

Lexical Stress in Speech Recognition

Master's thesis presentation

Rogier van Dalen

13th May 2005

Topics

- Objective
- What is lexical stress?
- Properties of lexical stress
- Model
- System
- Results

Objective

Can lexical stress be used in a speech recogniser to make it perform better?

Objective

Can lexical stress be used in a speech recogniser to make it perform better?

- Find properties of lexical stress

Objective

Can lexical stress be used in a speech recogniser to make it perform better?

- Find properties of lexical stress
- Model speech recogniser

Objective

Can lexical stress be used in a speech recogniser to make it perform better?

- Find properties of lexical stress
- Model speech recogniser
- Implement speech recogniser

Objective

Can lexical stress be used in a speech recogniser to make it perform better?

- Find properties of lexical stress
- Model speech recogniser
- Implement speech recogniser
- Test speech recogniser

Garden-variety speech recognition

Input modelled as a concatenation of phonemes

What is lexical stress?

/hɒːrɪkdaːrənkaːnɒn/

What is lexical stress?

/ho:ɾɪkda:ɾənka:nɔn/

Hoor ik daar een kanon?

What is lexical stress?

/ho:ɾɪkda:rənka:nɔn/

Hoor ik daar een kanon?

kanón 'gun' or *kánon* 'song'?

Use of lexical stress

- Minimal pairs

Use of lexical stress

- Minimal pairs
(a) subject – (to) subject

Use of lexical stress

- Minimal pairs
(a) subject – *(to) subject*
Du. *aanbod* ‘offer’ – *aan bod* ‘first in line’

Use of lexical stress

- Minimal pairs
(a) subject – (to) subject
Du. *aanbod* ‘offer’ – *aan bod* ‘first in line’
Du. *voorkomen* ‘prevent’ – *voorkomen*
‘happen’

Use of lexical stress

- Minimal pairs

(a) subject – (to) subject

Du. *aanbod* ‘offer’ – *aan bod* ‘first in line’

Du. *voorkomen* ‘prevent’ – *voorkomen*
‘happen’

Portuguese *falara* ‘I had spoken’ – *falará* ‘he
will speak’

Use of lexical stress

Use of lexical stress

- Word recognition

Use of lexical stress

- Word recognition
Du. *october* – *octopus*

Use of lexical stress

- Word recognition
Du. *october* – *octopus*
tigress – *digress*

Use of lexical stress

Use of lexical stress

- Segmentation

Use of lexical stress

- Segmentation
conduct ascends uphill

Use of lexical stress

- Segmentation
conduct ascends uphill
'a doctor sends a pill'?

Properties of lexical stress

Properties of lexical stress

- Stress works on the syllable level

A lexicon

<i>a</i>	/eɪ/
<i>are</i>	/ɑː/
<i>the</i>	/ðɪː/
<i>garden</i>	/gɑːdən/
<i>ordinary</i>	/ɔːdɪnəri/
<i>table</i>	/teɪbəl/
<i>variety</i>	/vəraɪəti/

A lexicon

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	
<i>the</i>	/ðɪː/	
<i>garden</i>	/gɑːdən/	
<i>ordinary</i>	/ɔːdɪnəri/	
<i>table</i>	/teɪbəl/	
<i>variety</i>	/vəraɪəti/	

A lexicon

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðiː/	
<i>garden</i>	/gɑːdən/	
<i>ordinary</i>	/ɔːdɪnəri/	
<i>table</i>	/teɪbəl/	
<i>variety</i>	/vəraɪəti/	

A lexicon

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðɪː/	[ðə]
<i>garden</i>	/gɑːdən/	
<i>ordinary</i>	/ɔːdɪnəri/	
<i>table</i>	/teɪbəl/	
<i>variety</i>	/vəraɪəti/	

A lexicon with stress marks

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðɪː/	[ðə]
<i>garden</i>	/'gɑːdən/	
<i>ordinary</i>	/ɔːdɪnəri/	
<i>table</i>	/teɪbəl/	
<i>variety</i>	/vəraɪəti/	

A lexicon with stress marks

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðɪː/	[ðə]
<i>garden</i>	/'gɑːdən/	['gɑːdŋ]
<i>ordinary</i>	/ɔːdɪnəri/	
<i>table</i>	/teɪbəl/	
<i>variety</i>	/vəraɪəti/	

A lexicon with stress marks

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðɪː/	[ðə]
<i>garden</i>	/'gɑːdən/	['gɑːdŋ]
<i>ordinary</i>	/'ɔːdɪnəri/	
<i>table</i>	/teɪbəl/	
<i>variety</i>	/vəraɪəti/	

A lexicon with stress marks

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðɪː/	[ðə]
<i>garden</i>	/'gɑːdən/	['gɑːdŋ]
<i>ordinary</i>	/'ɔːdɪnəri/	['ɔːdŋri]
<i>table</i>	/teɪbəl/	
<i>variety</i>	/vəraɪəti/	

A lexicon with stress marks

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðɪː/	[ðə]
<i>garden</i>	/'gɑːdən/	['gɑːdŋ]
<i>ordinary</i>	/'ɔːdɪnəri/	['ɔːdŋri]
<i>table</i>	/'teɪbəl/	
<i>variety</i>	/vəraɪəti/	

A lexicon with stress marks

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðɪː/	[ðə]
<i>garden</i>	/'gɑːdən/	['gɑːdŋ]
<i>ordinary</i>	/'ɔːdɪnəri/	['ɔːdŋri]
<i>table</i>	/'teɪbəl/	['t ^h eɪb]
<i>variety</i>	/vəraɪəti/	

A lexicon with stress marks

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðɪː/	[ðə]
<i>garden</i>	/'gɑːdən/	['gɑːdŋ̩]
<i>ordinary</i>	/'ɔːdɪnəri/	['ɔːdŋ̩ri]
<i>table</i>	/'teɪbəl/	['t ^h eɪb̩l]
<i>variety</i>	/və'raɪəti/	

A lexicon with stress marks

<i>a</i>	/eɪ/	[ə]
<i>are</i>	/ɑː/	[ə]
<i>the</i>	/ðɪː/	[ðə]
<i>garden</i>	/'gɑːdən/	['gɑːdŋ̩]
<i>ordinary</i>	/'ɔːdɪnəri/	['ɔːdŋ̩ri]
<i>table</i>	/'teɪbəl/	['t ^h eɪb̩l]
<i>variety</i>	/və'raɪəti/	[v̩'raɪəti]

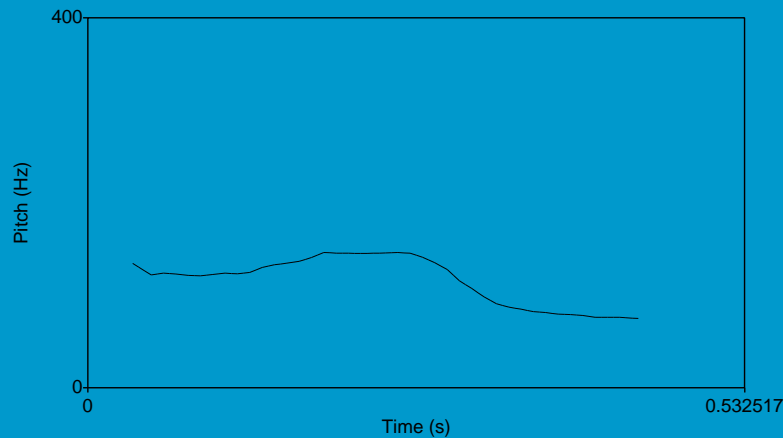
Properties of lexical stress

- Stress works on the syllable level

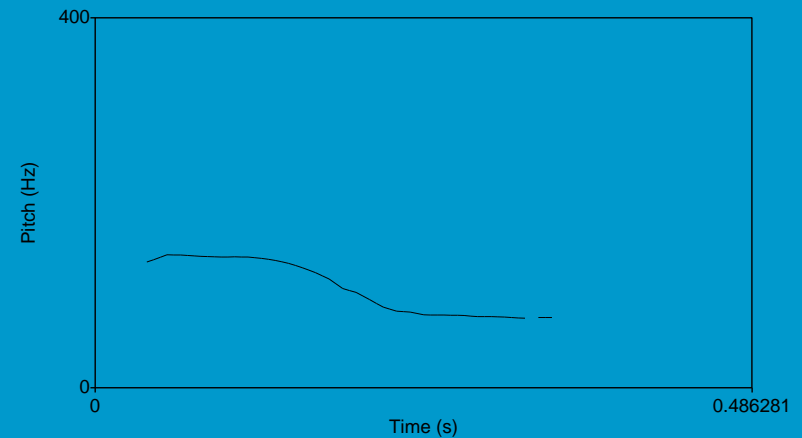
Properties of lexical stress

- Stress works on the syllable level
- Unstressed syllables are reduced

/ka:'nɔn/ 'gun' or /'ka:nɔn/ 'song'?

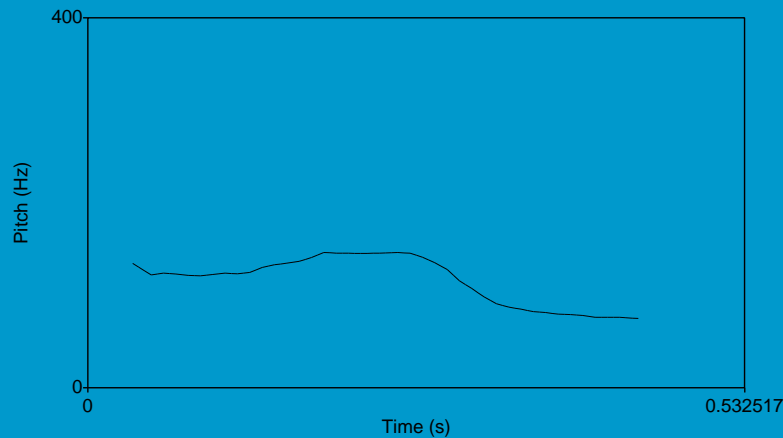


|k| a: | n | ɔ | n |

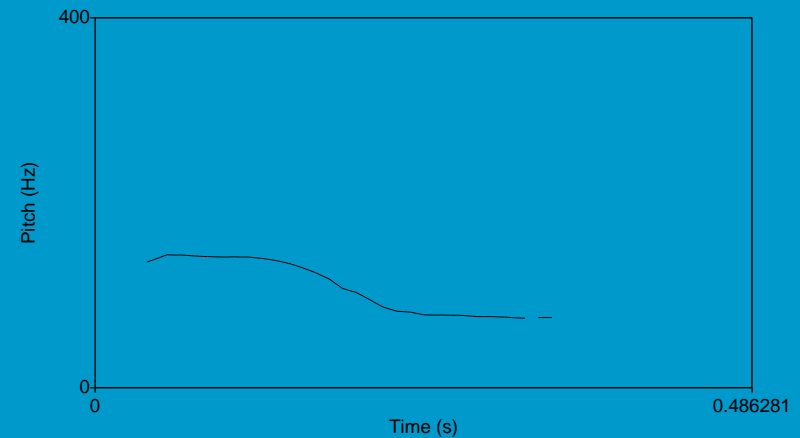


|k| a: | n|ɔ | n |

/ka:'nɔn/ 'gun' or /'ka:nɔn/ 'song'?

