SpeCT 2.0 – Speech Corpus Toolkit for Praat

Mietta Lennes
Department of Digital Humanities / FIN-CLARIN
University of Helsinki, Finland
mietta.lennes@helsinki.fi

Abstract

Praat (Boersma and Weenink, 2019) is a versatile, open-source platform that provides a multitude of features for annotating, processing, analyzing and manipulating speech and audio data. By using the built-in scripting language, Praat can be easily extended and adjusted for different purposes while reducing manual work. The Speech Corpus Toolkit for Praat (SpeCT) is a collection of Praat scripts that can be used to perform various small tasks when building, processing and analyzing a speech corpus. SpeCT can help both beginners and advanced users solve some common issues in, e.g., semi-automatic annotation or speech corpus management. This work describes some of the general functionalities in SpeCT. A selection of the scripts will also be made available via the Mylly service at the Language Bank of Finland, maintained by FIN-CLARIN.

1 Introduction

Praat (Boersma and Weenink, 2019) is a well-established and popular speech analysis platform not only for speech researchers but also for many others working with audio signals. Praat is based on open source code and the package is maintained on GitHub[1]. Praat is available for Windows, Mac and Unix/Linux and it can be downloaded and used free of charge.

Originally, Praat is "a system for doing phonetics by computer" (Boersma and Weenink, 2019). Indeed, it provides a huge number of useful functionalities for analyzing, visualizing and annotating speech signals. However, Praat is much more than a speech transcription workbench. For instance, it also provides many statistical analyses, an articulatory speech synthesizer (described in (Boersma, 1998)), features for doing source-filter synthesis, an editor for manipulating pitch and durations within sound files, and a system for creating graphics for publications.

Importantly, Praat also includes a full-scale scripting language that makes use of all the high-level menu commands of the graphical user interface in addition to general programming syntax. Praat scripts can be run in the graphical user interface or in batch mode, and they can be called by other programs. Nearly identical features, displays and file formats are available in all operating systems. Thus, Praat represents a speech data ecosystem that can be readily combined with other digital research environments and analysis pipelines.

The Speech Corpus Toolkit for Praat (SpeCT)[2] is a collection of Praat scripts that can be used to perform various small tasks when building, processing and analyzing a speech corpus. Earlier versions of many of the scripts have been available online since 2001. The repository was renamed as "the Speech Corpus Toolkit for Praat" in 2011 and moved to GitHub in 2017 (Lennes, 2017). From the start, the scripts have been well commented. Instructions for particular tasks are offered on supplementary web pages[3] in order to reduce the need for email support, since only one person has been responsible for maintaining the scripts. However, during the past few years, there has been an increasing need for updating the collection, for fixing bugs and for making the scripts more consistent and interoperable.

---

[1]Praat repository on GitHub: https://github.com/praat
[2]The Speech Corpus Toolkit for Praat on GitHub: https://github.com/lennes/spect
[3]SpeCT documentation: https://lennes.github.io/spect/

SpeCT was originally created in order to provide accessible and lightweight solutions to some frequently occurring practical problems and tasks in speech research, especially bearing in mind the "less technical" researchers and students. In addition to SpeCT, there currently exists a wealth of other online resources containing various kinds of Praat scripts and plugins that could potentially be used as components in customized workflows. However, for more experienced programmers it is useful to look into Python modules and libraries related to Praat, e.g., `pympi` (Lubbers and Torreira, 2013 2018), and/or interfaces between Praat and statistical analysis software, e.g., `praatR` (Albin, 2014).

2   Speech corpus workflow

After recording some audio material for a speech corpus, researchers, research assistants or students tend to face many practical problems. SpeCT can provide at least partial solutions to some typical issues in organizing and handling speech corpora.

2.1   Slicing sound files

Sound files often require some trimming and sometimes they need to be sliced into shorter excerpts. SpeCT contains scripts with different options for exporting annotated segments into individual sound files. The user can mark the portions either semi-automatically (by first letting the computer mark the quiet portions as silences) and/or manually according to visual and auditory inspection. This process reduces the manual work load, and the resulting files can be automatically named in a systematic way, thus minimizing the risk for errors.

2.2   Pause detection and semi-automatic alignment of transcripts

At some point, the original sound files usually need to be transcribed and/or annotated with other necessary details so as to eventually allow for the efficient searching and analysis of the complete corpus. However, manual annotation is typically very time-consuming, and for many languages and especially for casual conversational speech, fully automatic speech recognition tools might not produce satisfactory results. Therefore, any kind of semi-automatic aid may turn out to be useful during the annotation stage.

