

STATE-TYING FÜR KONTEXTABHÄNGIGE PHONEMMODELLE

K. Beulen, E. Bransch, M. Kramer, H. Ney

Lehrstuhl für Informatik VI, RWTH Aachen, D-52056 Aachen

KURZFASSUNG

Diese Arbeit beschreibt zwei Verfahren zur Parameterreduktion bei *Hidden-Markov-Modellen* (HMM) durch *State-Tying*: ein datengetriebenes Verfahren, das Triphonzustände mit einem Bottom-Up-Verfahren clustert [3, 8], und ein Top-Down-Verfahren, das Entscheidungsbäume für Triphonzustände erzeugt [2, 9]. Durch das Tying wird bei gleicher Parameterzahl eine bessere akustische Modellierung im Vergleich zu reinen Triphonmodellen erreicht. Untersucht werden die Verbesserungen der Fehlerrate durch das State-Tying, die Auswirkungen der verschiedenen Abstandsmaße und die Unterschiede zwischen dem datengetriebenen Verfahren und dem Verfahren mit Entscheidungsbäumen. Die Tests wurden auf dem Testkorpus für das 20 000-Wort-Vokabular des *North American Business Task* von 1994 (NAB'94) durchgeführt. Die erreichten Verbesserungen in der Fehlerrate durch State-Tying liegen zwischen 8% und 14% bei gleicher Parameterzahl im Vergleich zur Verwendung von reinen Triphonmodellen.

1. EINFÜHRUNG

Systeme zur automatischen Erkennung kontinuierlich gesprochener Sprache für großen Wortschatz modellieren die Lautfolgen der Worte des Vokabulars meist mit Phonemmodellen. Jedem Phonemmodell liegt dabei ein entsprechendes *Hidden-Markov-Modell* (HMM) zugrunde. Da der phonetische Kontext die akustische Realisierung eines Lautes stark beeinflusst, versucht man, diese Abhängigkeit durch kontextabhängige Phonemmodelle zu erfassen [6, 7]. Das am meisten verwendete kontextabhängige Phonemmodell ist das Phonemmodell im Triphon-Kontext, oft einfach „Triphon“ genannt. Die Modellierung eines derartigen Triphon-Phonems erfolgt in Abhängigkeit vom linken und rechten Nachbarphonem. Dadurch erhält man die gewünschte Modellierung abhängig vom akustischen Kontext. Ein Problem hierbei ist, daß die mögliche Anzahl der Modelle kubisch mit der Anzahl der Phoneme steigt. Wegen des meist begrenzten Trainingsmaterials ist es aber schwierig, diese Modelle ausreichend genau zu schätzen. Zur Lösung dieses Problems gibt es mehrere Ansätze wie z.B. Glättung der Parameter oder Verknüpfung von Parametern unterschiedlicher Modelle.

Das Verknüpfen der Parameter von akustischen Modellen, sogenanntes *Parameter-Tying*, ist auf verschiedenen Ebenen möglich, z.B. auf der Ebene der Triphone, der Zustände oder der Varianzen. Verknüpfte Modelle verwenden dieselben Modellparameter, so daß sich die Gesamtzahl der Modellparameter des Erkennungssystems reduziert und so die Parameter robuster geschätzt werden können. In dieser Arbeit wird das *State-Tying* betrachtet, bei dem die akustischen Modelle auf Zustandsebene verknüpft werden. Untersucht werden zwei Verfahren, um die Verknüpfungen der Zustände zu ermitteln: ein datengetriebenes Bottom-Up-Verfahren, das Clu-

ster aus Triphonzuständen bildet [3, 8], und ein Top-Down-Verfahren, das Entscheidungsbäume generiert [2, 9].

2. STATE-TYING

Das Ziel des State-Tying ist es, die Zahl der Parameter des Spracherkennungssystems deutlich zu verringern, ohne die Modellierung signifikant zu verschlechtern. Dazu werden die Zustände der im Training verwendeten Triphonmodelle, die ähnlich bezüglich eines Abstandsmaßes sind, verknüpft. Man ermittelt die in den Trainingsdaten vorkommenden Triphone und wählt aus dieser Menge eine geeignete Untermenge aus, deren Modelle verknüpft werden sollen. Da diese initiale Triphonmenge groß sein muß, um eine exakte Modellierung des akustischen Kontextes zu erreichen, verwendet man für diese Menge von Triphonzuständen einfache Modelle (eine Gauß-Verteilung mit diagonaler Kovarianzmatrix), die mit den vorhandenen Daten gut geschätzt werden können. Anhand einer guten Segmentierung der Daten wird zu jedem Triphonzustand X der Mittelwert $\hat{\mu}_{d,X}$ und die Varianz $\hat{\sigma}_{d,X}^2$ über die üblichen Formeln