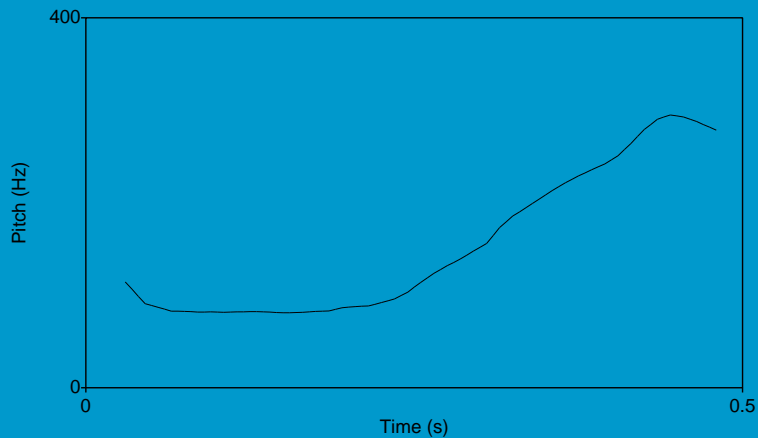


|k| a: | n | ɔ | n | |
kanón /ka:'non/ 'gun'

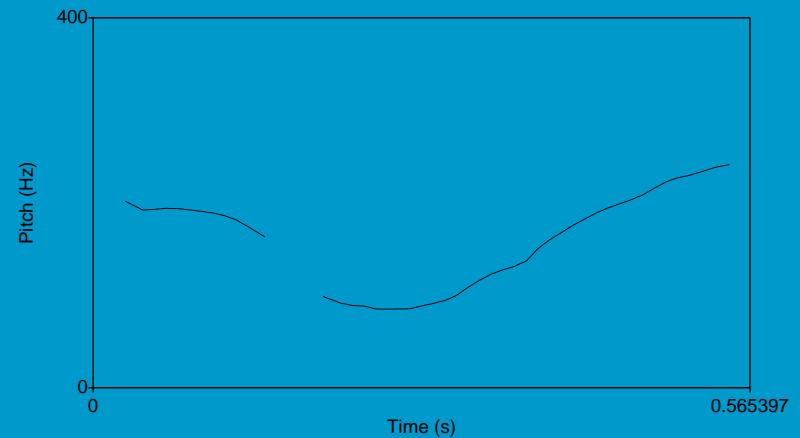


|k| a: | n | ɔ | n | |
kánon /'ka:nɔn/ 'song'

/ˈkɑːnɔn?/ ‘song?’ or /kɑːˈnɔn?/ ‘gun?’?

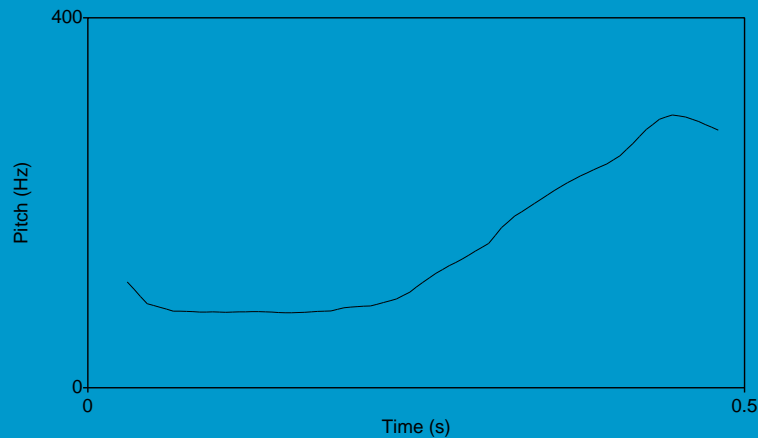


|k| aː | n | ɔ | n |

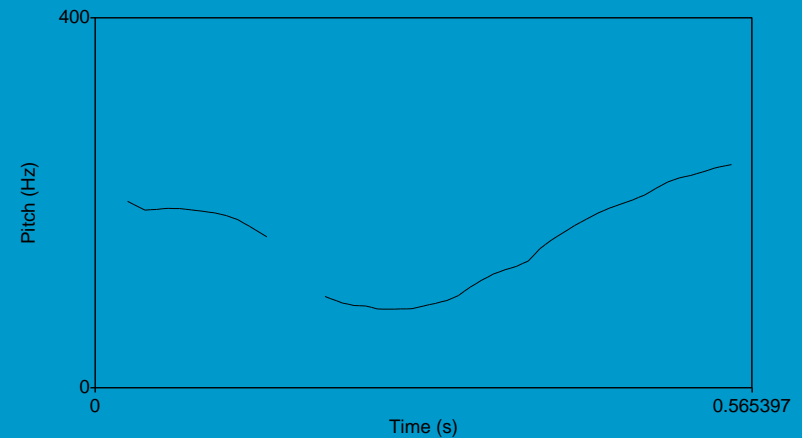


|k| aː | n | ɔ | n |

/'ka:nɔn? / 'song?' or /ka:'nɔn? / 'gun'??

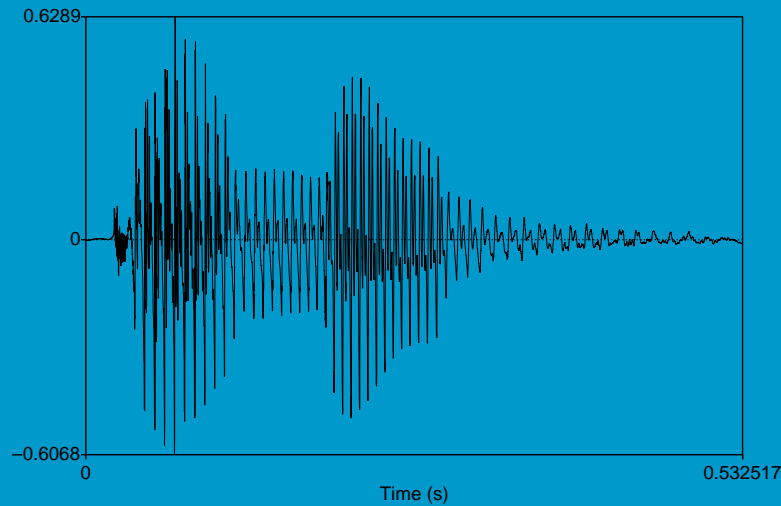


|k| a: | n | ɔ | n | |
kanón /ka:'non/ 'gun'

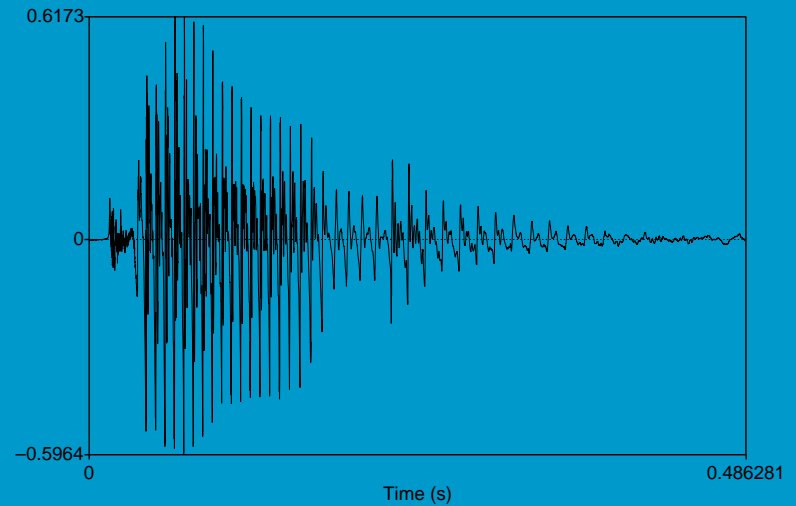


|k| a: | n | ɔ | n | |
kánon /'ka:nɔn/ 'song'

$/\text{'ka:n}\text{ɔn}/$ 'song' or $/\text{ka:}'\text{n}\text{ɔn}/$ 'gun'?

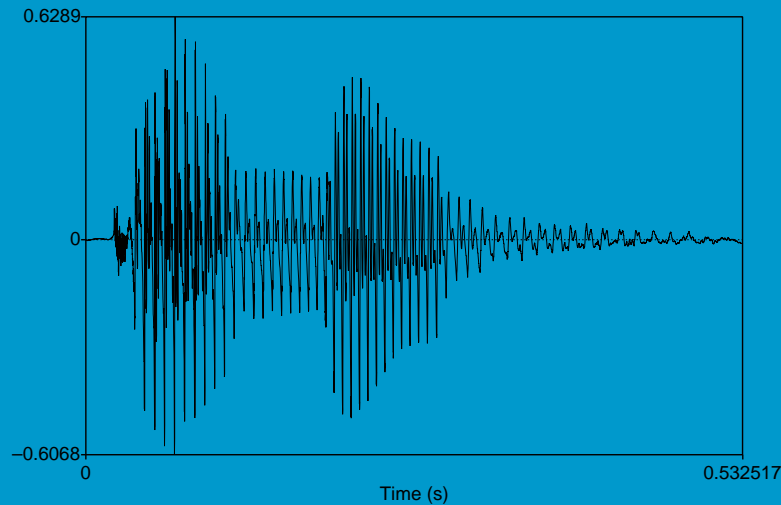


|k| a: | n | ɔ | n |

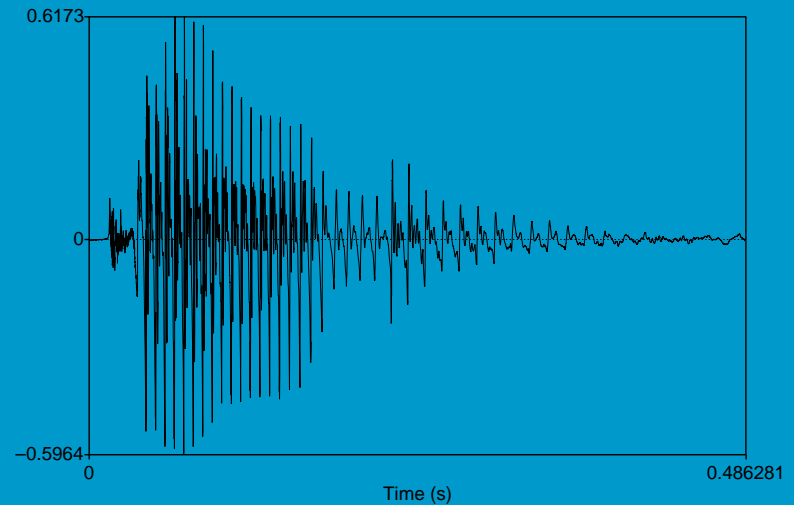


|k| a: | n|ɔ | n |

/'ka:nɔn/ 'song' or /ka:'nɔn/ 'gun'?



|k| a: | n | ɔ | n | |
kanón /ka:'non/ 'gun'



|k| a: | n | ɔ | n | |
kánon /'ka:nɔn/ 'song'

Properties of lexical stress

- Stress works on the syllable level
- Unstressed syllables are reduced

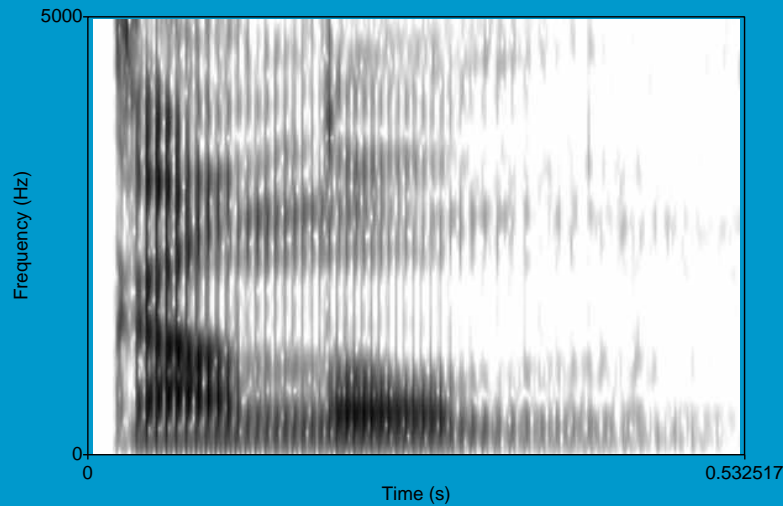
Properties of lexical stress

- Stress works on the syllable level
- Unstressed syllables are reduced
- Stressed syllables have longer durations

