genutzt. Die irrelevante Information wird bei der Analyse (genauer im Abstraktionsprozess) abgeführt und bei der Synthese (im Adjunktionsprozess) wieder zugeführt. Das Wesen der relevanten Information wird bereits recht gut verstanden. Die Trennung beider Informationsformen bietet nun auch die Möglichkeit, das Wesen der irrelevanten Information durch das Prinzip der „Analyse durch Synthese“ besser zu verstehen. Die Berücksichtigung von gespiegelten Kommunikationsprozessen führt damit zu einer strukturellen Annäherung an das biologische Vorbild [31] und verdeutlicht die Motivation, die zur Einbindung in das UASR-Konzept geführt hat.

Nach dem die bidirektionalen Signalwege nachvollziehbar begründet werden konnten, ist die Gewährleistung einer optimalen Kommunikation aber erst dann gegeben, wenn die bidirektionalen Signalpfade in jeder Hierarchieebene auch eine entsprechende Verknüpfung aufweisen. Nur dadurch können beide Kommunikationsteilnehmer tatsächlich von der Ausbildung eines inneren Modells profitieren. Für die Modellierung dieser Verknüpfung hat sich das Studium der Struktur der motorischen und sensorischen Hierarchie des Neocortex als fruchtbringend erwiesen.

S. Haykin verweist in diesem Zusammenhang auf das biologische Modell des Neocortex nach [31]. Dort sind die hierarchischen Strukturen durch einen bidirektionalen Informationsfluss und durch spezifische Arbeitsspeicher in den einzelnen Hierarchieebenen gekennzeichnet. Außerdem belegen Forschungsarbeiten auf dem Gebiet der Neurowissenschaften, dass der Neocortex nicht über eine Sammlung spezialisierter cortikaler Architekturen und Algorithmen verfügt, sondern eine ziemlich einfach organisierte hierarchische Struktur besitzt [59].

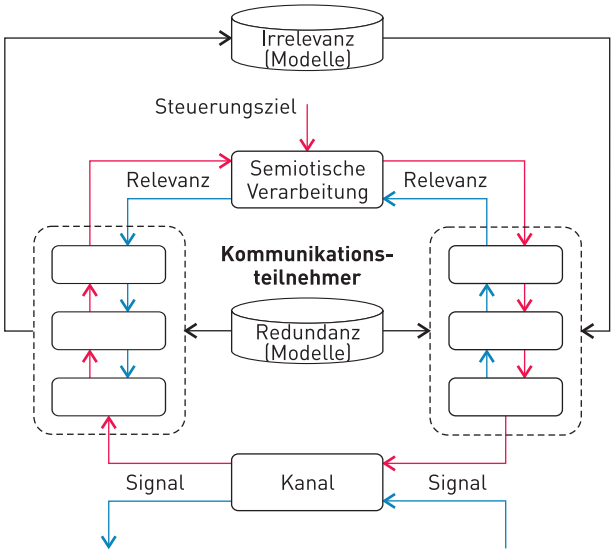

Bild 5: Ausbildung bidirektionaler Signalpfade beim Kommunikationsteilnehmer in Folge eines gespiegelten Kommunikationsprozesses. Die damit verbundene Annäherung an das biologische Vorbild nach [31] verdeutlicht die Motivation, die zur Einbindung des Cortikalen Algorithmus in das UASR-Konzept geführt hat.

Daher kann innerhalb der verschiedenen hierarchischen Ebenen der gleiche allgemeine Verarbeitungsprozess ablaufen. Findet außerdem eine einheitliche Einbindung der vertikalen Informationsflüsse aus den benachbarten Hierarchieebenen statt, dann wird ein solcher Prozess Cortikaler Algorithmus genannt. In [36] wurden mit den CBHMMs und Multiraten-CBHMMs zwei Strukturen vorgestellt, mit denen dieser Algorithmus nachgebildet werden kann. Dabei handelt es sich um Strukturen, die sowohl für die Analyse als auch für die Synthese geeignet sind. Eine Besonderheit liegt darin, dass bei der Analyse und der Synthese ein bidirektionaler Informationsfluss zur gleichen Zeit vorliegt. Der zusätzliche Informationsfluss wird bei der Analyse als Prädiktion der zu erwartenden Reaktion des Kommunikationsteilnehmers interpretiert, umgekehrt wird bei der Synthese der zusätzliche Informationsfluss als zu erwartende Interpretation des Kommunikationsteilnehmers verstanden. Die Verknüpfung beider Informationsflüsse kann sowohl bei der Analyse als auch bei der Synthese nach dem Bayesschen Prinzip erfolgen. In beiden Fällen liegen dann in jeder Hierarchieebene eine Messverteilung und eine Prädiktionsverteilung vor. Beide Verteilungen werden multiplikativ miteinander verknüpft und anschließend normalisiert, so dass schließlich eine a-posteriori Verteilung als Ergebnis einer Informationsfusion vorliegt.