$$\hat{\mu}_{d,X} = \frac{1}{n_X} \sum_{z \in X} z_d \quad (1)$$

$$\hat{\sigma}_{d,X}^2 = \frac{1}{n_X} \sum_{z \in X} z_d^2 - \hat{\mu}_{d,X}^2 \quad (2)$$

geschätzt, wobei z einen Beobachtungsvektor, der auf den Zustand X entfällt, n_X die Anzahl und $d = 1, \dots, D$ die Komponenten dieser Beobachtungsvektoren bezeichnet. Die Triphonzustände werden dann nach ihrem zentralen Phonem und ihrer Position im Phonemmodell geordnet. Bei 43 Phonemen und drei Positionen erhält man 129 Mengen von zu verknüpfenden Triphonzuständen. Innerhalb dieser Zustandsmengen werden anhand der Abstände der Modelle die günstigsten Verknüpfungen durchgeführt. Dabei muß sichergestellt werden, daß alle resultierenden Modelle eine Mindestzahl von Beobachtungen haben. Die verknüpften Modelle werden in einem Training reestimiert, wobei aufgrund der reduzierten Parameterzahl eine höhere akustische Auflösung erreicht werden kann.

3. DATENGETRIEBENES TYING

Das datengetriebene Tying verwendet zur Bestimmung der zu verknüpfenden Triphonzustände X ausschließlich die aufgrund der Trainingsdaten geschätzten Mittelwerte $\hat{\mu}_X$ und Varianzen $\hat{\sigma}_X^2$. Diese werden mit Hilfe eines Bottom-Up-Algorithmus geclustert, der die günstigsten Verknüpfungen bezüglich eines Abstandsmaßes bestimmt. In einem zweiten Schritt wird durch zusätzliche Verknüpfungen sichergestellt, daß alle Cluster eine Mindestzahl an Beobachtungen haben, so daß die entsprechenden Modelle im Training robust geschätzt werden können.

Tabelle 1: Bottom-Up-Clusteralgorithmus.

Erzeuge für jeden Triphonzustand ein Cluster	
	Bestimme das Clusterpaar (A, B) mit dem kleinsten Abstand
	Verschmelze A und B
Weiter, bis der Abstand für alle Paare (A, B) oberhalb der Schwelle liegt	
Für alle Cluster A mit Anzahl der Beobachtungen unterhalb der Schwelle	
	Suche nächsten Nachbarn B zu A
	Verschmelze A und B

Das Problem beim datengetriebenen Tying ist, daß für die im Training nicht gesehenen Triphone keine Zuordnung ihrer Zustände zu den verknüpften Verteilungen möglich ist. Daher werden im Training zusätzlich sogenannte *Backing-Off-Modelle* [6] trainiert, meist Monophone. Die nicht gesehenen Triphone werden auf diese Modelle abgebildet. Um eine ausreichende Trainingsdatenmenge für die Backing-Off-Modelle zu erhalten, teilt man die im Training gesehenen Triphone in zwei Mengen auf. Nur eine der Mengen wird für das Aussprachelexikon im Training verwendet, so daß die Daten der anderen Menge zum Trainieren der Backing-Off-Modelle benutzt werden können. Wir haben diese Einteilung über die Häufigkeit des Auftretens im Trainingstext des WSJ0+1-Korpus vorgenommen. Alle Triphone, die mehr als 50mal im geschlechtsunabhängigen Trainingskorpus bzw. 25mal im geschlechtsabhängigen Trainingskorpus gesehen wurden, wurden zum Aufbau des Trainingslexikons verwendet.

