Manuelle Segmentierung von Sprachkorpora: das Phon und die akustische Realität

Transkription in Sprachsynthese und -erkennung

Hauptseminar im Sommersemester 2004

Prof. Dr. Wolfgang Hess

Stefan Breuer, M.A.

Referentin: Anastasija Eifer

Gliederung

- Einführung
- Segmentierung und Annotation
- Konsistenz manueller Segmentierung und Transkription: klare und nicht klare Fälle
- Manuelle Segmentierung und Transkription in verschiedenen Sprachen
- Abschätzung der Qualität manueller und automatischer Segmentierung und Transkription
- Zusammenfassung

Einführung (1/3)

Beispiel für ein segmentiertes und annotiertes Sprachsignal:

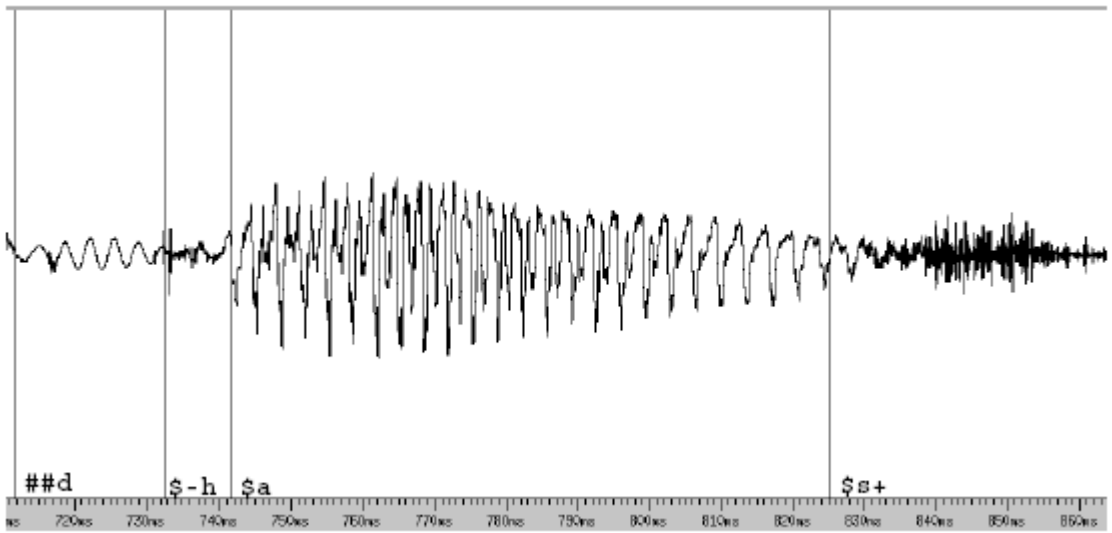


Abb.1: Die Segmentation und Etikettierung des Wortes *das* (*das Kiel Korpus*)

Einführung (2/3)

- manuelle Segmentierung
  - sehr zeitaufwändig
  - von geschulten Phonetikern
- automatische Segmentiermethoden
  - viele verschiedene Systeme
  - schnell
  - oft manuelle Korrektur der Labels

Einführung (3/3)

Warum Segmentierung und Annotation von Sprachkorpora?

- sprachtechnologische Zwecke
  - Trainingsmaterial für Spracherkennung und Sprachsynthese, Testen
- Sprachforschung

Segmentierung und Annotation (1/5)

Probleme :

- Kontinuität des Signals:

  Koartikulation - die Laute eines Wortes lassen sich nicht gegeneinander abgrenzen.

  [am], [um], [an] – man kann nicht den Vokal vom Nasal trennen

- Varianz des Signals:
  - gleicher Laut hört sich nicht immer gleich an,
  - gleiches Wort wird nicht immer gleich ausgesprochen

Segmentierung und Annotation (2/5)

Probleme :

- Segmentgrenzen: ein nicht trivialer Aspekt
  - Segmenten lassen sich nicht immer klar abgrenzen
  - Entscheidungen müssen willkürlich getroffen werden

Segmentierung und Annotation (3/5)

