Voice recognition in the LabTablet electronic laboratory notebook

[Short paper]

Susana Ventura
Ricardo Carvalho Amorim
Faculdade de Engenharia
Universidade do Porto
{ei12009, rcamorim}@fe.up.pt

João Rocha da Silva
Faculdade de Engenharia
Universidade do Porto / INESC-TEC
joaorosilva@gmail.com

Cristina Ribeiro
Faculdade de Engenharia
Universidade do Porto / INESC-TEC
mcr@fe.up.pt

ABSTRACT

Research institutions are considering data repositories to manage their outputs and ensure their visibility. In many domains, purpose-built tools can help collect data and metadata as they are created. LabTablet is such a tool, designed to provide the functions of a laboratory notebook, and being able to accompany users in either experimental sessions or field trips. In these contexts, the interaction with the device can be problematic, so we experimented with a speech recognition extension for two purposes: to provide commands, such as requesting readings from the built-in sensors, and to record observations such as a dictated note in a field trip.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.5 [**Information Interfaces and Presentation**]: User Interfaces; E.2 [**Data Storage Representations**]

1. INTRODUCTION

The management of research data is becoming a concern for both researchers and institutions. Data are collected or created in many forms, and they are not easy to organise, describe or share. Moreover, datasets generated in the context of research projects may contain sensitive data, and this also raises barriers to the adoption of emerging tools. Besides dealing with the variety of existing research domains, research data management is being regulated by mandates and guidelines from the funding agencies, pushing institutions to adopt data management platforms. These are digital repositories that tend to focus on two main high-level purposes, sometimes not addressed together: collaborative environments where researchers can manage their regular outputs and collaborate with their teams in the process; and preservation repositories, that often implement existing standards to ensure the long term preservation of the deposited data. In spite of the advantages of an integrated workflow to provide for both aspects, current solutions do not address them yet.

Due to their flexibility and overall availability, mobile devices can be used in research environments to perform tasks in the early stages of data production. Laboratory notebooks are a classical well-established device for recording all aspects of experimental work, and it is therefore natural for similar devices, supported by mobile devices, to appear. The problem with some of these solutions is the integration with the platforms where data are recorded, so typically the task of depositing the produced records is left to the researchers.

Mobile devices are evolving fast and becoming always connected. Speech recognition capabilities became accessible, supported by either cloud-based processing or custom-built libraries for off-line capabilities. In this work we consider LabTablet, a mobile application for collecting data and metadata in research environments, and experiment and evaluate speech recognition libraries. Some limitations were identified, but a working prototype is ready for testing beyond the development team.

2. MANAGING RESEARCH DATA WITH LABTABLET

The management of research data, specially in institutions with a large diversity of research domains, requires de definition of workflows for researchers, the description of datasets, their deposit in repositories where they can be searched and re-used.

Description, the creation of metadata for datasets, is a complex and specialised task, requiring the commitment of the researchers. LabTablet was developed as part of a suite of tools designed to make their task easier [1]. The principles behind LabTablet are that 1) using laboratory notebooks to keep record of research activities is a common task; and 2) automatic metadata can be collected in the research environment with the sensors from mobile devices.

2.1 LabTablet and Dendro

The motivation for LabTablet is therefore the automation of metadata creation. To accomplish this, LabTablet is coupled with Dendro, a research data management platform [3].

C3S2E'16 July 20 - 22 2016, Porto, Portugal
Copyright 2016 ACM ISBN 978-1-4503-4075-5/16/07 ...$15.00.
http://dx.doi.org/10.1145/2948992.2948998.

In Dendro, research teams can load metadata models, i.e. sets of descriptors appropriate for their domain. LabTablet loads profiles from Dendro, uses them in the production environment, and then synchronises back.

The LabTablet application relies on the Android platform to create an interface similar to what researchers already use—laboratory notebooks. With mobile devices, nonetheless, additional functionality for metadata capture is available; voice recordings and multi-media contents, for example, are invaluable as metadata to provide the context for a dataset.