Zum gegenwärtigen Zeitpunkt liegt eine mathematische Beschreibung des Cortikalen Algorithmus unter Verwendung von CBHMM vor [36]. Damit werden einerseits die zusätzlichen Prädiktionspfade für die Analyse und die Synthese berücksichtigt, so dass schließlich auf jeder Ebene eine Fusion beider Informationsflüsse nach dem Bayesschen Prinzip stattfinden kann. Andererseits kann durch die Verwendung von Multiraten-CBHMM ein Analyse-Synthese-System realisiert

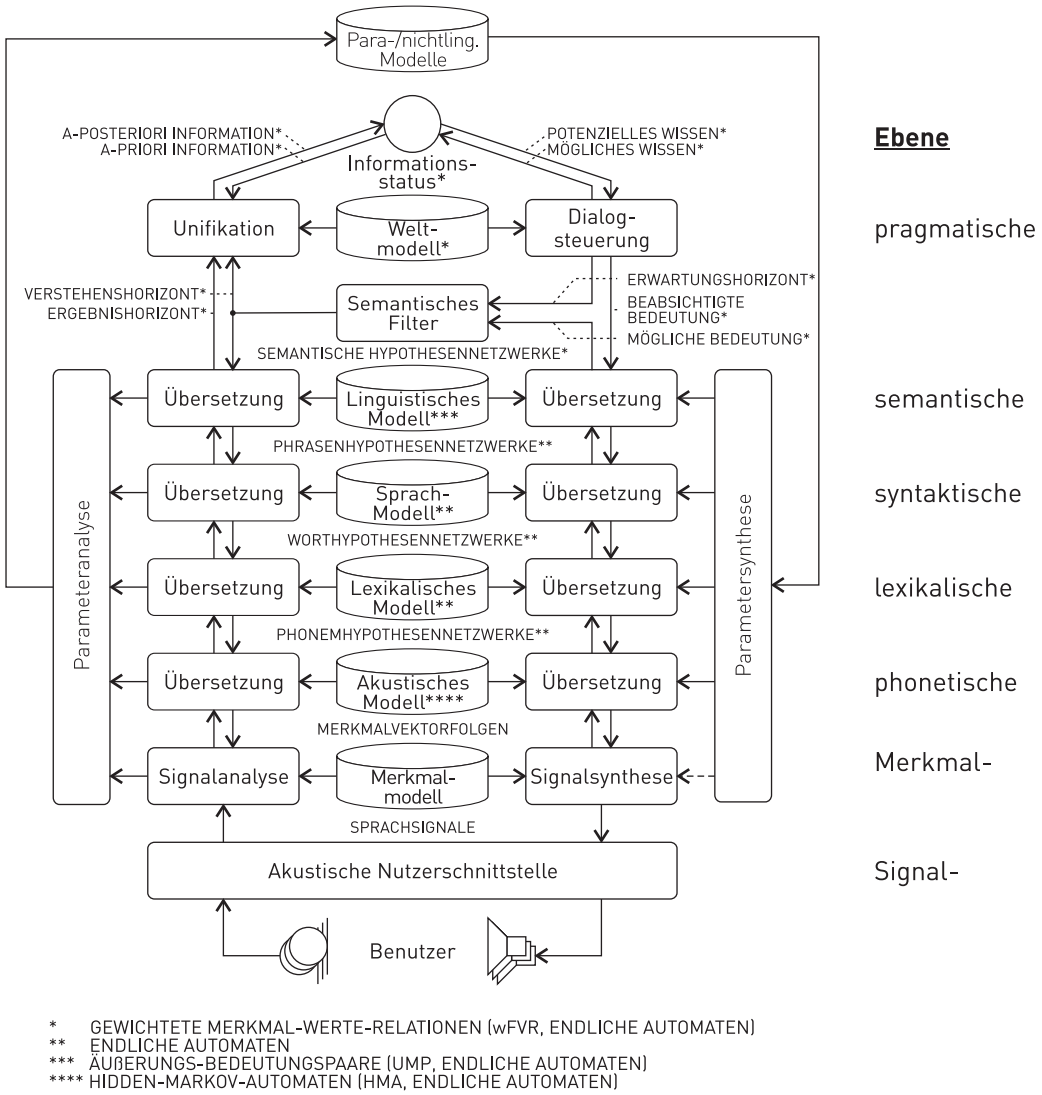

Bild 6: Aktuelles Konzept der Erweiterung des Systems UASR.