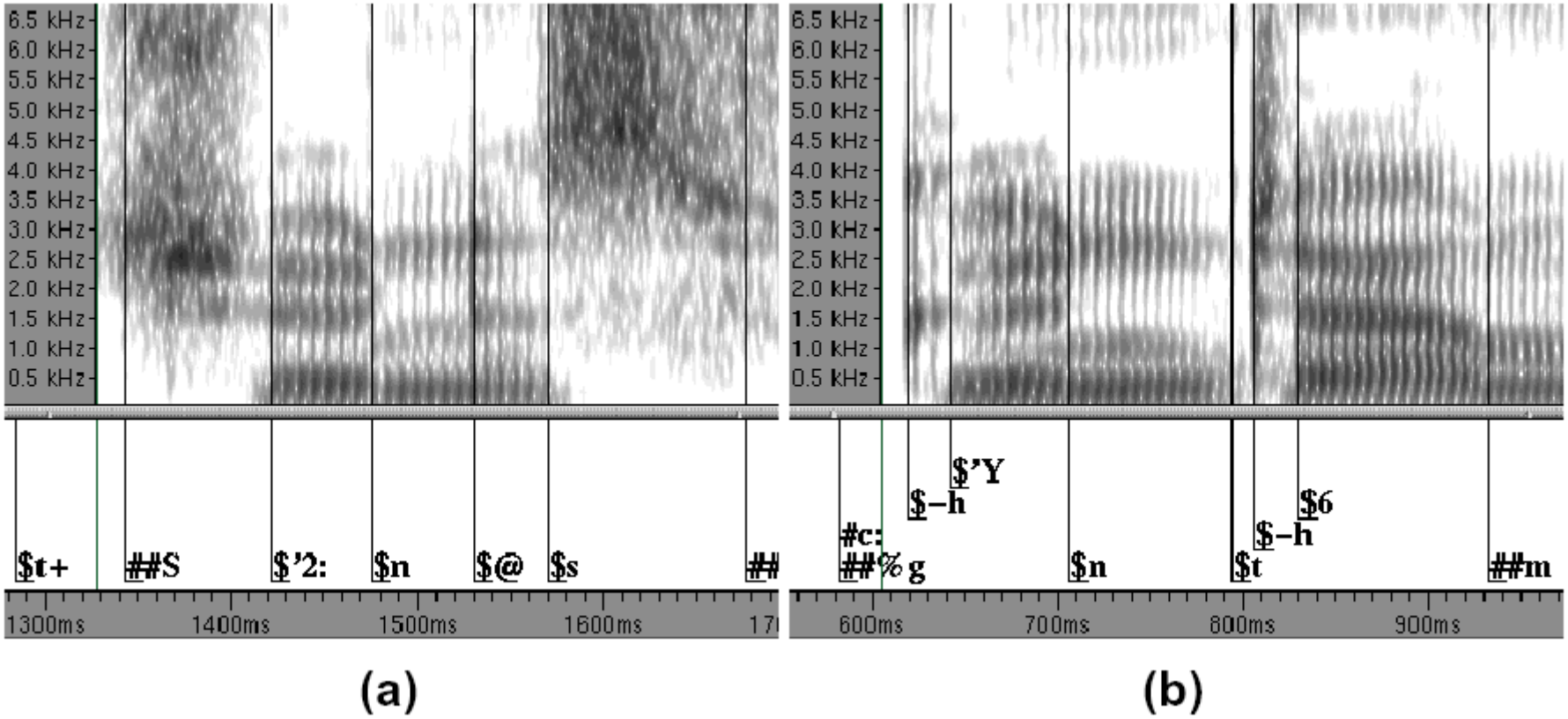


Abb.2: Segmentationen un Etikettierungen der Wörter (a) *schönes* und (b) *Günther* (*das Kiel Korpus*)

Segmentierung und Annotation (4/5)