The application has a strong focus on building metadata records that can be associated to the corresponding dataset. The process relies on Dendro, and on a process where researchers select the sets of descriptors appropriate for their data. When LabTablet is used in a selected domain (e.g. Biodiversity) the metadata profile is imported from Dendro, the descriptors become available to be filled in the LabTablet session, and in the end linked to the descriptors in Dendro.

2.2 Filling in descriptors

When using the LabTablet application, researchers can manage different open projects at the same time. At this point, each project is associated with a metadata record that can be incrementally filled by the researcher as soon as new metadata becomes available. The application has an interface specifically tailored to systematic metadata gathering—the Field Mode—where the researcher has access to a variety of sensors that depend on the device's capabilities. Along with these, LabTablet can also collect text-based records, voice recordings and other multi-media resources such as photos and videos. Moreover, if the researcher's activities involve field trips, the application can also track their position over time.

Figure 1a shows a metadata record for a specific project. Coupled with the appropriate metadata, each of these descriptors provides insights about aspects such as the dataset's origin. During a field session—that can also be an experimental session—, The LabTablet application resorts to a specifically designed interface to show the available inputs. These depend on the devices capabilities to harvest data such as geographic location, temperature and others. Besides that, other actions can also be carried out such as sketching, voice recording, text notes, photos and videos. In the end of each session, these are associated with specific descriptors.

3. SPEECH RECOGNITION IN MOBILE ENVIRONMENTS

From our experience in research contexts, and by analysing existing tools for electronic laboratory notebooks, we can observe that research activities often rely on the researcher to be promptly available to intervene, while at the same time they are expected to describe the process. This may prove to be a serious barrier to the adoption of digital platforms, as the researcher may find it easier to use a paper notebook instead. In these cases, interacting with the application through voice instructions can help overcome such limitations.

Google describes mobile voice search as a challenging problem due to the extensive amount of vocabulary, the unpredictability of voice input and the noise conditions that vary greatly, due to the different possible scenarios while using mobile devices. Nevertheless, speech recognition performance is already at a level where its users become repeat users [4].

Speech recognition gives users more freedom to use their tablet device in situations where lab work doesn't allow them to directly interact, such as having their hands occupied. On the other hand it also intends to make the process of collecting metadata descriptors easier, as described above.

To enable this kind of interaction, we evaluated existing voice recognition solutions. Our initial assumption was that we were selecting a single library, to be used off-line. The main concern was to allow researchers to do voice input on the most diverse conditions. However, the emerging tools that feature higher translation speed and accuracy take advantage of the Internet connectivity. Thus, we decided to also evaluate on-line speech recognition.

3.1 Evaluating existing solutions

The choices for off-line speech recognition libraries were scarce at the beginning. A preliminary evaluation showed us that several libraries were not prepared for mobile devices, either requiring large processing power or failing to recognise speech in noisy environments. Others, such as Shout[1] and Kaldi[2] were not compatible with the Android ecosystem at all.

After the initial survey, CMUSphinx[3] was chosen as the library for off-line recognition as it was the one with better results on the preliminary tests. CMUSphinx is also an open source library, with an active community improving and expanding it, which is also an important aspect when considering the evolution of our application. This library already couts with a trained language model, which allowed a quicker prototyped version.

The CMUSphinx library has several decoders available, one of them is PocketSphinx. It is CMU's fastest speech recognition system and it requires fewer resources when comparing it to the remaining solutions [2]. It is therefore suitable for mobile applications. This tool is also prepared to handle continuous speech recognition, which is an interesting feature for our application.

It was also clear that, when users have an active Internet connection, the built-in speech recognition tools are the most accurate in diverse scenarios. Since the involved companies have put a lot of effort in bringing this kind of interaction to their mobile ecosystems, the integration in the application's domain was expected to be easier. This tool was therefore selected to be used whenever the device is connected to the Internet.

