UNTERSUCHUNGEN ZUM CORTIKALEN ALGORITHMUS UNTER VERWENDUNG VON BIDIREKTIONALEN HMMS

Ronald Römer

Lehrstuhl Kommunikationstechnik, BTU Cottbus
Ronald.Roemer@tu-cottbus.de

Kurzfassung: Im diesjährigen Beitrag werden erste Untersuchungen zum Verhalten des Cortikalen Algorithmus in Analyse-Synthese-Systemen beschrieben. Obwohl der Algorithmus für hierarchische, probabilistische Analyse-Synthese-Systeme unter Verwendung von kaskadierten, bidirektionalen HMMs (CBHMMs) konzipiert wurde, kann der Einfluss der bidirektionalen Informationsverarbeitung zunächst für die einfacheren einschichtigen bidirektionalen HMMs untersucht werden. Das Experiment orientiert sich am Kanalmodell nach Shannon, dabei werden sendeseitig synthetische Daten erzeugt und empfangsseitig für verschiedene SNR decodiert. Die Verwendung von künstlich erzeugten Daten hat den Vorteil, dass das Training der Parameter von bidirektionalen HMMs zunächst vermieden werden kann. Trotzdem besteht frühzeitig die Möglichkeit, den Cortikalen Algorithmus in probabilistischen bidirektionalen Analyse-Synthese-Systemen systematisch zu untersuchen und den vorgestellten Algorithmus in einfacher Weise zu verifizieren.

1 Einführung

Im letztjährigen Beitrag wurden CBHMM in die Gruppe der probabilistischen hierarchischen Analyse-Synthese-Systeme eingeordnet [1]. Diese Systeme stellen einen wesentlichen Bestandteil von übergeordneten Systemen dar, die wie der Mensch ein zielgerichtetes Verhalten aufweisen und für die S. Haykin die Bezeichnung *kognitive dynamische Systeme* geprägt hat [2]. Da diese Systemklasse ihre Wurzeln in der regelungstechnischen Systemtheorie hat [3], stehen naturgemäß die auftretenden Interaktionen zwischen einem System und seiner Umgebung im Mittelpunkt der Betrachtungen, sie werden 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. Bezüglich der Struktur von Analyse und Synthese verweist S. Haykin auf die motorische und sensorische Hierarchie, wie sie im biologischen Modell des Neocortex nach Fuster angegeben wird [4].

Für die Untersuchung von hierarchisch organisierten Analyse-Synthese-Systemen wurde im vergangenen Jahrzehnt an der TU Dresden UASR (**U**nified **A**pproach for Speech **S**ynthesis and **R**ecognition) entwickelt [5]. Besonderer Wert wurde dabei auf den technologischen Aspekt gelegt, denn nur durch die konsequente Implementierung aller UASR-Systemkomponenten in der Technologie der **F**inite-**S**tate-**T**ransducer (FST) konnte die Forschung systematisch nach dem „Analyse-durch-Synthese“-Prinzip vorangetrieben werden [6].
Im Unterschied zum biologischen Vorbild fallen hier zwei formale Unterschiede auf: Zum einen ist der in dem Bottom-Up-Prozess verlaufende Abstraktionsvorgang mit einer schritt-

weisen Abtrennung irrelevanter Information verbunden. Im UASR werden diese Informationen in einem zusätzlichen Zweig gesammelt. Die dort enthaltene Modellinformation muss bei der Synthese im Top-down-Prozess wieder zugesetzt werden, um ein realistisches Synthesesignal zu erhalten. Zum anderen ist die Kopplung der Hierachieebenen im biologischen Modell durch einen bidirektionalen Informationsfluss gekennzeichnet. Diese biologische Eigenart kann durch die Ausbildung und Simulation eines inneren Modells vom Kommunikationspartner begründet werden [7], [1]. Die Einführung der bidirektionalen Signalpfade kann daher auf gespiegelte Kommunikationsprozesse zurückgeführt werden. Erste konzeptuelle Modellierungsvorschläge zur bidirektionalen Informationsverarbeitung in hierarchischen Systemen sind in [8] publiziert worden. Biologisch motivierte Modellierungen führen zum sogenannten Cortikalen Algorithmus [9], [10]. Ein eigenständiges Konzept zur Realisierung des Cortikalen Algorithmus unter Verwendung von kaskadierten bidirektionalen Hidden-Markov-Modellen (CBHMM) wurde in [11] vorgestellt.

