BAS Infrastrukturen zur Technischen Sprachverarbeitung (BITS)

Teilprojekt 8 (Doku 8/5e)

Conventions for segmentation

Content: Here are the complete conventions used in the BITS-segmentation group. These contain principles for transcription and segmentation with examples for difficult cases. The different classes of phonemes - plosives, affricates, fricatives, nasals, r-realisations, vowels and diphthongs - are discussed separately. Segmentation of sentences and logatomes are discussed separately. At the end of the document a complete list of the SAM-PA signs used in BITS can be found.

Author: Tania Ellbogen
Date: 03.11.2005
Version: 1.6

Exact segmentation of the sentences

I. Basic principles

1. The levels of labelling

The labelling of the utterance takes place on two levels.

Level I:
The phonemic transcription on the basis of the word forms produced by MAUS. The segments of this level are used as proposals for the second level.

Level II:
Segmentation and transcription of the actual spoken utterance in reference to the representation of phonemes of level I.

2. The principles of the reference

Level II is mapped non-ambiguous and completely to level I. Thus, there are four ways of mapping the segments to the phonemes created by MAUS.

1. Acceptance

A proposed element from level I is accepted on level II: the actual utterance corresponds with the representation of phonemes.

e.g.: /fYnf/ is realised as [fYnf]

2. Replacement

A proposed element from level I is realised differently. There is a discrepancy:

e.g.: /fYnf/ is realised as [fYmf]

3. Elision

An element from level I was not realised.

e.g.: /hat@n/ is realised as [hatn]

It is possible that more than one element is missing.

4. Insertion

In the given utterance, an additional element is existing which is not present on level I.

e.g.: /gans/ is realised as [gants]

The insertion can contain more than one element. In this case there is a segment for every single element.

II. Principles for transcription

GT1
The assignment of symbols for transcription is based primary on the auditory judgement of the utterance. The underlying period of the judgement should be at least the size of a syllable. No transcription of single elements!

GT2
A discrepancy of the proposed representation of phonemes on level I is annotated solely, if another category is perceived and if the assignment of another symbol of the given inventory is justifiable (e.g. /i:/ instead of /I/). Variants in consequence of coarticulation are not annotated.

GT3
The sample of symbols is constricted to the BITS-SAM-PA inventory. Other symbols are not allowed.

GT 4
The label '<p:>' (pause) is given if there are pauses within an utterance, that can not be interpreted as aspiration or silence prior a plosive. Pauses can be filled with noises or even glottal stops if the glottal stop does not belong obligatory to the preceding or following phoneme. The label '<br:>' (breathing) is given if there are clearly audible noises of breathing in a given utterance. A preceding or following pause is not labelled separately. The whole segment is labelled '<br:>'. Breathing preceding or following the sentence is not labelled. These parts are marked with '<p:>' as a principle.

GT 5
Discrepancies with the text can be: false, added or missed words or phonemes. In this case, the file is not segmented (enter „defect“ in the shell after quitting PRAAT). Consequently the file won't be segmented any further. It will be recorded again in correct manner.

III. Principles for segmentation

GS 1
Within the sentences every phoneme is segmented. Beginning and end of the sentence (ahead of the first phone respectively after the last phone) are marked with '<p:>'.

GS 2
The borderline for segments are always set at positive 0-crossings in the oscillogram.

GS 3
The setting of the borderline should be controlled by sonagram and oscillogram.

GS 4

At periods where both of two neighbouring phonemes can be heard together the border is set in the middle of this period. (Examples for this are fricative combinations /s-f/, /s-S/)

GS 5

Voiced (periodic) elements start with the first clear identifiable period.

GS 6

The border at signals with low intensity (especially /h/, aspiration) is set where the signal can be clearly distinguished from the background noise. To find out where exactly the border lies you have to zoom in the speech signal. The placing of the final border (e.g. aspirated plosives at the and of an utterance) results from the same principle. Noises of breathing - if recognised clearly - have to be cut off from the friction or aspiration.

