



## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

Triphone Tying

Decision Trees

Tree based  
State Tying

Outline

Building

Tree Clustering

Discussion

References

# Tree-Based State Tying for High Accuracy Acoustic Modelling

S.J. Young, J.J. Odell, P.C. Woodland 1994

Martin Raab

University of Karlsruhe

May 24, 2005



# Overview

## Tree-Based State Tying

Martin Raab

Introduction

Parameters  
Motivation

Basics

Triphone Tying  
Decision Trees

Tree based  
State Tying

Outline  
Building  
Tree Clustering

Discussion


References

**1** Introduction

**2** Basics

**3** Tree based State Tying

**4** Discussion

 Overview

- 1 Introduction
- 2 Basics
- 3 Tree based State Tying
- 4 Discussion

1. basically a technique to reduce parameters



# Co-Articulation

## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

Triphone Tying

Decision Trees

Tree based  
State Tying

Outline

Building

Tree Clustering

Discussion

References

## Co-Articulation Effect

- Pronunciation of a phone depends on its context
- Context independent HMMs do not model the context of a phone
- A HMM for every context has to be trained

## Co-Articulation Effect

- Pronunciation of a phone depends on its context
- Context independent HMMs do not model the context of a phone
- A HMM for every context has to be trained

1. It is a common problem in ASR that there are too many parameters and there is either too few data to estimate them correctly or the amount of storage is exceeded
2. due to the HMM assumption that every transition is only depending on the last state



# Parameter Example

## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

Triphone Tying

Decision Trees

Tree based  
State Tying

Outline

Building

Tree Clustering

Discussion

References

## How many Parameters?

- Phone alphabet  $P$  with 50 phones
- This makes  $P^3 = 125,000$  triphones
- 8 variate gaussian has 44 parameters
- 16.5 million parameters for gaussians of the states, each 32 Bit makes 528 Megabyte
- What about quintphones? (312 million of them ...)

## How many Parameters?

- Phone alphabet  $P$  with 50 phones
- This makes  $P^3 = 125,000$  triphones
- 8 variate gaussian has 44 parameters
- 16.5 million parameters for gaussians of the states, each 32 Bit makes 528 Megabyte
- What about quintphones? (312 million of them ...)

1. TODO : correct? slide from SMMK  $d + (d + 1) * d/2$  for d-variate gaussian
2. assume 3 states per triphone
3. The number of quintphones is 2500 times bigger than the number of triphones, 1320 Gigabyte needed



# Motivation

## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

Triphone Tying

Decision Trees

Tree based  
State Tying

Outline

Building

Tree Clustering

Discussion

References

## Recently used

“Most past work has focused on the use of triphone phonetic decision trees” (Chen, 2003)

---

“Decision tree is a popular clustering method which has been used in many research areas” (Hirsimaki, 2003)

---

“A new decision tree-based clustering technique called Phonetic, Dimensional and State Positional Decision Tree (PDS-DT) is proposed” (Zen, Tokuda, & Kitamura, 2003)





# Triphones

## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

**Triphone Tying**

Decision Trees

Tree based State Tying

Outline

Building

Tree Clustering

Discussion

References

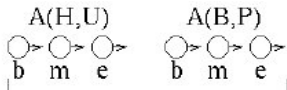


Figure: Two triphone models

- A phone is usually modelled by 3 HMM states
- Not the reason why a model is called triphone model
- Triphone model depends on left and right context

## Tree-Based State Tying

└ Basics

└└ Triphone Tying

└└└ Triphones



Figure: Two triphone models

- A phone is usually modeled by 3 HMM states
- Not the reason why a model is called triphone model
- Triphone model depends on left and right context

