LANGUAGE DOCUMENTATION & CONSERVATION

Vol. 2, No. 1 (June 2008), pp. 166-175
http://nflrc.hawaii.edu/ldc/

EMU Speech Database System

from The EMU Project

Reviewed by Briony Williams, *Bangor University, UK*

Document URI: http://hdl.handle.net/10125/1807

1. INTRODUCTION. EMU (http://emu.sourceforge.net) is, as described on its homepage:

> … a collection of software tools for the creation, manipulation and analysis of speech databases. At the core of EMU is a database search engine which allows the researcher to find various speech segments based on the sequential and hierarchical structure of the utterances in which they occur. EMU includes an interactive labeller which can display spectrograms and other speech waveforms, and which allows the creation of hierarchical, as well as sequential, labels for a speech utterance.

EMU is closely integrated with the free "R" statistics software (http://www.r-project.org), and can also be used with the proprietary "Splus" statistics software. Its particular features include the following:

- Perform database queries at several linguistic levels simultaneously, using a "wizard" to construct database queries easily.
- Pass the results of database queries seamlessly to the "Splus" or "R" statistical software.
- Perform semi-automatic annotation at higher linguistic levels, based on a user-specified context-free grammar of units at lower levels.

EMU also includes other features commonly found in speech analysis software, such as Praat (http://www.praat.org) or WaveSurfer (http://www.speech.kth.se/wavesurfer/):

- Display waveform, spectrogram, and F0.
- Playback all or selected parts of the utterance.
- Annotate utterance at one or more levels.
- Create and display spectral slices.
- Calculate and display formant frequencies and bandwidths.

Licensed under Creative Commons
Attribution Non-Commercial No Derivative License

E-ISSN 1934-5275

However, EMU has strengths and weaknesses that are complementary to Praat, WaveSurfer, etc., and would be best used as part of a comprehensive suite of tools incorporating the strengths of all such software.

2. OVERVIEW.

2.1. FUNCTIONALITY. EMU has two main views of the data. The initial view on opening EMU is the "signal view," showing time-aligned annotations together with the speech waveform and spectrogram. An example is seen in figure 1 below, where the annotation (in this case, a syllable-level annotation) has been created manually.

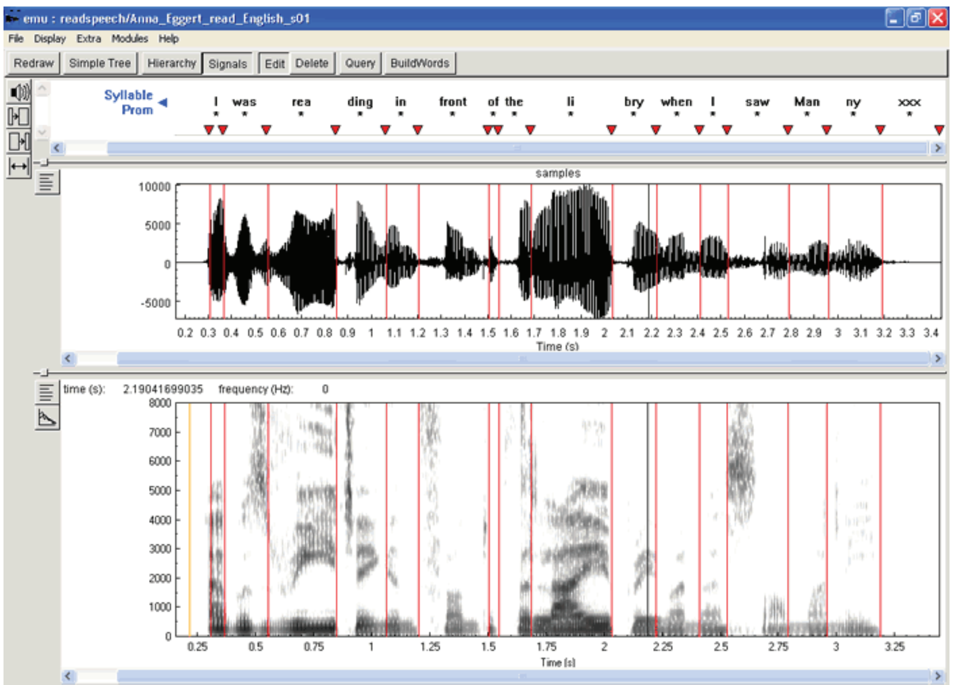

FIGURE 1. Signal view showing time-aligned annotation, waveform and spectrogram.

