Teil-Bericht zum 6. Projektabschnitt

Andreas Kitzig

September 2011

Bericht zum Vorhaben „Robuste Spracherkennung in gestörter Umgebung durch die Kombination einer robusten Merkmalsextraktion und einer Adaption der Referenzmuster“, gefördert durch das BMBF im Rahmen des Förderprogramms FHProfUnd an der Hochschule Niederrhein unter Leitung von Prof. Dr. Hirsch, Förderkennzeichen 1762X08

Zusammenfassung

Die Leistungsfähigkeit der in diesem Projekt verwendeten und entwickelten Erkennungssysteme „HGH adapt“ und „HGH robust“ wurde in den vorherigen Projektberichten bereits ausführlich dargestellt. Hierbei wurde in Tests eine Vielzahl an unterschiedlichen Störumgebungen eingesetzt, um einen vollständigen Überblick über die Leistungsfähigkeit der Systeme zu vermitteln. Abschließend soll in diesem Projektbericht die Verwendbarkeit der robusten Systeme „HGH adapt“ und „HGH robust“ auf Rechnersystemen mit beschränkten Hardware-Ressourcen abgeschätzt werden. Dazu wurden die beiden robusten Erkennungssysteme mit einem Profiling-Tool untersucht, um die Anzahl der benötigten Instruktionen der einzelnen Funktionsblöcke pro Sekunde zu ermitteln. Dieser Wert wird in Form von W-MIPSSP (Whetstone Million Instructions per Second Speechdata) dargestellt. Anhand des W-MIPSSP Wertes kann anschließend der Rechenaufwand der robusten Erkennungssysteme vom Test-Rechnersystem unabhängig dargestellt und auf Rechnersystemen mit beschränkten Hardware-Ressourcen übertragen werden. Im vorliegenden Text werden die Durchführung und die Ergebnisse der Untersuchung dokumentiert. Nach einer kurzen Einleitung werden im folgenden Abschnitt die Vorgehensweise und die verwendeten Testdaten und Verfahren dargestellt. Abschließend werden die erzielten Ergebnisse ausgewertet und diskutiert.

Einleitung

In den vorherigen Projektberichten wurde der Aufbau und die Erkennungsleistung der einzelnen HGH-Erkennungssysteme ausführlich erläutert und verglichen (siehe dazu Projektbericht[1] 1 bis 6). In diesem Projektabschnitt wurde ermittelt, wie sich die HGH-Systeme auf Rechnersystemen mit beschränkten Hardware-Ressourcen einsetzen lassen. Dies sind zum Beispiel Tablett-PCs, Smartphones oder herkömmliche Mobiltelefone. Die Betrachtung mobiler Endgeräte beschränkt sich in dieser Untersuchung jedoch auf die Bestimmung der Laufzeit der Funktionsaufrufe, die die einzelnen Verarbeitungsblöcke des jeweiligen Erkennungssystems benötigen sowie auf eine Bestimmung der Anzahl der Instruktion pro Sekunde in W-MIPS (Whetstone Million Instructions per Second) des Rechner-Testsystems. Dazu werden ein Profiling- und ein Benchmarktool verwendet, die im folgenden Abschnitt näher beschrieben werden. Aus den Ergebnissen wird anschließend ein vom Rechner-Testsystem entkoppelter und verallgemeinerter Wert zum Vergleich mit anderen Systemen berechnet.
Als Testdaten werden die bereits in den vorherigen Berichten dargestellten Daten der Aurora5 / TIDigits Datenbank verwendet. Das verwendete Testset enthält 8700 Testfiles, die verarbeitet werden.

Rechner-Testsystem und Benchmark

In diesem Abschnitt werden zuerst die Eigenschaften des für den Test verwendeten Rechner-Testsystems aufgelistet. Anschließend erfolgt eine Darstellung des verwendeten Benchmark-Tools.
Die Erkennungsexperimente wurde auf einer SUN Workstation mit folgenden Eigenschaften durchgeführt:

- E8600 Intel Pentium Doppelkern-Prozessor (3,3GHz)
- 4GB Arbeitsspeicher
- nVidia Quadro FX 1700 Grafikkarte
- System: openSUSE 11.1 (x86_64)
- KDE: 3.5.10 ”release 21.11”
- Kernel 2.6.27.7-9-default x86_64

[1]Projektseite unter www.dnt.kr.hsnr.de / → Forschung → FHProfUnd Projekt

Um die mit dem Rechner-Testsystem ermittelten Werte mit anderen Systemen vergleichen und unabhängig darstellen zu können, wurde ein Maß für die Leistung des Rechner-Testsystems bestimmt. Dazu wurde ein Benchmark mit der Software "Berkeley Open Infrastructure for Network Computing" (kurz BOINC)[2] durchgeführt. Auf dieses Benchmark beziehen sich die im weiteren verwendeten Ergebnisse.