In SpeCT, there are scripts for detecting relatively silent portions of the sound signal ("pauses"), and a similar feature is also built in Praat. Moreover, in case the user already has a text file containing a transcript of the conversation, it is possible to use a script in SpeCT for semi-automatically aligning the transcribed utterances or turns with the audio file. The script works by suggesting the start and end boundaries for one utterance or turn at a time, and the user needs to modify the suggested boundary locations if required. The suggestions are based on detecting silent portions as well as the number of characters in the transcribed line of text. Each speaker's turns are marked in a separate annotation tier. Naturally, the method is quite crude and intervention is required from the user, but since a language-independent heuristic is used, the script has turned out to be quite helpful in some cases.

2.3   Search and look-up from a number of corpus files

Sometimes systematic transcription errors are spotted while the annotation project is still ongoing. Using a Praat script, it is possible to look up a given portion of a specific annotated sound file in a large corpus, without remembering the name of the file. If certain transcriptions need to be changed in the entire corpus, the replacement can be done automatically by using a script.

If the annotation project is large, another script can be used in order to obtain an overview of which files have been annotated already, what tiers they include, how many annotation items they contain, or which files lack particular annotation tiers.

2.4   Measuring and extracting samples for specific purposes

In the analysis stage, the researcher may require a tool for measuring and collecting acoustic-phonetic parameters (e.g., durations, pitch, formant frequencies etc.) or complex linguistic properties that may combine time-synchronous data from several annotation tiers as well as the original audio signal in the annotated corpus. The analysis procedure is likely to be quite specific for each research question,

so a general-purpose search tool may not suffice. Several different scripts are available for exporting measurements and other data from Praat to spreadsheet documents or to other interoperable data formats, or for plotting measured data in figures.

Many more speech corpus issues can be addressed by using the Praat scripts already included in SpeCT. It is also relatively straightforward to adapt the scripts in order to match other similar purposes.

3   Future work

During 2019, a second generation of SpeCT will be created by modifying and updating the existing Praat scripts according to more consistent design principles. The script collection will continue to be publicly available via GitHub. In addition, as part of the updating process of SpeCT, a sample of the Praat scripts will also be installed for test use in the researchers' toolbox called *Mylly*, 'the Mill' (Lennes et al., 2017), one of the services available at the Language Bank of Finland.[4] In Mylly, the user can log in to the online service, upload files to a personal workspace and process them with various tools by simply selecting the desired commands from the menu. Mylly allows users to build and save workflows for reuse.

For SpeCT, the general development goal is to ensure that each script can be conveniently used either in isolation or as part of a longer workflow that might include processing steps outside the Praat environment. In order to fulfil the development goal, each script should be self-contained and well instructed. For input and output, the scripts in SpeCT should rely on native Praat objects or, if necessary, on simple and generic data formats. By complying to these principles, it will be even easier to modify the scripts or to combine them with other similar resources. In the conference, we will demonstrate some of the main tools provided by SpeCT for the different stages of the speech corpus processing workflow.

References

Aaron Albin. 2014. PraatR: An architecture for controlling the phonetics software "Praat" with the R programming language. *Journal of the Acoustical Society of America*, 135(4):2198.

Paul Boersma and David Weenink. 2019. Praat: doing phonetics by computer (Version 6.1.03). [Computer program]. https://www.praat.org/.

Paul Boersma. 1998. *Functional phonology. Formalizing the interactions between articulatory and perceptual drives*. Ph.D. thesis, University of Amsterdam.

Mietta Lennes, Jussi Piitulainen, and Martin Matthiesen. 2017. Mylly — The Mill: A new platform for processing speech and text corpora easily and efficiently. In *Interspeech 2017: Show & Tell Contribution, Stockholm, Sweden*, pages 829–830.

Mietta Lennes. 2017. SpeCT — The Speech Corpus Toolkit for Praat (v1.0.0). First release on GitHub. Zenodo. http://doi.org/10.5281/zenodo.375923.

Mart Lubbers and Francisco Torreira. 2013-2018. pympi-ling: a Python module for processing ELANs EAF and Praats TextGrid annotation files. https://pypi.python.org/pypi/pympi-ling. Version 1.69.

---

[4]Kielipankki, The Language Bank of Finland (https://www.kielipankki.fi) is a collection of services maintained by FIN-CLARIN.