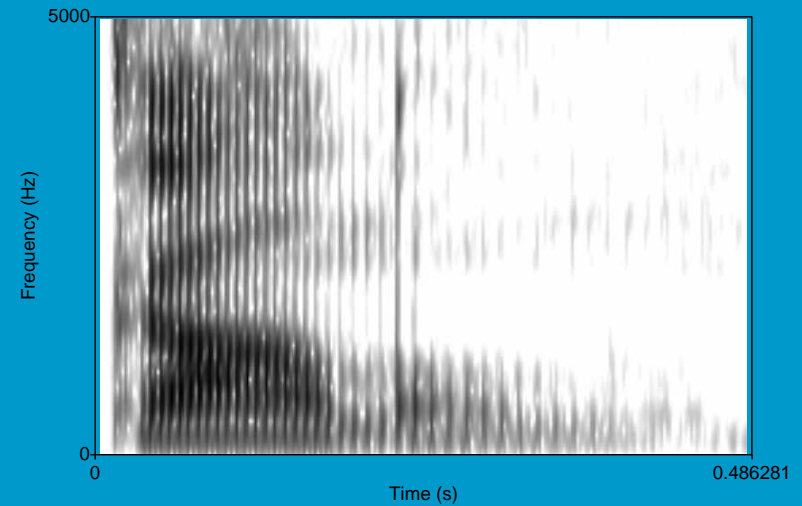
Properties of lexical stress

- Stress works on the syllable level
- Unstressed syllables are reduced
- Stressed syllables have longer durations
- Stressed syllables are louder

/ka:'nɔn/ 'gun' or /'ka:nɔn/ 'song'?

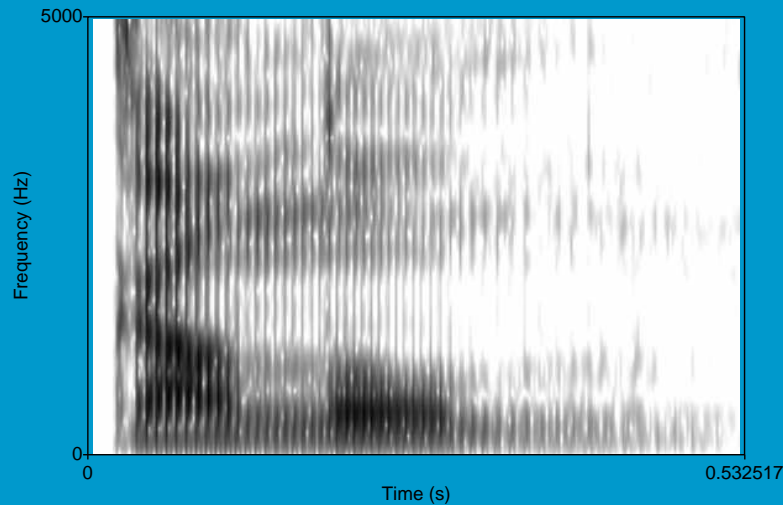


|k| a: | n | ɔ | n |

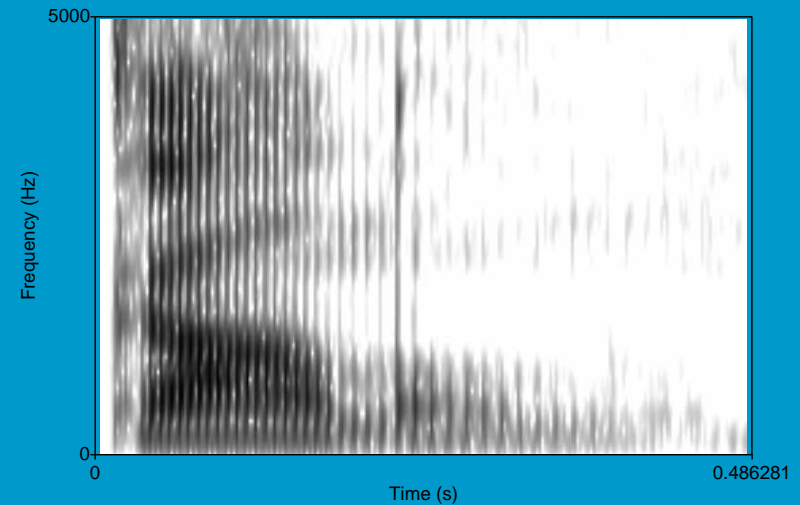


|k| a: | n|ɔ | n |

/ka:'nɔn/ 'gun' or /'ka:nɔn/ 'song'?

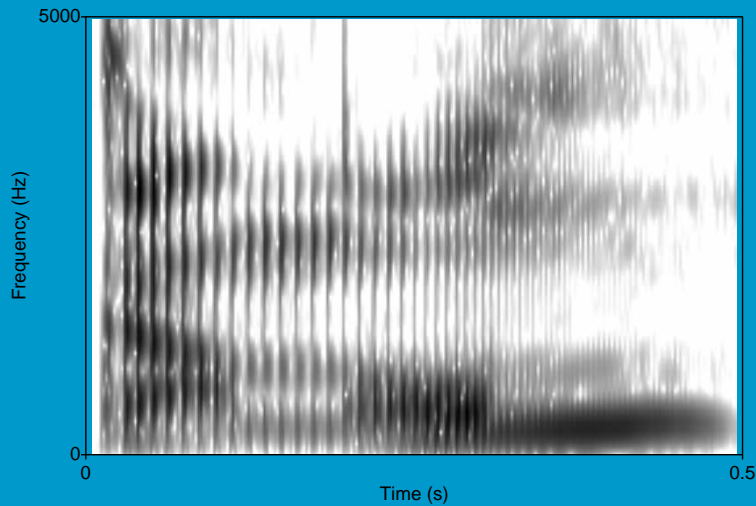


|k| a: | n | ɔ | n |
kanón /ka:'non/ 'gun'

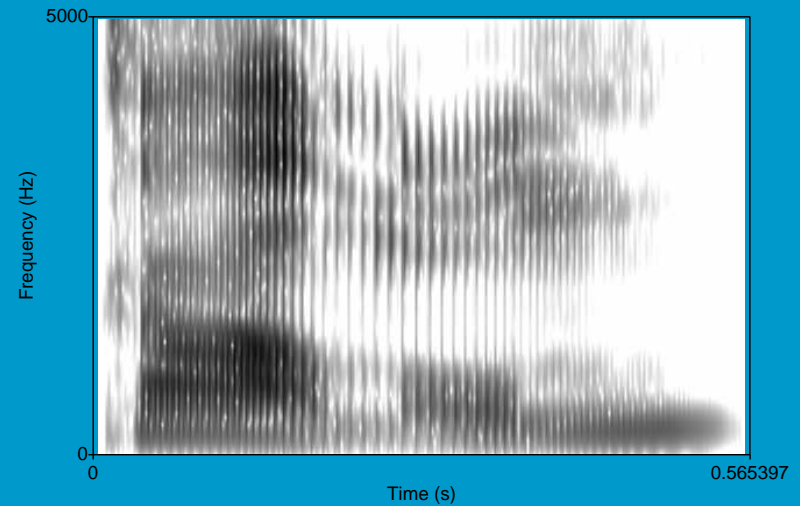


|k| a: | n | ɔ | n |
kánon /'ka:nɔn/ 'song'

/ˈkɑːnɔn?/ ‘song?’ or /kɑːˈnɔn?/ ‘gun?’?

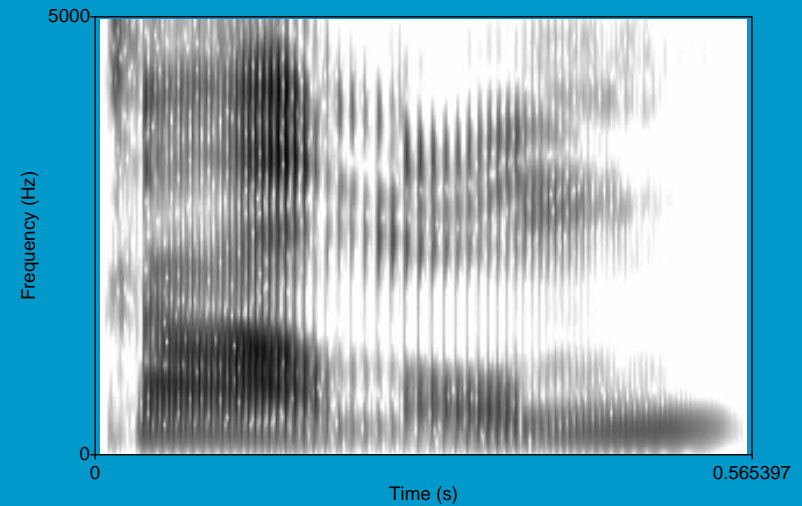
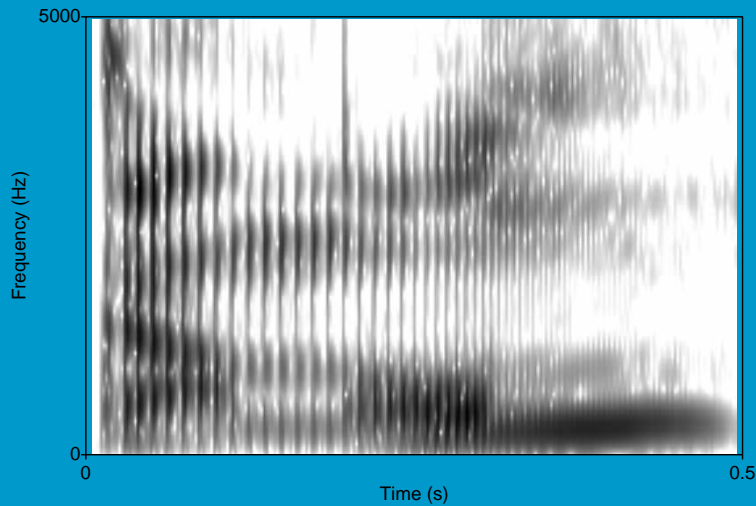


|k| aː | n | ɔ | n |



|k| aː | n | ɔ | n |

***/'ka:nɔn?* / ‘song?’ or */ka:'nɔn?* / ‘gun?’?**



ka | n | ɔ | n |
kanón /ka:'non/ ‘gun’

k | a | n | ɔ | n |
kánon /'ka:nɔn/ ‘song’