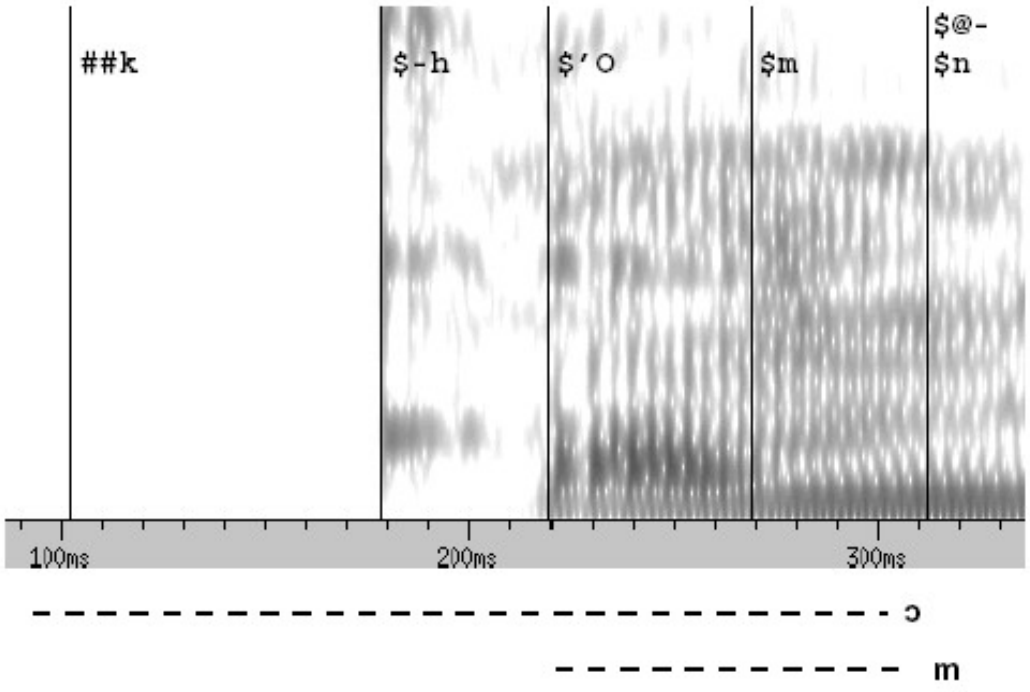


Abb.3: Sonagramm des Wortes *kommen* samt Segmentation (vertikale Striche) und Etikettierung. Die horizontalen gestrichelten Linien zeigen die mögliche zeitliche Ausdehnung der phonetischen Korrelate der phonologischen Elemente ɔ und m (*das Kiel Korpus*)

Segmentierung und Annotation (5/5)

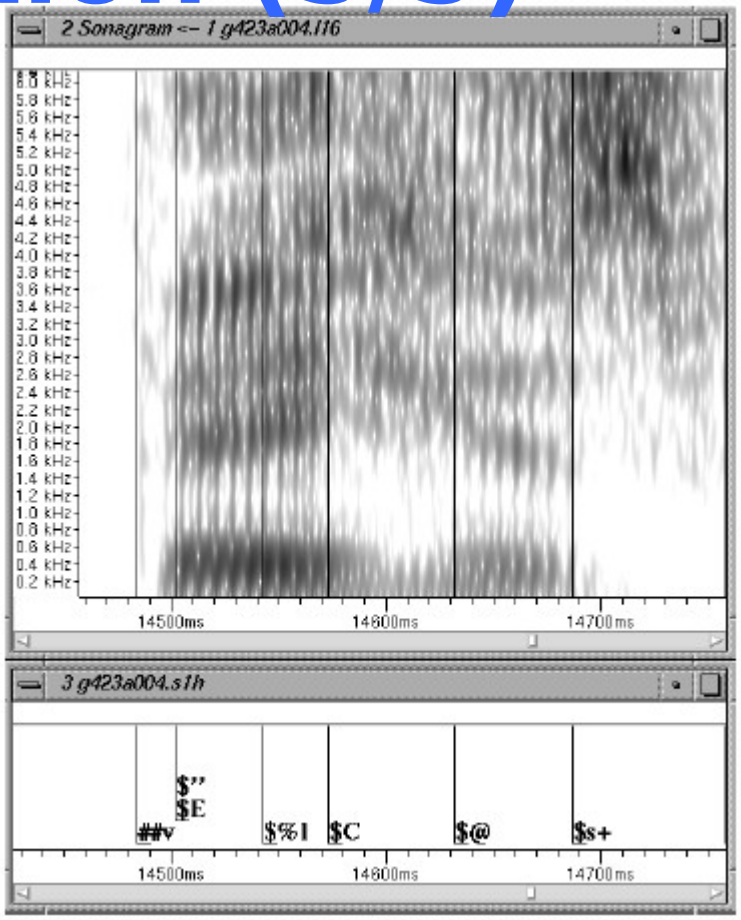


Abb.4: Sonagramm und Etikettierung des Wortes *welches* (*das Kiel Korpus*)

Konsistenz manueller Segmentierung und Transkription: klare und nicht klare Fälle (1/8)

B.Eisen, H.G.Tilmann

Institut für Phonetik und Sprachliche Kommunikation der Universität München

Das Ziel :

- Untersuchung der Konsistenz phonetischer Segmentierung und Transkription
  - interindividuelle Konsistenz
  - intraindividuelle Konsistenz

Konsistenz manueller Segmentierung und Transkription: klare und nicht klare Fälle (2/8)

Sprachdaten:

- 100 Sätze, gelesen von sechs deutschen Sprechern (insgesamt 600 Sätze)
- manuell segmentiert und etikettiert von vier Transkribenten
- resegmentiert nach ca.10-12 Monaten