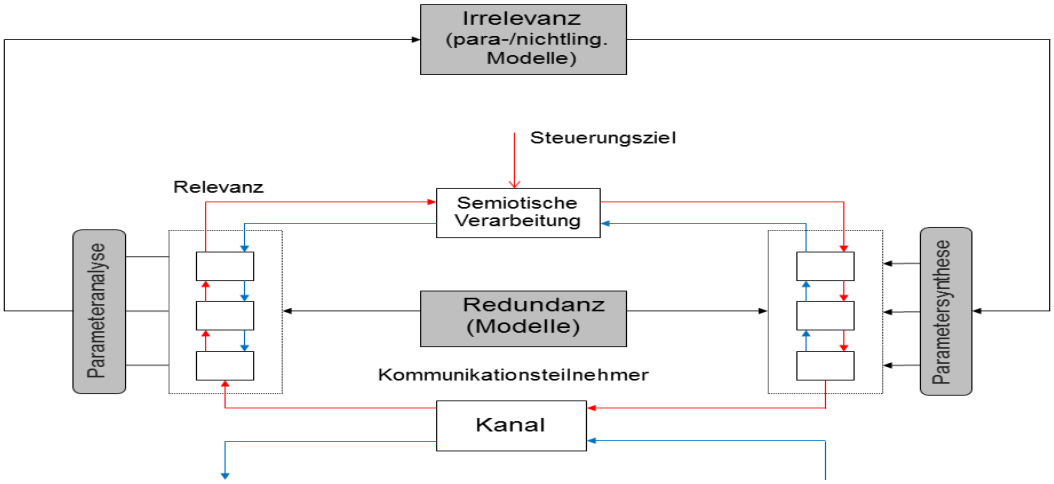

Abbildung 1 – Modellvorstellung für Analyse-Synthese-Systeme unter Berücksichtigung des bidirektionalen Informationsfluss zwischen den hierarchischen Ebenen. In der Modellvorstellung wird zwischen relevanter und irrelvanter Information und Redundanz unterschieden.

Die vorangegangenen Überlegungen führen schließlich zu einer Modellvorstellung von Analyse-Synthese-Systemen, nach denen sich ein erweitertes UASR skizzieren lässt (Abbildung 1). Weitergehende Überlegungen zur Erweiterung von UASR beziehen sich auf die Berücksichtigung der semantischen- und der pragmatischen Ebene sowie auf die Ebene der Zielvorstellungen [3].

Im Folgenden wird der Frage nachgegangen, welchen Einfluss die bidirektionale Signalverarbeitung auf ein probabilistisches Analyse-Synthese-Systeme hat. Der Beitrag gliedert sich wie folgt: In Abschnitt 2 wird zunächst die Struktur der CBHMM und dem damit verbundenen Cortikalen Algorithmus vorgestellt, anschließend erfolgt die Beschreibung des im Experiment verwendeten bidirektionalen Analyse-Synthese-Systems. In den folgenden Abschnitten wird auf den experimentellen Aufbau sowie die Durchführung und Auswertung des Experiments eingegangen. Abschließend werden die Ergebnisse diskutiert und ein Ausblick gegeben.

2 CBHMM

Die auch heute noch verwendete Notation für HMM wurde 1993 von Rabiner eingeführt [12], darüber hinaus können HMM aber auch in einer Notation angegeben werden, die sich an der Zustandsdarstellung für dynamische Systeme orientiert [13]. Mit dieser Notation ist einerseits der Zusammenhang zu den E/A-Automaten gegeben, andererseits kann die Darstellung von kaskadierten HMMs vereinfacht werden (vgl. Abbildung 2).

$$\begin{aligned} \mathbf{p}[s(t+1)] &= \mathbf{A} \cdot \mathbf{p}[s(t)] + \pi\delta(t) \\ \mathbf{q}(t) &= \mathbf{C} \cdot \mathbf{p}[s(t)] \end{aligned} \tag{1}$$

Aus der Sicht der Synthese modelliert die obere Gleichung die Zustandsübergange, wobei die Matrix **A** Übergangswahrscheinlichkeiten der gleichen Qualität (Zustand) für die Transitionen $s_{i,j}$ definiert. Das Eingangssymbol $\delta(t)$ ist per Definition nur zum Zeitpunkt *t=0* aktiv und wird nur zu diesem Zeitpunkt mit der Startverteilung $\boldsymbol{\pi}$ multipliziert. Die Ausgabeverteilung wird durch die Matrix $\mathbf{C}$ bestimmt, sie bestimmt die Übergangswahrscheinlichkeiten $c_{i,k}$ zwischen unterschiedlichen Qualitäten (Zustand, Gaußklasse). Mit dieser Matrix kann die Stärke der Kopplung zwischen Zustand und Gaußklasse bestimmt werden. Durch die Zustandsgleichung wird jedoch nicht die Emission eines Merkmalvektors $\mathbf{x}_t$ berücksichtigt. Hierfür muss gemäß der Ausgabeverteilung $q(t) = p_t(\omega)$ eine der *k* Gaußverteilungen ausgewählt werden, anschließend kann ein Merkmalvektor nach Gleichung 2 emittiert werden.

$$p(\mathbf{x}_t \mid \omega_k) = N(\mathbf{x}_t \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{2}$$

Der damit verbundene zweistufige Zufallsprozess endet, wenn ein Endzustand erreicht wurde. Aus Analysesicht vollzieht sich der umgekehrte Prozess. Hier wird der Merkmalvektor auf die Ausgabeverteilung abgebildet (Soft-Vektorquantisierung), Gleichung 2 wird dabei als Likelihood interpretiert.

$$p(\omega_k \mid \mathbf{x}_t) = \frac{p(\mathbf{x}_t \mid \omega_k) \cdot p(\omega_k)}{\sum_K p(\mathbf{x}_t \mid \omega_k) \cdot p(\omega_k)} \quad \text{bzw.} \quad p(\omega_k \mid \mathbf{x}_t) \approx \frac{p(\mathbf{x}_t \mid \omega_k)}{\sum_K p(\mathbf{x}_t \mid \omega_k)} \tag{3}$$