Die Software-Plattform BOINC wurde ursprünglich mit dem Ziel erstellt, verteiltes Rechnen auf vielen Einzelsystemen zu ermöglichen. Sie dient zur Bearbeitung von rechenintensiven wissenschaftlichen Projekten unter Ausnutzung der Rechenleistung von Computern angemeldeter Nutzer. Auf den offiziellen Web-Seiten[3] des Projekts findet sich eine Angabe von zur Zeit 290604 aktiven Benutzern, die ihre Rechenleistung zur Verfügung stellen. Ein aktuell durchgeführtes Projekt ist zum Beispiel die Erstellung eines genauen 3D-Modells der Milchstraße.

Weiterhin enthält die Software auch das zuvor erwähnte Benchmark Tool. Als Ergebnis liefert das Tool einen Wert, mit der sich die Geschwindigkeit und Effizienz von Computern vergleichbar messen lässt. Dieser Wert liegt in Form von „Million Whetstone-Instruktionen pro Sekunde“, kurz W-MIPS, vor und bezieht sich auf einen 1976 im National Physical Laboratory entwickelten Benchmark, der für seine Messungen Gleitkommazahl-Operationen, aber auch Ganzzahl-Arithmetik und Zugriffe auf Feld-Elemente, nutzt. Dieser Benchmark lässt sich auf nahezu jedem Rechnersystem ausführen und stellt eine weit verbreitete Möglichkeit zum Vergleich der Leistungsfähigkeit von Rechnersystemen dar. Für das verwendete Rechner-Testsystem ergibt sich folgender mittlerer Benchmark- Wert in W-MIPS (Whetstone Million Instructions per second) pro CPU:

Whetstone MIPS
3912

Tabelle 1: Ergebnisse des Benchmarks in W-MIPS

In Worten ausgedrückt bedeutet dies, dass es mit dem vorgestellten Testsystem möglich ist, pro Sekunde 3912 Millionen Gleitkomma-Instruktionen (Whetstone) auf einer CPU durchzuführen. Um einen Vergleich bezüglich der W-MIPS Werte zu ermöglichen, sind in Tabelle 2 die in einem Internet-Forum[4] aufgelisteten Benchmark-Werte für zwei weitere Systeme mit aktuellen Intel und AMD CPUs dargestellt:

[2] http://boinc.berkeley.edu/
[3] http://boincstats.com/
[4] http://www.meisterkuehler.de/forum/benchmarks/23400-benchmark-boinc-benchmark-version-6-x.html

	Testsystem SUN Workstation @ 3,3 GHz	Vergleichs-system Intel i7-2600k @ 4,5 GHz	Vergleichs-system Athlon2 x2 260 @ 3,2GHz
W-MIPS	3912	4237	2570

Tabelle 2: Vergleichssysteme

Aus der Betrachtung der Werte in Tabelle 2 wird ersichtlich, dass die verwendete Workstation im Vergleich mit aktuellen Systemen eine mittlere Systemleistung aufweist.

Laufzeitmessung

In diesem Abschnitt wird das zur Messung der Laufzeit verwendete Profiling-Utility vorgestellt. Die Darstellung der Ergebnisse der Laufzeitmessung erfolgt im nächsten Abschnitt.

Zur Messung der Laufzeiten der einzelnen Funktionen wurde ein Profiling-Utility mit dem Namen ”gprof” verwendet. Das Utility liefert ausführliche Statistiken zum Ablauf des zu überprüfenden Test-Programms, zum Beispiel, wie oft und wo eine Funktion aufgerufen wird sowie die Laufzeit für jede einzelne Funktion usw. Ein Teil der Ausgabe ist als Beispiel in Abbildung 1 dargestellt:

Flat profile:

Each sample counts as 0.01 seconds.