Konsistenz manueller Segmentierung und Transkription: klare und nicht klare Fälle (3/8)

Struktur der Dateien:

- eine Liste von
  - Segmentgrenzen mit den entsprechenden IPA - Etiketten
  - Gruppierung der orthographischen Darstellung des Wortes zu einem String von phonetischen Segmenten

Konsistenz manueller Segmentierung und Transkription: klare und nicht klare Fälle (4/8)

Datenverwaltung mit Prolog:

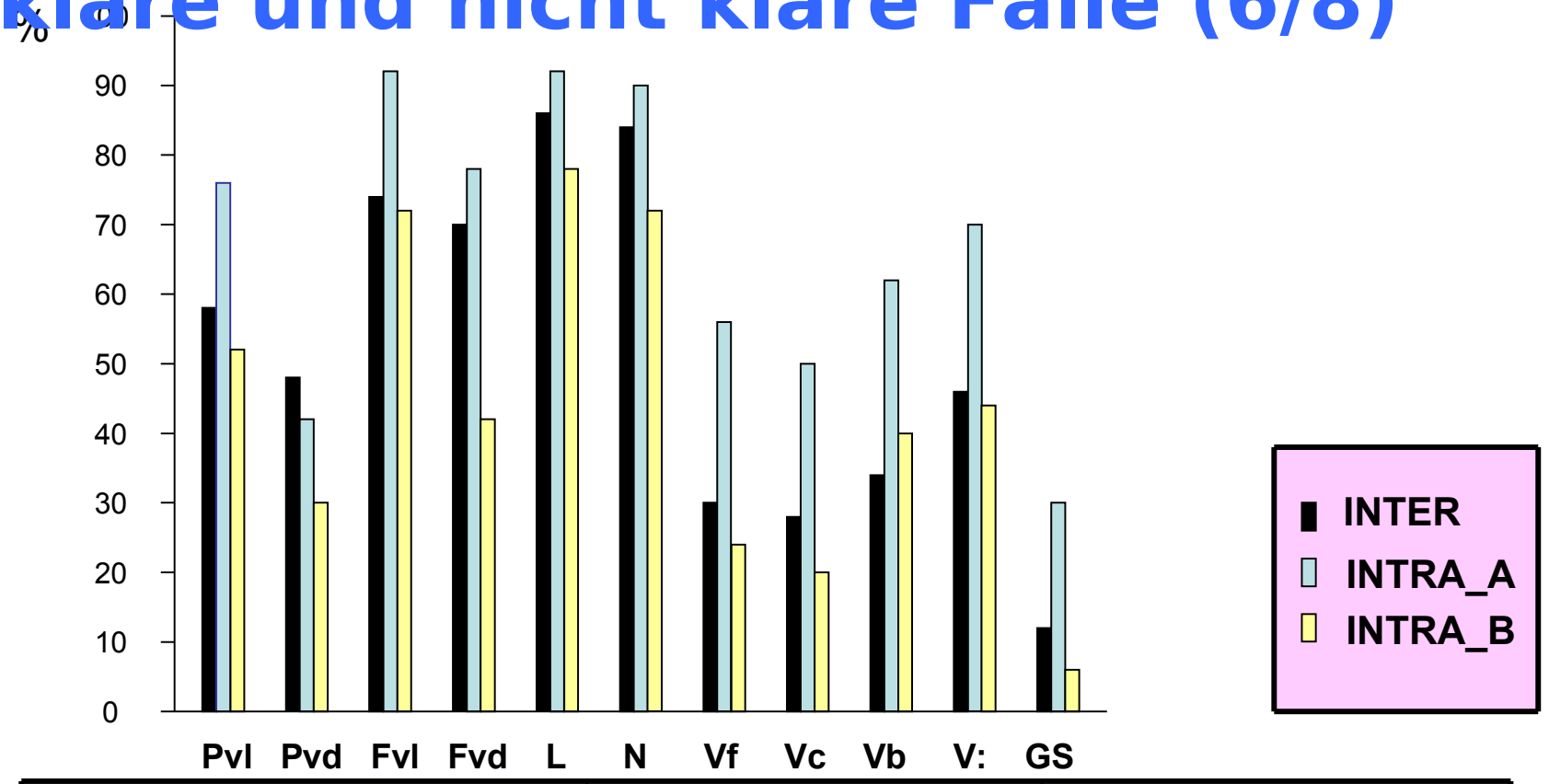
- Fakten: Segment- und Labeldateien
- Prädikate für :
  - Trankriptionsabbildung
  - Identifizierung der Sätze des Korpus und ihre kanonische Formen
  - orthographische Darstellung des Wortes mit der Variantendarstellung
- Regeln: Beziehungen zwischen Segmenten, Transkribenten und Wörtern