Den direkten Zusammenhang zwischen Beobachtungsvektor und der Zustandsverteilung erhält man schließlich durch die Anwendung der Transitionen $c_{i,k}$ zwischen Gaußklasse und Zustand:

$$p(s_i \mid \mathbf{x}_t) = \sum_K c_{i,k}\, p(\omega_k \mid \mathbf{x}_t) \tag{4}$$

Der letzte Analyseschritt kann als Soft-Interpretation verstanden werden [1]. Damit wurde nun sowohl die Analyserichtung - vom Merkmalvektor zur Zustandsverteilung - als auch die Syntheserichtung - von der Zustandsverteilung zum Merkmalvektor- für einen Zeitschritt beschrieben. Anschließend werden die Zustandsübergange nach Gleichung 1 durchgeführt, wobei ab Zeitschritt *t=1* der zweite Term verschwindet. D.h. Analyse und Synthese können einheitlich durch den Forward-Algorithmus [12] beschrieben werden. Die bidirektionale Erweiterung bei CBHMM erfolgt dadurch, dass bei der Analyse (Bottom-Up) eine Vorhersage in Top-Down Richtung erfolgt. Umgekehrt bei der Synthese (Top-Down), erfolgt die Vorhersage in Bottom-Up Richtung. Somit liegen in jeder symbolischen Hierachieebene $d \in \{0,1,...,D-1\}$ drei Terme vor, die in Gleichung 5 nach dem Bayesschen Prinzip fusioniert werden [11]. In der Fusionsgleichung wurde nun die Forward-Notation verwendet. Die Kopplung zwischen den symbolischen Hierarchieebenen wird durch die $\mathbf{B}$-Matrix realisiert, wobei der Übergang zur subsymbolischen Schicht $d = D$ von der $\mathbf{C}$-Matrix erfasst wird.

$$\gamma_j^d(k) = \underbrace{\left[\sum_l \beta_l^{d-1}(k) \cdot b_{j,l}^{d-1}\right]}_{Top-Down-Kopplung} \cdot \underbrace{\left[\sum_i \gamma_i^d(k-1) \cdot a_{i,j}^d\right]}_{zeitl.\ Kopplung} \cdot \underbrace{\left[\sum_r \alpha_r^{d+1}(k) \cdot b_{r,j}^d\right]}_{Bottom-Up-Kopplung} \tag{5}$$

Die Struktur des CBHMM und das Prinzip des Cortikalen Algorithmus ist im linken Teil der Abbildung 2 gemäß der in Abschnitt 1 angegebenen Modellvorstellung veranschaulicht worden. Die Darstellung orientiert sich dabei an der Zustandsdarstellung von HMM nach Gleichung 1. Daraus lässt sich erkennen, dass die mathematische Struktur der Kopplung für alle Schichten identisch ist.

3 Bidirektionales Analyse-Synthese-System

Nach [2] sind Analyse-Synthese-Systeme in den Perzeptions-Aktions-Zyklus eingebunden, wobei die bidirektionalen Signalpfade durch Spiegelung des Kommunikationspartners begründet werden können [1]. Die Aktionen des Kommunikationssubjekts zur Erreichung des Steuerungsziels werden von der Semiotischen Verarbeitung nach einer bestimmten Strategie ausgewählt und an das Kommunikationsobjekt übertragen. Außerdem ist es ein Kennzeichen kognitiver Systeme, dass es die Konsequenzen der ausgewählten Aktion vorhersagen kann. Dazu benötigt das System ein inneres Modell vom Kommunikationspartner, welches gemäß dem Reafferenzprinzip als Eingabe die reinterpretierte Aktion erhält und Vorhersagen liefert, mit denen die Wahrnehmung vorgebahnt wird. Über den Kanal wird schließlich die tatsächliche Reaktion des Kommunikationspartners auf die Sensorik des Kommunikationssubjekts zurückgekoppelt. Damit ist der Kommunikationskreislauf geschlossen und das Kommunikationssubjekt kann überprüfen, ob es sich der Zielvorstellung angenähert hat. Dieser Zyklus wiederholt sich solange bis die Zielvorstellung erreicht wurde.

In Abbildung 2 ist der Perzeptions-Aktions-Zyklus unter Verwendung von CBHMM für das Kommunikationssubjekt dargestellt. Um nun den Einfluss der bidirektionalen Signalverarbeitung auf die Kommunikation zu untersuchen, können zwei Möglichkeiten der Einspeisung in die Kopplungsmatrix $\mathbf{B}$, entweder vor der Fusion (Abbildung 2, links) oder nach der Fusion (Abbildung 2, rechts), betrachtet werden. Dies wirkt sich direkt auf die Fusionsgleichung aus, da die Bottom-Up Kopplung nun von $\gamma_r^{d+1}(k)$ verursacht wird.

$$\gamma_j^d(k) = \underbrace{\left[\sum_l \beta_l^{d-1}(k) \cdot b_{j,l}^{d-1}\right]}_{Top-Down-Kopplung} \cdot \underbrace{\left[\sum_i \gamma_i^d(k-1) \cdot a_{i,j}^d\right]}_{zeitl.\ Kopplung} \cdot \underbrace{\left[\sum_r \gamma_r^{d+1}(k) \cdot b_{r,j}^d\right]}_{Bottom-Up-Kopplung} \tag{6}$$