GS 7

If a smack (or technical noise) can be heard in the utterance, this has to be indicated with a '§' (without blank) in the concerning segment.

GS8

The single words of a sentence are marked with brackets '(', ')'. If the last phoneme of word is the same as the first phoneme of the following word, this phoneme is part of both words and is therefore marked as beginning as well as ending of a word. e.g.: “hat den” --> /(h/ /a/ /(t)/ /e:/ /n)/.
If the according phoneme is a plosive the phase of silence is the common segment. If voiced and voiceless plosive come together, then as a principle, the first phoneme of the second word is labelled, e.g. “hat den” --> /(h/ /a/ /(d_s)/ /d_b/ /e:/ /n)/.
If between two words an affricate is following a plosive, the common segment is the phase of silence of the affricate, e.g. “wird zum” --> /(v/ /I/ /R/ /(ts_s)/ /ts_b/ /U/ /m)/.

IV. Handling of difficult cases

In the following typical difficult cases will be exemplified.

1. Plosives

a) Plosives are separated into two segments. The first segment contains the occlusion. The second segment contains the burst and possibly an aspiration. To distinguish the two segments they are labelled e.g. |t_s| and |t_b|, where 's' stands for 'silence' and 'b' stands for 'burst'.

b) The borderline of plosives at the beginning of an utterance gets an occlusion arbitrary set at 20 - 40ms.

c) After pauses plosives are treated like plosives at the beginning of an utterance.

d) The occlusion of a voiced plosive with voicing lead in between vowels starts after the last identifiable period of the vowel. The occlusion can be recognised by a break-in of the energy of the higher formants and in a damped sinus like signal.

e) Plosives at the end of an utterance end with the burst respectively after decay of the aspiration (see signal). Possible breathing noise has to be cut off from the segment.

f) After nasals the start of voiced plosives (activity of the velum) often can not be identified clearly. In this case the decreasing phase of the nasal is part of the occlusion. Often the burst can just be noticed as a irregularity in the following period. This is part of the plosive, too.

g) Plosives with an incomplete occlusion are noted as complete plosives if the auditory impression suggests an occlusion. There should be a clear noticeable reduction of energy during the phase of occlusion. In other cases the segment has to be labelled with a equivalent fricative if necessary.

h) The proposition of MAUS with the discrimination of voiced/voiceless is not adopted if a change of categories is evident.
Example: /p, t, k/ is realised with voicing lead
/b, d, g/ is realised aspirated and voiceless in the beginning of a syllable.

i) Glottal stops are in principle treated like plosives. There is a arbitrary first borderline (20 - 40ms) with a glottal stop at the beginning of an utterance. If the occlusion is missing completely, only 'Q' is segmented (without |_s| and |_b|). The borderline of 'Q' at the beginning of an utterance gets an occlusion arbitrary set at 20 – 40 ms.

j) If instead of a glottal stop only a creaky phoneme can be heard, this phoneme is labelled with 'q' after the SAM-PA sign, e.g. 'aq'. The preceding phoneme (before the expected glottal stop) should stay unmodified if possible.

2. Affricates

Affricates (ts, tS, pf) are treated as one phoneme. Like plosives they are divided into two segments: the first segment is the phase of occlusion, the second segment contains burst and fricative, e.g. |pf_s| and |pf_b|.

3. Fricatives

If two fricatives with the same point of articulation follow each other (e.g. 'auffallen') two segments are transcribed solely if they are clearly distinguishable.

4. Nasals

a) Syllabic nasals after nasals are segmented if they are perceived as two segments (e.g. long duration or internal structuring).

b) Voiceless nasals are not labelled in particular. The label proposed by MAUS is kept if every other parameter is realised adequate.

5. R-Realisations

The symbol /R/ stands for:

uvular trill
alveolar trill
uvular fricative (voiced/voiceless)
velar fricative.