The second view is the "Hierarchy view," showing both the lowest level displayed above, and also higher linguistic levels in a tree structure. This is the point where EMU 's functionality begins to diverge from that of comparable software. A simple example is seen in figure 2 below. Given the lowest (time-aligned) level, the higher levels can be annotated manually in this window, or can be annotated semi-automatically if the user has defined the relevant rules.

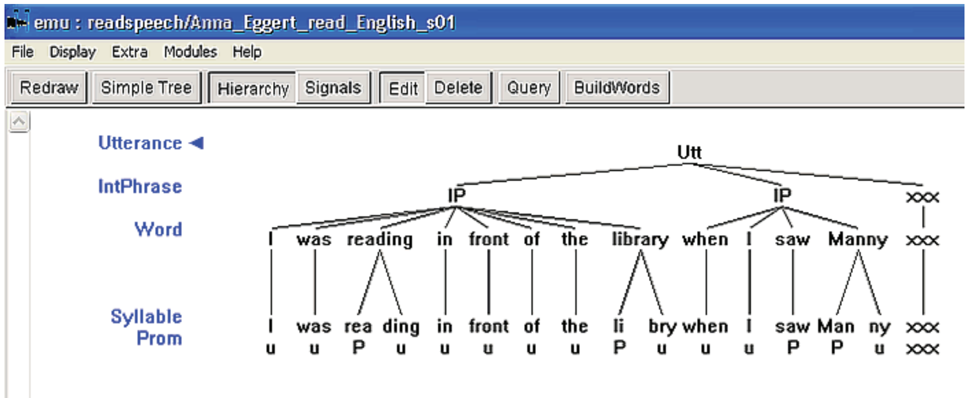

FIGURE 2. Hierarchy view showing all linguistic levels used.

2.2 RESOURCES. EMU has a significant amount of documentation. As well as the extensive help files online at http://emu.sourceforge.net, there is also a book (Harrington and Cassidy 1999) and several papers (Harrington et al. 1993; Cassidy and Harrington 1996; Cassidy 1999; Cassidy and Harrington 2001; Harrington et al. 2003; Bombien et al. 2006). In addition, there is a (very low-traffic) mailing list, "emu-users" (https://lists.sourceforge.net/lists/listinfo/emu-users), which is moderated by Steve Cassidy, EMU 's chief software developer.

3. USING EMU . As well as the two major "views" described above, EMU contains several other tools and utilities (including a utility that converts label files to and from Praat TextGrid format). Some of these tools are alluded to below, but this review cannot cover the full breadth of functionality available.

3.1 TEMPLATE FILE. The interface between EMU and a database of speech files is the "template file," a key component in setting up EMU for use with a specific speech corpus. The template file specifies the various linguistic levels to be used and the relations between them, together with the labels that are permitted at each level, and the location of each type of file (speech waveform, F0 and formant track files, and all annotation files). A template wizard within EMU simplifies the process of creating a new template, while the demo database supplied with EMU contains an example template file for possible copying. Even with this assistance, however, setting up a new template file is sometimes not completely straightforward, and this represents a major part of the learning curve needed to master EMU .

3.2 SEGMENTAL LABELING. Segmental labeling by hand (or time-aligned labeling at any linguistic level) is carried out in the Signal view window, and is done in a way similar to other comparable software (Praat, WaveSurfer). The display contains all necessary components (waveform, spectrogram, playback facility, selection by cursors). A particularly helpful feature of this display is the fact that, having created a new segment by left-clicking at the desired time, the researcher can then right-click on the new place-holder and display a menu of all permitted segments at this level. This feature (shared with WaveSurfer) has the following advantages:

- Speed: "Right-click, select symbol" is usually much quicker than pausing to type in the desired symbol.

- Avoiding human error: The provision of a closed set of permitted labels minimizes the scope for human error when entering a new label.