werden, welches auf unterschiedlichen Zeitskalen arbeitet. Da solche Systeme in den jeweiligen Ebenen über unterschiedlich ausgedehnte Kontextweiten verfügen, ergibt sich die Möglichkeit, Kontextwissen aus höheren Ebenen auszunutzen und damit die Informationsverarbeitung bei der Analyse in Top-Down Richtung zu steuern. Umgekehrt kann nun auch die Synthese in Bottom-Up Richtung gesteuert werden, indem von bereits synthetisierten Daten ausgehend, eine Prädiktion in die höheren Hierarchieebenen propagiert wird. Die Steuerung der Synthese entspricht damit einem Kontrollmechanismus, der in der Kybernetik als Reafferenzprinzip bekannt ist [60].

Neben der Berücksichtigung von strukturellen Aspekten hierarchischer Analyse-Synthese-Systeme konnte mit der Unterscheidung von deskriptiver und selektiver Information [61] der Cortikale Algorithmus auch informationstheoretisch gedeutet werden. Dabei wurde gezeigt, dass die Informationsverarbeitung entlang der hierarchischen Achse auf Bayessche Inferenzmechanismen zurückgeführt werden kann, was bei der Fusion der verschiedenen Informationsflüsse besonders deutlich wird [58], [62].

Zusammenfassend kann schließlich festgehalten werden, dass mit der Nachbildung des Cortikalen Algorithmus durch CBHMM die Voraussetzung für die Realisierung eines hierarchisch organisierten bidirektionalen Analyse-Synthese-Systems nach biologischem Vorbild geschaffen werden konnte.

III. Forschungsaufgaben

A. Weiterentwicklung von UASR (Sprachsignalverarbeitung)

Auch wenn UASR offenbar gut dazu geeignet ist, auch allgemeinere Fragen einer hierarchischen Signalverarbeitung zu studieren, sollte es vorrangig als sprachverarbeitendes System weiterentwickelt werden, da diese Anwendung typisch, am anspruchsvollsten und für weitere Applikationen beispielgebend ist. Dabei ergeben sich die folgenden Schwerpunkte:

- Im *Analysezweig* des Modells sollte der Schwerpunkt des algorithmischen Ausbaus in einer Verbesserung der Top-down- und Bottom-up-Interaktion zwischen den Ebenen bestehen. Ein aussichtsreicher Vorschlag für eine Verbesserung der statistischen Modellierung wurde in [36] for-

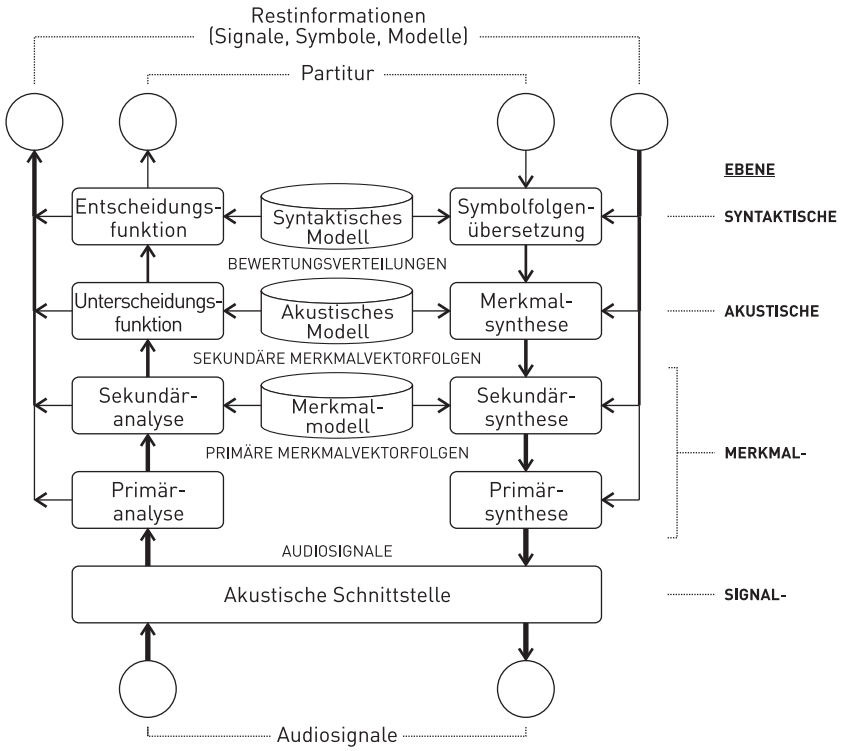

Bild 7: Intelligentes System zur (akustischen) Signalverarbeitung [21].

muliert. Interessant ist, dass sich auch die Neurobiologie dieser Fragestellung zu widmen beginnt [37].

