MAT: a tool for L2 pronunciation errors annotation

Renlong Ai, Marcela Charfuelan

DFKI GmbH, Language Technology Lab
Alt-Moabit 91c, 10559, Berlin, Germany
firstname.lastname@dfki.de

Abstract

In the area of Computer Assisted Language Learning(CALL), second language (L2) learners' spoken data is an important resource for analysing and annotating typical L2 pronunciation errors. The annotation of L2 pronunciation errors in spoken data is not an easy task though, normally it requires manual annotation from trained linguists or phoneticians. In order to facilitate this task, in this paper, we present MAT a web-based tool intended to facilitate the annotation of L2 learners' pronunciation errors at various levels. The tool has been designed taking into account recent studies on error detection in pronunciation training. It also aims at providing an easy and fast annotation process via a comprehensive and friendly user interface. The tool is based on the MARY TTS open source platform, from which it uses the components: text analyser (tokeniser, syllabifier, phonemiser), phonetic aligner and speech signal processor. Annotation results at sentence, word, syllable and phoneme levels are stored in XML format. The tool is currently under evaluation with a L2 learners' spoken corpus recorded in the SPRINTER (Language Technology for Interactive, Multi-Media Online Language Learning) project.

Keywords: CALL, phonetic annotation, L2 pronunciation errors

1. Introduction

In this paper we introduce MAT (MARY Annotation Tool) a tool for annotation of second non-native (L2) learners' pronunciation errors at phoneme, syllable, word and sentence level. The tool is based on MARY TTS (Schröder et al., 2011), which is a flexible tool for research, development and teaching in the domain of text-to-speech (TTS) synthesis.

MARY TTS includes among others, tools for text analysis (tokeniser, syllabifier, phonemiser), phonetic alignment to speech via the EHMM force alignment tool[1], speech signal processing, text and acoustic resources ready to use and available in several languages like English, German, Italian, etc., (see latest version in MARY TTS repository[2]).

One of the tasks in the SPRINTER[3] (Language Technology for Interactive, Multi-Media Online Language Learning) project (Ai et al., 2014), is to provide feedback about pronunciation errors to L2 learners of a language. In order to address this task, we have started to analyse available tools that allow us to annotate these type of errors in learner's pronunciation recordings.

We have found that most of the state of the art tools that can be used in this task, like SPPAS (Bigi and Hirst, 2012), EasyAlign (Goldman, 2011), Train&Align (Brognaux et al., 2012), have in common components like a text analyser and a speech aligner, which we have available in MARY TTS. Also we found, that these tools are mainly intended to perform phonetic alignment or prosody analysis in general. To the best of our knowledge there is no tool available to annotate, in particular, L2 pronunciation errors at various levels. Therefore we have designed our own tool based on MARY TTS, covering annotation of errors not only at phoneme and syllable level but also at word and sentence level.

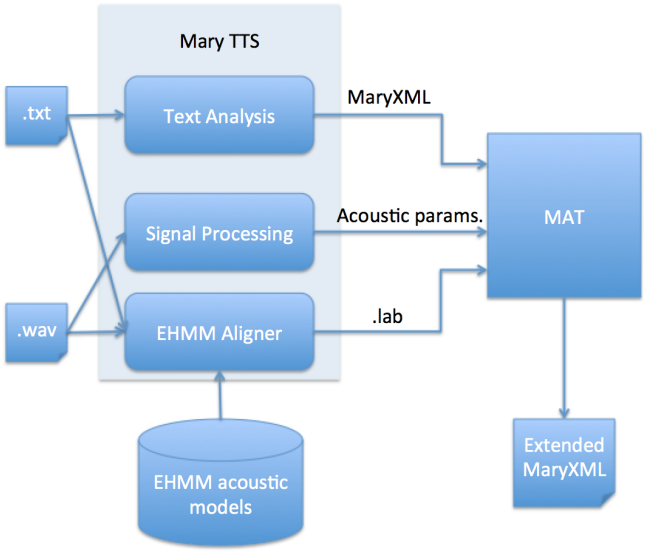

Figure 1: MAT: NLP components.