4. SPEECH IN LABTABLET

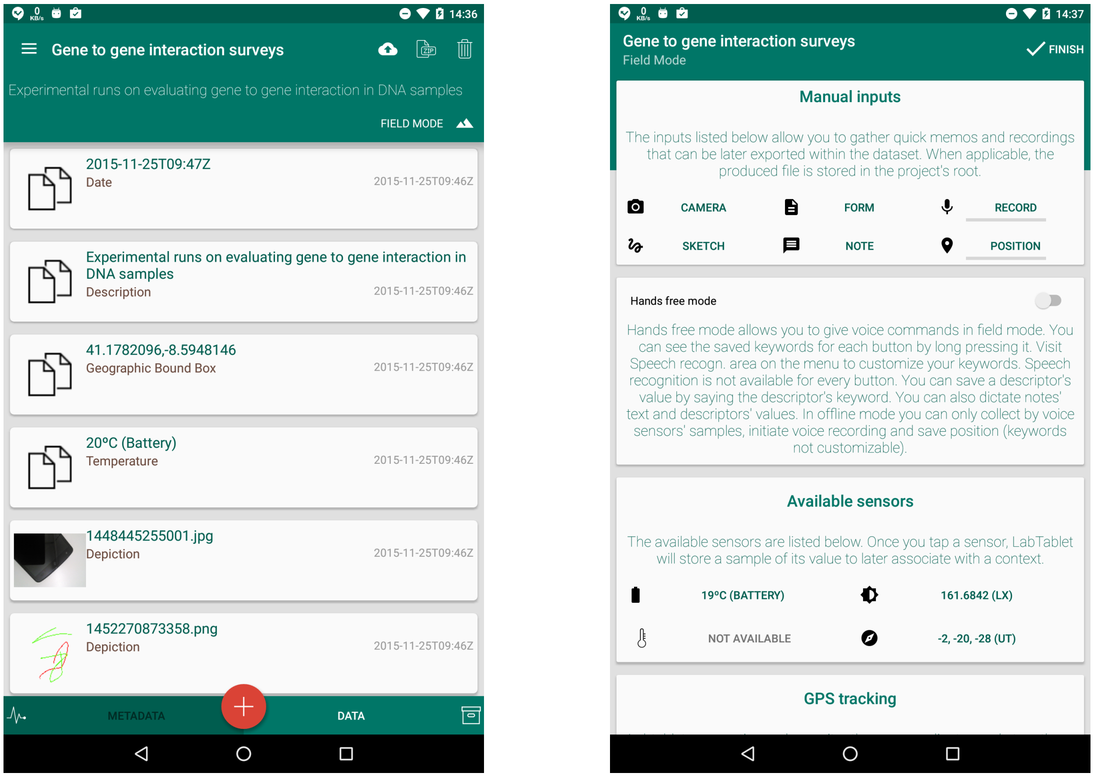
One of the key aspects of getting the speech recogniser to work is creating an adequate Language Model.

A Language model is used to achieve faster execution times and better accuracy in the recognition of words. Language Models work by limiting the number of possible words that need to be considered at any given time during the search, either by listing a subset of possible expansions (grammar) or by using a statistical model to calculate the proba-

[1] `http://shout-toolkit.sourceforge.net/index.html`

[2] `http://kaldi-asr.org/`

[3] `http://cmusphinx.sourceforge.net/`



Gene to gene interaction surveys

Experimental runs on evaluating gene to gene interaction in DNA samples

2015-11-25T09:47Z
Date
2015-11-25T09:46Z

Experimental runs on evaluating gene to gene interaction in DNA samples
Description
2015-11-25T09:46Z

41.1782096,-8.5948146
Geographic Bound Box
2015-11-25T09:46Z

20ºC (Battery)
Temperature
2015-11-25T09:46Z

1448445255001.jpg
Depiction
2015-11-25T09:46Z

1452270873358.png
Depiction
2015-11-25T09:46Z

(a) List of gathered metadata

Gene to gene interaction surveys
Field Mode

Manual inputs

The inputs listed below allow you to gather quick memos and recordings that can be later exported within the dataset. When applicable, the produced file is stored in the project's root.