- Im *Synthesezweig* des Modells scheint die weitere Verfolgung der sog. HMM-Synthese noch sehr viele Anregungen zu bieten, so dass nicht nur eine weitere Verbesserung der Qualität der TTS-Systeme, sondern auch erweiterte Möglichkeiten für AbS-basierte Untersuchungen zu erwarten sind [12].
- Am *oberen Ende* des Modells steht bei UASR ein Text, da die herkömmliche Sprachverarbeitung dort endet. Allgemein anerkannt ist, dass ein tiefgreifender Fortschritt der Sprachtechnologie zu erwarten ist, wenn die Systeme „verstehen“, was sie erkennen bzw. synthetisieren. Dies verlangt eine Fortsetzung durch mindestens eine semantische Ebene. Erfahrungen mit einer solchen Erweiterung wurden in der Vergangenheit speziell für Übersetzungssysteme gewonnen [5]. Vielversprechend ist auch die Einbeziehung einer Dialogmodellierung, wie sie in [38] vorgeschlagen wurde.
- Am *unteren Ende* des Modells steht bei UASR das Sprachsignal, dessen Aufnahme bzw. Abstrahlung in „geeigneter“ Form vorausgesetzt wird. Realistischer ist der Ansatz, ein allgemeineres Interface zur Umgebung (sog. akustisches Frontend) zu schaffen, mit dessen Hilfe ein internes Modell der Umgebung entsteht (akustische Szenenanalyse), das eine optimierte Informationsgewinnung bzw. –abgabe ermöglicht.
- Der in Bild 1 als *Klammer* ausgeführte Prosodiezweig bietet Raum für Erweiterungen zur Behandlung para– und nichtlinguistischer Information (nach der Definition von H. Fujisaki, [47]). Die Struktur von UASR muss dazu gegenüber Bild 1 verfeinert werden [39]. In diesem Zusammenhang spielt die Verbindung mit anderen Interaktionsmodi eine wichtige Rolle, die in jüngster Zeit sehr erfolgreich im Rahmen der COST-Action 2102 (Cross Modal Analysis of Verbal and Nonverbal Communication) behandelt wurde [40].
- Weiterhin muss auf eine Komponente verwiesen werden, die in Bild 1 nicht erkennbar ist, da es sich um einen zeitlichen Schnappschuss handelt. Die Entwicklung des Sprachsignals in der Zeit wird in der Sprachverarbeitung vorwiegend auf die Entstehung einer Symbolfolge reduziert; lediglich bei der Sprachsynthese ist man gezwungen, den Lautsymbolen explizit Dauern zuzuordnen. Diese Vorgehensweise vernachlässigt die Bedeutung suprasegmentaler zeitlicher Strukturen (Rhythmus). Ihre Modellierung stellt ein schwieriges, teilweise umstrittenes, zweifellos aber für die Zukunft wichtiges Forschungsgebiet dar [41].
- Schließlich steht eine vertiefende *informationstheoretische Betrachtung* aus. So fällt auf, dass der von uns verwendete Begriff *Irrelevanz* (im Ggs. zu Relevanz) nicht Irrelevanz im Shannonschen Sinne bedeuten kann, da es sich dabei nicht um eine Störentropie handelt. Eher liegen Makrozustände auf verschiedenen Hierarchiestufen des Systems vor, was einen Entropiebegriff nach Boltzmann nahelegt. Es stellt sich also die Frage, ob für den horizontalen Informationsfluss die „Shannon“-Entropie, jedoch für den vertikalen die „Boltzmann“-Entropie die jeweils angemessene Vorstellung ist (vgl. z. B. [42]).

Aus dem Forderungskatalog wurde das in Bild 6 dargestellte erweiterte Konzept entwickelt, welches den „Masterplan“ für

unsere zukünftige Forschungsarbeit darstellt.