In beiden Fällen wird der Ausgang des HMM nicht direkt auf den Eingang des benachbarten HMM geführt, sondern der Ausgang des HMM wird über die Kopplungsmatrix mit der Zustandsverteilung des benachbarten HMM verknüpft. Aus dieser Betrachtung wird deutlich, dass die HMM-Kaskade nicht durch eine Hintereinanderschaltung im Sinne von - Ausgabe des einen Teilsystems ist Eingabe des folgenden Teilsystems - realisiert wird. Die Zustände von CBHMM werden also nicht nur vom Eingang her, sondern auch direkt durch Einkopplung über die $\mathbf{B}$-Matrix angesteuert. Damit liegt aber keine der typischen Verknüpfungen von Teilsystemen vor, so wie sie in der Systemtheorie behandelt werden. Insbesondere darf somit nicht von einer Komposition benachbarter HMM im strengen Sinne ausgegangen werden. Andererseits kann festgestellt werden, dass der linke Teil von Abbildung 2 auch nach der Fusion die Zustandsgleichung 1 berücksichtigt. Nun übernimmt die Kopplungsmatrix $\mathbf{B}$ die Rolle der Ausgabematrix $\mathbf{C}$, so dass wie oben gezeigt wurde, die Ausgabe mit der Zustandsverteilung des benachbarten HMM verknüpft wird.

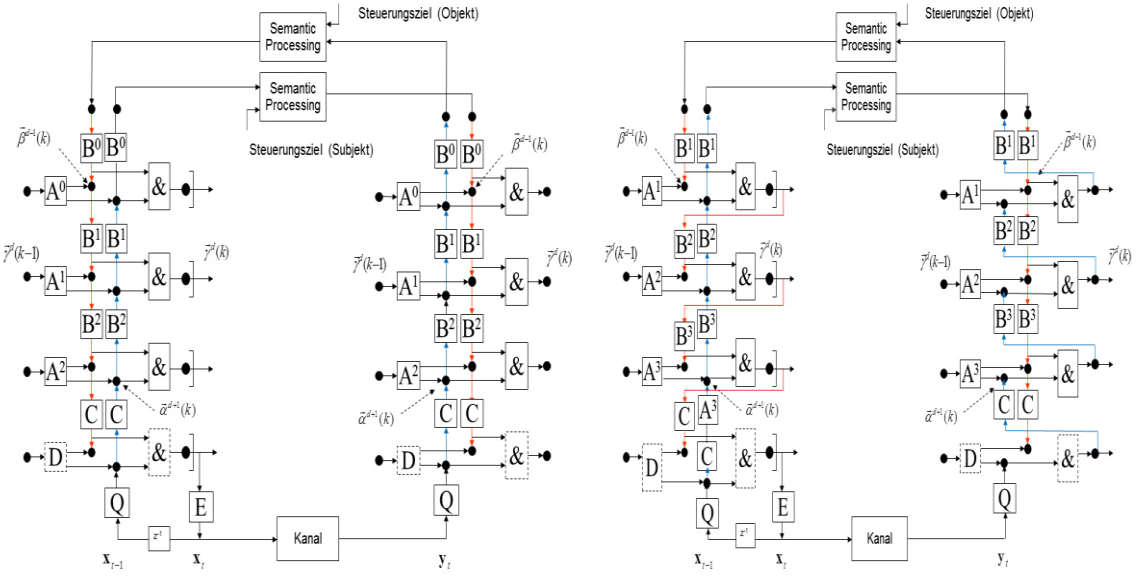

Abbildung 2 – Bidirektionales Analyse-Synthese-System. Links) Die Realisierung eines Analyse-bei Einspeisung der Kopplung vor der Fusion. Rechts) Alternativ kann die Einspeisung der Kopplung nach der Fusion erfolgen. In beiden Fällen wird nicht in den Eingang des benachbarten HMM eingekoppelt sondern auf die Zustandsverteilung.

4 Beschreibung des Experiments

Bereits in Abschnitt 3 wurde darauf hingewiesen, dass die Kopplung zwischen den Schichten über zwei verschiedene Wege erfolgen kann. Um den Algorithmus für bidirektionale Synthese und Analyse dahingehend zu untersuchen, wird zunächst mit einschichtigen HMM experimentiert. Dadurch sollen vor allem sekundäre Einflüsse ausgeschlossen werden, die eine Interpretation der Ergebnisse erschweren würden. Um die Untersuchung auf das Wesentliche zu beschränken, wird von der Quelle nur ein Modell verwendet, so dass immer das gleiche Symbol erzeugt wird. Dabei durchläuft das HMM verschiedene Zustände und emittiert in jedem Zeitschritt einen Merkmalvektor. Die Aufgabe des Empfängers besteht darin, ausgehend von der Merkmalvektorfolge die entsprechende Zustandsfolge zu decodieren. Die Anzahl der vom Viterbi-Decoder korrekt decodierten Zustände kennzeichnet somit die Qualität der Übertragung und wird im Folgenden als Erkennrate bezeichnet .