In level I /R/ in the appropriate positions is transcribed as a vowel and offered for segmentation as R-diphthong like in /h a m b U 6 k/ (Hamburg).
If /R/ is realised as trill or fricative ([h a m b U R k]) the diphthong has to be replaced by the appropriate vowel and /R/ has to be inserted. If instead of a diphthong only a vowel is realised (e.g. [d E:] instead of [d e: 6]) the diphthong has to be replaced by the vowel.
Also possible is the realisation with R-diphthong + /R/, e.g. in /s E6 R b_s b_b m/ (Serben).

6. Vowels

a) Long vowels get the sign of duration (':'), e.g. /a:/. Exclusively the BITS-SAM-PA signs are allowed, e.g. no /O:/ in “small talk”. Aberrations from the canonical duration are noted if a change of categories is perceived.

b) Aberrations of the vowel quality are noted if a change of categories is perceived.

c) If a diphthong clearly is perceived instead of a vowel, the segment can be labelled with one of the diphthongs /aI/, /OY/ or /aU/ instead of the vowel.

d) Whisper or voiceless parts are not marked in particular.

7. Diphthongs

a) Apart from the diphthongs /aI/, /OY/ and /aU/ sixteen different R-realisations are noted as diphthongs in the sentences.

b) Aberrations from the canonical form have to be noted. This is also true for R-realisations.

c) If an aberration in vowel quality is perceived it is noted solely if the segment can be labelled with another diphthong from the inventory. Otherwise the proposal given by MAUS has to be accepted. New combinations (e.g. /Ui:/) are not allowed.

Rough segmentation of the sentences

The principles and rules stay the same as in the exact segmentation. There is only one exception: the boundaries do not have to be placed at positive 0-crossings. With this exception a noticeable saving of time should be achieved. Zooming in PRAAT is no longer necessary and furthermore the placing of boundaries at positive 0-crossings is not necessary for a good speech synthesis.

Segmentation of the logatomes

I. Basic principles

The labelling of the diphones takes place by forced alignment on the basis of the canonical form. Only the segmentation of the diphone is given. The SAM-PA sings must not be changed. The rest of the logatome is out of interest and is not worked on.

II. Principles for segmentation

GS1

Within the logatomes only the accordant diphone is segmented. The rest of the logatome is out of interest. Beginning and end of the diphone (ahead of the first phoneme respectively after the last phoneme) are marked with '<p:>'.

GS 2

The borderline for segments are always set on positive 0-crossings in the oscillogram.

GS 3

The setting of the borderline should be controlled by sonagram and oscillogram.

GS 4

At periods where both of two neighbouring phonemes can be heard together the border is set in the middle of this period (Examples for this are fricative combinations /s-f/, /s-S/).

GS 5

Voiced (periodic) elements start with the first clear identifiable period.

GS 6

The border at signals with low intensity (especially /h/, aspiration) is set where the signal can be clearly distinguished from the background noise. To find out where exactly the border lies you have to zoom in the speech signal. The placing of the final border (e.g. aspirated plosives at the and of an utterance) results from the same principle. Noises of breathing - if recognised clearly - have to be cut off from the friction or aspiration.

GS7

If a smack (or a technical noise) occurs in a logatome there are two alternatives:

a) the smack (or a technical noise) is on the concerning diphone
In this case the segmentation is discarded. At the monitoring in the shell „defect“ is entered so that the logatome will be recorded again.

b) the smack (or a technical noise) is outside the diphone
In this case it can be ignored because within the logatomes only the diphone is important.

III. Handling of difficult cases

In the following typical difficult cases will be exemplified.

1. Plosives

a) All plosives (including glottal stop) are separated into two segments. The first segment contains the occlusion. The second segment contains the burst and possibly an aspiration. To distinguish the two segments they are labelled e.g. |t_s| and |t_b|, where 's' stands for 'silence' and 'b' stands for 'burst'.

b) The borderline of plosives at the beginning of an utterance gets an occlusion arbitrary set at 20 - 40ms.

c) After pauses plosives are treated like plosives at the beginning of an utterance.

d) The occlusion of a voiced plosive with voicing lead in between vowels starts after the last identifiable period of the vowel. The occlusion can be recognised by a break-in of the energy of the higher formants and in a damped sinus like signal.

e) Plosives at the end of an utterance end with the burst respectively after decay of the aspiration (see signal). Possible breathing noise has to be cut off from the segment.

f) After nasals the start of voiced plosives (activity of the velum) often can not be identified clearly. In this case the decreasing phase of the nasal is counted for the occlusion. Often the burst can just be noticed as a irregularity in the following period. This is counted for the plosive, too.