Hands free mode

Hands free mode allows you to give voice commands in field mode. You can see the saved keywords for each button by long pressing it. Visit Speech recogn. area on the menu to customize your keywords. Speech recognition is not available for every button. You can save a descriptor's value by saying the descriptor's keyword. You can also dictate notes' text and descriptors' values. In offline mode you can only collect by voice sensors' samples, initiate voice recording and save position (keywords not customizable).

Available sensors

The available sensors are listed below. Once you tap a sensor, LabTablet will store a sample of its value to later associate with a context.

19ºC (BATTERY)
161.6842 (LX)
NOT AVAILABLE
-2, -20, -28 (UT)

GPS tracking

(b) Application's field mode

Figure 1: LabTablet's screenshots showing an opened project in both project view (1a) and field mode 1b
.

bility for each possible successor word. The latter method is more useful for free-form input, such as dictation, whereas the use of a grammar requires that every possible legal work sequence is defined a priori. As it stands, when training the recogniser, higher quantities of data often yield better result accuracy.

For the LabTablet application, two scenarios were considered for the voice interaction in a open field session: 1) navigate through the available options in the interface and 2) recording voice notes.

To navigate the options, each of the active sources of data (sensors that are available in the device) is associated with a specific keyword that triggers their function. As an example of such interaction, when the researcher says "temperature" the device automatically saves the current temperature reading.

To record voice notes, we take advantage of speech recognition to transcribe voice recordings that the researcher wants to associate to the dataset. These are often related to informal descriptions of the procedures being executed, or personal notes to address future work. As in some cases researchers may be available to directly use the application, both scenarios can be deactivated.

4.1 Off-line interactions

Due to the resources allocated to the project, building our own language model and dictionary was not an option. The training required to achieve an acceptable accuracy was beyond the available time. We decided to use the existing English dictionary, which is already tuned for best performance, and adapt it to our specific needs, creating a grammar with the commands we needed such as "luminosity" for the luminosity sensor or "battery" for the battery temperature sensor.

In off-line mode the capabilities are limited to basic word recognition, so we decided to implement keywords that can trigger each button from the available sensors. After several tests we concluded that the free speech recognition would not produce the desired level of accuracy without a deeper adaptation of the existing model, so this functionality was disabled. Furthermore, the English dictionary does not contain certain keywords used by LabTablet users and researchers in general such as "descriptor".

The available commands are thus related to the features in the Field Mode, allowing the user to select some of those sources without directly tapping on the screen. This functionality is a simpler version of the one that it is available for the on-line mode which allows the selection of any of the sources.

4.2 On-line usage

As expected, the *Google*'s Speech Recognition API, shows a high performance due to its access to larger computing resources and the use of improved mechanisms, such as voice sampling to improve the overall translation quality [4]. In addition to what the off-line mode implements, on-line users can also save voice recordings that are automatically transcribed, and navigate through the descriptor selection interfaces.

Having an always-on, continuous speech recognition can compromise the application's performance. Although *Google* allows its users to enable this kind of interaction with the "OK Google" trigger, it limits the possible words to a smaller set to avoid the impact on the operative system[4]. For this reason, the continuous speech recognition is only available to record text notes and is automatically deactivated whenever the user finishes recording each note.

Additionally LabTablet allows users to customize the on-line keywords in both Portuguese and English to use in the on-line mode[5].

4.3 Issuing simple commands

On the field mode the user can turn on the speech recognition by enabling the hands free mode, as seen in Figure 1b. Speech recognition will be available until user gives a "shut down" command—or taps on the switch again. LabTablet can recognise simple commands to read the available sensors (visible in this interface, such as temperature, GPS location, battery temperature or luminosity). In both cases, the device gives audible feedback about the success or failure of the readings, allowing the user to work without needing to look at the device's screen.