Der Aufbau des Experiments orientiert sich am Kanalmodell nach Shannon und verwendet als Quelle ein einschichtiges HMM mit 5 Zuständen und 5 Gaußklassen für 2-dimensionale Merkmalvektoren. Die Matrizen **A** und **C** werden vorgegeben und auch auf der Empfangsseite verwendet. Das Kanalmodell setzt auf der Ebene der Merkmalvektoren auf und erzeugt unkorreliertes 2-dimensionales Gaußsches Rauschen, welches dem Sendevektor bei gegebenem SNR überlagert wird.

$$SNR = 10\log_{10}\left[\frac{1}{N}\cdot\sum_{n=1}^{N}\frac{\sigma_{S,n}^2}{\sigma_{W,n}^2}\right] \tag{7}$$

Das SNR wird im Experiment vorgegeben, daher muss der Verstärkungsfaktor für jede einzelne Gaußkomponente des Rauschvektors wie folgt berechnet werden:

$$g = \frac{\sigma_S}{\sigma_W}\cdot 10^{-SNR/20} \tag{8}$$

Sowohl Synthese als auch Analyse können unidirektional bzw. bidirektional betrieben werden. Um die Prädiktionen in Top-Down Richtung bzw. in Bottom-Up Richtung für das einfache Kanalmodell berechnen zu können, muss die semantische Verarbeitung aus Abbildung 2 durch ein Signalmodell ersetzt werden. Bei der Synthese wird hierfür ein um einen Takt verzögerter Merkmalvektor quantisiert und die Quantisierungsverteilung mit der transponierten $\mathbf{C}$-Matrix multipliziert. Anschließend liegt eine Zustandsverteilung vor, mit der schließlich die vorhergesagte Zustandsverteilung über die $\mathbf{A}$-Matrix berechnet werden kann. Bei der Analyse wird eine Top-Down Prädiktion benötigt. Hier wird nun der Zustandsvektor um einen Takt verzögert und anschließend ebenfalls über die A-Matrix eine Zustandsverteilung für den aktuellen Zeitschritt prädiziert.

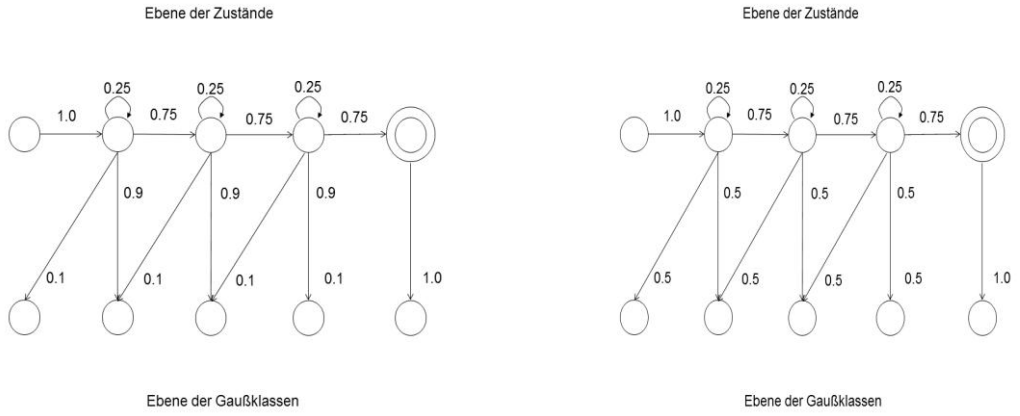

Abbildung 3 – Darstellung der HMM Parameter für starke und schwache Kopplung. Oben sind die Zustandsübergänge definiert, nach unten erfolgt die Kopplung zu jeweils zwei Gaußklassen.

Die Untersuchungen zum Einfluss der bidirektionalen Signalverarbeitung konzentrieren sich auf zwei Aspekte. Einerseits kann die Qualität der Übertragung für die beiden Ausprägungen der Fusionsgleichung gemessen werden, so dass eine Strukturentscheidung fallen kann. Darüber hinaus kann aber auch die Parametrierung des HMM unterschiedlich gewählt werden. Im vorliegenden Experiment wurde daher die $\mathbf{C}$-Matrix für zwei Extremfälle untersucht. Bei der schwachen Kopplung kann ein Zustand verschiedene Gaußkurven auswählen. Dagegen ist die Anzahl der auszuwählenden Gaußkurven bei der starken Kopplung stark reduziert. Die $\mathbf{A}$-Matrix wurde im Experiment nicht verändert. Die Gaußkurven wurden hinsichtlich Mittelwert und Varianz so gewählt, dass die Überlappungen im ungestörten Fall vernachlässigbar sind. Die konkrete Parametrierung für die starke und schwache Kopplung ist in Abbildung 3 ersichtlich.

5 Durchführung und Auswertung

Das Experiment wurde zuerst für die starke Kopplung durchgeführt. Bereits bei der Synthese konnte ein interessanter Effekt beobachtet werden. Im Vergleich zur unidirektionalen Synthese konnte unabhängig von der Anzahl der gesendeten Symbole bei der bidirektionalen Synthese ungefähr eine Verdopplung der Zeitdauer beobachtet werden, die zum Durchlaufen der HMM-Zustandsfolge benötigt wird. Dieses Verhalten kann auf ein leichtes Nachlaufen der Prädiktionsverteilung zurückgeführt werden und hat in der Konsequenz eine Erhöhung der Selfloop-Wahrscheinlichkeiten zur Folge. Diese Beobachtung konnte für beide Ausprägungen der Fusionsgleichung festgestellt werden und kann als Einfügen von Redundanz interpretiert werden, welche vom Empfänger bei der Decodierung ausgenutzt werden könnte. Im nächsten Schritt wurde für beide Ausprägungen die ungestörte Übertragung überprüft. Danach wurde

das SNR im Bereich von 20 dB bis -20 dB variiert und die Anzahl der korrekt decodierten Zustände gemessen. Die Ergebnisse sind in Abbildung 4 dargestellt. Bei guten und mittleren SNR können leicht bessere Ergebnisse erzielt werden, wenn Fusionsgleichung 5 zur Anwendung kommt. Auch bei schlechten SNR ist diese Ausprägung der Fusionsgleichung robuster.

