

# A European accent map

Lee Harkis and Martin Cooke

Department of Computer Science, University of Sheffield, UK

E-mail: harkis@bcs.org.uk, m.cooke@dcs.shef.ac.uk

## ABSTRACT

The European Accent Map project aims to provide an internet resource for the collection and dissemination of regional speech data. This paper describes the methodological and technical issues raised by such a proposal, amongst which are the questions of what speech material to collect, how to allow contributors to adequately self-record their linguistic background, and how to capture audio across the web in a secure and efficient fashion. The current prototype solution employs a distributed architecture based on Java, XML and Unicode technologies. The tool is illustrated via a series of screenshots. The process of localising the tool to European languages is described. The European Accent Map currently stands as a proof-of-concept and future development will focus on larger-scale data collection activities.

## 1. INTRODUCTION

The European Accent Map (EAM) project aims to provide an internet resource for the collection and dissemination of regional speech data. The principal objective is to develop a teaching tool for both phonetics and advanced foreign language learners. However, information contained in such data is also of value for linguistic research into accent variation [1]. Further, it is increasingly important to reflect accent diversity in speech technology, both through synthesis of regional accents and in the training of speech recognisers. Diversity, quantity and quality are measures of the usefulness of an accent resource. The larger the sample range, the broader the potential uses of the data. Web technologies for collection and dissemination can be used to gain the widest exposure, providing access to a large group of potential data providers.

The development of such a facility raises several interesting methodological and technical issues. What target material should be collected from the contributing speakers? What background information is needed to make full use of their accent sample? How should the tool be internationalised and utilized in each country or accent region? How can audio material be collected securely over the internet while guaranteeing privacy for contributors? How should the information collected be disseminated via the web? Section 2 discusses these issues in more detail.

This paper describes a tool which attempts to provide the functionality outlined above. The tool has been designed to allow European linguists to fully localize the tool for their region. Amongst other localization activities, linguists can

define word sets and phrases relevant to the study of accents in their region. They can also upload maps which are used by contributing speakers to identify their place of birth and subsequent places of residence. Speakers provide their accent samples by responding to the list of words and phrases defined by the linguist. Users (as opposed to contributing speakers) of the tool are able to browse the map and listen to accent samples. Section 3 depicts the functionality of the tool through a series of screenshots. Architecture and implementation are described in section 4.

## 2. METHODOLOGICAL AND TECHNICAL ISSUES

### 2.1 Target speech material

Accents can vary depending on the information being conveyed and its context. To gather a wide range of vocalizations, the forms that are used in EAM are isolated words, phrases (eg fig. 1) and free-speech. The latter provides an opportunity for a contributor to express themselves language without the constraint of a script. The combination of all three forms gives an accent sample covering differing speech styles. Samples collected must be of adequate quality for analysis purposes but not so excessively large that transfer over the 'net is unacceptably slow.

The information to be gathered can be defined in word and phrase sets created and maintained by linguists. This permits customisation to specific collection needs.

Larry bought a new car because somebody crashed into his old one.

Did you do that drawing yourself? I didn't know you were into painting!

Mary was hurting his feelings so he told her to get lost.

I put all those things away but the house still looks very cluttered

Will you get it open or shall I cut it out ?

You told me to shut up so I think I better go home now

Did someone send you a love letter?

They'll have to do to several exams before getting a pass in that unit

**Figure 1:** Example phrase set for British English, specialised for northern vs southern accent contrasts.

## 2.2 Speaker background

To aid analysis of captured samples, recordings must be associated with data on the social and geographical background of the speaker. Social data such as date of birth, occupation, gender and level of education can be provided directly by the contributor using simple web forms. In addition the speaker can be asked to describe their accent, the place they consider to be their home area and from where they think their accent originates. To capture the geographical background, speakers can click on a map and indicate the period for which the location was their residence. This makes for a detailed, accurate and easy to use interface, which also allows accent samples to be presented visually. Figures 5 and 9 illustrate the map-based interface.