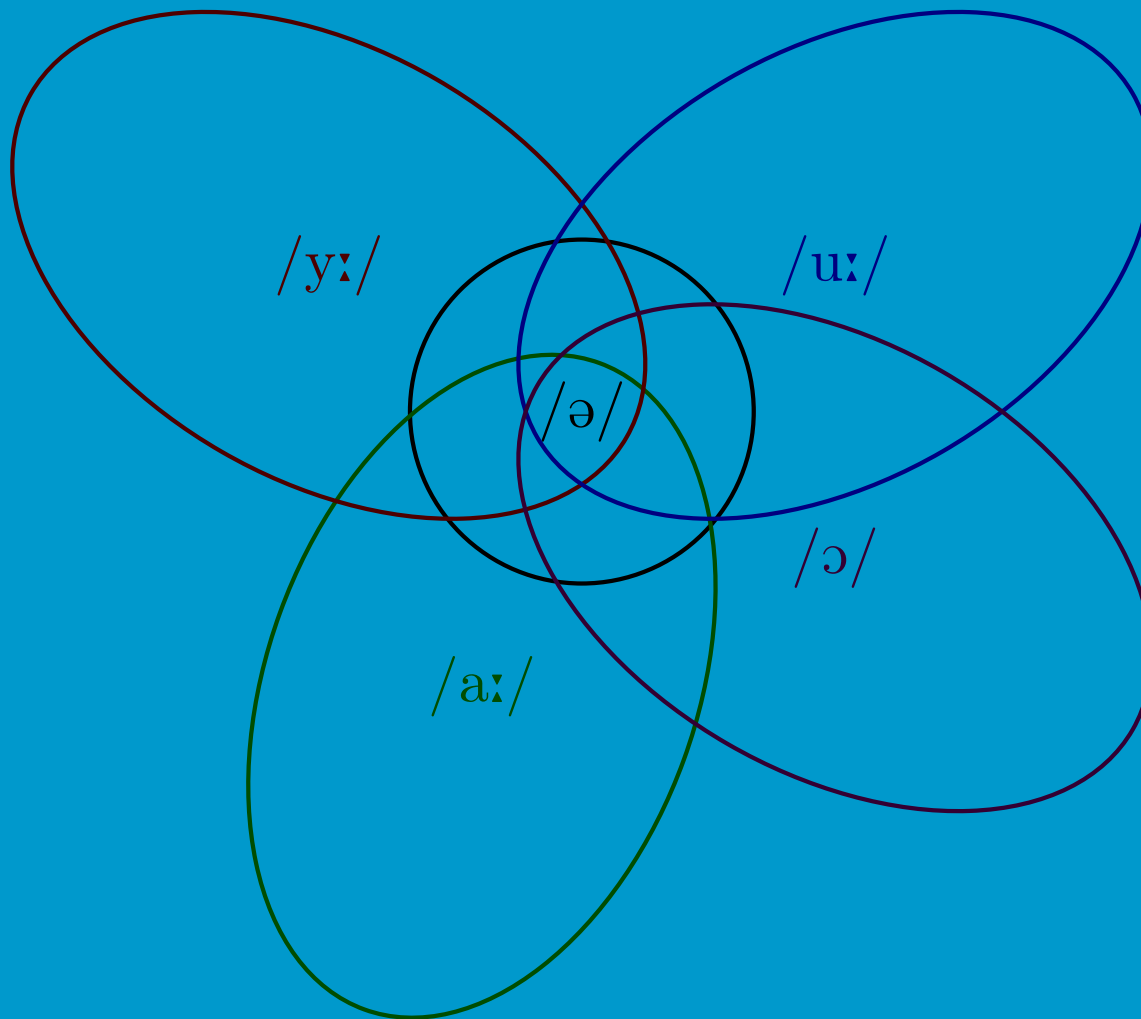
Properties of lexical stress

- Stress works on the syllable level
- Unstressed syllables are reduced
- Stressed syllables have longer durations
- Stressed syllables are louder

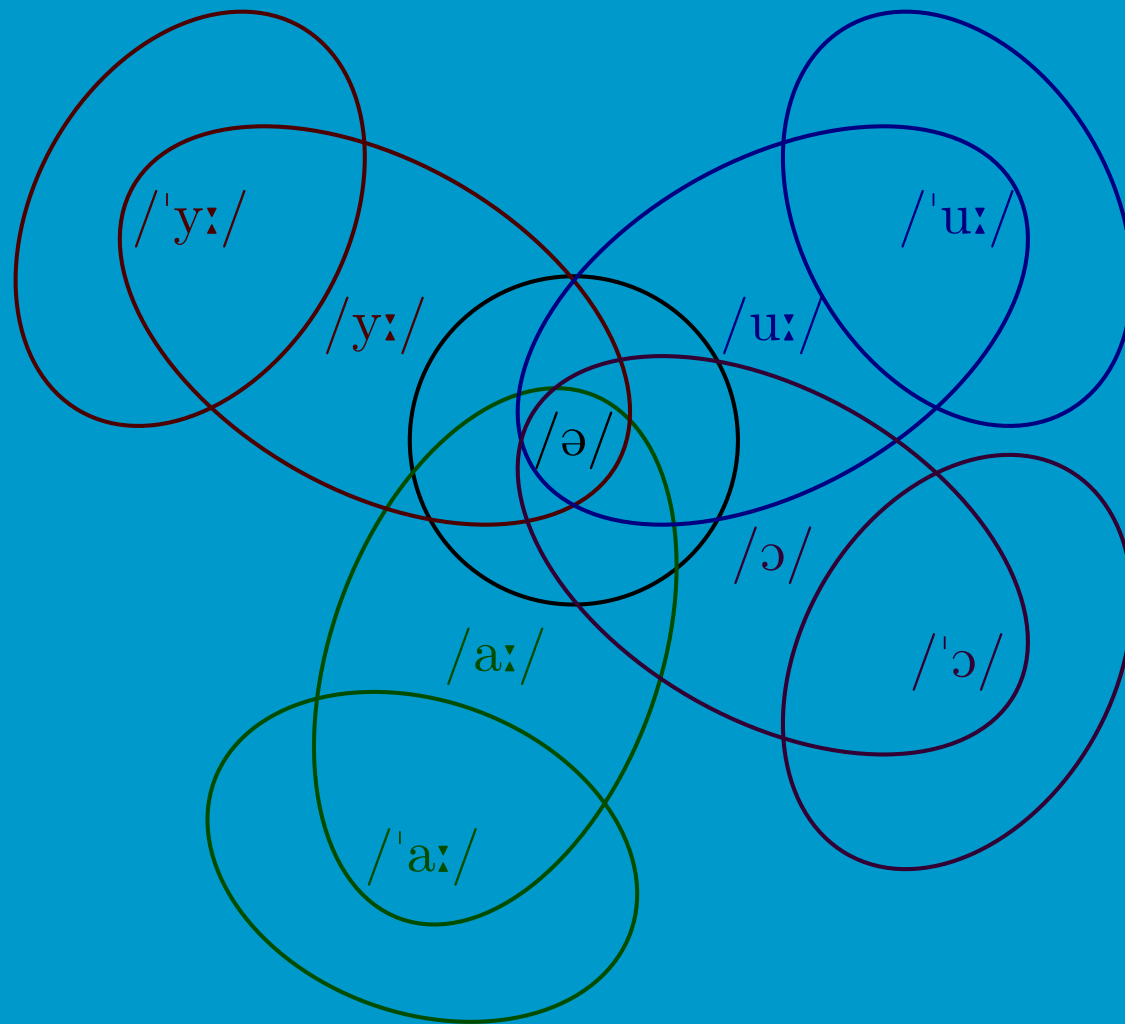
Properties of lexical stress

- Stress works on the syllable level
- Unstressed syllables are reduced
- Stressed syllables have longer durations
- Stressed syllables are louder
- Stressed syllables have more high frequencies

Distinguishing phonemes



Distinguishing phonemes



Integration in a speech recogniser

- /ɑ a: p t ɔ v u/

Integration in a speech recogniser

- /ɑ aː p t ɔ v u/
- Stressed and unstressed versions of phonemes
/ɑ 'ɑ aː 'aː p 'p t 't ɔ 'ɔ v 'v u 'u/

Integration in the lexicon

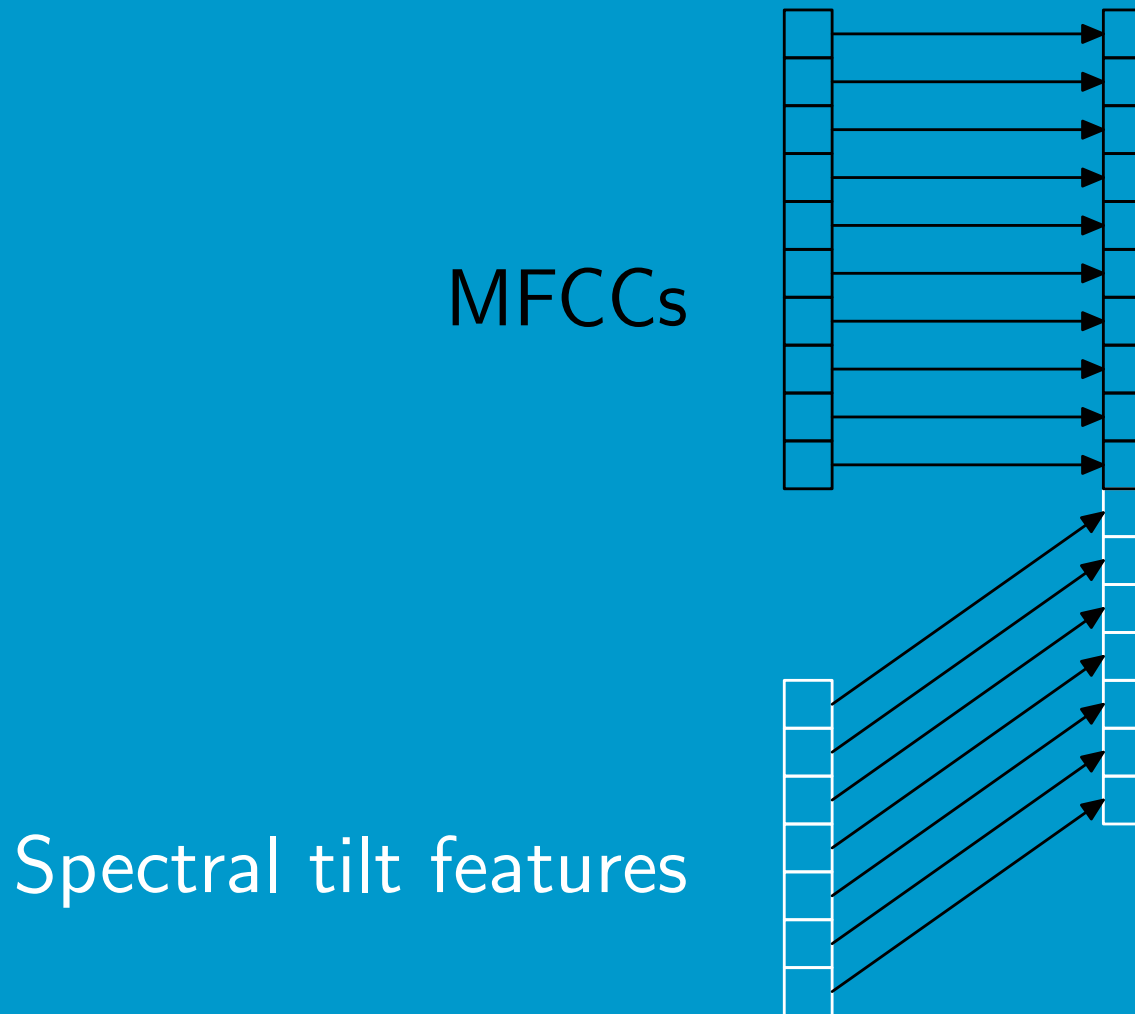
<i>are</i>	ə
<i>a</i>	ə
<i>the</i>	ð ə
<i>garden</i>	gɑːdn
<i>ordinary</i>	ɔːdnri
<i>table</i>	teɪbl
<i>variety</i>	vəriəti