When adding specific descriptors, the user can also navigate through the existing profile to add metadata with a step-by-step procedure, accompanied by appropriate feedback. The process is illustrated in Figure 2.

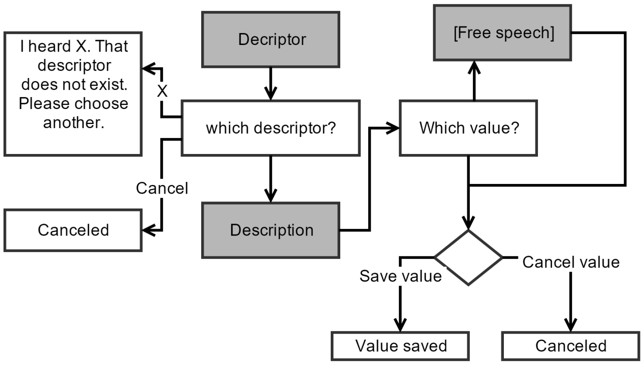


Figure 2: Descriptors gathering process flow

The grey rectangles indicate voice input and the white rectangles indicate feedback that the device gives to the user. On this example the user wants to select the descriptor "Description" and assign a value to it. To do so, the user asks the device to listen for the descriptor he wants by saying the reserved keyword "Descriptor". All feedback is given by the device through speech using Android's text-to-speech (TTS) Voice capabilities. In the case the device does not recognize the dictated descriptor it waits for new voice input until the user indicates an existing descriptor or chooses to cancel the process.

If a descriptor is recognized the device then proceeds to ask for a value to assign to that descriptor. It then listens to free speech and writes the dictated value on a text field.

5. CONCLUSIONS

Research data management tools are being proposed to deal with the publication of datasets. One of their main limitations is the need for good quality metadata records, and the scarce time researchers have to create them.

Enabling the LabTablet with speech recognition makes it more likely for records, which would only be available in paper, to be generated in digital form from the start. Together with Dendro, one of the research data management platforms Labtablet can synchronize with, an environment where more datasets get to the publication phase is in place. Voice recognition is seeing increasing usage and its use is making many routine tasks easier. In the context of metadata collection, though, it can facilitate the researchers interaction with tools to create valuable standards-compliant metadata records.

The voice interface in LabTablet is preliminary, but robust enough to be put to test with a panel of researchers, continuing previous test of the LabTablet and Dendro tools. The implementation also provided insights on emerging libraries, such as the MARF framework[6] can also be easily included to take advantage of additional voice recognition modules for both on-line and off-line work. The resulting implementation allows the users to switch between the installed voice recognition platforms, allowing the application to adapt itself to specific scenarios.

6. ACKNOWLEDGMENTS

This work is financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme, and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project POCI-01-0145-FEDER-006961.

7. REFERENCES

[1] R. C. Amorim, J. A. Castro, J. Rocha da Silva, and C. Ribeiro. Labtablet: Semantic metadata collection on a multi-domain laboratory notebook. In *Metadata and Semantics Research Conference Proceedings*, pages 193–205. Springer International Publishing, 2014.

[2] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, A. Rudnicky, et al. Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 1, pages I–I. IEEE, 2006.

[3] J. Rocha da Silva, J. A. Castro, C. Ribeiro, and J. Correia Lopes. Dendro: collaborative research data management built on linked open data. 2014.

[4] J. Schalkwyk, D. Beeferman, F. Beaufays, B. Byrne, C. Chelba, M. Cohen, M. Kamvar, and B. Strope. "your word is my command": Google search by voice: A case study. In *Advances in Speech Recognition*, pages 61–90. Springer, 2010.

[4] High-end devices can also have a dedicated processor that allows decreasing the power consumption while maintaining the overall responsiveness

[5] Wider support is also an ongoing task.

[6] `http://marf.sourceforge.net/`