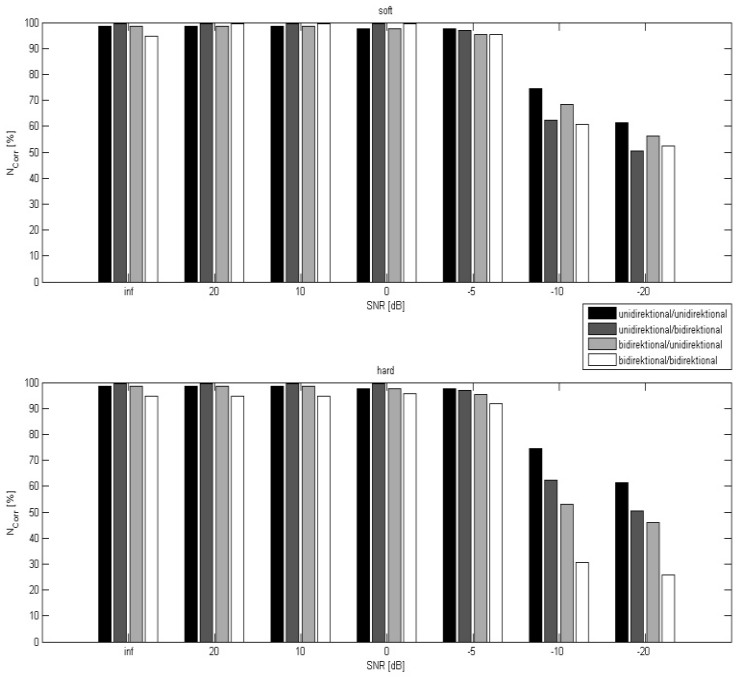

Abbildung 4 – Anzahl der korrekten Zustandserkennungen in Abhängigkeit vom SNR bei **starker Kopplung** zwischen Zustand und Klasse. Die in der Legende angegebene Richtung des Informationsflusses folgt dem Schema Synthese/Analyse. Oben erfolgt die Einspeisung der Kopplung vor der Fusion (5), unten dagegen nach der Fusion (6).

Anschließend wurde das Experiment ebenfalls für beide Ausprägungen der Fusionsgleichung nun aber bei schwacher Kopplung wiederholt. Abbildung 5 zeigt die Ergebnisse für die schwache Kopplung. Hier werden die Verbesserungen noch etwas deutlicher und auch bei der schwachen Kopplung zeigt Fusionsgleichung 5 ein robusteres Verhalten. Insgesamt zeigen die ersten Experimente ermunternde Ergebnisse, allerdings müssen weitere Messungen mit verschiedenen Parametrierungen durchgeführt werden, um aussagekräftige Statistiken zu erhalten. Darüber hinaus muss bei niedrigen SNR geklärt werden, welche Ursachen für die Abnahme der Decodierbarkeit verantwortlich sind. Vor allem der starke Abfall für bidirektionale Synthese und bidirektionale Analyse lässt zunächst nur den Schluss zu, dass die hinzugefügte Redundanz von der Analyse nicht ausgenutzt werden kann.

Die Hoffnung, dass der Decoder die von der Synthese eingefügte Redundanz bei schlechten SNR ausnutzen kann, hat sich nicht erfüllt. Möglicherweise beruht dieses Verhalten darauf, dass die Prädiktion auf der einen Zeitschritt zuvor gemessenen Zustandsverteilung basiert, welche mit geringer werdendem SNR zunehmend fehlerhaft wird. Die auf dem Empfangssignal basierte Prädiktion hat in diesem Fall scheinbar eine störende Wirkung. Daher sollen in einem nächsten Schritt die Grenzen der Prädiktionsfähigkeit bei variierendem SNR ausgelotet werden.