Integration in the lexicon

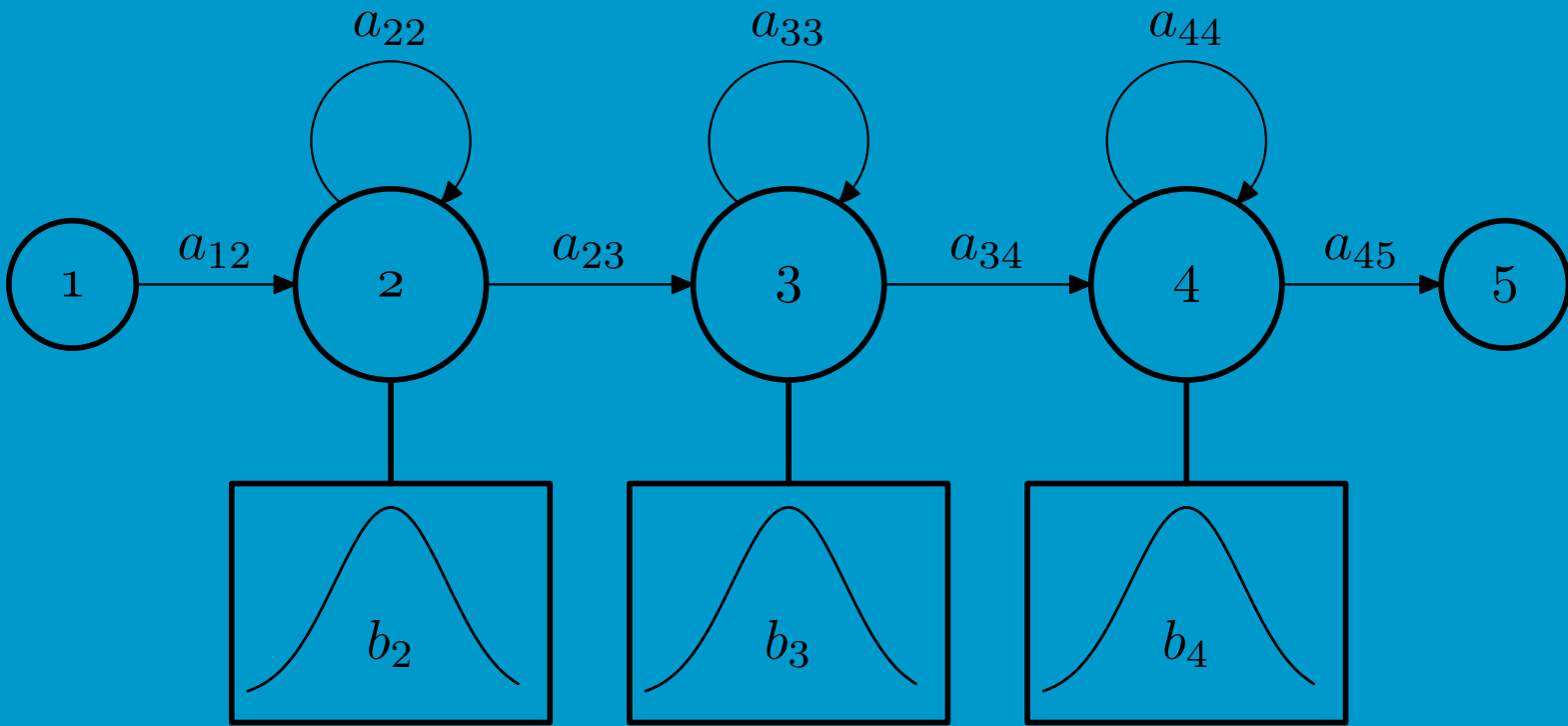
<i>are</i>	ə
<i>a</i>	ə
<i>the</i>	ð ə
<i>garden</i>	'g 'ɑː d n
<i>ordinary</i>	'ɔː d n r i
<i>table</i>	't 'eɪ b l
<i>variety</i>	v 'r 'aɪ ə t i