3.1. Bottom-Up-Clusteralgorithmus

Der Algorithmus [3, 8] startet mit einem Triphonzustand pro Cluster. Diese Cluster werden dann iterativ zu größeren Clustern verschmolzen (siehe Tabelle 1). Der Algorithmus arbeitet in zwei Schritten. Im ersten Schritt werden die Cluster mit einem Bottom-Up-Verfahren verschmolzen, bis der Abstand aller Clusterpaare eine vordefinierte Schwelle überschreitet. Im zweiten Schritt werden die Cluster, die zu wenig Beobachtungen enthalten, mit anderen verschmolzen, um eine Mindestzahl von Beobachtungen pro Cluster zu garantieren. Die Schwelle für die Mindestzahl an Beobachtungen betrug beim datengetriebenen Tying 100. Die Triphonzustände eines Clusters werden verknüpft und in einem anschließenden Training mit einer höheren akustischen Auflösung neu geschätzt.

Die Komplexität des datengetriebenen Tying ist kubisch in der Anzahl der Triphonzustände. Der Algorithmus gehört zu den *Greedy*-Verfahren, ist also nur lokal, nicht global optimal.

3.2. Abstandsmaße

Das Abstandsmaß ist das Kriterium, anhand dessen entschieden wird, welches Paar von Clustern verschmolzen wird. In jedem Clusterschritt wird das Paar (\hat{A}, \hat{B}) verschmolzen, für das gilt:

$$(\hat{A}, \hat{B}) = \underset{(A, B)}{\operatorname{argmin}} D(A, B) \quad (3)$$

wobei $D(A, B)$ der Abstand der beiden Cluster (A, B) ist. Für das datengetriebene Tying haben wir folgende Abstandsmaße getestet:

- die *Furthest-Neighbour-Distanz*

$$D(A, B) = \max_{X \in A, Y \in B} d(X, Y) \quad (4)$$

wobei die folgenden Distanzen $d(X, Y)$ zwischen den Elementen X und Y der Cluster verwendet wurden:

- die l_1 -Norm der Differenz zwischen den Mittelwertsvektoren $\hat{\mu}_X$ und $\hat{\mu}_Y$ der Triphonzustände X und Y [3]

$$d(X, Y) = \sum_{d=1}^D |\hat{\mu}_{d,X} - \hat{\mu}_{d,Y}| \quad (5)$$

- die approximative Divergenz [8] zwischen zwei Triphonzuständen X und Y mit Mittelwertsvektoren $\hat{\mu}_X$ und $\hat{\mu}_Y$ und Varianzvektoren $\hat{\sigma}_X$ und $\hat{\sigma}_Y$:

$$d(X, Y) = \left[\sum_{d=1}^D \frac{(\hat{\mu}_{d,X} - \hat{\mu}_{d,Y})^2}{\sqrt{\hat{\sigma}_{d,X}^2 \hat{\sigma}_{d,Y}^2}} \right]^{\frac{1}{2}} \quad (6)$$

mit D als der Dimension des Vektorraums.

- die Differenz zwischen der negativen log-Likelihood $LL(AB)$ des verschmolzenen Clusters AB und der Summe der negativen log-Likelihoods der beiden Cluster A und B , $LL(A)$ und $LL(B)$. Dazu müssen die Parameter des Gauß-Modells für das verschmolzene Cluster AB neu geschätzt werden:

$$\hat{\mu}_{d,AB} = \frac{1}{n_A + n_B} \left(\sum_{z \in A} z_d + \sum_{z \in B} z_d \right) \quad (7)$$

$$\hat{\sigma}_{d,AB}^2 = \frac{1}{n_A + n_B} \left(\sum_{z \in A} z_d^2 + \sum_{z \in B} z_d^2 \right) - \hat{\mu}_{d,AB}^2 \quad (8)$$

n_A und n_B bezeichnen die Anzahl der Beobachtungsvektoren im jeweiligen Cluster. Als log-Likelihood-Differenz erhält man:

$$D(A, B) = LL(AB) - (LL(A) + LL(B)) \quad (9)$$

$$= -\frac{1}{2} \left(n_A \sum_{d=1}^D \log \left[\frac{\hat{\sigma}_{d,AB}}{\hat{\sigma}_{d,A}} \right]^2 + n_B \sum_{d=1}^D \log \left[\frac{\hat{\sigma}_{d,AB}}{\hat{\sigma}_{d,B}} \right]^2 \right)$$