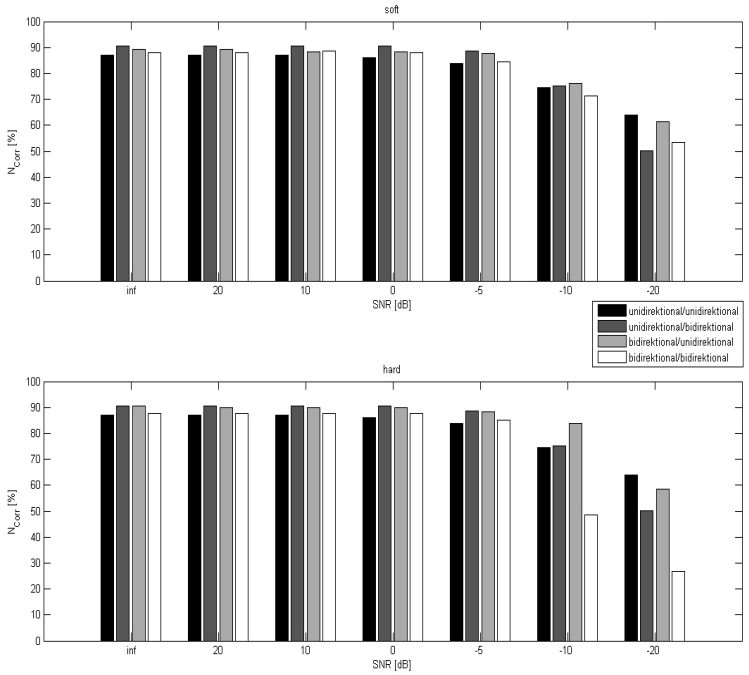

Abbildung 5 – Anzahl der korrekten Zustandserkennungen in Abhängigkeit vom SNR bei schwacher Kopplung zwischen Zustand und Klasse. Die in der Legende angegebene Richtung des Informationsflusses folgt dem Schema Synthese/Analyse. Oben erfolgt die Einspeisung der Kopplung vor der Fusion (5), unten dagegen nach der Fusion (6).

Zu diesem Zweck wurden dem Decoder anstelle der auf den Empfangsvektoren basierenden Prädiktionen direkt die Prädiktionen von der Synthese zugeführt. Damit liegen für den Decoder ungestörte Prädiktionen vor, mit denen sich empirisch eine obere Grenze für die Erkennrate ermitteln läss. Die Abbildungen 6 und 7 zeigen die Anzahl der Zustandserkennungen wiederum für starke und schwache Kopplung nun aber bei idealen bzw. ungestörten Prädiktionen. Für beide Fälle treten die erwarteten Beobachtungen ein. Zum einen erhält man nun die besten Ergebnisse für die „matched“ Situationen, in denen sowohl die Synthese als auch die Analyse bidirektional ausgeführt wird. Dabei zeigt sich über den gesamten SNR-Bereich ein robustes Verhalten, dies gilt in besonderem Maß für die schwache Kopplung. Zum anderen fällt die Erkennrate in den „unmatched“ Situationen gegenüber den „matched“ Situationen ab.

Bei bidirektionaler Synthese können keine gestörten Prädiktionen auftreten, die dabei eingeführte Redundanz kann nun auch bei unidirektionaler Decodierung ausgenutzt werden und führt bei schlechten SNR zu besseren Erkennraten als bei unidirektionaler Synthese und unidirektionaler Analyse. Dagegen kann die bidirektionale Analyse keine Verbesserungen erzielen, wenn lediglich unidirektional synthetisiert wird. Insgesamt kann abschließend festgehalten werden, dass die Ergebnisse des Experiments mit idealer Prädiktion für alle vier Analyse-Synthese Situationen plausibel erklärt werden können.

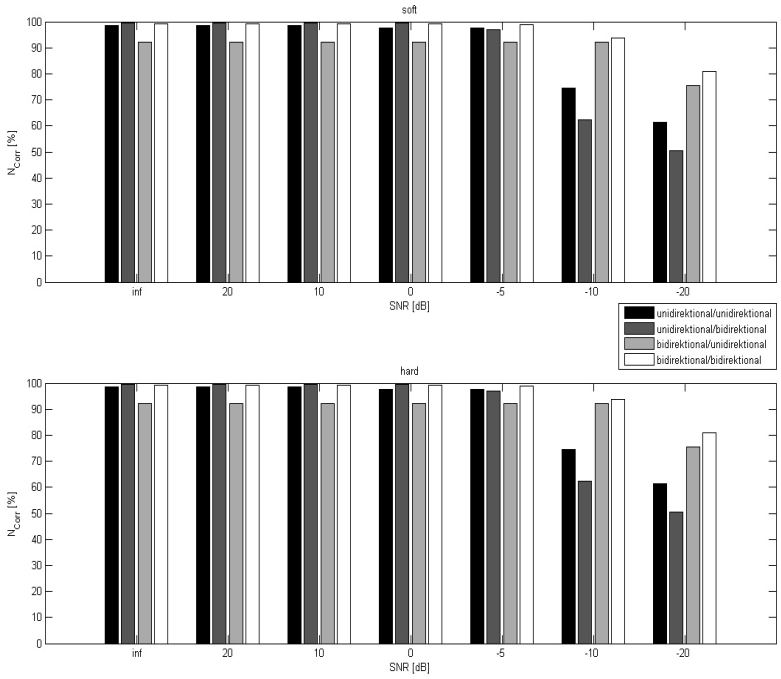

Abbildung 6 – Anzahl der korrekten Zustandserkennungen in Abhängigkeit vom SNR bei **starker Kopplung** zwischen Zustand und Klasse nun aber bei idealer Prädiktion. Die in der Legende angegebene Richtung des Informationsflusses folgt dem Schema Synthese/Analyse. Oben erfolgt die Einspeisung der Kopplung vor der Fusion (5), unten dagegen nach der Fusion (6).