3.3 ACOUSTIC PARAMETER FILES. EMU contains a utility to run the former ESPS pitch and formant tracker code over the speech data, producing data files containing the F0, formant frequencies and bandwidths, intensity, etc. These data files can later be queried using the built-in EMU database query tool. The output of these queries can then be passed seamlessly to the "R" statistics software, to yield infomation on the acoustic parameters of various linguistic entities (e.g., to find out the mean F1 and F2 of a stressed /a/ that is syllable-final in a content word; or to test for any significant difference in the mean intensity of a stressed versus an unstressed vowel in non-word-final syllables).

3.4 HIERARCHICAL ANNOTATION. In the Hierarchy view, the higher-level annotation may be carried out manually in tree format. This view likewise has the right-click menu of permitted labels at each level, as specified in the template file for the given database. The annotation is carried out as follows:

- Left-click on a level to create a new unit at that level.

- Right click on the new unit to select a label from the set of permitted labels.

- Left-click and drag from a unit at one level to one or more units at the next level down, to indicate the dominance relations.

The "hlb" ("higher-level bracketing") file thus produced contains information on the hierarchical structure of these higher-level annotations. In conjunction with the time-aligned annotation file, EMU can derive segment start and end times for units at all linguistic levels.

It is also possible to carry out higher-level annotation automatically or semi-automatically (i.e., with manual post-editing). This can be done by writing a set of context-free rules to describe the relationship of units on two adjacent levels (e.g., word and syllable). These rules are then used by EMU 's "AutoBuild" facility to build up a segment of the tree structure automatically. This feature is best used when the data derive from several speakers uttering the same stimuli (isolated sentences or a set passage), since in this case the same rules can be used for all speakers. As the task of setting up the rules is non-trivial, it is worth doing so only when the ruleset is to be re-used for more than one speaker. In such a case, however, it can make possible a very significant saving of time compared to manual high-level annotation. This is a unique feature of EMU that helps to take it beyond a simple transcription tool to a full speech database management system.

3.5 QUERY TOOL. Once the speech database has been annotated at all linguistic levels, it can be queried. EMU contains a database query wizard to help in the formulation of queries, so it is not necessary to learn the query language. Queries can be of two kinds:

- Simple queries focus on only one level of annotation. They can specify preceding and following segments, and can use variables.

- Complex queries can address two or more annotation levels simultaneously, targeting segments according to their parent or child segments, as well as their context on the same level.

The output of queries is a set of segments with their filenames and locations within each file. This output can be used together with the track data files (intensity, F0, etc.) to extract the relevant acoustic parameters. The result can then be imported into other tools, such as R or Excel.

3.6 INTEGRATION WITH STATISTICAL SOFTWARE. The EMU distribution includes special-purpose functions for use with “R” or Splus, which enable these statistical packages to accept the results of EMU database queries and perform statistical analyses on them. Splus is a commercial product that runs on Unix and Windows systems. R (http://www.r-project.org) is a freely available clone of Splus, and runs on Unix, Mac, and Windows, providing an environment for statistical computation and graphics. EMU ’s close integration with these packages is another feature that makes it especially suitable as a speech database management system.

4. STRONG POINTS. Certain features of EMU (e.g., segmental annotation by hand, creation of data files) are comparable to the features offered by other software, such as Praat. However, other features of EMU represent particular strengths, and these are described below.

4.1 HIERARCHICAL ANNOTATION. The tree-structured annotation format is unique to EMU , and enables it to encapsulate relations between units that cannot be shown in other software. For example, EMU permits not only one-to-many relationships across adjacent levels, but also many-to-many relationships. Thus two or more consecutive units at a given level may map to just one unit at the next level down. So, for example, two underlying short syllables on the syllable level might be mapped to one long vowel on the phonemic level, or two underlying phonemes might be mapped to one acoustic unit at the phonetic level. This many-to-one mapping is not possible in the Praat TextGrid format. An example in EMU is seen in figure 3 below in the case of an utterance in the Welsh language: the red boxes highlight the points where two underlying phonemes have been mapped to one surface phone. This shows the following cases:

- A word-final voiceless alveolar stop /t/ is adjacent to a word-initial voiceless alveolar stop.

- A word-final voiced alveolar stop /d/ has been deleted before a voiceless lateral fricative.

- A word-final voiced alveolar tap has been deleted, and the preceding segment (/t/) is identical to the following word-initial segment.

