INTERSPEECH 2019: Show & Tell Contribution
September 15–19, 2019, Graz, Austria

A User-Friendly and Adaptable Re-Implementation of an Acoustic Prominence Detection and Annotation Tool

Jana Voße[1], *Petra Wagner*[1,2]

[1]Phonetics and Phonology Work Group, Bielefeld University, Germany
[2]Center of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Germany
j.vosse@uni-bielefeld.de, petra.wagner@uni-bielefeld.de

Abstract

Annotating prominence phenomena in speech corpora is a challenging task, as it requires many resources. Therefore, several approaches have emerged in the past decades to automatise the process of detecting and annotating prominence. Among these, [1] propose a fully automatically operating acoustic prominence detection and annotation tool that yields promising results. The present work aims at making this tool accessible to a broader community and more inviting in the manipulation of features. To do so, we re-implemented the prominence annotation approach of [1] in the programming language of the software Praat [2], which is commonly used for speech analysis purposes within several areas of research. By implementing a user-friendly interface, the Praat-based prominence detection and annotation tool can be controlled without any source code interaction, which makes it accessible to users with differing levels of programming experience. More experienced users have the option to directly work with the comprehensively commented and documented source code to manipulate or add features within the prominence detection and annotation process. Providing a more accessible and easier to manipulate re-implementation of the tool of [1], we want to contribute to further developments in the area of automatic prominence detection and annotation.

Index Terms: automatic speech processing, prominence, automatic prominence annotation, acoustic prominence modeling

1. Introduction

The linguistic phenomenon of prominence describes the highlighting of single units such as syllables or words within their environment. This highlighting is used strategically by humans within communication, e.g. to structure the information delivered in an utterance [1], to disambiguate words [3] or to express phonetic variation [3]. To let a unit appear prominent, several features on the acoustic and the linguistic level are involved, which makes prominence an inherently complex phenomenon. Due to its complexity, the detection and annotation of prominence in corpora is not trivial. To capture it accurately, prominence is commonly annotated manually. This process is time-consuming and costly, especially for the annotation of large corpora. To facilitate the process of prominence annotation, a variety of annotation tools have been developed during the past decades (e.g., [4],[5]). However, many of these annotation tools require manual preprocessing of the data or operate only semi-automatically, which reduces, but not solves the issue of costly prominence annotations.

To tackle this issue, [1] suggest a fully automatic annotation approach. Their prominence tagger, which was introduced by [6] and later adapted to different languages, detects prominent units on information derived exclusively from the speech signal and assigns them a prominence value, expressing the units' "degree of standing out of [their] environment" ([1], p. 1809). Although the tool's restriction to signal-based features implicates that only a subset of prominence-relevant features is considered, it delivers good predictions [1].

2. Contribution of the present work

The prominence detection and annotation tool proposed by [1] has already been employed successfully in different studies (e.g., [7],[8]). The present work re-implements this tool in the Praat scripting language. With Praat being the "de facto standard speech analysis program" ([9], p. 47), we aim at contributing with our work to the further development of acoustic prominence modeling in two ways: First, we want to make the prominence tagger of [1] more accessible to the community by re-implementing it in a commonly used speech analysis software. We further add features to the tool that simplify its usage, so it can be controlled by users with little to no programming experience. Second, we want to make the prominence tagger by [1] more inviting in the manipulation of features. By preparing a well-documented source code, interested users can easily follow the structure of the tool and adapt it to their own needs. This feature is of special interest for prominence analyses, as the prosodic structure of prominence phenomena is highly language-specific [10].

3. Usage

To perform prominence annotations with our tool, the users open the source code in the Praat menu and run it. The 'run' command invokes a graphical user interface (GUI) (see Figure 1), which guides the users through the prominence annotation process. The GUI asks the users specifically for the information it requires to perform the prominence detection and annotation, so there is no interaction necessary between the users and the source code.

The information the Praat-based prominence detection and annotation tool requires is little: The users must provide the file path to the files which shall be annotated, a few information about the internal structure of the files as well as the kind of information they would like the output file to contain. By submitting this information to the tool, it will generate a table file containing the information specified by the users (see Figure 2). This table file is available over the Praat menu and can be saved in different formats, e.g. as a tab- or comma-separated file.

The files which shall be annotated for prominence require only a small amount of preprocessing: The data must be annotated on syllable level and on phone level. Both can be obtained by additional speech processing tools such as WebMAUS [11], so no manual preparation of the data is required.

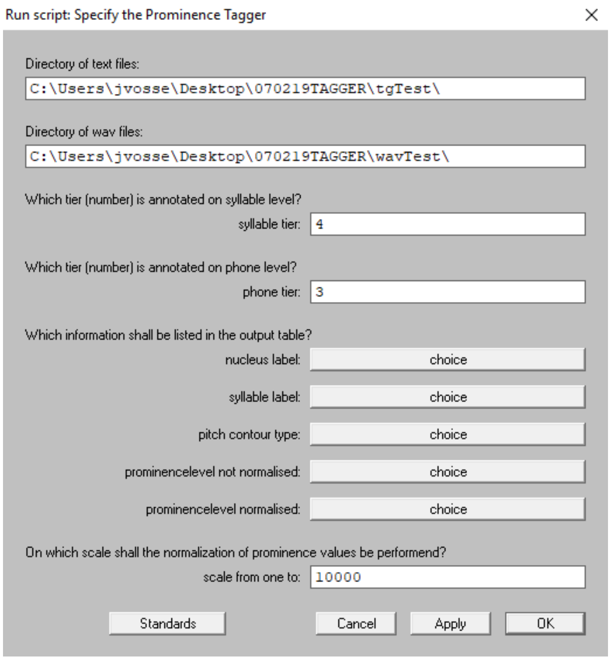

Figure 1: *Graphical user interface (GUI) of Praat-based prominence annotation tool.*