Konsistenz manueller Segmentierung und Transkription: klare und nicht klare Fälle (5/8)

Resultierende Datenbasis:

- individuelle etikettierte Dateien für jeden Satz des Korpus und für jede Version der Segmentierung
- File Header:
  - Sprecher-Information
  - Satz-ID
  - orthographische Repräsentation des Satzes
  - Wortformen
  - Name des Transkribenten
  - Version der Segmentierung

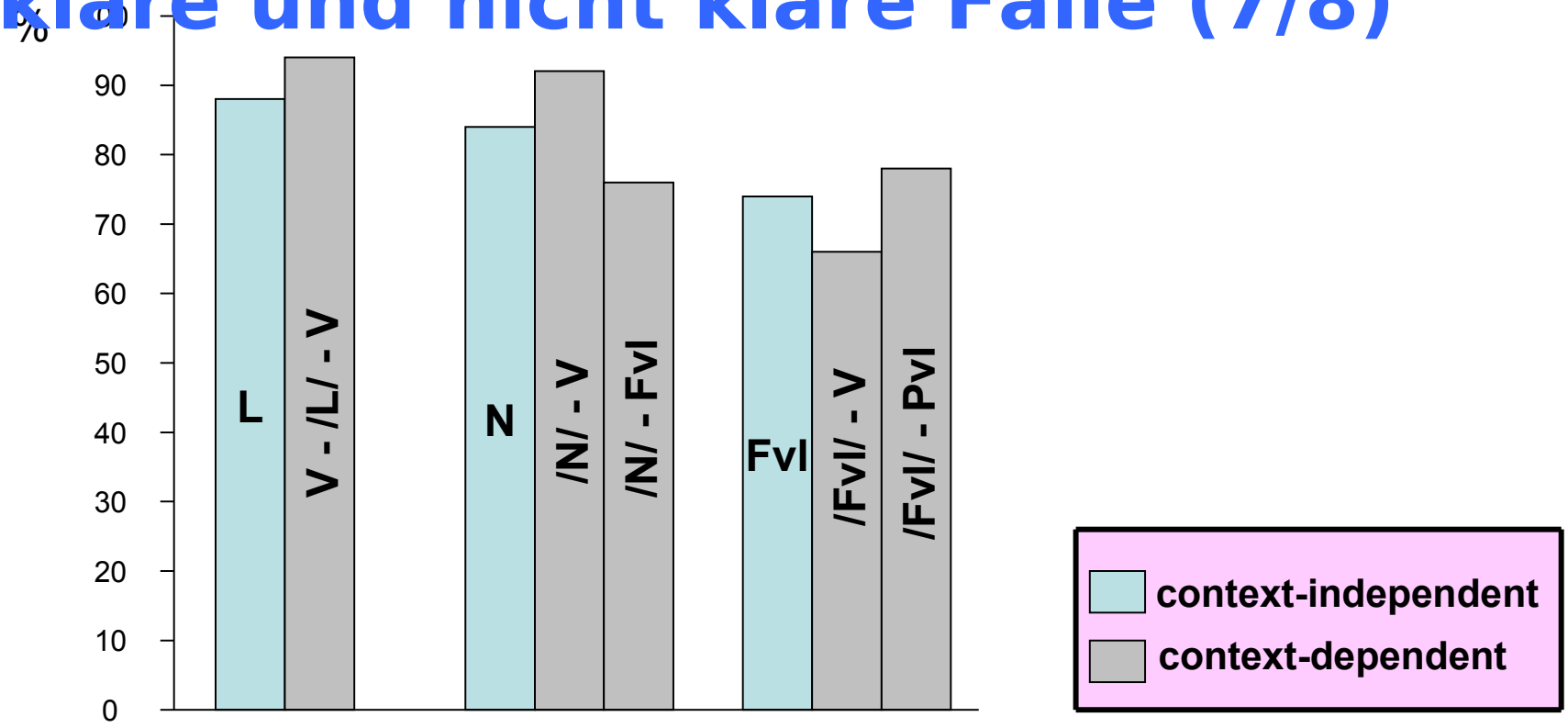
Konsistenz manueller Segmentierung und Transkription: klare und nicht klare Fälle (6/8)



Pvl	voiceless plosives	L	laterals	V:	long vowels
Pvd	voiced plosives	N	nasals	Vf	front vowels
Fvl	voiceless frikatives	GS	glottal stops	Vc	central vowels
Fvd	voiced frikatives	V	vowels	Vb	back vowels