Vergleicht man die erste mit der aktuellen UASR-Version (also Bild 1 mit Bild 6), fallen vermutlich nicht nur die neue semantische Komponente und die bidirektionalen Interaktionen als Unterschied auf, sondern es zeigt sich auch in der Bezeichnung der Funktionsblöcke eine deutliche Vereinheitlichung. Dies ist eine Auswirkung der konsequenten Implementierung der UASR-Komponenten in der Technologie der *finite state transducers* (FST), auf die bereits unter I-B hingewiesen wurde und die für den Erfolg des Systems essentiell gewesen ist. Aus diesem Grund soll sich in der Weiterentwicklung die konsequente Anwendung der FST-Technologie auch auf die neu entstehenden Systemkomponenten erstrecken. Dies ist eine theoretisch anspruchsvolle Aufgabe, deren Lösung in Zusammenarbeit mit Informatikern und Mathematikern angestrebt wird.

B. Verallgemeinerung von UASR (Sensorsignalverarbeitung)

Es wurde bereits erwähnt, dass die Anwendung struktureller Erkennungsverfahren auf unterschiedliche nichtsprachliche Signale, darunter auch Musik, unter Nutzung von UASR-Komponenten sehr erfolgreich verlaufen ist. Daraus ergeben sich Entwicklungsmöglichkeiten in Richtung eines verallgemeinerten, hierarchischen Schemas der Sensorsignalverarbeitung. Insbesondere ist die Trennung in relevante und irrelevante Information, die unter II-B für Sprachsignale diskutiert wurde, verallgemeinerungsfähig. Dazu wurde in [21] unter dem Namen „Intelligentes Audiosignalverarbeitungssystem“ eine klare Konzeptionen entwickelt. Der Ansatz geht von der folgenden Modellvorstellung aus:

> Audiosignale bestehen aus akustischen Elementarereignissen, die nach einer bestimmten Vorschrift neben- und nacheinander angeordnet sind.

In [21, Abschn. 1] wird die Plausibilität dieser Vorstellung anhand einiger Beispiele verdeutlicht. Es sei ausdrücklich darauf hingewiesen, dass die Anordnungsvorschrift („Partitur“ des Signals) im allgemeinen Fall hierarchisch organisiert ist (z. B. Noten/Akkorde – Takte – Phrasen – Sätze – usw.) Die Funktion eines intelligenten Systems zur Audiosignalverarbeitung besteht in der Dekomposition von Signalen in eine strukturelle Darstellung (Partitur) und nicht strukturelle Restinformationen sowie in der Komposition von Signalen aus diesen Informationen. Bild 7 zeigt die grundsätzliche Struktur eines solchen Systems. In [21, Abschn. 7.2] wird beschrieben, wie daraus die verschiedenen Applikationen entstehen. Die Erweiterung zu einem kognitiven dynamischen System erfolgt durch Anschließen eines (problemabhängigem und daher hier nicht näher auszuführenden) Rückkopplungsmoduls am oberen Ende der Struktur.

SCHLUSSBEMERKUNG

Wir haben ein Konzept für hierarchische kognitive dynamische Systeme mit folgenden Eigenschaften vorgestellt:

- Realisierung des kybernetischen Zyklus: Umwelt (Kanal) – Sensorik (Analyse) – Rückkopplung (semiotische Verarbeitung) – Motorik (Synthese) – Umwelt,
- hierarchische Struktur mit korrespondierenden Verabeitungseinheiten in Analyse– und Synthesezweig,
- bidirektionaler Informationsfluss zwischen den Verarbeitungsebenen (realisiert z. B. durch den Cortikalen Algorithmus), sowie
- Zielvorstellung, ohne die planmäßiges Handeln nicht möglich ist.

Dieses Konzept wurde für die Anwendung auf Sprache in Form des Unified Approch for Speech Synthesis and Recognition (UASR) bereits in wesentlichen Teilen (Signal– bis syntaktische Ebene mit unidirektionaler Verarbeitung) realisiert und erfolgreich praktisch erprobt. Weiterhin wurde der Analysezweig des Systems in vielen Anwendungen für die akustische Mustererkennung von Musik-, Bio- und technischen Signalen eingesetzt.

Unsere künftige Forschung zielt auf eine weitere Annäherung an das biologische Vorbild nach Fuster [31]: Einführung einer bidirektionalen Verarbeitung zur „inneren“ Modellierung des Kommunikationspartners, semiotische Verarbeitung mit Hilfe gewichteter Merkmal-Werte-Relationen sowie Modellierung und Verarbeitung para– und nichtlinguistischer Informationen inklusive Rhythmus.