1. for begin, middle, end



# Tying - Two Approaches

## Tree-Based State Tying

Martin Raab

### Introduction

Parameters

Motivation

### Basics

#### Triphone Tying

Decision Trees

#### Tree based State Tying

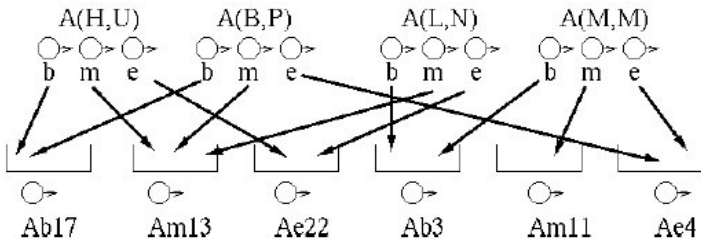
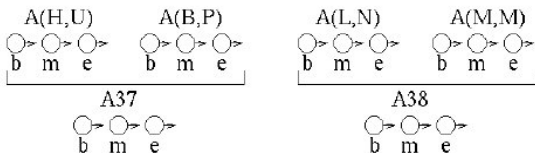
Outline

Building

Tree Clustering

### Discussion

### References

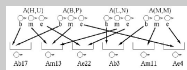


## Tree-Based State Tying

## Basics

## Triphone Tying

## Tying - Two Approaches



1. model based tying, only complete models are tied
2. Traditional approach is the model tying, these triphones are called generalised triphones
3. state based tying, each state can be tied separately
4. no matter what is tied, they are tied to either continuous or discrete probability functions



# Decision Tree

## Tree-Based State Tying

Martin Raab

### Introduction

Parameters

Motivation

### Basics

Triphone Tying

Decision Trees

### Tree based State Tying

Outline

Building

Tree Clustering

### Discussion

### References

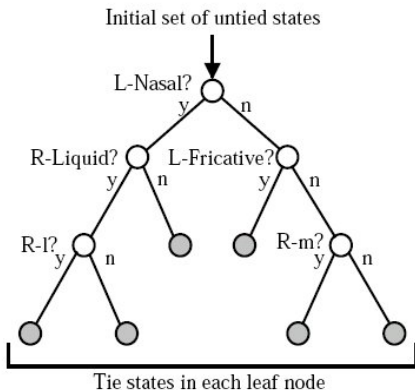


Figure: Example of a phonetic decision tree

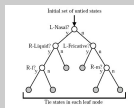


Figure: Example of a phonetic decision tree

1. In each node a question is asked
2. For each possible answer to this question there is an outgoing arc
3. In SR usually binary decision trees are covered, this means, that the answer to the question is either yes or no
4. It should be obvious, how such a tree works once it is constructed, however the construction usually depends on the knowledge of all utterances (or probabilities) to set up the tree that each question provides a maximum incline of separation of similar phones
5. Liquid phones : Fließlaute, r und Laterale(L-Laute), Nasal : m,n, Frikativ : Reibelaute, s,h,x



# Outline

## Tree-Based State Tying

Martin Raab

Introduction

Parameters  
Motivation

Basics

Triphone Tying  
Decision Trees

Tree based  
State Tying

**Outline**  
Building  
Tree Clustering

Discussion

References

## Aim of this paper

- State tying has already been used with a data driven agglomerative clustering procedure
- This paper worked out how state tying can be done with decision trees
- The main advantage is the possible construction of unseen triphones



# Building Process

## Tree-Based State Tying

Martin Raab

### Introduction

Parameters  
Motivation

### Basics

Triphone Tying  
Decision Trees

### Tree based State Tying

Outline  
**Building**  
Tree Clustering

### Discussion

### References

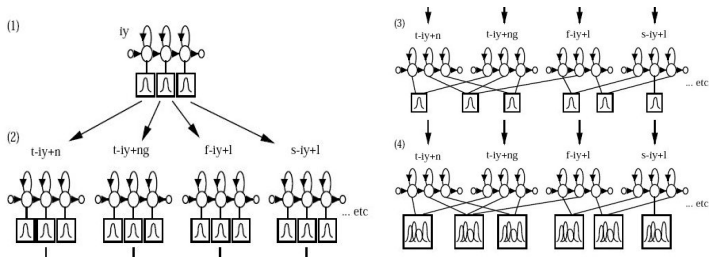


Figure: The tied-state HMM system build-procedure

⇒ This paper wants to improve step 3.