% time	cumulative seconds	self seconds	self calls	total ms/call	ms/call	name
78.87	741.62	741.62	6582220	0.11	0.11	calc_local_prob
3.31	772.71	31.09	20781358	0.00	0.00	cos_t
2.06	792.09	19.38	46767	0.41	1.82	adapt_hmm_all
2.04	811.30	19.21				spline1_c
2.03	830.37	19.07	779759407	0.00	0.00	d_max_arg
1.64	845.81	15.44	6582220	0.00	0.12	viterbi_syn_calc
1.62	861.01	15.21	1522431	0.01	0.02	anal_cep_frame
1.37	873.88	12.87	1522431	0.01	0.01	rfft
1.13	884.53	10.65	1705816	0.01	0.01	calc_deltas_ada
1.09	894.80	10.27	46767	0.22	0.22	extract_c0_ref
1.02	904.37	9.57	46767	0.20	0.94	adapt_hmm_delta
0.88	912.63	8.26	46767	0.18	0.18	viterbi_syn_set
0.86	920.72	8.09	5249630	0.00	0.00	icos_t
0.80	928.24	7.52	274381	0.03	0.03	calc_weight_coeff
0.33	931.36	3.12	6582220	0.00	0.00	calc_node_prob
0.29	934.12	2.76	1522431	0.00	0.00	estimate_noise
...						

Abbildung 1: Ausschnitt der Profiling Informationen

Die Information über die Laufzeit liegen, wie in Abbildung 1 dargestellt wurde, unter anderem als „Flat Profile“ vor. In diesem Profil sind folgende Werte angegeben:

- der prozentuale Anteil der Laufzeit („% time“) der einzelnen Funktion an der Gesamtlaufzeit
- die aufsummierte Laufzeit („cumulative seconds“)
- die einzelne Laufzeit der jeweiligen Funktion („self seconds“)
- die Anzahl der Aufrufe („self calls“)
- sowie Informationen über den Zeitbedarf pro Aufruf und der Name der aufgerufenen Funktion

Falls in dem Profile-Ergebnis die Information zu der Anzahl der Aufrufe sowie den Informationen über den Zeitbedarf pro Aufruf nicht angegeben sind, wurde dieser Funktionsaufruf bei der Kompilierung nicht mit in den Profiler eingebunden. Da aber für alle Aufrufe, die in die Auswertung eingehen, die benötigte Laufzeit angegeben ist, ist dieser Punkt in der Betrachtung zu vernachlässigen. Für die weitere Auswertung wurden die Angaben zu der einzelnen Laufzeit der jeweiligen Funktion („self seconds“) verwendet.

Berechnung W-MIPSSP

In diesem Abschnitt wird aus dem in den Benchmarks ermittelten W-MIPS Wert für das Rechner-Testsystem und den mit dem Profiling-Utility ermittelten Laufzeiten der einzelnen Funktionsblöcke ein systemunabhängiger Wert mit der Bezeichnung W-MIPSSP (Whetstone Million Instructions per Second Speechdata) berechnet. Dieser stellt einen vom verwendeten Rechnersystem unabhängigen Wert dar und repräsentiert die zur Verarbeitung einer Sekunde Sprachdaten benötigte Anzahl an Instruktionen. Der Begriff „eine Sekunde Sprachdaten“ bedeutet in diesem Zusammenhang, dass der Ausschnitt aus dem Sprachsignal entweder Sprache, eine Sprachpause oder eine Kombination von beidem enthalten kann.

Für die Durchführung der Laufzeittests wurde eine C-Implementierung verwendet. Diese beinhaltet beide robusten Erkennungssysteme „HGH adapt“ und „HGH robust“, die als Erkennungsmodus ausgewählt werden können. Dem Erkennungssystem werden dabei als Parameter das zu verarbeitende Testfile, eine Liste mit Referenzmustern, ein Syntaxfile und eine Angabe über das robuste Erkennungssystem, das verwendet werden soll, übergeben.

Für die Erkennungsexperimente werden in jedem Experiment insgesamt 23 Referenzmodelle mit je 42 Cepstral-Koeffizienten verwendet. Diese sind unterteilt in 22 geschlechtsspezifische Wort-Modelle, die die englischen Ziffern von 1 bis 9 inkl. „zero“ und „oh“ repräsentieren. Jedes Wort-Modell beinhaltet 16 States und zwei Gauss- Mischverteilungen pro State. Weiterhin wird ein Pausenmodell verwendet, dieses beinhaltet einen State und acht Gauss- Mischverteilungen.