Integration in feature vectors

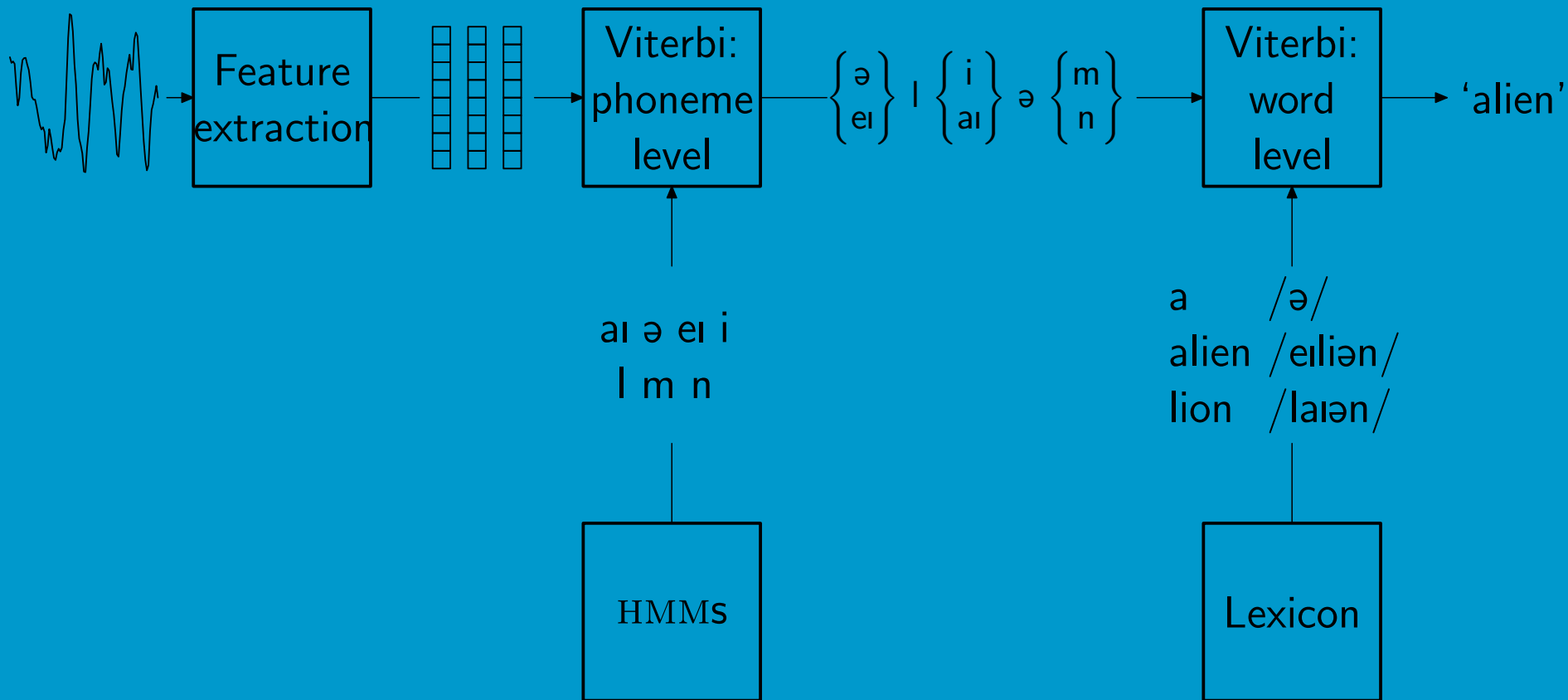
Integration in feature vectors



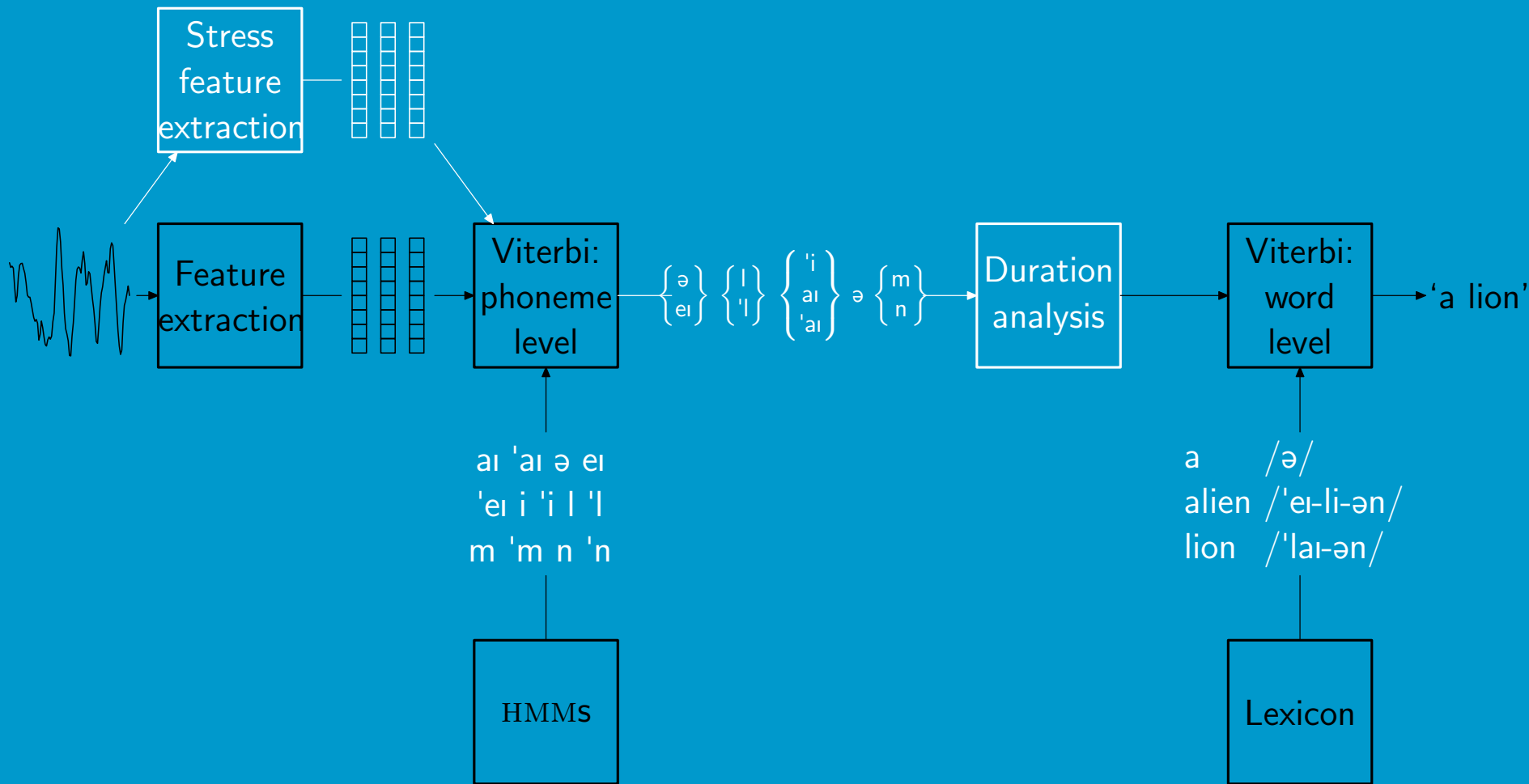
Modelling duration



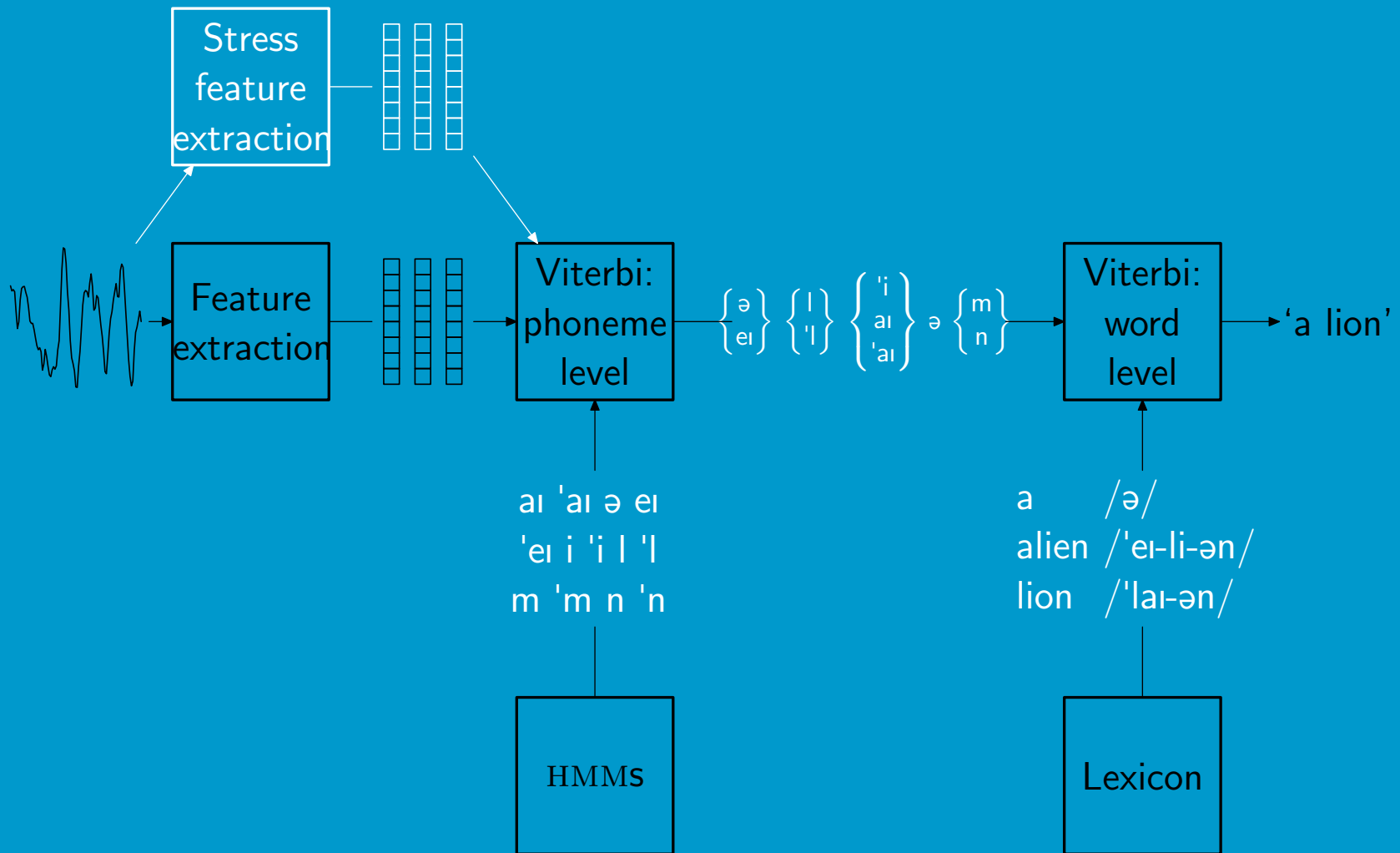
Model — baseline



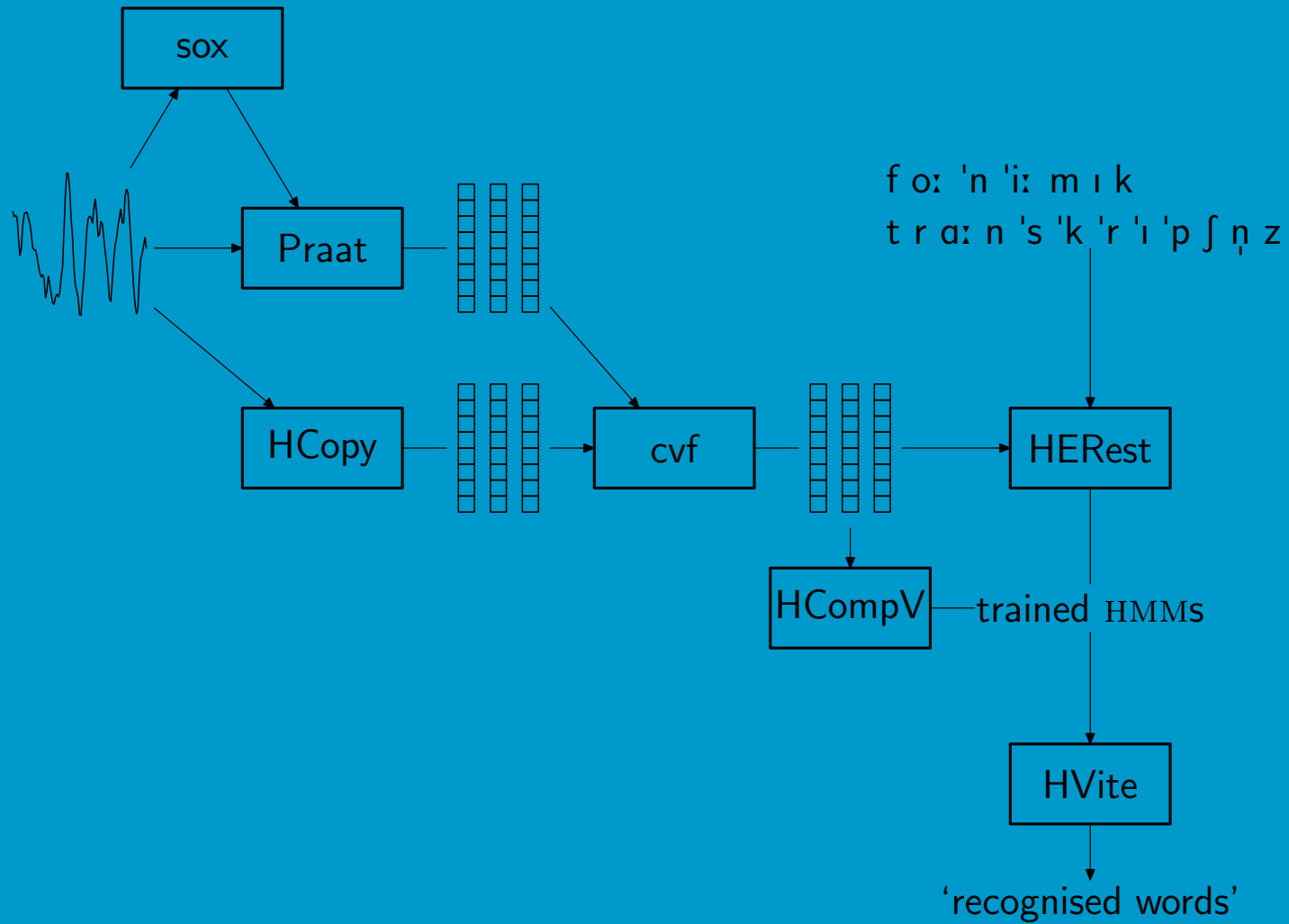
Model — stress-enabled



Model — implemented



System



System

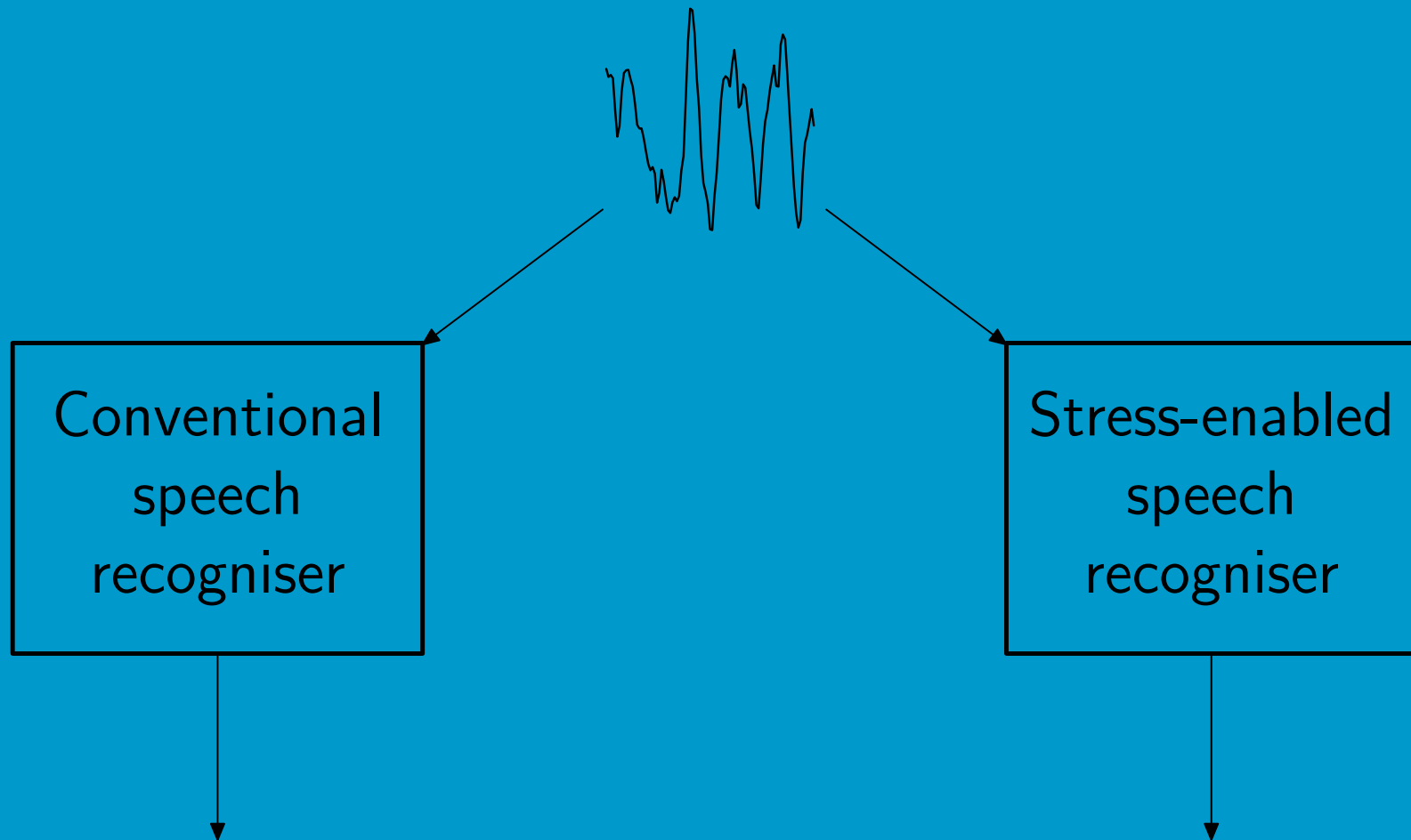
- Hidden Markov Toolkit
- Corpus Gesproken Nederlands
- 772 recordings
- 54 842 files
- 775 034 words
- 53 hours