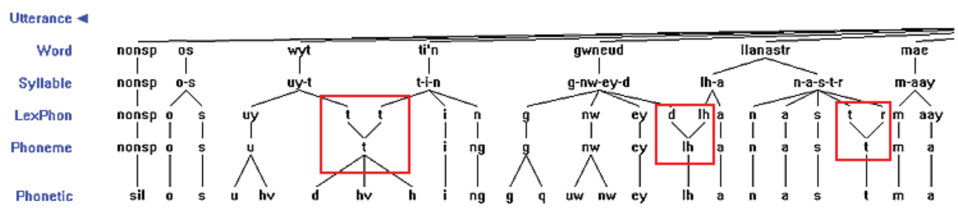

FIGURE 3. Example of two units at a higher level mapping to one unit at a lower level in EMU .

For purposes of comparison, this hierarchical label file was converted to a Praat TextGrid, using EMU 's label conversion utility. The result can be seen in figure 4 below, where the red boxes mark the portions of interest. In each case of many-to-one mapping, the Praat tier has converted the "many" symbols to a single composite symbol, formed by concatenating the symbols involved and separating them with an underscore character (e.g. "t_t"). This means that, although the time relations are retained, it is no longer possible to derive the underlying units from the unitary symbols available, and hence some information has been lost compared to the representation in EMU.

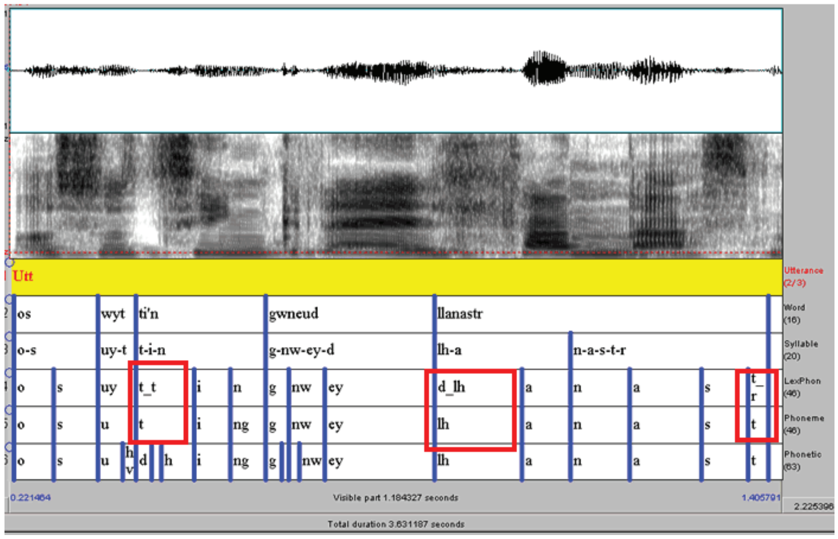

FIGURE 4. Praat TextGrid version of the data in Figure 3.

4.2 DATABASE QUERY TOOL. The EMU database query tool is another advantage. A query "wizard" forms part of EMU , so that it is not necessary to learn the query language (alternatively, it is possible to enter queries manually). The query tool exploits EMU's hierarchical structure to the fullest extent, since it makes it possible to formulate queries both horizontally (on the same level as the target segment) and vertically (across several levels simultaneously). It is possible that a Praat script may exist that will do this for a given configuration of levels and possible segments. However, EMU has a more generalized database query facility that can be used across all configurations.

4.3 SEMI-AUTOMATIC ANNOTATION. The semi-automatic annotation at higher levels is a significant strong point for a database consisting of speech that has been read. Since the same text has been used by several speakers, the same across-level mapping rules can be used for all speakers (for mapping, e.g., from syllables to the words of the text). It would be less relevant in the case of a database of spontaneous speech, where the data would differ completely across speakers. However, for speech that has been read, it can offer a very significant saving of time in annotation, since usually only checking and light editing will be needed.

4.4 PRACTICAL ADVANTAGES. EMU has the following practical advantages:

- Multi-platform (Windows, Mac, Unix/Linux).

- Open-source (and hence free of charge, and open to improvement from the developer community).

- Ongoing support by the chief software developer via an email discussion list.