Da im Rahmen des Profiling die Laufzeit jeder verwendeten Funktion ermittelt wird, werden einzelne Funktionen zu Verarbeitungsblöcken zusammengefasst. Hierbei lassen sich die beiden robusten Erkennungsysteme HGH adapt und HGH robust auf die in Abbildung 2 dargestellten, essentiellen Verarbeitungsblöcke zerlegen:

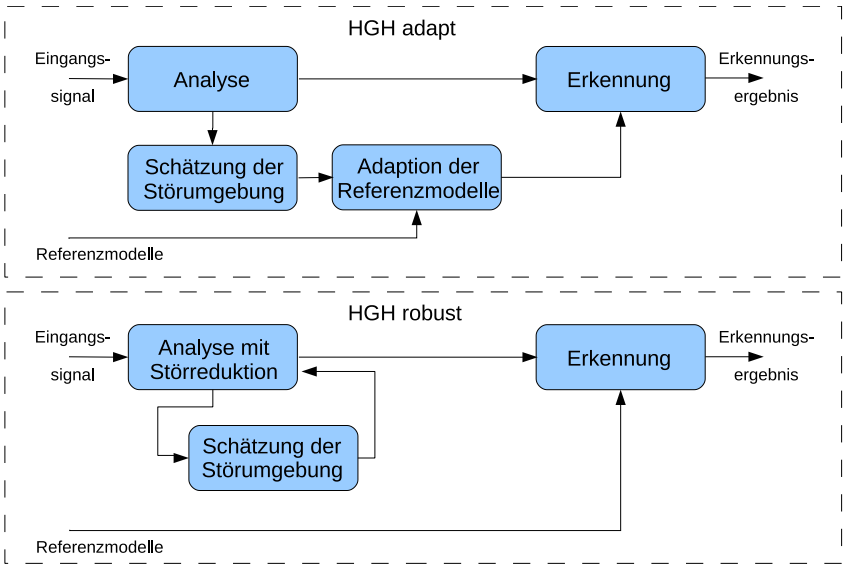

Abbildung 2: Blockschaltbilder der Erkennungsysteme

Um nun einen vergleichbaren Wert über die Anzahl der Instruktionen pro Sekunde Sprachdaten, die ein Funktionsblock benötigt, angeben zu können, wird aus der aufsummierten Laufzeit eines Funktionsblocks in Sekunden, der Angabe über die W-MIPS und der Gesamtlänge aller verarbeiteten Äußerungen ein systemunabhängiger Wert berechnet. Dieser Wert wird im Folgenden als W-MIPSSP (Whetstone Million Instructions per Second Speechdata) bezeichnet:

$$WMIPSSP = \frac{\sum_{i=1}^{M} t_i \cdot MIPS}{N}$$

$$mit\ t = Dauer\ eines\ Funktionsaufrufes$$

$$M = Anzahl\ aller\ Funktionsaufrufe\ im\ Verarbeitungsblock$$

$$MIPS = Anzahl\ der\ Instruktionen\ pro\ Sekunde$$

$$N = Gesamtlänge\ der\ Äußerungen\ in\ Sekunden$$

In die Betrachtung, welche Funktionsaufrufe zu einem Funktionsblock zusammengefasst werden, fließen nur Funktionsaufrufe ein, die in dem Profiling Ergebnis mehr als 0,2% Anteil an der Gesamtlaufzeit des Programms besitzen, alle anderen Funktionsaufrufe werden vernachlässigt. Anhand der Berechnung ergibt sich die in Tabelle 3 angegebene Anzahl an Instruktionen für die einzelnen Funktionsblöcke des jeweiligen Erkennungssystems:

	HGH adapt		HGH robust	
	Gesamt-Laufzeit über 15399s Sprachdaten	W-MIPSSP für eine Sekunde Sprachdaten	Gesamt-Laufzeit über 15399s Sprachdaten	W-MIPSSP für eine Sekunde Sprachdaten
Analyse	39s	10	201s	51
Stör-schätzung	3s	1	18s	5
Erkennung	-	-	623s	158
Erkennung + Adaption	893s	227	-	-
Summe	935s	238	842s	214