Abb.5: Prozentsatz identischer Transkriptionen (inter- und intraindividuelle Konsistenz)

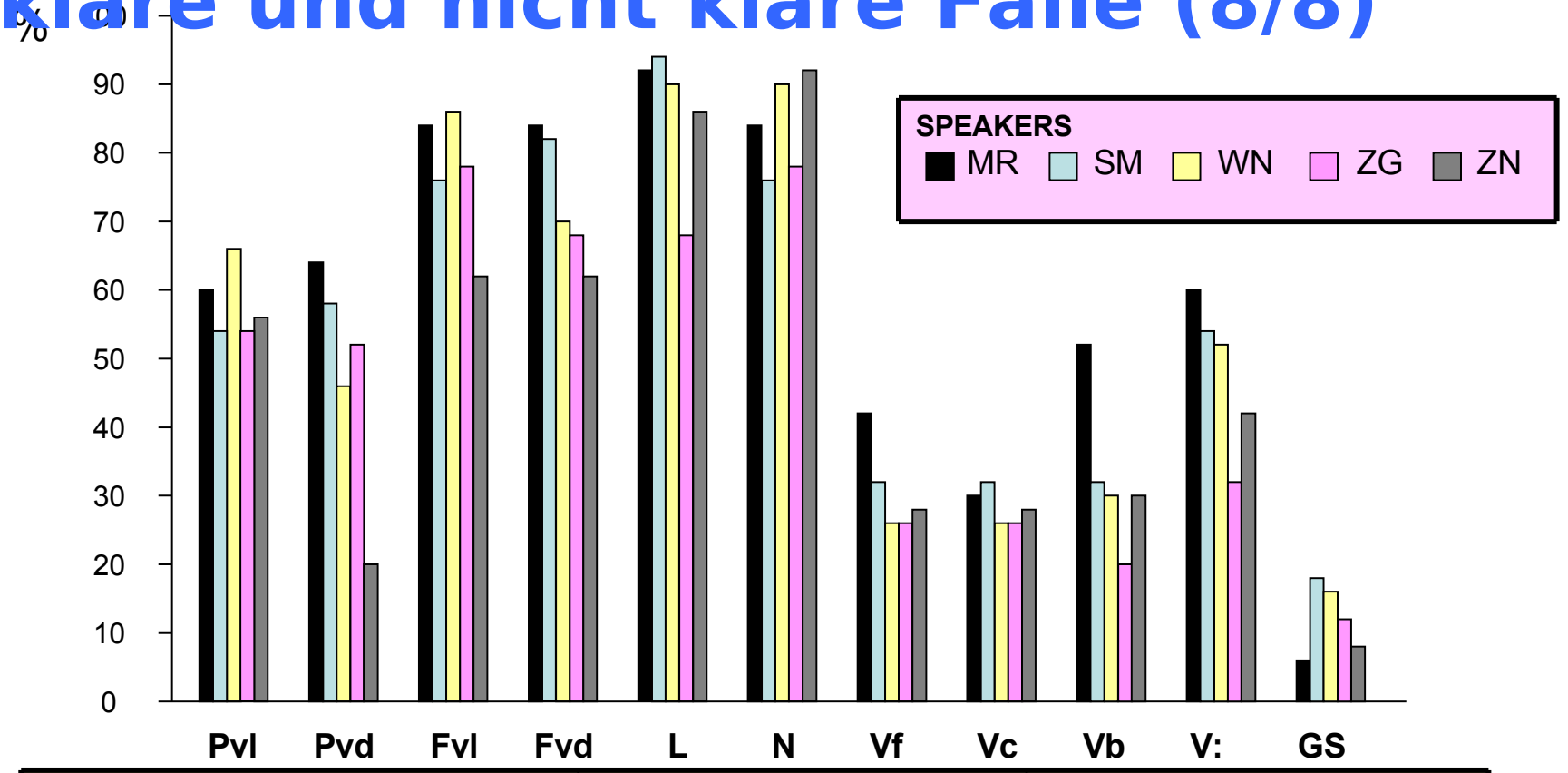
Konsistenz manueller Segmentierung und Transkription: klare und nicht klare Fälle (7/8)



Pvl	voiceless plosives	L	laterals	V:	long vowels
Pvd	voiced plosives	N	nasals	Vf	front vowels
Fvl	voiceless fricatives	GS	glottal stops	Vc	central vowels
Fvd	voiced fricatives	V	vowels	Vb	back vowels