Time

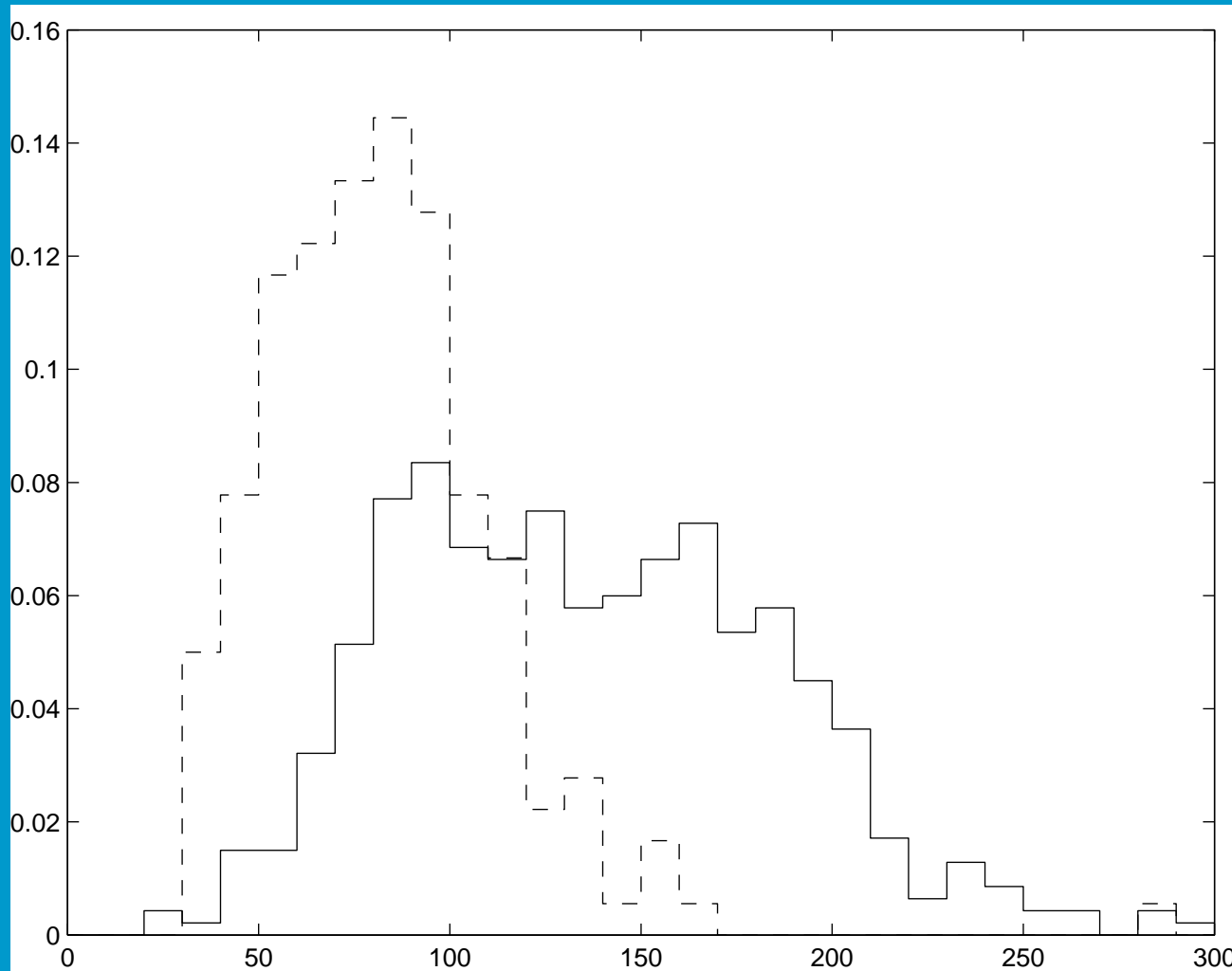
Using 6 to 8 computers:

- Training: 1 – 8 hours per iteration
- Evaluation: 4 – 10 hours per iteration
- 60 training iterations for 2 recognisers

