UvA-DARE (Digital Academic Repository)

Prominence. Acoustic and lexical/syntactic correlates

Streefkerk, B.M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
# Table of Content

1 Introduction

1.1 Notion of prominence
   1.1.1 General viewpoint
   1.1.2 Phonetic viewpoint
   1.1.3 Linguistic viewpoint

1.2 Topic of investigation

1.3 Usefulness of acoustic and lexical / syntactic correlates of prominence

1.4 Research method

1.5 Outline of this study

2 Prominence marking by naive listeners

2.1 Introduction
   2.1.1 Literature survey
   2.1.2 Our approach

2.2 Speech material
   2.2.1 Recordings
   2.2.2 Material
   2.2.3 Speakers
   2.2.4 Selected speech material

2.3 Two pilot experiments to define prominence
   2.3.1 Method
   2.3.2 Results
   2.3.3 Conclusion

2.4 Main experiment on assigning prominence
   2.4.1 First set marked by ten listeners
      2.4.1.1 Resulting prominence marks
      2.4.1.2 Differences within and between listeners
      2.4.1.3 Clustering of the cumulative prominence marks
   2.4.2 Second set marked by one ‘optimal’ listener
      2.4.2.1 Prominence assignment

2.5 Concluding remarks

3 Lexical and syntactic correlates of prominence

3.1 Introduction
   3.1.1 Relevant studies
# TABLE OF CONTENT

3.1.1 Data driven studies 40  
3.1.2 Application-oriented studies 41  
3.2 Pilot study to find lexical / syntactic correlates 44  
3.3 Main experiment on lexical / syntactic correlates of prominence 45  
3.3.1 Assigning lexical and syntactic features 46  
3.3.1.1 Description and evaluation of the automatically derived Part-of-Speech (POS) tags 48  
3.3.2 Relationship between word class and prominence 49  
3.3.3 Relationship between word length and prominence 53  
3.3.4 Relationship between the position of a word in a sentence and prominence 56  
3.3.5 Adjective-Noun combinations and prominence 56  
3.3.6 Algorithm to predict prominence on lexical / syntactic input for prominence prediction 58  
3.4 Independent test of the prominence assignment rules 67  
3.5 Discussion and conclusion 68  

4 Acoustical correlates of prominence 71  
4.1 Introduction 72  
4.1.1 General description of possible acoustic correlates 72  
4.1.2 Relevant studies on automatic feature extraction 74  
4.2 Feature extraction 77  
4.2.1 Segmentation and labeling of the speech material 77  
4.2.1.1 Training of the HMM-recognizer 78  
4.2.1.2 Resulting segmentation 79  
4.2.1.3 Accuracy of the automatic segmentation 80  
4.2.2 Acoustical correlates 81  
4.2.2.1 Unit selection 81  
4.2.2.2 $F_0$ features 82  
4.2.2.2.1 Measuring $F_0$ 84  
4.2.2.2.2 Extracting $F_0$ features 86  
4.2.2.3 Duration features 89  
4.2.2.4 Intensity features 96  
4.2.2.4.1 Measuring intensity 96  
4.3 Summary and conclusion 100  

5 Neural net classification of prominence with acoustic input features 101  
5.1 Introduction 102  
5.1.1 How feed-forward networks work 102  
5.1.1.1 General training procedure 103  
5.1.2 Distributions of prominence classes and the relationship to applications 105  
5.2 Prominence recognition with neural networks 105  
5.2.1 Acoustical input features 105  
5.2.2 Pre-processing of the input features 107
### Table of Content

5.2.2.1 Correlation 107  
5.2.3 Design of the training and testing data 108  
5.2.4 Binary prominence classification 110  
5.2.4.1 Testing with the Independent test set 113  
5.2.4.2 Summary and conclusion 114  
5.2.5 Gradient prominence prediction 115  
5.2.5.1 Results 116  
5.2.6 Analyses of individual features 119  
5.2.6.1 Analyzing the neural network 119  
5.2.6.2 Analyzing the performance of the individual feature 123  
5.2.6.3 Summary and conclusion 125  
5.2.7 Analyzing combinations of features 125  
5.2.8 Prominence classification with an ‘optimal’ feature combination 126  
5.3 Discussion and conclusion 127

6 General conclusion and discussion 129  
6.1 Introduction 130  
6.2 Prominence assignment by naive listeners 130  
6.2.1 Word or syllable prominence marking 130  
6.2.2 Binary or gradient prominence marking 131  
6.2.3 Consistency and reliability 132  
6.2.4 Concluding remarks 132  
6.3 Lexical / syntactic correlates of prominence 132  
6.3.1 Individual features 133  
6.3.2 Prominence prediction on textual input 133  
6.3.3 Discussion about lexical / syntactic correlates 134  
6.3.3.1 Comparison to the literature 134  
6.3.3.2 Method used 134  
6.3.3.3 Useful for Text-to-Speech 135  
6.4 Acoustical correlates of prominence 136  
6.4.1 Individual features 136  
6.4.2 Prominence prediction on acoustic input 136  
6.4.2.1 Complexity 137  
6.4.3 Discussion about acoustic correlates 137  
6.4.3.1 Comparison to the literature 137  
6.4.3.2 Method used 138  
6.4.3.3 HMM-alignment 138  
6.4.3.4 Normalizations 138  
6.4.3.5 Separate use of linguistic and acoustic features 139  
6.5 Future research 139  
6.5.1 Promising features of prominence 139  
6.5.2 Speech-technology applications 139  
6.5.3 Combining linguistic and acoustical features 140
### Table of Content

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 2.1</td>
<td>Correspondence matrix assigning prominence without pitch movements</td>
<td>143</td>
</tr>
<tr>
<td>A 2.2</td>
<td>NIST header</td>
<td>144</td>
</tr>
<tr>
<td>A 2.3</td>
<td>Demographic data of the training set</td>
<td>145</td>
</tr>
<tr>
<td>A 2.4</td>
<td>Demographic data of the test set</td>
<td>146</td>
</tr>
<tr>
<td>A 2.5</td>
<td>Instruction for the listening experiment</td>
<td>147</td>
</tr>
<tr>
<td>A 3.1</td>
<td>Overview of predicted prominence marks and lexical and syntactic correlates</td>
<td>148</td>
</tr>
<tr>
<td>A 4.1</td>
<td>SAMPA symbols</td>
<td>149</td>
</tr>
<tr>
<td>A 5.1</td>
<td>Scaling values</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td><strong>Bibliography</strong></td>
<td>151</td>
</tr>
<tr>
<td></td>
<td><strong>Summary</strong></td>
<td>161</td>
</tr>
<tr>
<td></td>
<td><strong>Samenvatting</strong></td>
<td>165</td>
</tr>
<tr>
<td></td>
<td><strong>Curriculum vitae</strong></td>
<td>169</td>
</tr>
</tbody>
</table>