4. PHONETISCHE ENTSCHEIDUNGSBÄUME

Ein phonetischer Entscheidungsbaum ist ein binärer Baum, an dessen inneren Knoten im Fall des State-Tying Fragen an den phonetischen Kontext eines Triphonzustands stehen. Ein Beispiel für eine solche Frage ist: „Ist der linke Kontext ein Vokal?“. Der linke Teilbaum des inneren Knotens enthält die Triphonzustände, für die die Frage mit *Ja* beantwortet wird, der rechte Teilbaum die, für die die Frage mit *Nein* beantwortet wird. Die Blätter des Baums enthalten die verknüpften Triphonzustände und repräsentieren damit die Emissionsverteilungen des Erkennungssystems. Die Knoten des Entscheidungsbaums entsprechen damit den Clustern beim datengetriebenen Tying. Will man zu einem Triphonzustand die zugehörige Emissionsverteilung ermitteln, stellt man, am Wurzelknoten des entsprechenden Entscheidungsbaums beginnend, die jeweilige Frage an den Triphonzustand und verzweigt dann entweder nach links (*Ja*) oder rechts (*Nein*), bis man an einem Blatt angekommen ist.

Tabelle 2: Top-Down-Clusteralgorithmus.

Bilde einen Knoten mit allen Triphonzuständen	
Für alle Knoten A	
Bestimme die Frage Q, für die nach dem Spalten	
<ul style="list-style-type: none"> • der log-Likelihood-Gewinn maximal wird • die Anzahl der Beobachtungen für beide Knoten oberhalb einer Schwelle liegt 	
Falls der log-Likelihood-Gewinn über einer Schwelle liegt, spalte den Knoten A mit der Frage Q auf	
Weiter, bis kein Knoten mehr gespalten wird	

Ein Beispiel für einen solchen Baum ist in Abbildung 1 zu sehen. Die inneren Knoten sind mit den Fragen beschriftet, die Blätter mit der Anzahl der Triphone und der Anzahl der Beobachtungen. Die Rechtslastigkeit des Baums hängt damit zusammen, daß die meisten Fragen nur wenige Phoneme enthalten und dadurch der größere Teil der Triphone durch den rechten Teilbaum modelliert werden. Auffällig ist auch, daß das am weitesten rechts stehende Blatt relativ viele Triphone enthält. Da für dieses Blatt alle Fragen mit *Nein* beantwortet wurden, kann man dieses Modell als eine Art Backing-Off-Modell interpretieren, da darauf alle Triphone abgebildet werden, für die keine der Eigenschaften, die erfragt wurden, paßt.

Zu jedem Phonem und zu jedem Zustand wird ein Entscheidungsbaum konstruiert. Der Konstruktionsprozeß versucht, eine möglichst günstige Unterteilung der Triphonzustände zu ermitteln, gleichzeitig aber eine Mindestzahl von Beobachtungen für jedes Modell zu garantieren. Dazu wird ein Top-Down-Algorithmus verwendet. Dieser Algorithmus verwendet im Gegensatz zu dem rein datengetriebenen Algorithmus zusätzliches Wissen in Form der phonetischen Fragen, die eine Unterteilung der Phoneme in Kategorien vornehmen. Bei unseren Tests haben wir 44 Fragen nach allgemeinen phonetischen Kategorien, die aus [5] entnommen und um einige zusätzliche Fragen erweitert wurden, 43 Fragen nach speziellen Phonemen und eine Frage nach der Wortgrenze benutzt, insgesamt 88 Fragen. Die Fragen wurden ohne Einschränkung sowohl an den linken als auch an den rechten Kontext gestellt.

4.1. Top-Down-Clusteralgorithmus

Der Algorithmus [2, 9] startet mit einem Knoten, der alle zu verknüpfenden Triphonzustände enthält. Mit den phonetischen Fragen werden die Knoten iterativ aufgespalten, wobei die Differenz zwischen der log-Likelihood des Knotens vor dem Spalten und log-Likelihoods der beiden Knoten nach dem Spalten über alle Fragen und beide Kontexte maximiert wird (siehe Formel 9). Der Knoten wird nicht gespalten, wenn der Gewinn in der log-Likelihood für alle möglichen Aufspaltungen des Knotens unter eine Schwelle sinkt oder eine Mindestzahl von Beobachtungen für die entstehenden Knoten unterschritten wird (siehe Tabelle 2). Als Schwelle für die Mindestzahl an Beobachtungen wurde bei den Tests 500 verwendet. Kann kein Knoten mehr gespalten werden, stoppt der Algorithmus. Das Ergebnis ist ein Entscheidungsbaum, der jeden Triphonzustand einem Blatt zuordnet [2]. Wie beim datengetriebenen Tying werden die Triphonzustände eines Blattes verknüpft und in einem anschließenden Training reestimiert. Ein Vorteil gegenüber dem datengetriebenen Verfahren ist, daß auch ungesehene Triphone zugeordnet werden können, es werden keine Backing-Off-Modelle benötigt.