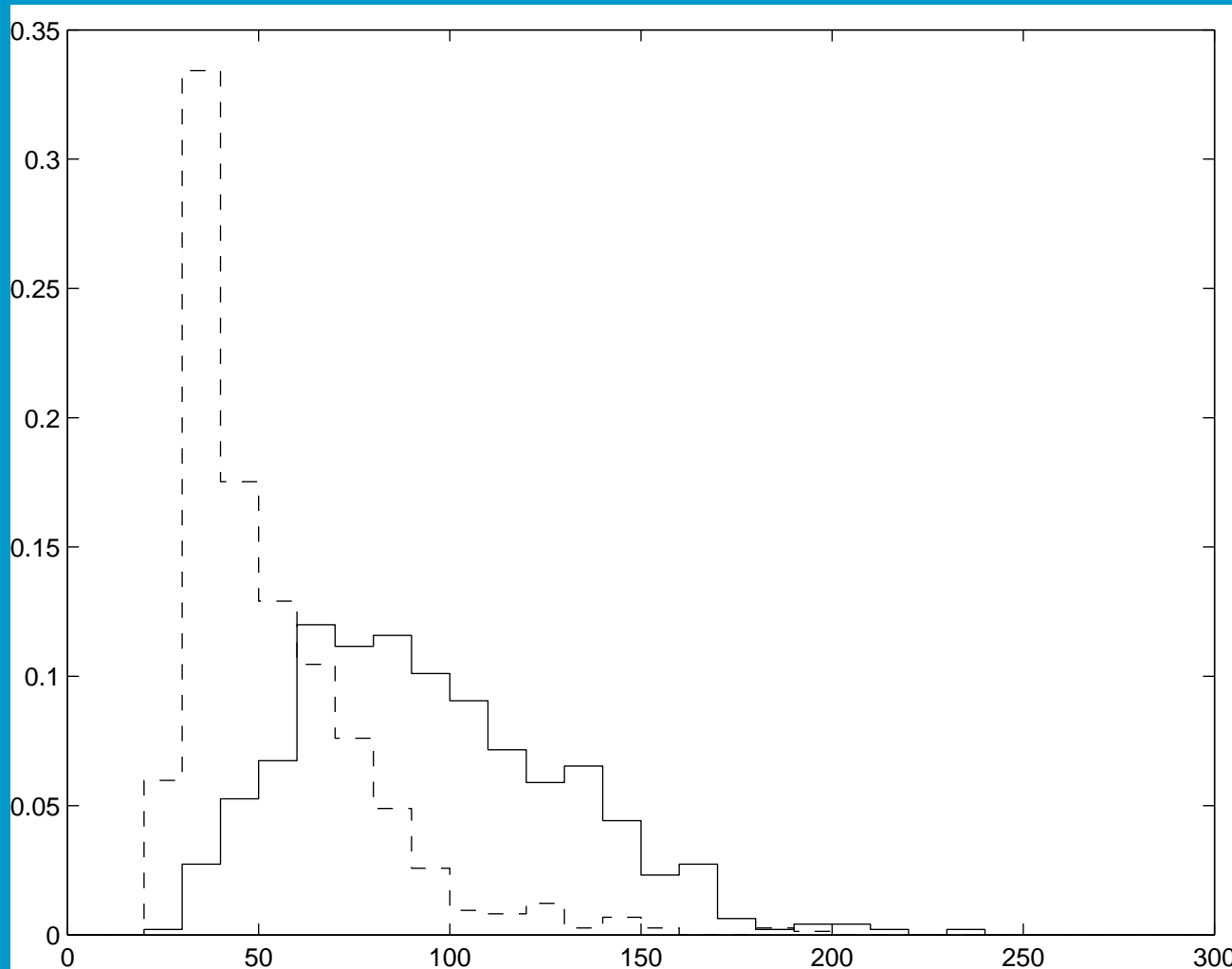
Experimental set-up



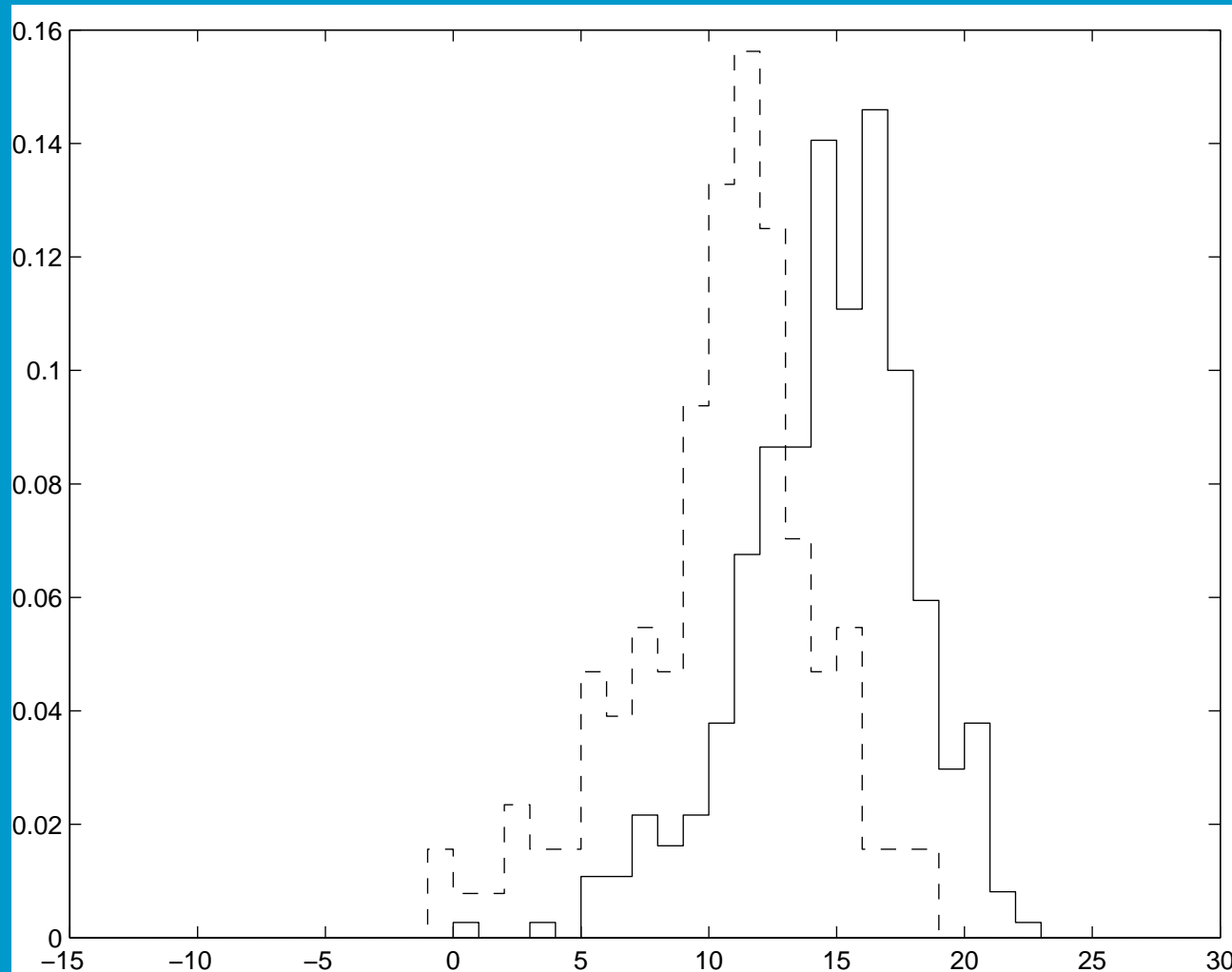
Results — duration /'i:/-/'i:/



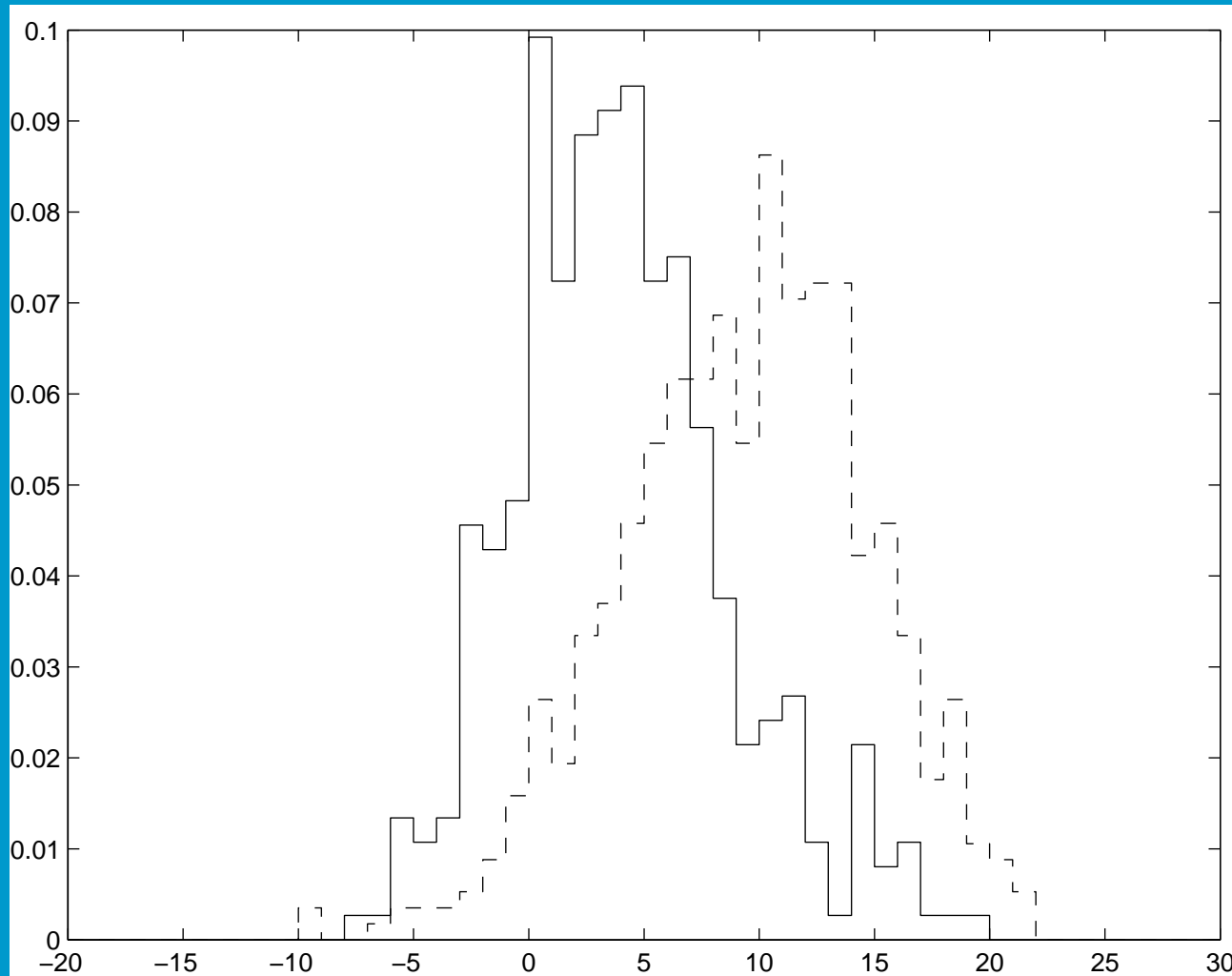
Results — duration /'n/-/n/



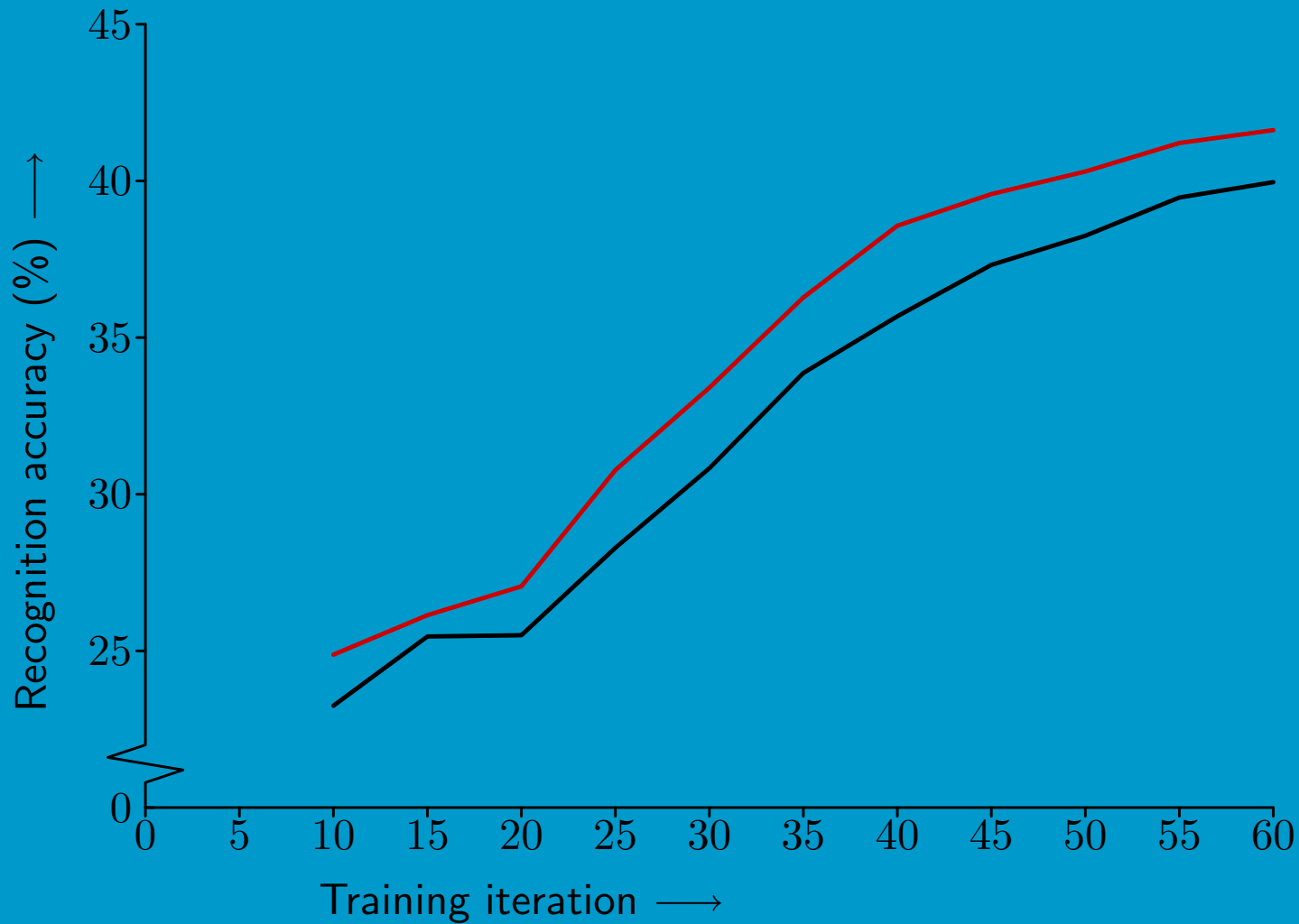
Results — spectral tilt /'a:/-/a:/'



Results — spectral tilt /'d/-/d/

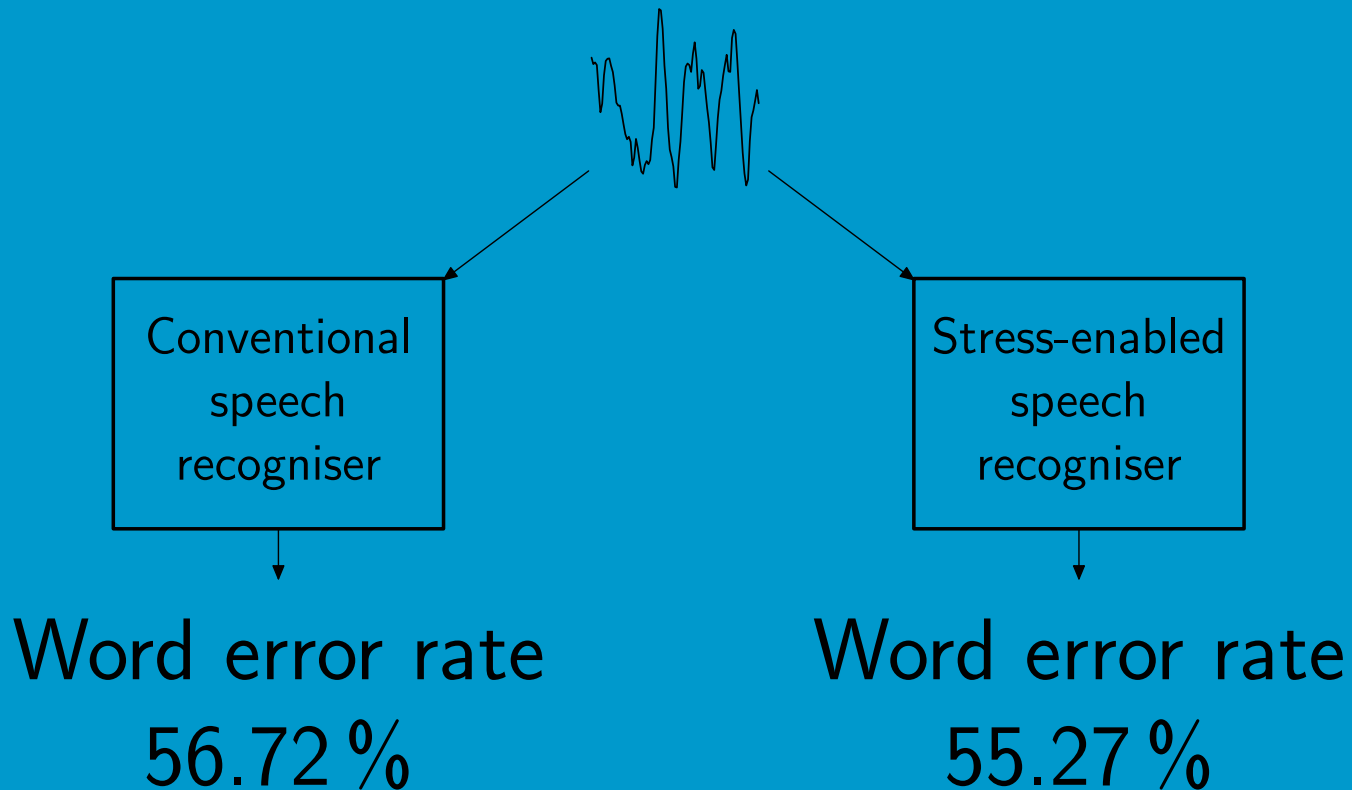


Results — training



Results — Recognition improvement

Results — Recognition improvement



A 2.6 % relative improvement.

Conclusion

Conclusion

Using lexical stress in an automatic speech recogniser for continuous, large-vocabulary speech can improve the recognition rate.

Conclusion

Using lexical stress in an automatic speech recogniser for continuous, large-vocabulary speech can improve the recognition rate.

- Consonants

Future work

- Model duration
- Model phrasal stress