5. WEAKER POINTS. Certain features of EMU are less strong, such as the following:

- The syntax of the database template file is complex, and can be a source of errors. Although the built-in template wizard assists a great deal, creating a functioning template file can still sometimes be less than straightforward. For example, one small irritation when running under Windows concerns the format of the "path" specification in the template file. Although Windows path names use backward slashes, in the template file these need to be converted to forward slashes when specifying the location of database files.

- Compared to Praat, EMU offers less scope for creating and running user-written scripts. EMU 's user-scripting facilities are confined to the AutoBuild feature, for building up levels of annotation automatically. There is no equivalent to the more extensive soundfile manipulation features of Praat, for such operations as splitting a long file into sub-files and saving them, or applying transformations to the duration or F0 of the data.

- A steep learning curve is involved on initial exposure to EMU. It is not possible to use EMU without some kind of database template file, however simple, and so this feature must be mastered before continuing. By comparison, Praat can be used at a basic level very soon after installation, with a minimum of learning of the software.

- EMU does not appear able to handle Unicode characters (unlike Praat, which has support for Unicode).

6. CONCLUSIONS. Overall, EMU is a strong candidate for a speech database management system. It has strengths that are complementary to those of Praat and WaveSurfer, and would form a valuable part of a comprehensive suite of software for creating and analyzing a speech database. Praat displays great flexibility in user-written scripting facilities, and a very wide range of sound file manipulation features. EMU, on the other hand, shows less flexibility, but is more focussed on a particular set of tasks. These include such functions as: semi-automatic annotation, user-friendly database querying, and integration with statistical software. For at least one of these tasks (semi-automatic annotation), EMU is the only tool currently available; and for all of them, EMU is already provided with facilities that in Praat would require the creation or identification of extra scripts.

The best strategy is to play to the strengths of the particular software that is used. Hence, for practical speech database research, the following suggestions can be made:

- Praat is best used for: manual annotation at the lower (time-aligned) level, sound file manipulation and synthesis, and work that requires the use of Unicode characters.

• EMU is best used for: automatic annotation at higher levels, database queries, and integration with statistical software. In addition, EMU can be used to handle cases of many-to-many mapping across levels that cannot be handled by other software.

Pros:	- Built-in database query language, with easy interface. - Tree-structured annotations. - Semi-automatic annotation at higher levels. - Integration with Splus/R statistics software.
Cons:	- Steep learning curve. - Unforgiving template syntax.
Primary function:	Speech database annotation at several levels, and subsequent statistical analysis of its acoustic parameters.
Platforms:	Windows, Mac, Unix/Linux
Open Source?:	Yes: Source code (and binaries) downloadable from http://emu.sourceforge.net/
Proprietary?:	No: EMU is freeware.
Reviewed version:	EMU 2.0
Application size:	35.3 MB for entire directory, containing all tools.
Documentation:	http://emu.sourceforge.net

References

Bombien, Lasse, Steve Cassidy, Jonathan Harrington, Tina John, and Sallyanne Palethorpe. 2006. Recent developments in the EMU speech database system. *Proceedings of the Australian Speech Science and Technology Conference*, December, 2006, Auckland, Australia, 313–316.

Cassidy, Steve, and Jonathan Harrington. 1996. EMU : An enhanced hierarchical speech data management system. Paper presented at the Speech Science and Technology Conference, December 4–7, 1996, Adelaide, Australia.

Cassidy, Steve and Jonathan Harrington. 2001. Multi-level annotation in the EMU speech database management system. *Speech Communication* 33(1–2):61–77.

Cassidy, Steve. 1999. Compiling multi-tiered speech databases into the relational model: Experiments with the EMU system. In *Proceedings of the Sixth European Conference on Speech Communication and Technology (Eurospeech '99)*, September 5–9, 1999, Budapest, Hungary, 2239–2242.

Harrington, Jonathan, and Steve Cassidy. 1999. *Techniques in speech acoustics*. Dordrecht: Kluwer Academic Press.

Harrington, Jonathan, Steve Cassidy, Janet Fletcher, and Andrew McVeigh. 1993. The Mu+ speech database system. *Computer Speech and Language* 7(4):305–331.

Harrington, Jonathan, Steve Cassidy, Tina John, and Michel Scheffers. 2003. Building an interface between EMU and Praat: A modular approach to speech database analysis. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPHS 2003)*, August, 2003, Barcelona, Spain, 355–358.

Briony Williams
b.williams@bangor.ac.uk