2. Affricates

Affricates (ts, tS, pf) are treated as one phoneme. They are divided into two segments: the first segment is the phase of occlusion, the second segment contains burst and fricative, e.g. |pf_s| and |pf_b|.

3. Fricatives

If two fricatives with the same point of articulation follow each other (e.g. 'auffallen') two segments are transcribed solely if they are clearly distinguishable.

4. R-Realisations

The symbol /R/ stands for:

uvular trill
alveolar trill
uvular fricative (voiced/voiceless)
velar fricative.

5. Vowels

a) Long vowels get the sign of duration ':'. Exclusively the signs of the BITS-SAM-PA list are allowed! e.g. no /A:/.

b) Aberrations of vowel quality in logatomes are not accepted. The prompt has to be recorded again.

c) Whisper or voiceless parts in logatomes are not segmented. The prompt has to be recorded again.

SAM-PA-list of all used signs and examples:

SAM-PA-sign	**e.g. orthographically**	**e.g. transcribed**
vowels:		
I	Sitz	zits
E	Gesetz	g@zEts

SAM-PA-sign	e.g. orthographically	e.g. transcribed
a	Satz	zats
O	Trotz	tROts
U	Schutz	SUts
Y	hübsch	hYpS
9	plötzlich	pl9tslIC
i:	Lied	li:t
e:	Beet	be:t
E:	spät	SpE:t
a:	Tat	ta:t
o:	rot	Ro:t
u:	Blut	blu:t
y:	süß	zy:s
2:	blöd	bl2:t
diphthongs:		
aI	Eis	aIs
aU	Haus	haUs
OY	Kreuz	kROYts
unstressed „schwa“ vowels:		
@	bitte	bIt@
6	besser	bEs6
glottal stop:		
Q	Verein	fE6QaIn
consonants:		
p	Pein	paIn
b	Bein	baIn
t	Teich	taIC
d	Deich	daIC
k	Kunst	kUnst
g	Gunst	gUnst
f	fast	fast
v	was	vas
s	Tasse	tas@

SAM-PA-sign	e.g. orthographically	e.g. transcribed
z	Hase	ha:z@
S	waschen	vaS@n
Z	Genie	Ze:ni:
C	sicher	zIC6
j	Jahr	ja:6
x	Buch	bu:x
h	Hand	hant
m	mein	maIn
n	nein	naIn
N	Ding	dIN
l	Leim	laIm
R	Reim	RaIm
affricates:		
pf	Pfahl	pfa:l
ts	Zahl	tsa:l
tS	deutsch	dOYtS
additional english phonemes:		
EI	raise	rEIz
@U	nose	n@Uz
T	thin	TIn
D	this	DIs
r	wrong	rON
L	long	LON
w	wasp	wOsp
additional french phonemes:		
E~	vin	vE~
a~	vent	va~
o~	bon	bo~
6-phoneme combinations:		
6	besser	bEs6
i:6	Tier	ti:6
I6	Wirt	vI6t

SAM-PA-sign	e.g. orthographically	e.g. transcribed
y:6	Tür	ty:6
Y6	Türke	TY6k@
e:6	schwer	Sve:6
E6	Berg	bE6k
E:6	Bär	bE:6
2:6	Föhr	f2:6
96	Wörter	v96t6
a:6	Haar	ha:6
a6	hart	ha6t
u:6	Kur	ku:6
U6	kurz	kU6ts
o:6	Ohr	o:6
O6	dort	dO6t
special character:		
*	*for silence previous of after a phoneme (in the beginning resp. after a logatome)*	