The paper is organised as follows. In Section 2. we describe how the tool was designed, taking into account previous studies in error detection in pronunciation training (Witt, 2012; Strik et al., 2009). In section 3. we explain the main features of the MAT tool and one scenario of annotation. We conclude in Section 4. summarising main features of the tool and presenting some ideas for future work.

2. MAT Purpose and Design

One of the objectives in the SPRINTER project is to provide automatic feedback to learners of a second non-native (L2) language. One of the recent techniques to provide automatic feedback about pronunciation errors is to detect

[1] http://festvox.org
[2] https://github.com/marytts
[3] http://sprinter.dfki.de/

Level	Errors	Description
Phoneme	Deletion	The phoneme is deleted in the learner's utterance
	Insertion	A phoneme is inserted after the phoneme
	Distortion	The phoneme is distorted in the learner's utterance
	Substitution	The learner substituted the phoneme with another phoneme
	Actually spoken	In case of phoneme insertion and/or substitution, the annotator can optionally write the inserted or substituted phoneme.
	Foreign accent	The phoneme is pronounced with a foreign accent.
Syllable	Stress	The stress is misplaced by the learner.
Word	Foreign accent	The whole word is pronounced with a foreign accent.
	Long/short pause before/after word	L2 learners sometimes make long pauses in their pronunciation because of hesitation. Actually long pauses might no be considered as errors, but we would like to have them annotated to study their effect/correlation with alignment errors, intonation problems, etc.
Sentence	Rhythm	The rhythm of the whole sentence is not smooth.
	Intonation	The sentence has problem with intonation.
	Score	A score (1-10, 10 is the best) is decided taking into account the previous errors and having as a reference the teacher's recordings, if available, or synthesised speech; the use of synthetic speech will be experimental and subject to evaluation.

Table 1: MAT: pronunciation errors at various levels.

them using trained statistical models (Strik et al., 2009; Eskenazi, 2009). In order to train those models it is necessary to have annotated data, which might be difficult to obtain and laborious to generate. One database available, annotated in terms of word and phone level pronunciation errors is the one obtained in the ISLE project (Menzel et al., 2000). Although we might be able to use this data, we will generate more data in the SPRINTER project which we would like to annotate taking into account recent studies of error detection in pronunciation training. For example in (Witt, 2012) there is an excellent review about types of pronunciation errors that we have used as a base to design the levels of annotation in MAT. The levels of pronunciation errors included in the first version of MAT are presented in Table1. Further refinement of this list of errors will be done during the evaluation of the tool.

3. MAT Description

3.1. Components

As shown in Figure 1, MAT has as input the result of the text analysis performed by MARY TTS and the phonetic alignment result generated with the EHMM Aligner (.lab file). Acoustic parameters extracted from the audio signal are also used in MAT to display spectrum, pitch or energy graphs for further analysis. The output of MAT is an extended version of MaryXML, the internal XML-based representation language in MARY TTS. We have extended this representation in terms of pronunciation error properties at phoneme, syllable, word and sentence level, see Figure 3 and explanation below.

One advantage of using MARY TTS, is that it provides easy tutorials and tools to train acoustic models using the EHMM tool, also there are several speech databases freely available that can be used to train these acoustic models. Another advantage, is that it is possible to create acoustic models tuned to the type of data that is going to be aligned. As pointed out in (Brognaux et al., 2012), it is important to be able to train the acoustic models with the corpus to align, so the quality of the alignment improves. The possibility to use the many languages available in MARY TTS is another interesting feature, currently it supports more than seven languages.

3.2. Scenario of annotation

We tried to minimise the work of the annotator while annotating by using check boxes. As shown in Figure 2, check boxes are used in MAT for the exist-or-not errors like phoneme deletion, insertion, etc. Text fields are scarcely used, that is, just in cases that requires textual input, e.g. to annotate the actually pronounced phoneme by error phoneme substitution. A typical scenario of annotation is the following:

1. The annotator opens the learner folder which contains sub folders with audio files (.wav) and text files (.txt). The file names should be listed on the left side.