## 2.3 Localisation

Localisation describes the process by which a piece of software is made available for a given language/region. Software which supports localisation is said to be internationalised. It is clear that EAM should support localisation at several levels. First, as for any tool, all instructions/dialogs presented to users should be stored in a system dictionary which can be tailored to the local language. More importantly, since the core purpose of the tool is to handle linguistic diversity, it should be possible for linguists to add word sets and regional maps for any European language. By providing central longitude, latitude and scale references for the latter, samples can be translated on to other maps and the collected data used in a wider context. Multiple maps for the same geographical area allow fine control of the level of information provided.

## 2.4 Security

A critical issue for contributing speakers is security. Any software designed to collect acoustic data remotely must deal with the threat or suspicion of internet ‘bugging’. Consequently, software libraries which allow access to microphone resources on a remote machine typically require the user to set certain audio permissions before recording can take place. An alternative is the use of digital certificates, which provide a measure of trust in the origin and purpose of a piece of software. In the current implementation, audio permissions are used. Consequently, collection of accent data is best achieved in a controlled setting (e.g. a university lab class). Such restrictions do not apply to sound output for users browsing the accent map.

## 2.5 Usability

The contributors and audience for this project are expected to come from diverse backgrounds so it is important that the tool is easy to use and that it does not alienate any potential user groups. Localisation helps by presenting the interface in the language of the contributor. Where possible, it is preferable to detect this automatically, but any tool should allow the selection of a different language.

Usability can be promoted through the use of a familiar idiom. Here, the map concept is used throughout to promote both rapid collection of a contributor’s linguistic history and in the presentation of information to users.

## 3. A TOUR OF EAM

The following screenshots give an overview of the interface and functionality of the tool as it currently stands. A range of languages are used in these displays to illustrate the multilingual nature of the tool.



**Figure 2:** Language selection. The local language is automatically selected and presented, but the option of using other languages is available.



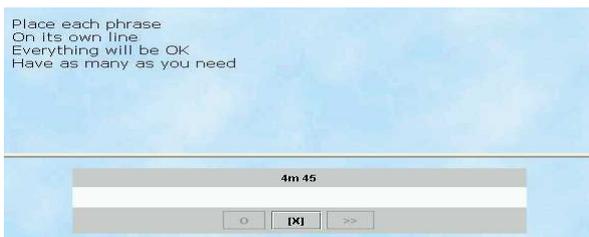
**Figure 3:** A choice of word sets available for a given language is presented to the speaker. Each language can have multiple wordsets available associated with different maps.



**Figure 4:** The contributor is presented with a simple form on which to detail their social background.



**Figure 5:** Geographical background. By clicking on a map, the speaker can indicate the areas in which they have lived, and the length of time in each.



**Figure 6:** Recording an accent sample (in this case, a list of phrases). A Java applet is used to control the audio recording. The time indicated is the remaining time allowed for recording.



**Figure 7:** Searching for samples. Complex queries can be constructed by searching speaker background, using multiple fields.



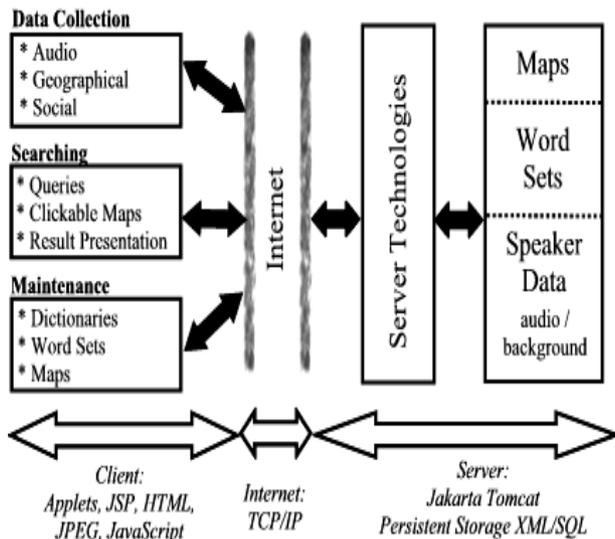
**Figure 8:** Search result. From a list of contributions matching a query, a linguist selects a sample and is presented with a screen which allows access to each audio sample, the speaker background, the set of words used and a map showing the period for which the speaker lived in each point on the map.