Tabelle 3: Anzahl der Instruktionen für die einzelnen Funktionsblöcke

Für die Erstellung der Ergebnisse wurden Erkennungsexperimente mit den Aurora5 Testdaten „G712 CarNoise05dB Handsfree“ durchgeführt. Die Gesamtlänge der 8700 verarbeiteten Testfiles beträgt 15399s.
Für das Verfahren „HGH adapt“ werden die Laufzeit und die benötigten MIPS für die Verarbeitungsblöcke „Erkennung“ und „Adaption“ zusammengefasst, da diese Punkte, bedingt durch den Aufbau des Adaptionsverfahrens, miteinander verknüpft sind. So werden beispielsweise zur Schätzung der Nachhallzeit die Funktionen zur Erkennung ebenfalls verwendet, jedoch mit anderen Parametern und tragen damit zu zusätzlichen Funktionsaufrufen der für die Erkennung verwendeten Funktionen bei. Aus diesem Grund lässt sich die benötigte Zeit und die benötigte Anzahl an MIPSSP für die reine Erkennung nicht genau festlegen. Prinzipiell sollte sich jedoch ein ähnlicher Wert wie bei der Erkennung

mit „HGH robust“ ergeben, da sich die Erkennungsfunktionen gleichen. Weiterhin ist bei den Ergebnissen anzumerken, dass es sich um experimentell ermittelte Daten handelt, die eine gewisse Abhängigkeit im Bezug zu dem verwendeten Rechner-System aufweisen, obwohl die MIPSSP-Werte als systemunabhängig angegeben werden. Aus diesem Grund kann es vorkommen, dass sich die Daten nicht vollständig reproduzieren lassen.

Auswertung

Werden die Werte in Tabelle 3 miteinander verglichen, kann festgestellt werden, dass die beiden robusten Verfahren eine ähnliche Gesamtlaufzeit aufweisen. Das robuste Verfahren „HGH adapt“ arbeitet geringfügig langsamer. Dies ist auf den etwas komplexeren Aufbau des Verfahrens „HGH adapt“ im Bezug auf die Anzahl der verwendeten Funktion und deren Rechenzeit zurückzuführen.
Aus der Betrachtung der Laufzeiten der einzelnen Funktionsblöcke ist deutlich der zusätzliche Rechenaufwand des jeweiligen robusten Systems im Vergleich zu einem nicht robusten System zu erkennen. Die Anzahl der benötigten Instruktionen für ein nicht robustes Erkennungssystem kann ungefähr aus den ermittelten Werten der robusten Systeme abgeschätzt werden. Diese sind in Tabelle 4 zusammen mit den W-MIPSSP Werten für die robusten Systeme dargestellt:

	HGH adapt	HGH robust	nicht robustes System
	W-MIPSSP für eine Sekunde Sprachdaten	W-MIPSSP für eine Sekunde Sprachdaten	W-MIPSSP für eine Sekunde Sprachdaten
Analyse	10	51	10
Stör-schätzung	1	5	-
Erkennung	-	158	158
Erkennung + Adaption	227	-	-
Summe	238	214	168

Tabelle 4: Vergleich der Anzahl der Instruktionen

Wird der Berechnungs- und Zeitaufwand der einzelnen Funktionen miteinander verglichen, wird deutlich, dass die Analyse der robusten Merkmalsextraktion etwa fünfmal mehr W-MIPSSP im Vergleich zu der nicht-robusten Analyse von „HGH Adapt“ benötigt. Im Gegenzug dazu benötigt die Erkennung mit Adaption von „HGH Adapt“ etwa 1,4 mal mehr W-MIPSSP als die Erkennung ohne Adaption von „HGH robust“.

Anhand der benötigten W-MIPSSP können nun Betrachtungen im Bezug zu einem System mit geringeren Ressourcen durchgeführt werden. Auf einer Website von Texas Instruments[5], kurz TI, sind die Benchmarkergebnisse für verschiedene Prozessoren dieser Marke aufgelistet. Als Betriebssystem für die Prozessoren sieht TI das von der „Open Handset Alliance“ entwickelte System „Android“ vor. Die TI Prozessoren werden zum Beispiel in Tablett-PCs mit Android als Betriebssystem verwendet.

Im Folgenden soll nun an einem Beispiel erörtert werden, ob eine Anwendung der robusten Erkennungssystem auf einem mobilen Endgerät möglich ist. Da von aktuellen Endgeräten so gut wie keine Information über Benchmarkergebnisse existieren, wird als Beispiel nur der Prozessor „OMAP3530“ mit 720Mhz als Vergleichssystem betrachtet. In der Referenzliste in Fußnote 5 ist für den Prozessor ein Wert von 333.3 W-MIPS angegeben. Alle anderen Einflüsse wie z.B. Speicher, Betriebssystem etc. sind zu vernachlässigen.