## Tree-Based State Tying

## Tree based State Tying

## Building a Tied State HMM

## Building Process



Figure: The tied-state HMM system build-procedure

→ This paper wants to improve step 3.

1. Standard model for a monophone with 3 states, the output probability is modelled by a single gaussian
2. The state output distributions are cloned to initialise a set of untied triphones. → Triphones are trained with Baum Welch  
→ The clone step has guaranteed training for rare combinations of phones
3. Triphones derived from the same monophone are clustered. A typical state is chosen, others are clustered to it → **THIS is the step this paper wants to improve**
4. Mixture components are increased and retrained to gain maximum performance
5. Left Out: Why using a single gaussian?
6. Left Out: Bootstrap, Baum Welch



# Tree Based Clustering

## Tree-Based State Tying

Martin Raab

### Introduction

Parameters

Motivation

### Basics

Triphone Tying

Decision Trees

### Tree based State Tying

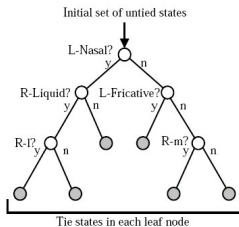
Outline

Building

**Tree Clustering**

Discussion

References



## What questions to ask?

- What question can be asked to separate the states efficiently?
- → Typically done with questions designed by experts
- A pool of about 200 questions from broad phonetic classes (nasal, vowel) till other phones like r or l

## Tree-Based State Tying

- └ Tree based State Tying
  - └ Tree Based Clustering
    - └ Tree Based Clustering



## What questions to ask?

- What question can be asked to separate the states efficiently?
- — Typically done with questions designed by experts
- A pool of about 200 questions from broad phonetic classes (nasal, vowel) till other phones like r or l

1. Tie states in each leaf node
2. One tree for each state of each phone



# Tree Based Clustering II

## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

Triphone Tying

Decision Trees

Tree based  
State Tying

Outline

Building

**Tree Clustering**

Discussion

References

## Which question to ask?

- Put all triphones of one phone in the root node
- Calculate the log likelihood of the training data being generated
- Split this node into two that the maximum increase of log likelihood is achieved
- Stop : threshold and/or minimal occupation

# Tree-Based State Tying

- Tree based State Tying
  - Tree Based Clustering
    - Tree Based Clustering II

**Which question to ask?**

- Put all triphones of one phone in the root node
- Calculate the log likelihood of the training data being generated
- Split this node into two that the maximum increase of log likelihood is achieved
- Stop : threshold and/or minimal occupation

1. Calculate the log likelihood that the training data was generated by the assumption that all states in this node are tied
2. Split this node into two by finding the question, that gives two nodes in a way that the maximum increase of log likelihood is achieved
3. This is continued till a threshold or a minimal occupation count stops this process



# Tree Based Clustering III

## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

Triphone Tying

Decision Trees

Tree based State Tying

Outline

Building

Tree Clustering

Discussion

References

## Which question to ask?

- $S$  is a set of HMM states
- $L(S)$  the log likelihood that  $S$  generated the set of training frames  $F$
- $L(S) = \sum_{s \in S} \sum_{f \in F} \log Pr(o_f; \text{Distribution}) * \gamma_s(o_f)$
- $L(S) = -\frac{1}{2}(\log[(2\pi)^n | \sum S|] + n) \sum_{s \in S} \sum_{f \in F} \gamma_s(o_f)$
- $\Delta L_q = L(S_y(q)) + L(S_n(q)) - L(S)$  with question  $q$  and new nodes  $y, n$
- Stop : threshold, minimal occupation count

## Tree-Based State Tying

## └ Tree based State Tying

## └└ Tree Based Clustering

## └└└ Tree Based Clustering III

## Which question to ask?

- $S$  is a set of HMM states
- $L(S)$  the log likelihood that  $S$  generated the set of training frames  $F$
- $L(S) = \sum_{\alpha \in S} \sum_{f \in F} \log Pr(\alpha; \text{Distribution}) + \gamma_s(\alpha)$
- $L(S) = -\frac{1}{2}(\log((2\pi)^n \sum S) + n) \sum_{\alpha \in S} \sum_{f \in F} \gamma_s(\alpha)$
- $\Delta L_q = L(S_q(q)) + L(S_s(q)) - L(S)$  with question  $q$  and new nodes  $s, n$
- Stop : threshold, minimal occupation count