Abb.6: Prozentsatz identischer Transkriptionen in verschiedenen Kontexten

Konsistenz manueller Segmentierung und Transkription: klare und nicht klare Fälle (8/8)



Pvl	voiceless plosives	L	laterals	V:	long vowels
Pvd	voiced plosives	N	nasals	Vf	front vowels
Fvl	voiceless fricatives	GS	glottal stops	Vc	central vowels
Fvd	voiced fricatives	V	vowels	Vb	back vowels

Abb.7: Prozentsatz identischer Transkriptionen (verschiedene Sprecher)

Manuelle Segmentierung und Transkription in verschiedenen Sprachen (1/9)

Ronald Cole, Beatrice T.Oshika, Mike Noel, Terri Lander, Mark Fanty

Center for Spoken Language Understanding

Oregon Graduate Institute of Sience and Technologie, USA

- **Experiment 1:**

  Englisch, Deutsch, Mandarin und Spanisch, segmentiert und etikettiert von Linguisten, die diese Sprachen fließend sprechen

- **Experiment 2:**

  Deutsch und Hindi, segmentiert und etikettiert von Linguisten, die diese Sprachen nicht sprechen

Manuelle Segmentierung und Transkription in verschiedenen Sprachen (2/9)

Sprachdaten:

- Quelle: *OGI Multi-language Telephone Speech corpus*
- der Korpus enthält 50 sec.-Segmente von kontinuierlicher Telefonaten in verschiedenen Sprachen, sog. „*stories*“
- von jeder Sprache wurden 10 *stories* ausgewählt (insgesamt ca. 30 min.)
- jede *story* wurde von zwei Linguisten bearbeitet

Manuelle Segmentierung und Transkription in verschiedenen Sprachen (3/9)

- die Analyse wurde auf drei verschiedenen Ebenen durchgeführt:
  - **full** – full label set
  - **base** – reduced symbol set (ohne Diakritika)
  - **broad categories** – Vokale, Verschlußlaute, Plosive, Frikative, Semivokale, Nasale und nichtsprachliche Laute

Manuelle Segmentierung und Transkription in verschiedenen Sprachen (4/9)

Experiment 1:

	Full	Base	Broad	Segments
Englisch	69,67	70,79	89,06	512
Deutsch	60,98	64,69	80,78	533
Mandarin	65,61	77,90	86,75	410
Spanisch	73,81	81,77	90,13	523

Tab.1: Ergebnisse der Transkriptionsanalyse

Manuelle Segmentierung und Transkription in verschiedenen Sprachen (5/9)

Experiment 1:

milliseconds

	< 2	< 4	< 6	< 11
Englisch	29%	55%	67%	79%
Deutsch	21%	46%	63%	79%
Mandarin	32%	58%	71%	83%
Spanisch	20%	40%	53%	71%

Tab.2: Ergebnisse der Analyse von Segmentgrenzen (*broad categories*)

Manuelle Segmentierung und Transkription in verschiedenen Sprachen (6/9)

Experiment 2:

	Full	Base	Broad	Segments
Deutsch	34,79	40,50	77,52	25
Hindi	34,03	42,22	82,87	26

Tab.3: Ergebnisse der Transkriptionsanalyse

Manuelle Segmentierung und Transkription in verschiedenen Sprachen (7/9)

Experiment 2:

milliseconds

	< 2	< 4	< 6	< 11
Deutsch	**32%**	**59%**	**69%**	**81%**
Hindi	**27%**	**56%**	**67%**	**79%**

Tab.4: Ergebnisse der Analyse von Segmentgrenzen (*broad categories*)

Manuelle Segmentierung und Transkription in verschiedenen Sprachen (8/9)

- **Experiment 1:**
  - Ergebnisse der Transkriptionsanalyse:
    durchschnitt. 67,5% (*full label set*),
    73,79% (ohne Diakritika) und
    86,68% (*broad*)
  - Analyse der Segmentgrenzen:
    durchschnitt. 78%

- **Experiment 2:**
  - Ergebnisse der Transkriptionsanalyse:
    durchschnitt. 34,41% (*full label set*),
    41,36% (ohne Diakritika) und
    80,2% (*broad*)
  - Analyse der Segmentgrenzen:
    durchschnitt. 80%

Manuelle Segmentierung und Transkription in verschiedenen Sprachen (9/9)

	count	correct
vowel	3118	59%
nasal	937	89%
semi-vowel	1125	79%
plosive	1293	90%
closure	1225	86%
fricative	1501	82%
nonspeech	1123	78%

Tab.5: Ergebnisse der Analyse von Englisch (Experiment 1, *broad categories*)