Ein weiterer Schwerpunkt wird die angesprochene Verallgemeinerung auf nichtsprachliche Signalverarbeitung sein, welche besonders durch die breite Vielfalt möglicher Anwendungen interessant ist. Unter theoretischen Gesichtspunkten scheint hier unter anderem die Untersuchung der Synthese von nichtsprachlichen Signalen, beispielsweise zum Zwecke der Modellverifikation durch das AbS-Prinzip, wünschenswert.

Natürlich sind auch in diesem Aufsatz nicht berücksichtigte Aspekte wie beispielsweise das (autonome) Lernen zu berücksichtigen, welches ebenfalls eine notwendige Voraussetzung für die Konstruktion von „kognitiven“ Systemen ist.

LITERATUR

[1] C. G. Bell; H. Fujisaki; J. M. Heinz; K. N. Stevens; A. S. House: Reduction of speech spectra by analysis-by-synthesis techniques. The Journal of the Acoustical Society of America 33 (1961) 12, 1725–1736.

[2] K. N. Stevens: Toward a model for speech recognition. The Journal of the Acoustical Society of America 32 (1960) 1, 47–55.

[3] S. E. Levinson: Structural methods in automatic speech recognition. Proceedings of the IEEE 73 (1985) 11, 1625–1650.

[4] J. N. Holmes: Speech Synthesis and Recognition. London: Van Norstrand Reinhold 1988

[5] W. Wahlster: Verbmobil – Foundations of Speech-to-Speech Translation. Berlin etc.: Springer 2000.

[6] M. Eichner; M. Wolff; R. Hoffmann: A unified approach for speech synthesis and speech recognition using stochastic Markov graphs. Proc. ICSLP, Beijing 2000, vol. 1, 701–704.

[7] R. Hoffmann; M. Eichner; M. Wolff: Analysis of verbal and nonverbal acoustic signals with the Dresden UASR system. In: A. Esposito et al. (Eds.): Verbal and Nonverbal Communication Behaviours. Berlin etc.: Springer 2007 (LNAI vol. 4775), 200–218.

[8] R. Hoffmann: Der automatentheoretische Zugang zur Spracherkennung. TU Dresden, Informationen ET-ITA-01-1996, 16 S.

[9] S. Werner; M. Eichner; M. Wolff; R. Hoffmann: Towards spontaneous speech synthesis - Utilizing language model information in TTS. IEEE Trans. on Speech and Audio Processing 12 (2004) 4, 436–445.

[10] 2nd One Day Meeting on Unified Models for Speech Recognition and Synthesis. The University of Birmingham, 30 March 2009, unveröffentlicht.

[11] F. Duckhorn; G. Strecha; M. Wolff; R. Hoffmann: Ein Sprachdialogsystem mit begrenzten Hardwareressourcen. In: R. Hoffmann (Hrsg.): Elektronische Sprachsignalverarbeitung, Dresden, 21.–23. September 2009, Bd. 1 = Studientexte zur Sprachkommunikation, Bd. 53, 88–93.

[12] G. Strecha; M. Wolff; F. Duckhorn; S. Wittenberg; C. Tschöpe: The HMM synthesis algorithm of an embedded unified speech recognizer and synthesizer. Proc. Interspeech 2009, Brighton, Sept. 2009.

[13] M. Wolff; U. Kordon; H. Hussein; M. Eichner; C. Tschöpe; R. Hoffmann: Auscultatory blood pressure measurement using HMMs. Proc. IEEE ICASSP, Honolulu 2007, vol. 1, 405–408.

[14] U. Kordon; M. Wolff; C. Tschöpe: Mustererkennung für Sensorsignale. In: Gerlach, G.; Hoffmann, R. (Hrsg.): Neue Entwicklungen in der Elektroakustik und elektromechanischen Messtechnik. Dresden: TUDpress 2009 = Dresdner Beiträge zur Sensorik, Bd. 40, 69–78.

[15] Maschinendiagnose - Grundlagen, Konzepte, Visionen. ERS-Workshop, RWTH Aachen. 8. Juli 2009, unveröffentlicht.

[16] C. Tschöpe; M. Wolff: Statistical classifiers for structural health monitoring. IEEE Sensors Journal 9 (2009) 11, 1567–1576.

[17] M. Wolff: Automatisches Lernen von Aussprachewörterbüchern. Dresden: w.e.b. Universitätsverlag 2004 = Studientexte zur Sprachkommunikation, Bd. 32.

