Prosit: a Praat plug-in for search and inspection of corpora of annotated audio files

Introduction

The paper presents the software *Prosit*, a plug-in for Praat (Boersma & Weenink 1992-2017), that is one of the most renowned program for doing phonetic research.
The Prosit plug-in is designed to help researchers in:

- making a flexible search on a corpus of annotated sound files. Information referring to the sounds can be stored both as Praat TextGrid annotations and DC metadata[1];
- building their own corpus by editing and managing sound files, textgrids and DC metadata;
- listening, visualizing, inspecting, analyzing sounds that match the search criteria.

Motivation

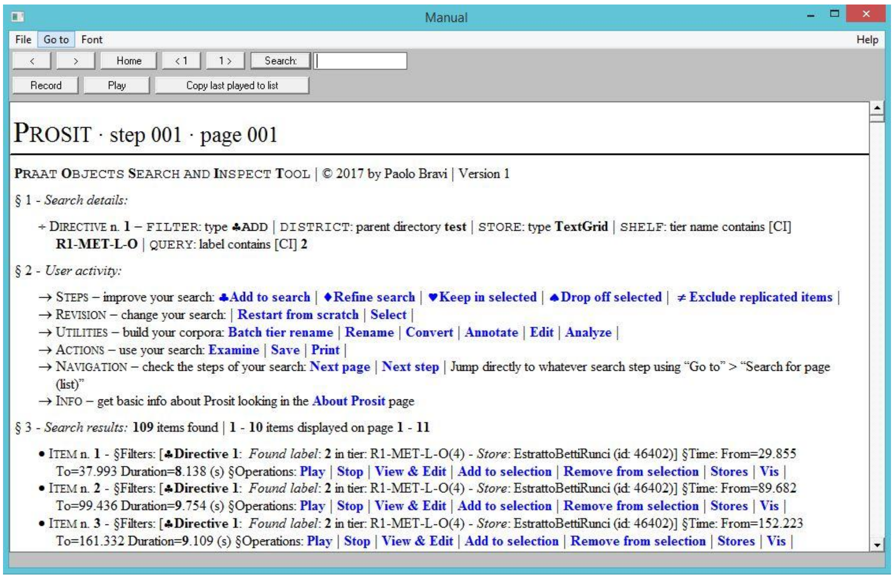
TextGrid is a widely diffused format for files storing information related to audio files. In many cases audio files analysis needs a previous annotation (that can be either manual or automatic) of the sound files, so that eventually the corpus wherein analysis is carried out consists of a number of pairs of sounds and TextGrids thereof.
That being the case, a typical analytical batch procedure consists of iterating some kind of analytical procedures throughout all files comprising the corpus, which are usually stored in one directory and have a consistent annotation. But what if the "corpus" – or a pre-version of it – is not (still) built using a consistent annotation scheme, or if we simply want to explore - perhaps as a preliminary step towards the creation of a well-organized corpus - within possibly large and non-homogeneous groups of annotated sound files?
Prosit is designed to face this kind of need: while from one side it is a search-engine (without external dependencies) that allows a step-wise flexible research within possibly large and non-homogeneous groups of TextGrids, from the other side it allows to either listening to, editing, visualize and analyze the audio files that match the TextGrid interval(s) requested and save the relevant output.

Searching the annotations

Searching intervals in Prosit comes about in three phases. For ease of conceptualization, these phases are identified using the metaphor (and the relevant names) of a shopping trip. In phase 1, the user sets what is called the *district*, viz a directory-mother, and a *store type*, viz a file type (with extension starting with ".TextGrid" or ".DC") that is able to stores different kind of information – either annotations or Dublin Core metadata – relevant to each audio file. In phase 2, the user sets the *shelf* (or the shelves) that Prosit has to explore. If the store type chosen is "TextGrid", this means identifying the tier(s) that will be search through. A spectrum of possibility is offered to this aim. In phase 3, the user sets the *item*(s) s/he is interested in.
Search results, together with a description of the research carried out, is made available to the user by means of dynamically created "manpages"[2] that allow the user to listen to the

[1] Search on DC metadata is not addressed at all in this abstract. For reference, see http://dublincore.org/

specified part of the sound, visualize it through specific animation devices (and also to play it at a speed lower than the original) and make a number of batch operations on all found items.



PROSIT · step 001 · page 001

PRAAT OBJECTS SEARCH AND INSPECT TOOL | © 2017 by Paolo Bravi | Version 1

§ 1 - *Search details:*

÷ DIRECTIVE n. **1** – FILTER: type ♣**ADD** | DISTRICT: parent directory **test** | STORE: type **TextGrid** | SHELF: tier name contains [CI] **R1-MET-L-O** | QUERY: label contains [CI] **2**