Abschätzung der Qualität manueller und automatischer Segmentierung und Transkription

Maria-Barbara Wesenick, Andreas Kipp

Institut für Phonetik und Sprachliche Kommunikation (IPSK)
Ludwig-Maximilians-Universität München

- manuelle und automatische Segmentierung und Transkription:
  - **Untersuchung von Segmentlabels (von Konsonanten)**
  - **Untersuchung von Segmentgrenzen**

Abschätzung der Qualität manueller und automatischer Segmentierung und Transkription

Sprachdaten: Phondat-II Datenbank des Deutschen

- manuelle Segmentierung:
  - insgesamt 10 Sprecher und 10 Linguisten
  - von jedem Sprecher jeweils 64 Sätze
  - die Daten von jedem Sprecher im Schnitt von drei Linguisten segmentiert
- automatische Segmentierung: MAUS
  – Munich AUutomatic Segmentation System

Abschätzung der Qualität manueller und automatischer Segmentierung und Transkription

	labels	a) manual transcriptions	b) automatic transcriptions
Stops	p	93.8	76.4
	b	97.8	82.5
	t	92.5	80.2
	d	79.6	75.1
	k	92.1	89.2
	g	85.9	72.1
	Q	86.6	78.3
	all stops	89.9	80.2
Fricatives	f	99.2	99.6
	v	96.5	88.2
	s	98.5	95.1
	z	92.9	98.6
	S	99.2	94.0
	C	98.3	94.3
	j	96.4	97.5
	x	99.4	92.9
	h	92.3	71.5
	all fric.	98.0	93.6
Nasals	m	98.2	97
	n	97.9	94.9
	N	93.4	83.5
	all nas.	97.5	94.4
	l	98.0	64.1
	r	96.0	99.0
	all consonants	94.8	88.4

Tab.6: Prozentsatz identischer Labels für Konsonanten in a)manuellen Transkriptionen und b)automatischen Transkriptionen

Abschätzung der Qualität manueller und automatischer Segmentierung und Transkription

labels	a) manual transcriptions	b) automatic transcriptions
p	b – 5.2% v – 2.6%	b – 22%
b	p – 0.8%	p – 9% v – 4%
t	d – 4.4%	d – 10%
d	t – 11.4%	t – 10% p – 2% b – 2% Q – 1%
k	g – 5.6%	g – 5% Q – 1%
g	k – 8.7%	k – 11%
v	f – 2.1% b – 10%	f – 9.5%
s	z – 0.8%	z – 3.3%
N	n – 4.3%	n – 4%

Tab.7: Verwechslungen von Labels in a)manuellen Transkriptionen und b)automatischen Transkriptionen

Abschätzung der Qualität manueller und automatischer Segmentierung und Transkription

Segment boundary	a) manual segmentations	b) automatic segmentations
N-N	16	43
N-Fvd	11	34
V-L	14	36
V-Fvd	9	31
Fvl-Fvd	12	28
Fvl-Pvl	5	21
Fvl-Pvd	7	19
V-N	9	19
N-V	8	18
L-V	8	17
L-Pvd	12	19
Pvd-V	6	12
V-V	15	20
N-Pvd	11	15
Fvl-Fvl	11	13
Fvl-N	13	14
V-Fvl	7	8
N-Fvl	6	7
Fvd-V	12	12
N-Pvl	10	10
Pvd-N	9	9
V-Pvl	12	11
Pvl-Fvl	11	10
Fvl-V	7	6
Pvl-N	12	7

Tab.8: Mittlere Abweichung von Segmentgrenzen in ms für a)manuelle Segmentierung und b)automatische Segmentierung

Abschätzung der Qualität manueller und automatischer Segmentierung und Transkription

time range	a) manual segmentations	b) automatic segmentations
= 0 ms	63%	1% (< 0.5 ms: 15%)
< 5 ms	73%	36%
< 10 ms	87%	61%
< 15 ms	91%	76%
< 20 ms	96%	84%
< 32 ms	99%	90%
< 64 ms	100%	95%

Tab.9: Ergebnisse der Analyse von Segmentgrenzen für a)manuelle Segmentierung und b)automatische Segmentierung

Zusammenfassung

- Qualität phonetischer Segmentierung und Transkription ist unter anderem wichtig für automatische Spracherkennung- und Sprachsynthesesysteme
- Es gibt keine „einzig richtige“ Transkription, Abweichungen sind möglich
- Bestimmte phonetische Kategorien lassen sich leichter segmentieren
- Dieser Prozess ist kontext- und sprecherabhängig