[18] M. Eichner: Sprachsynthese und Spracherkennung mit gemeinsamen Datenbasen - Akustische Analyse und Modellierung. Dresden: TUDpress 2007 = Studientexte zur Sprachkommunikation, Bd. 43.

[19] S. Werner: Sprachsynthese und Spracherkennung mit gemeinsamen Datenbasen - Sprachmodell und Aussprachemodellierung. Dresden: TUDpress 2008 = Studientexte zur Sprachkommunikation, Bd. 48.

[20] C. Tschöpe: Akustische zerstörungsfreie Prüfung mit Hidden-Markov-Modellen. Dresden: TUDpress, erscheint 2012 = Studientexte zur Sprachkommunikation, Bd. 60.

[21] M. Wolff: Akustische Mustererkennung. Habilitationsschrift. Dresden: TUDpress 2011 = Studientexte zur Sprachkommunikation, Bd. 57.

[22] K. Küpfmüller: Die Systemtheorie der elektrischen Nachrichtenübertragung. Stuttgart: S. Hirzel 1949; 2., erweiterte Auflage 1952.

[23] G. Wunsch: Geschichte der Systemtheorie. Dynamische Systeme und Prozesse. Berlin: Akademie-Verlag 1985 = Wissenschaftliche Taschenbücher Bd. 296.

[24] G. Wunsch: Grundlagen der Prozesstheorie. Struktur und Verhalten dynamischer Systeme in Technik und Naturwissenschaft. Stuttgart, Leipzig, Wiesbaden: B. G. Teubner 2000.

[25] H. Kindler: Grundlagen der Informationstheorie, anschaulich dargestellt. In: H.-J. Scharf (Hrsg.): Informatik. Leipzig: J. A. Barth 1972 = Nova Acta Leopoldina, N. F., Bd. 37/1, Nr. 206, 113–127.

[26] S. Haykin: Foundations of cognitive dynamic systems. IEEE Lecture, Queens University, 29 January 2009, http://soma.mcmaster.ca/papers/Slides_Haykin_Queens.pdf

[27] S. Haykin: Cognitive radar. IEEE Signal Processing Magazine 23 (2006) 1, 30–40.

[28] J. Mitola III; G.Q. Maguire Jr.: Cognitive radio: making software radios more personal. IEEE Personal Communications Magazine 6 (1999) 4, 13–18.

[29] S. Haykin: Cognitive Radio: Brain-Empowered Wireless Communications. IEEE Journal on Selected Areas in Communications 23 (2005) 2, 201–220.

[30] M. G. di Benedetto; Y. Hua; T. Kaiser; X. Wang (Guest Editors): Special Section - Cognitive Radio Technology. IEEE Signal Processing Magazine 25 (2008) 6, 10–93.

[31] J. M. Fuster: Cortex and Mind - Unifying Cognition. New York: Oxford University Press 2003.

[32] M. D. Mesarović; D. Macko; Y. Takahara: Theory of Hierarchical, Multilevel, Systems. New York and London: Academic Press 1970 = Mathematics in Science and Engineering, vol. 68.

[33] W. Hilberg: The Unexpected Fundamental Influence of Mathematics upon Language. Glottometrics, 5, pp. 29–50, 2002.

[34] W. Hilberg: Some results of quantitative linguistics derived from a structural language model. Glottometrics, 7, pp. 1–24, 2004.

[35] S. Haykin: Cognitive dynamic systems. Vortrag (ungedruckt), Joint COST 2102 and 2102 International Conference, Madrid, Sept. 2009.

[36] R. Römer; T. Herbig: Konzeptionelle Beschreibung des cortikalen Algorithmus und seine Anwendung in der Automatischen Sprachverarbeitung. In: R. Hoffmann (Hrsg.): Elektronische Sprachsignalverarbeitung, Dresden, 21.–23. September 2009, Bd. 1 = Studientexte zur Sprachkommunikation, Bd. 53, 33–40.

[37] S. J. Kiebel; K. v. Kriegstein; J. Daunizeau; K. J. Friston: Recognizing sequences of sequences. PLoS Computational Biology 5 (2009) 5, e1000464, 13 S.

[38] M. Huber; C. Kölbl; R. Lorenz; R. Römer; G. Wirsching: Semantische Dialogmodellierung mit gewichteten Merkmal-Werte-Relationen. In: R. Hoffmann (Hrsg.): Elektronische Sprachsignalverarbeitung, Dresden, 21.–23. September 2009, Bd. 1 = Studientexte zur Sprachkommunikation, Bd. 53, 25–32.