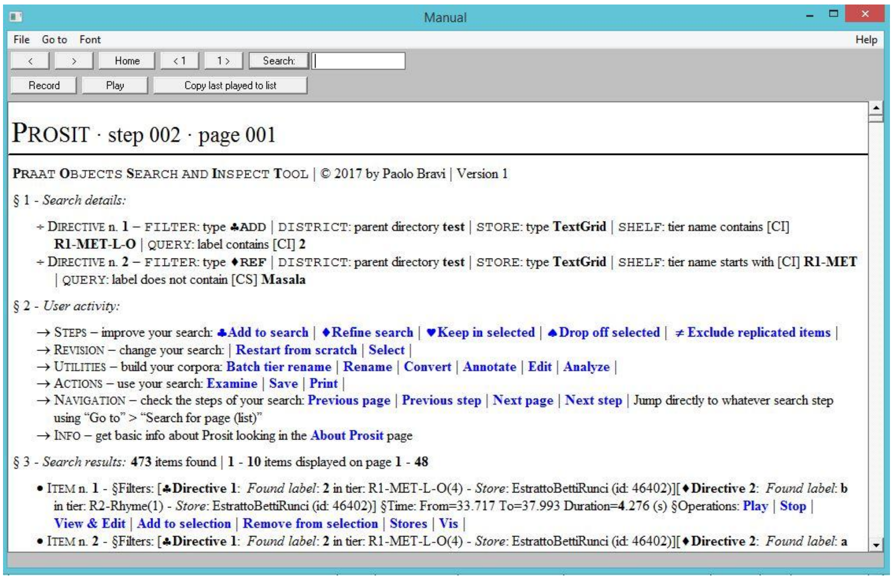
§ 2 - *User activity:*

→ STEPS – improve your search: **♣Add to search** | **♦Refine search** | **♥Keep in selected** | **♠Drop off selected** | **≠ Exclude replicated items** |
→ REVISION – change your search: | **Restart from scratch** | **Select** |
→ UTILITIES – build your corpora: **Batch tier rename** | **Rename** | **Convert** | **Annotate** | **Edit** | **Analyze** |
→ ACTIONS – use your search: **Examine** | **Save** | **Print** |
→ NAVIGATION – check the steps of your search: **Next page** | **Next step** | Jump directly to whatever search step using "Go to" > "Search for page (list)"
→ INFO – get basic info about Prosit looking in the **About Prosit** page

§ 3 - *Search results:* **109** items found | **1** - **10** items displayed on page **1** - **11**

- ITEM n. **1** - §Filters: [♣**Directive 1**: *Found label*: **2** in tier: R1-MET-L-O(4) - *Store*: EstrattoBettiRunci (id: 46402)] §Time: From=29.855 To=37.993 Duration=**8**.138 (s) §Operations: **Play** | **Stop** | **View & Edit** | **Add to selection** | **Remove from selection** | **Stores** | **Vis** |
- ITEM n. **2** - §Filters: [♣**Directive 1**: *Found label*: **2** in tier: R1-MET-L-O(4) - *Store*: EstrattoBettiRunci (id: 46402)] §Time: From=89.682 To=99.436 Duration=**9**.754 (s) §Operations: **Play** | **Stop** | **View & Edit** | **Add to selection** | **Remove from selection** | **Stores** | **Vis** |
- ITEM n. **3** - §Filters: [♣**Directive 1**: *Found label*: **2** in tier: R1-MET-L-O(4) - *Store*: EstrattoBettiRunci (id: 46402)] §Time: From=152.223 To=161.332 Duration=**9**.109 (s) §Operations: **Play** | **Stop** | **View & Edit** | **Add to selection** | **Remove from selection** | **Stores** | **Vis** |

Figure 1 – Example of a two-step search. Step 1: the search directive 1 looks up into a parent directory (in this case, the built-in "test" one) for TextGrid comprising tier names that contains the string "R1-MET-L-O" and look in these tiers searching for the label "2". 109 items are found.



PROSIT · step 002 · page 001

PRAAT OBJECTS SEARCH AND INSPECT TOOL | © 2017 by Paolo Bravi | Version 1

§ 1 - *Search details:*

÷ DIRECTIVE n. **1** – FILTER: type ♣**ADD** | DISTRICT: parent directory **test** | STORE: type **TextGrid** | SHELF: tier name contains [CI] **R1-MET-L-O** | QUERY: label contains [CI] **2**
÷ DIRECTIVE n. **2** – FILTER: type ♦**REF** | DISTRICT: parent directory **test** | STORE: type **TextGrid** | SHELF: tier name starts with [CI] **R1-MET** | QUERY: label does not contain [CS] **Masala**

§ 2 - *User activity:*

→ STEPS – improve your search: **♣Add to search** | **♦Refine search** | **♥Keep in selected** | **♠Drop off selected** | **≠ Exclude replicated items** |
→ REVISION – change your search: | **Restart from scratch** | **Select** |
→ UTILITIES – build your corpora: **Batch tier rename** | **Rename** | **Convert** | **Annotate** | **Edit** | **Analyze** |
→ ACTIONS – use your search: **Examine** | **Save** | **Print** |
→ NAVIGATION – check the steps of your search: **Previous page** | **Previous step** | **Next page** | **Next step** | Jump directly to whatever search step using "Go to" > "Search for page (list)"
→ INFO – get basic info about Prosit looking in the **About Prosit** page

§ 3 - *Search results:* **473** items found | **1** - **10** items displayed on page **1** - **48**

- ITEM n. **1** - §Filters: [♣**Directive 1**: *Found label*: **2** in tier: R1-MET-L-O(4) - *Store*: EstrattoBettiRunci (id: 46402)][♦**Directive 2**: *Found label*: **b** in tier: R2-Rhyme(1) - *Store*: EstrattoBettiRunci (id: 46402)] §Time: From=33.717 To=37.993 Duration=**4**.276 (s) §Operations: **Play** | **Stop** | **View & Edit** | **Add to selection** | **Remove from selection** | **Stores** | **Vis** |
- ITEM n. **2** - §Filters: [♣**Directive 1**: *Found label*: **2** in tier: R1-MET-L-O(4) - *Store*: EstrattoBettiRunci (id: 46402)][♦**Directive 2**: *Found label*: **a**

Figure 2 – Example of a two-step search. Step 2: the directive 2 refines the search adding a second condition: labels that do not contain the string "Masala" are searched in tiers whose names starts with "R1-MET". 473 items are found.

[2] "Manpages" are files that Praat can read and which acts as hypertext (see http://www.fon.hum.uva.nl/praat/manual/ManPages.html)