**Figure 9:** A speaker's geographical history. Each circle indicates the time in years the speaker lived in that location.

#### 4. IMPLEMENTATION

This section gives a brief outline of the technologies applied in the construction of EAM. Figure 10 illustrates its overall architecture, which is based around a 3-tier design [2]. This allows components to be interchanged as necessary. For example, it is possible to replace the database system used independently of other components, an approach which speeds development and reduces maintenance costs.



**Figure 10:** Architecture of the European Accent Map.

Java was chosen as the main technology for implementation for a number of reasons. First, it is virtually platform-independent, allowing a wide range of end users. Second, it has extensive function libraries and unparalleled support for large web applications. Crucially, it has the ability to record and playback audio remotely via applets. Finally, it supports internationalisation more fully than any other current option.

JavaSound, part of the Java Media Framework [3], provides standard libraries to capture, process and playback audio samples. Its integration into Java 1.3.1 allows JavaSound to be used in a Java applet (see figure 6). An applet is a program running within a web browser, and a combination of these technologies allows recording from the internet, independent of browser type and operating system, giving a huge potential audience. Currently, this is constrained by forcing each client to restrict access to recording in Java applets, but this issue can be resolved through the use of digital certificates to establish the trust of the client.

Java implements internationalisation via ‘locales’ [4]. A locale makes it straightforward to adapt the entire presentation layer to the local language. This has been used to adhere to recognised standards for international presentation (number formats, dates, etc). Unicode [5] is supported by each component in the system, leading to near universal character coverage.

XML (eXtensible Mark-up Language) [6] was chosen as a portable data format to store local language dictionaries. This allows editing of the dictionaries directly from data source if required.

English is used as the interlingua of the system, providing a common reference for all language entries. It is also the language used for administrative, searching and linguist functions at present.

## 5. FUTURE WORK

EAM is currently a proof of concept and awaits large-scale data collection. The technologies required for its foundation have been applied and a basis for future development is in place. There are recognised limitations and scope for much future work.

When collecting geographical information from the speaker, there is only the option to indicate residence on one map. Ideally, this would be replaced by a system of maps allowing the precise location to be defined by zooming into or out of related maps. There is a similar restriction on searching for samples. A better solution would be to use an existing map data resource such MapInfo, MultiMap.com or a similar GIS product integrated into the tool.

Since the tool is totally web-based, it is not ideal for collection of accent material via laptop in remote areas not connected to the net. A useful facility would be to allow local storage of material for later uploading.

At present, no acoustic analysis of the samples has been attempted. In principle, accent samples could be segmented at the phone level via forced alignment, and subsequent analyses used to measure relevant features such as formant frequencies, permitting a richer visualization of accent variety and variation.

Finally, no claims are made as to how representative any given accent in EAM is. It is clear that due to an uneven demographic distribution of both access and desire to use the internet, there is likely to exist significant social biases in the material collected.

## ACKNOWLEDGEMENTS

Drs María Luisa García Lecumberri, Gerry Docherty and Joan Beal provided useful guidance at various stages of the project.

## REFERENCES

- [1] G. Docherty and P. Foulkes (eds), *Urban Voices: Accent Studies in the British Isles*, Arnold, 1999.
- [2] A. Chaffee, “One, two, three, or n tiers?”, JavaWorld, available at <http://www.javaworld.com>, 2000.
- [3] Sun Microsystems, The Java Media Framework, see <http://java.sun.com/products/java-media/jmf/>, 2002.
- [4] A. Deitsch and D.A. Czarnecki, *Java Internationalisation*, O’Reilly, 2001.
- [5] <http://www.unicode.org>
- [6] <http://www.w3.org/XML>