1. There can't be a presentation at Karlsruhe without a sum sign
2.  $o_f$  is the observed distribution in training frame  $f$
3.  $\gamma_s$  is the a posteriori probability that  $o_f$  is generated by state  $s$ , this is hard to know if not in training, usually approximated with the Bayes rule, but this shows that this algorithm cannot be improved after the training session (without further improvements)
4.  $\gamma_s$  is known from preceding passes of optimization, more specific the Baum Welch algorithm done in step2 from the building process, see previous slides
5. if the output PDF (probability density functions) are gaussians, then point is the specified version of the more general formula
6. this is only depending on the pooled state variance  $\sum S$  and the total state occupancy of the pool  $\sum_{s \in S} \sum_{f \in F} \gamma_s$ , this is all known



# Evaluation

## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

Triphone Tying

Decision Trees

Tree based  
State Tying

Outline

Building

Tree Clustering

Discussion

References

## Data Driven vs. Tree Based Clustering

System	Feb 89	Oct 89	Feb 91	Sep 92
Add D-D	4.10	4.84	3.78	8.05
Tree	3.87	4.99	3.74	7.31

Word error rates on Resource Management system (1000 Words)

## State vs. Model-based clustering

System	Nov 92	si_dt_s6	si_dt_05	Nov 93
Model	7.17	10.61	12.17	11.22
State	5.90	10.33	10.73	9.89

Word error rates on Wall Street Journal (5000 Words)



Data Driven vs. Tree Based Clustering

System	Feb 89	Oct 89	Feb 91	Sep 92
Add D-D	4.10	4.84	3.78	8.05
Tree	3.87	4.99	3.74	7.31

Word error rates on Resource Management system (1000 Words)

State vs. Model-based clustering

System	Nov 92	si_dfl20f	si_dfl20S	Nov 93
Model	7.17	10.61	12.13	11.22
State	5.90	10.33	10.73	9.89

Word error rates on Wall Street Journal (5000 Words)

1. Left Out: What kind of questions?
2. Similar performance, but advantage of constructing unseen triphones
3. State model systems are more performant than model tying systems with an average of 14%



# Conclusion

## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

Triphone Tying

Decision Trees

Tree based  
State Tying

Outline

Building

Tree Clustering

Discussion

References

## Original Conclusions

- Phonetic decision trees are as effective for clustering as data driven methods
- State tying systems are more performant than model tying systems
- Overall performance was at 1994 state of the art

## Current Conclusions

- State tying is today a standard technique
- State tying offers the key advantage of modelling unseen triphones

## Original Conclusions

- Phonetic decision trees are as effective for clustering as data driven methods
- State tying systems are more performant than model tying systems
- Overall performance was at 1994 state of the art

## Current Conclusions

- State tying is today a standard technique
- State tying offers the key advantage of modelling unseen triphones

1. Left Out : It is important to use continous HMM ... as this is not discussed in this presentation



# References

## Tree-Based State Tying

Martin Raab

Introduction

Parameters

Motivation

Basics

Triphone Tying

Decision Trees

Tree based State Tying

Outline

Building

Tree Clustering

Discussion

References

- Chen, S. (2003). Compiling large-context phonetic decision trees into finite-state transducers. *Eurospeech 2003*.
- Hirsimaki, T. (2003). A review: Decision trees in speech recognition. *Helsinki University of Technology*.
- Young, S., Odell, J., & Woodland, P. (1994). Tree-based state tying for high accuracy acoustic modelling. *Cambridge University*.
- Zen, H., Tokuda, K., & Kitamura, T. (2003). Decision tree-based simultaneous clustering of phonetic contexts, dimensions, and state positions for acoustic modeling. *Nagoya Institute of Technology*.

---

Some figures have been taken from the lecture “Sprachliche Mensch Maschine Kommunikation” at the University of Karlsruhe.