Für die Konstruktion der Entscheidungsbäume wurde ausschließlich die log-Likelihood-Differenz als Abstandsmaß ver-

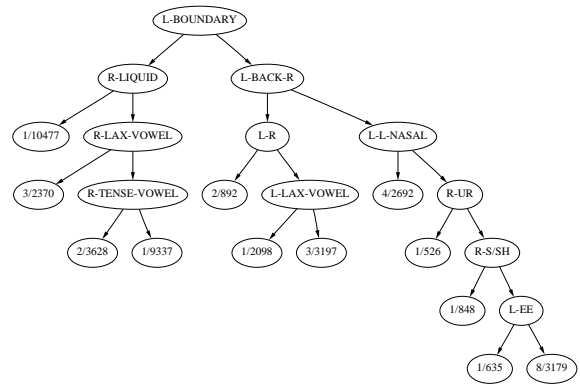


Abbildung 1: Entscheidungsbaum für Phonem *th*, Zustand 1.

wendet (siehe Gleichung 9). Der Top-Down-Verfahren ist wie das Bottom-Up-Verfahren nur lokal optimal.

5. ERGEBNISSE

Die Ergebnisse wurden mit dem in [1, 3] beschriebenen Spracherkennungssystem für großen Wortschatz erzielt. Die wichtigsten Eigenschaften dieses Systems sind:

- Alle 10ms 30 Filterbank-Ausgänge + Energie + 1. + 2. Ableitungen = 63 Komponenten,
- Merkmalstransformation und -reduktion mit LDA auf 35 Komponenten [4],
- kontinuierliche HMM mit Mischverteilungen,
- Laplacesche Dichten,
- ein zustandsunabhängiger Varianzvektor für alle Verteilungen,
- Viterbi-Approximation beim Training,
- One-Pass-Algorithmus mit Bigramm-Sprachmodell bei der Erkennung.

Das Training wurde auf dem *Wall Street Journal*-Trainingskorporus (WSJ0+1) durchgeführt. Die Ergebnisse für das 20 000-Wort-Vokabular (sprecherunabhängig) wurden auf dem *North American Business* (NAB'94) Korpus ermittelt. Dabei gab es folgende Randbedingungen:

- datengetriebenes Tying, approximative Divergenz und log-Likelihood-Differenz:
 - Reduzierte Triphonlisten mit 4100 Triphonen für Frauen und 4111 Triphonen für Männer,
 - 43 Monophonmodelle für ungesehene Triphone,
 - geschlechtsabhängig verknüpfte Modelle,
 - eine neu berechnete LDA-Matrix für jede durch das Tying ermittelten Modelldefinitionen.
- Tying mit Entscheidungsbäumen bzw. datengetriebenes Tying, l_1 -Norm:
 - eine gemeinsame reduzierte Triphonliste mit 4087 bzw. 4644 Triphonen für Frauen und Männer,
 - geschlechtsunabhängig verknüpfte Modelle,
 - eine universelle LDA-Matrix.

Tabelle 3: Wortfehlerrate (WER) für den H1-Development-Korpus, NAB'94 (10 Sprecher, 310 Sätze, 7387 gesprochene Wörter)

Methode	Kriterium	Anzahl der Zustände		Dichten	Suchaufwand in Hypothesen	DEL-INS	WER[%]
		vor Tying	nach Tying				
Kein Tying	-	4495	-	2*220k	26787	187 – 222	18.3
Datengetriebenes Tying	l_1 -Norm	14062	4623	2*294k	26489	179 – 195	16.5
	Appr. Divergenz	12448	4551	2*297k	27176	182 – 182	16.3
	log-Likelihood	12448	4416	2*295k	24580	159 – 202	16.6
Tying mit Entscheidungsbäumen	log-Likelihood	12261	4396	2*280k	19042	190 – 200	16.5

Tabelle 4: Wortfehlerrate (WER) für den H1-Evaluation-Korpus, NAB'94 (10 Sprecher, 316 Sätze, 8186 gesprochene Wörter)