2. A default configuration file will be generated in the working folder. It contains the pronunciation errors presented in Table 1. By opening the config panel via clicking the config button, the annotator can select all these error categories or select just the errors he wants to annotate. There is also the possibility of adding a new error type at any level (phoneme, word, etc) and assign it as check box or text field.

3. It is suggested to have as reference a native version (gold standard) of each sentence that is going to be annotated. If this is the case, the directory where this data is located can be set using the Teacher folder open button.

4. The annotator can then start opening the utterances one by one by choosing from the list on the left. Phones, syllables and words are well aligned to the speech signal and presented in different colour in the table. The annotator could:

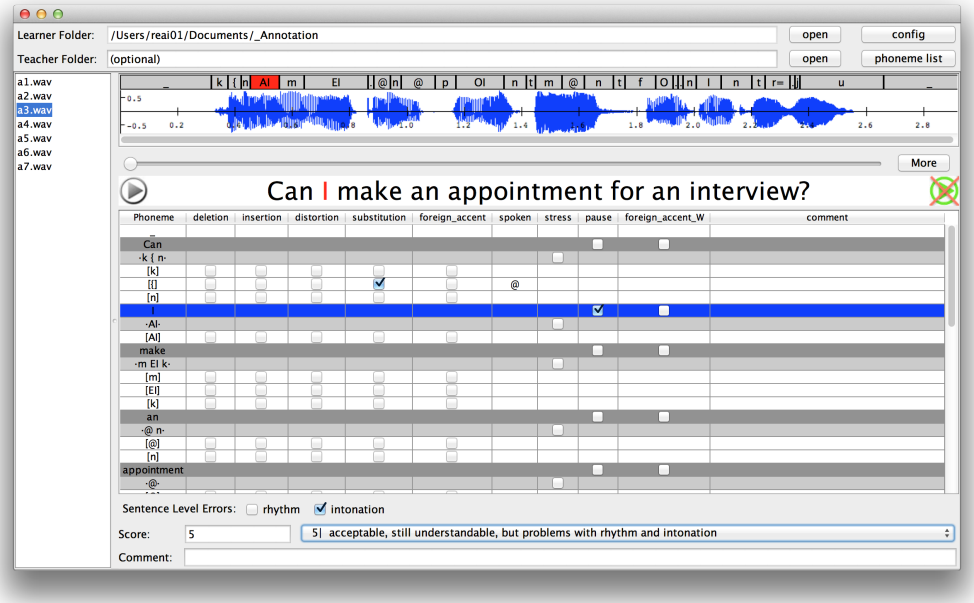

Figure 2: MAT: pronunciation errors annotation GUI.

- play a single word, syllable or phoneme by selecting a row in the table and hit space key;
- play a clip of the audio by choosing a range via mouse drag in the waveform and hit space key;
- play the whole sentence from the learner's recording, or from the corresponding audio from native speaker as a reference;
- toggle to the signal processing view to have a close look at the spectrogram, pitch contours and energy, Figure 4;
- open the phoneme list if he needs to check what token is used for the phoneme that the learner has actual spoken and enter this in the "spoken" column;
- even modify the alignment by dragging the bars separating the phonemes, in case he finds the time alignment is wrong for certain phonemes.

One example of annotation output in an extended MaryXML file is presented in Figure 3, it shows an excerpt from the annotation output corresponding to the sentence in Figure 2. The extended MaryXML includes the different levels, word, syllable and phoneme; we can see that the phoneme /{/ in the word "Can" was annotated because the learner pronounced it like a /@/, also a pause after the word 'I' has been annotated. These annotations can also be seen graphically in Figure 2. Besides, there is also annotated a problem with the sentence intonation. Taken into account these errors and other problems that are not shown in the XML excerpt, the annotator gives a score of 5 for the sentence (score from 1-10, 10 is very good).
It is important to notice that the annotator can compare the intonation and rhythm of the learner's sentence with the same sentence recorded by the teacher (gold standard). If recordings of a native speaker are not available, the MARY TTS synthesiser, or any high quality speech synthesiser,

```
<?xml version="1.0" encoding="UTF-8" >
<maryxml xmlns="http://mary.dfki.de/2002/MaryXML"
 version="0.5" xml:lang="en-US">
<p>
<s comment="" intonation="true" opened="1" score="55">
<phrase>
 <t g2p_method="lexicon" ph="' k { n"  pos="MD">
  Can
  <syllable  ph="k { n">
   <ph p="k"/>
   <ph p="{" spoken="@" substitution="true"/>
   <ph p="n"/>
  </syllable>
 </t>
 <t comment="pause after word"  g2p_method="lexicon"
  pause="true"  ph="' AI" pos="PRP" stress="true">
  I
  <syllable ph="AI">
   <ph p="AI"/>
  </syllable>
 </t>
 <t accent="H*" g2p_method="lexicon" ph="' m EI k"
  pos="VB">
  make
  <syllable accent="H*" ph="m EI k" stress="1">
    <ph p="m"/>
    <ph p="EI"/>
    <ph p="k"/>
  </syllable>
  ...
</phrase></s></p>
</maryxml>
```