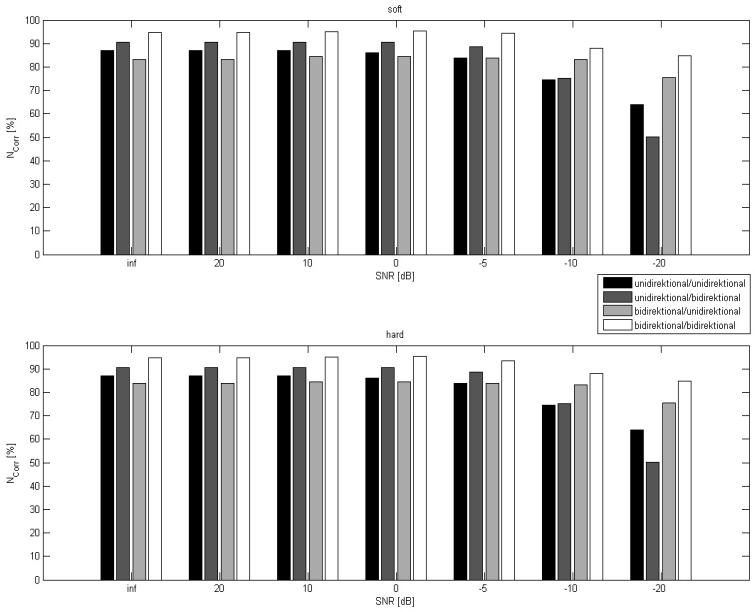

Abbildung 7 – Anzahl der korrekten Zustandserkennungen in Abhängigkeit vom SNR bei **schwacher Kopplung** zwischen Zustand und Klasse nun aber bei idealer Prädiktion. Die in der Legende angegebene Richtung des Informationsflusses folgt dem Schema Synthese/Analyse. Oben erfolgt die Einspeisung der Kopplung vor der Fusion (5), unten dagegen nach der Fusion (6).

6 Schlussbemerkung und Ausblick

In diesem Beitrag wurde der Einfluss der bidirektionalen Signalverarbeitung auf die Decodierbarkeit von HMM Zuständen untersucht. Der Aufbau des Experiments orientierte sich dabei am Shannonschen Kanalmodell. Für die Untersuchungen zur bidirektionalen Signalverarbeitung wurde auf zwei Prädiktionsmethoden zurückgegriffen. Zunächst zeigten die bidirektionalen HMM mit signalbasierten Prädiktionen bei guten und mittleren SNR gegenüber den unidirektionalen HMM eine leicht verbesserte Decodierbarkeit. Dieser Effekt wird insbesondere bei schwacher Kopplung deutlich. Die Hoffnung, dass der Decoder die von der Synthese eingefügte Redundanz bei schlechten SNR ausnutzen kann, hat sich bei den signalbasierten Prädiktionen allerdings nicht erfüllt. Daher wurden in einem zweiten Schritt ideale Prädiktionen verwendet, mit denen ein „proof of concept“ durchgeführt wurde. Hier zeigte sich nun das erwartete Verhalten für alle vier möglichen Analyse-Synthese Situationen. Die ermutigenden Ergebnisse des zweiten Experiments zeigen damit auch die Richtung der nächsten Forschungsarbeiten an, die das Ziel verfolgen müssen, dem Decoder weitgehend ungestörte Prädiktionen zur Verfügung stellen zu können.

7 Danksagung

Für die Unterstützung und kritischen Diskussionen gebührt mein Dank sowohl Tobias Herbig (Nuance Communications GmbH) als auch Matthias Wolff (BTU Cottbus).

8 Literatur

[1] 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.
[2] Haykin, S.: Cognitive Dynamic Systems. Cambridge University Press, 2012.
[3] Wolff, M.; Römer, R. und Hoffmann, R.: Hierarchische Kognitive Systeme zur Sprach- und Signalverarbeitung, 23. Konferenz Elektronische Sprachsignalverarbeitung ESSV-2012, TUDpress Dresden 2012.
[4] Fuster, J.: Cortex and Mind, Unifying Cognition, Oxford Press 2003.
[5] Hoffmann, R.; M. Eichner, M. and M. Wolff, M.: Analysis of verbal and nonverbal acoustic signals with the Dresden UASR system, in Verbal and Nonverbal Communication Behaviours, A. Esposito *et al.*, Eds. Berlin etc.: Springer, 2007, pp. 200–218 (LNAI vol. 4775).
[6] Wolff, M.: Akustische Mustererkennung. Habilitationsschrift, TUDpress Dresden, 2011.
[7] Goldman, A.I.: Simulating Minds. The Philosophy, Psychology and Neuroscience of Mindreading. Oxford University Press, Oxford 2006.
[8] Lee, T.S.; Mumford, D.: Hierarchical Bayesian Inference in the Visual Cortex. JOSA Vol.20, No.7/July 2003.
[9] George, D.; Hawkins, J.: Invariant Pattern Recognition using Bayesian Inference on Hierarchical Sequences. RNI Technical Report 2005.
[10] George, D.; Hawkins, J.: A Hierarchical Bayesian Model of Invariant Pattern Recognition in the Visual Cortex. International Joint Conference on Neural Networks 2006.
[11] Römer, R.; Herbig,T.: Konzeptionelle Beschreibung des Cortikalen Algorithmus und seine Anwendung in der automatischen Spracherkennung, 20. Konferenz der ESSV, Dresden September 2009.
[12] Lawrence R. Rabiner, L.R.; Juang. B.H.: Fundamentals of Speech Recognition, Prentice Hall 1993.
[13] Moon,T. K.; Stirling, W. C.: Stirling, Mathematical Methods and Algorithms for Signal Processing. Addison Wesley, September 1999.