[39] http://www.ias.et.tu-dresden.de/ias/index.php?id=443

[40] A. Esposito, R. Hoffmann, et al. (Eds.): Cognitive Behavioural Systems. COST 2102 Final Conference, February 2011, Dresden, Program and Abstracts. (Tagungsband in Vorbereitung)

[41] E. Keller: A phonetician's view of signal generation for speech synthesis. In: R. Vích (Ed.): Electronic Speech Signal Processing, Prague, September 26–28, 2005 = Studientexte zur Sprachkommunikation, Bd. 36, 13–20.

[42] R. Frigg; C. Werndl: Entropy – A Guide for the Perplexed. In Probabilities in Physics; Beisbart C. and Hartmann, S. Eds; Oxford University Press, Oxford, 2010

[43] S. Young: Cognitive User Interfaces. IEEE Signal Processing Magazine 27 (2010) 3, 128–140.

[44] C. Kölbl; M. Huber; G. Wirsching: Endliche gewichtete Transduktoren als semantischer Träger. 22. Konf. Elektronische Sprachsignalverarbeitung, Aachen, 28.–30. 9. 2011, Tagungsband = Studientexte zur Sprachkommunikation, Bd. 61.

[45] G. Wirsching; C. Kölbl; M. Huber: Zur Logik von Bestenlisten in der Dialogmodellierung. 22. Konf. Elektronische Sprachsignalverarbeitung, Aachen, 28.–30. 9. 2011, Tagungsband = Studientexte zur Sprachkommunikation, Bd. 61.

[46] G. Wirsching; M. Huber; C. Kölbl: The confidence-probability semiring. Technical Report No. 2010-4, Institute of Computer Science, University of Augsburg 2010.

[47] H. Fujisaki: In Search of Models: A Review of the Author's Research over half a Century, Keynote auf der 21. Konferenz „Elektronische Sprachsignalverarbeitung“ (ESSV) 2010, Berlin.

[48] Marko, H.: Die Theorie der bidirektionalen Kommunikation und ihre Anwendung auf die Informationsübermittlung zwischen Menschen, Kybernetik 3 1966, pp. 128–136.

[49] Nauta, D.: The Meaning of Information, Mouton, The Hague, Paris 1970.

[50] Flückiger, D.F.: Beiträge zur Entwicklung eines vereinheitlichten Informations-begriffs, Inauguraldissertation der Philosophisch-Naturwissenschaftlichen Fakultät an der Universität Bern, 1995.

[51] Lee, T.S.; Mumford, D.: Hierarchical Bayesian Inference in the Visual Cortex, JO-SA Vol.20, No.7/July 2003.

[52] George, D.; Hawkins, J.: Invariant Pattern Recognition using Bayesian Inference on Hierarchical Sequences. RNI Technical Report 2005.

[53] George, D.; Hawkins, J.: A Hierarchical Bayesian Model of Invariant Pattern Rec-ognition in the Visual Cortex, International Joint Conference on Neural Networks 2006.

[54] Blaylock, N.; Allen, J.: Fast Hierarchical Goal Schema Recognition, Proceedings of AAAI-06, Boston, July 2006.

[55] Goldman, A.I.: Simulating Minds. The Philosophy, Psychology and Neuroscience of Mindreading. Oxford University Press, Oxford 2006.

[56] Rizzolatti, G.: Premotor cortex and the recognition of motor actions, Cognitive Brain Research 1996.

[57] Rizzolatti, G.; Fogassi, L.; Gallese, V.: Mirrors in the Mind, Scientific American Band 295, Nr. 5, November 2006, pp. 30-37.

[58] Römer, R.: Beschreibung von Analyse-Synthese-Systemen unter Verwendung von CBHMM's. 22. Konferenz Elektronische Sprachsignalverarbeitung ESSV-2011, Aachen, pp. 67- 76, TUDpress Dresden 2011.

[59] Mountcastle, V.: An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System, The Mindful Brain (Gerald M. Edelman and Vernon B. Mountcastle, eds.) Cambridge, MA: MIT Press, 1978.

[60] E. von Holst und H. Mittelstaedt: Das Reafferenzprinzip, Die Naturwissenschaften 1950, 37.

[61] MacKay, D.M.: Information, Mechanism and Meaning, M.I.T. Press 1969, Second printing, June 1972.

[62] Römer, R.: A Cortical Approach based on Cascaded Bidirectional Hidden Markov Models. To appear in: Lecture Notes in Computer Science, 2012.