Using the BOSS label format for corpus annotation

Stefan Breuer
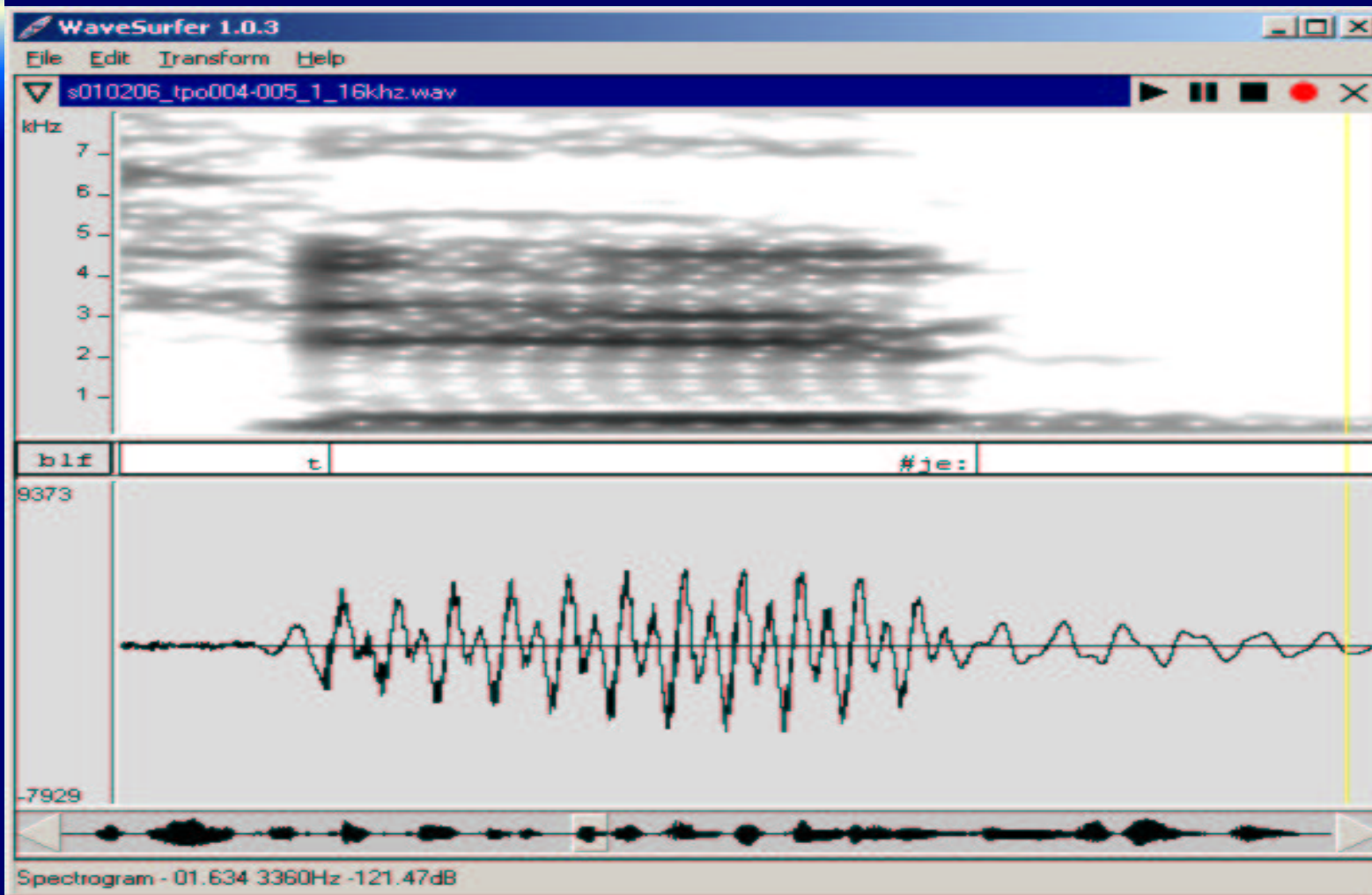
IKP

University of Bonn

Motivation

- Certain sounds, especially pre–vocalic glides are hard to separate from neighbouring phones in the spectrogram

- Forcing a segmentation at these points may result in units that are unsuitable for speech synthesis (based on phone concatenation).

Example: [j]+[e:]

The (simple) solution

- No attempt at segmentation is made at the critical unit 'boundaries':

- Consonants that have proven to be problematic are joined with succeeding vowels and diphthongs to form new units.

The resulting units

j		i:	
v / ?		?	
l	**+**	a:	**=**
?		?	
?/?/r/?		a?	
h		?	
...		...	

In SAMPA notation:

ji:, vI, la:, ?O, raU, hE...

8 onsets x 27 vowels/diphthongs = 216 units

Coda–/r/

- All variants of coda–/r/ are represented by [6]

- They are joined with preceding vowels or combined consonant–vowel labels to form new units:

$$ja: + 6 = ja:6$$

This yields an additional 8 x 22 (monophthongs only) = 176 units

Syllabic nasals and laterals

- /ʔn/, /ʔm/, and /ʔl/ morphemes are mostly realised as syllabic consonants but can also contain residual vowel articulations.

- Each of these forms a label of its own, grouping together the different realisations.

Advantages of BLF

- Faster manual correction or placement of segment boundaries
- Fewer points of concatenation –> fewer distortions –> better quality

but...

Drawbacks

- significantly greater number of unit classes (357 additional types) and possible 'triphone' contexts –> larger corpora / too few instances in the corpus

- difficulties in the adjustment of syllable structure during manual segmentation

- limited usability of corpora for phonetic research(?)

- fewer training instances for automatic prosodic prediction etc.

Workarounds for unit selection synthesis

- unit selection in BOSS makes use of context classes during pre–selection of units. Example:

- An instance of [Ia:] in the sequence [Ia:**Ia:**Ia:] is to be selected from the corpus for the second [Ia:].

- The context [Ia:] is mapped to different symbols, depending on its position relative to the unit in question:

Symbol	As right context	As left context
Ia:	I	a:

Context classes

Thus, the pre−selection mechanism searches for the following triphone context:

Left context	Symbol	Right context
a:	Ia:	I

Apart from simply splitting up the symbols into their constituent sounds, the context class mechanism can also be used to define other groupings, such as [m]/[n] for a class NASALS.

Summary

- The BOSS label format provides a way to avoid certain issues that occur during corpus segmentation and also speeds up the process of labelling

- This comes at the cost of a greater amount of unit types, which must be taken into account when designing a corpus.