Methode	Kriterium	Anzahl der Zustände		Dichten	Suchaufwand in Hypothesen	DEL-INS	WER[%]
		vor Tying	nach Tying				
Kein Tying	-	4495	-	2*220k	28012	239 – 244	19.2
Datengetriebenes Tying	l_1 -Norm	14062	4623	2*294k	27712	223 – 221	17.1
	Appr. Divergenz	12448	4551	2*297k	27734	234 – 223	17.7
	log-Likelihood	12448	4392	2*295k	25660	191 – 234	16.7
Tying mit Entscheidungsbäumen	log-Likelihood	12261	4396	2*280k	25947	222 – 221	16.9

Tabelle 3 enthält die Ergebnisse, die auf dem Development-Korpus (7387 gesprochene Wörter) erzielt wurden, Tabelle 4 die Ergebnisse für den Evaluation-Korpus (8186 gesprochene Wörter).

Die Ergebnisse zeigen, daß durch State-Tying deutliche Verbesserungen der akustischen Modelle zu erreichen sind. Sowohl das datengetriebene Tying als auch das Tying mit Entscheidungsbäumen verbessern die Wortfehlerrate (WER) um 8 bis 14%. Berücksichtigt man die Geschlechtsunabhängigkeit beim Verknüpfen der Modelle und die unspezifische LDA-Matrix beim Tying mit Entscheidungsbäumen kann man hier noch Verbesserungen für das Tying mit Entscheidungsbäumen erwarten. Außerdem ist zu sehen, daß das log-Likelihood-Abstandsmaß in der Summe für das datengetriebene Tying etwas besser als die approximative Divergenz und die l_1 -Norm ist.

6. ZUSAMMENFASSUNG

In dieser Arbeit haben wir zwei Verfahren zum State-Tying vorgestellt, ein datengetriebenes Verfahren und ein Verfahren, das mit Hilfe von phonetischen Fragen Entscheidungsbäume konstruiert. Für das datengetriebene Verfahren haben wir drei unterschiedliche Abstandsmaße untersucht. Die Untersuchungen haben folgende Ergebnisse gebracht:

- State-Tying verbessert die Fehlerrate auf den Testkorpora um 8 bis 14% im Vergleich zur Verwendung von reinen Triphonmodellen.
- Das log-Likelihood-Abstandsmaß erzielt etwas bessere Ergebnisse als die l_1 -Norm und die approximative Divergenz.

Literatur

1. X. Aubert, C. Dugast, H. Ney, V. Steinbiss: Large Vocabulary, Continuous Speech Recognition of Wall Street Journal Corpus. Proc. Int. Conf. on Acoustics, Speech and Signal Processing, Adelaide, Australia, Vol. II,

pp. 129-132, April 1994.

2. L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, "Classification and Regression Trees", *The Wadsworth Statistics/Probability Series*, Wadsworth Statistics/Probability Series, Belmont, CA, 1984.
3. C. Dugast, R. Kneser, X. Aubert, S. Ortmanms, K. Beulen, H. Ney, "Continuous Speech Recognition Tests and Results for the NAB'94 Corpus", *Proc. ARPA Spoken Language Technology Workshop*, Austin, TX, pp. 156-161, Januar 1995.
4. R. Haeb-Umbach, H. Ney, "Linear Discriminant Analysis for improved Large Vocabulary Continuous Speech Recognition" *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, San Francisco, CA, März 1992.
5. H.-W. Hon, "Vocabulary-Independent Speech Recognition: The VOCIND System", Ph.D. Thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1992.
6. H. Ney, "Acoustic Modelling of Phoneme Units for Continuous Speech Recognition", *Proc. Fifth Europ. Signal Processing Conf.*, Barcelona, pp. 65-72, September 1990.
7. R. Schwartz, Y. Chow, O. Kimball, S. Roucos, U. Krasner, J. Makhoul, "Context-Dependent Modelling for Acoustic-Phonetic Recognition of Continuous Speech", *Proc. 1985 Int. Conf. on Acoustics, Speech and Signal Processing*, Tampa, FL, pp. 1205-1208, März/April 1985.
8. S.J. Young, P.C. Woodland, "The Use of State Tying in Continuous Speech Recognition", *Proc. EURO-SPEECH*, Berlin, pp. 2203-2206, September 1993.
9. S.J. Young, J.J. Odell, P.C. Woodland, "Tree-Based State Tying for High Accuracy Acoustic Modelling", *Proc. ARPA Human Language Technology Workshop*, Plainsboro, NJ, pp. 405-410, Morgan Kaufmann, März 1994.