Werden die Ergebnisse in Tabelle 3 auf die Maximal zur Verfügung stehende Anzahl von 333.3 W-MIPS des Prozessors bezogen, wird deutlich, dass beide robusten Systeme auf einem Gerät betrieben werden können, in dem ein Prozessor diesen Typs arbeitet. Da noch circa 100 W-MIPS ungenutzt sind, kann davon ausgegangen werden, dass beide Erkennungssysteme ohne Probleme funktionieren sollten, auch ohne das Komponenten wie z.B. das Betriebssystem, das ebenfalls Rechenleistung benötigt, berücksichtigt wurden.

Da es sich bei dem Prozessor um ein aktuelles Modell handelt, soll ein weiterer Vergleich mit einem etwas älteren System angestellt werden. Dazu wird ein älteres Smartphone, das HTC Sprint HERO betrachtet. Die Markteinführung des Telefons war im Sommer 2009, in einem XDA-Entwickler Forum[6] findet sich dazu der Hinweis, dass das Gerät ungefähr 167 W-MIPS leistet. Dieses Gerät wäre für die Verwendung der robusten Erkennungssysteme nicht geeignet. Selbst eine Reduktion des Erkennungssystems auf die zuvor beschriebene nicht robuste Version würde ca. 168 W-MIPS (siehe Tabelle 4) benötigen. Somit wäre das Telefon bei der Anwendung der robusten Systeme überlastet. Selbst bei der Anwendung der nicht robusten Version des Erkennungssystems wäre es vollständig ausgelastet, was höchstwahrscheinlich auch zu Problemen während des Betriebs führen würde.

Zusammenfassung und Ausblick

In dem vorliegenden Text wurde der rechnerische Aufwand in W-MIPSSP für die beiden robusten Erkennungssysteme „HGH adapt“ und „HGH robust“ ermittelt und auf Geräte mit geringen Systemressourcen bezogen. Dabei konnte gezeigt werden, dass die Anwendung der beiden Verfahren auf aktuellen mobilen Geräten prinzipiell möglich ist.

[5] http://processors.wiki.ti.com/index.php/Android_Comparative_Benchmarks
[6] http://forum.xda-developers.com/showthread.php?p=6590232

Bei älteren Geräten oder weniger leistungsstarken Geräten würde die Anwendung der Erkennungssysteme jedoch zu Komplikationen führen.

An dieser Stelle muss jedoch angemerkt werden, dass sich die in diesem Bericht durchgeführten Tests und Betrachtung auf eine Implementierung der Verfahren bezieht, die für die Anwendung auf Rechnersystemen, die ähnlich dem oben beschriebenen Testsystem sind, entwickelt wurde. Der Einsatz auf mobilen Endgeräten wurde bei dieser Implementierung der beide Verfahren „HGH adapt“ und „HGH robust“ nicht vorgesehen. Eine Optimierung der Software mit dem Fokus auf Systeme mit geringen Ressourcen würde hierbei Abhilfe schaffen und den Einsatz in mobilen Endgeräten ermöglichen. Dennoch ist auch bei einer Optimierung davon auszugehen, dass eine gewisse Anforderung an die Rechenleistung des mobilen Systems bestehen bleibt.

Ein weiterer Punkt, der noch zu den durchgeführten Experimenten angemerkt werden sollte, ist die Abhängigkeit der ermittelten Laufzeiten von den verwendeten Referenz-Modellen. Durch die Anzahl der Gauss- Mischverteilungen pro State und die Anzahl der States in den Referenz-Modellen wird die Dauer der Berechnung für den Verarbeitungsblock „Erkennung“ festgelegt. Das bedeutet, dass mit steigender Anzahl der Gauss-Mischverteilungen und der States auch die benötigte Zeit zur Berechnung ansteigt. Bei der Verwendung von Referenz-Modellen zur Durchführung der Experimente, die eine andere Anzahl an States und Gauss- Mischverteilungen aufweisen als die bereits beschriebenen Modelle würden andere Laufzeiten und dadurch auch andere W-MIPSSP Werte resultieren. Dieser Punkt wird in der vorgestellten Arbeit nicht berücksichtigt, daher sind alle Laufzeiten und daraus berechneten W-MIPSSP Werte nur im Bezug zu den verwendeten Referenz-Modellen zu betrachten. Um absolut unabhängige Werte zu erhalten, müsste diese Abhängigkeit in der Berechnung der W-MIPSSP mit berücksichtigt werden.