	1	2	3
row	Nucleus	Syllable	Prominencelevel(normalised)
61	I	Iz	9.134959231409713
62	aI	aIn	320.12105352137024
63	a	fax	32.2493639879127
64	u:	zu:	357.34403936962616
65	6	p6	2476.230743274844
66	I	vIC	48.797370617168376
67	I	tIC	165.95440803655094

Figure 2: *Table output of of Praat-based prominence annotation tool.*

To experiment with features in the prominence detection and annotation process or to adapt them to the needs of specific investigations, users have the option to directly manipulate the source code. We commented the source code in a comprehensive manner to ensure a high level of guidance for the users throughout the code. The comments include hints to possible manipulation points and advice on manipulation strategies. For example, the normalization formula in the source code is a potential point of manipulation, if the users prefer different normalization strategies (see Figure 3). In this way, the users can easily manipulate existing or even add new features to the tool and experiment with their contribution to the overall prominence annotation performance.

4. Outlook

The performance of the prominence detection and annotation tool by [1] has been evaluated on read speech data of German, which yielded promising results. For the near future, we plan to refine this evaluation by testing the re-implemented tool on data expressing several speaking styles, e.g. motivating speech, charismatic speech and emotional speech to gain further insight into the tool's performance quality.

```
for value from 1 to (numberOfRows-1)

  currentValue = Get value... value Prominencelevel(not_normalised)

; -> MANIPULATION POINT: The following line performs the normalization.
;                        You can adapt it to your own needs, e.g. by
;                        changing the formula.

  normValue=((currentValue-minProm)/(maxProm-minProm))*norm_scale

  Set numeric value: value, "Prominencelevel(normalised)", normValue

endfor
```

Figure 3: *Excerpt of source code with hints to manipulation points.*

We further plan to improve the accuracy of certain operations within the tool to provide more precise results. For example, it is planned to refine the selection of candidates for prominent units. Currently the Praat-based prominence detection and annotation tool considers syllable nuclei as potential prominent instances for which the tool exclusively detects vowels. However, also other sonorant instances such as nasals or liquids may form the syllable nucleus. Removing this restriction by enabling the tool to consider different phones within the detection of prominent syllable nuclei would be a valuable improvement of the tool's performance.

Finally, implementing more manipulation options within the GUI will increase the usability of our re-implemented tool further. Here, it is planned for the near future to offer different normalization formulae the users can choose from and to have the opportunity to adapt the tool to different languages by the manipulation of language-specific weight coefficients.

5. References

[1] F. Tamburini and P. Wagner, "On automatic prominence detection for german," in *Eighth Annual Conference of the International Speech Communication Association*, 2007, pp. 1809–1812.

[2] P. Boersma, "Praat: doing phonetics by computer," *http://www.praat. org/*, 2019.

[3] G. Abete, F. Cutugno, B. Ludusan, and A. Origlia, "Pitch behavior detection for automatic prominence recognition," in *Speech Prosody 2010-Fifth International Conference*, 2010.

[4] T. Portele, B. Heuft, C. Widera, P. Wagner, and M. Wolters, "Perceptual prominence," in *Speech and Signals. Aspects of Speech Synthesis and Automatic Speech Recognition. Festschrift dedicated to Wolfgang Hess on his 60th birthday*, 2000, vol. 69, pp. 97–115.

[5] P. Wagner, S. Breuer, and K. Stöber, "Automatische prominenzetikettierung einer datenbank für die korpusbasierte sprachsynthese," in *Fortschritte der Akustik-DAGA 2000*, 2000.

[6] F. Tamburini, "Reliable prominence identification in english spontaneous speech," *Proc. Speech Prosody 2006*, 2006.

[7] Z. Malisz, H. Berthelsen, J. Beskow, and J. Gustafson, "Controlling prominence realisation in parametric dnn-based speech synthesis." in *INTERSPEECH*, 2017, pp. 1079–1083.

[8] B. Samlowski, P. Wagner, and B. Möbius, "Disentangling lexical, morphological, syntactic and semantic influences on german prominence-evidence from a production study," in *INTERSPEECH*, 2012, pp. 2406–2409.

[9] P. Martin, "Multi methods pitch tracking," in *Speech Prosody 2012-Sixth International Conference*, 2012, pp. 47–50.

[10] A. Rosenberg, E. Cooper, R. Levitan, and J. Hirschberg, "Cross-language prominence detection," in *Speech Prosody 2012-Sixth International Conference*, 2012, pp. 278–281.

[11] T. Kisler, F. Schiel, and H. Sloetjes, "Signal processing via web services: the use case webmaus," in *Digital Humanities Conference 2012*, 2012, pp. 30–34.