Figure 3: MAT: example of annotation output in an extended MaryXML file.

can be used instead to synthesise the sentence and use it as a reference. Nowadays the level of speech synthesis has reached such a level that it has been already incorporated in L2 learning activities (Handley, 2009).

4. Conclusions

In this paper we have presented the design and development of MAT, a tool for L2 pronunciation errors annotation at phoneme, syllable, word and sentence level. The tool will be evaluated during the annotation of L2 learners record-

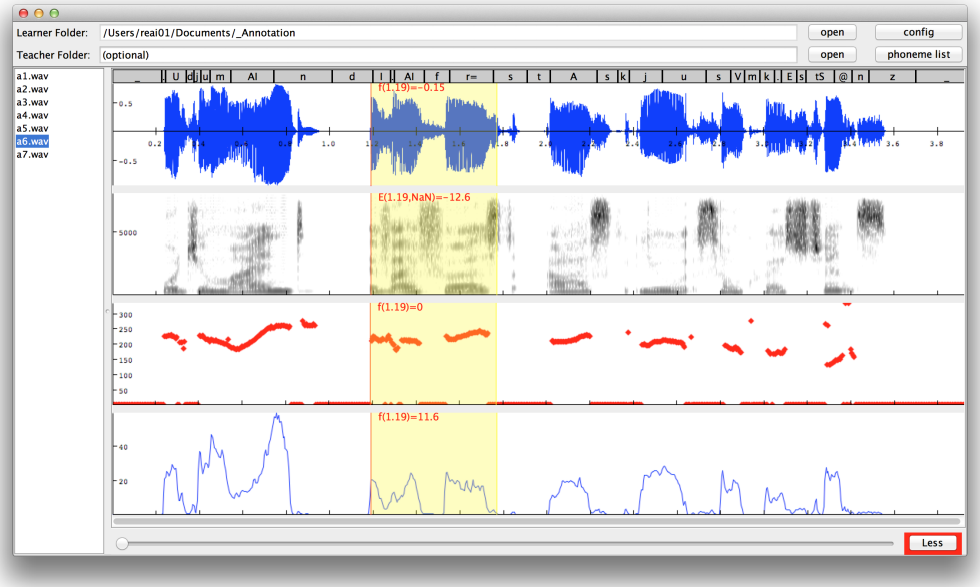

Figure 4: MAT: signal processing view.

ings to be collected in the SPRINTER project. The main features of the MAT tool are the following:

- Automatic segmentation of sentences into words, syllables and phonemes, with the possibility to play them separately or in sequence upon demand.
- Possibility to configure the type of errors to annotate in each level.
- Web-based and implemented in Java, hence accessible everywhere and independent from OS.
- Waveform alignment at different levels, display of spectrum, pitch contour and energy graph for further analysis (Figure 4) and also possibility to play teacher's audio or synthesised audio for comparison with learner's audio.

Different from other tools like EasyAlign, that presents alignment and annotation on Praat TextGrid tiers, MAT presents alignment and annotation in separate views. This is possible because alignment and annotation in MAT are stored in different files allowing to organise the GUI in a more user-friendly way. We are considering to port MAT's annotation result to Praat TextGrid format, so that would benefit the linguists who are used to Praat.
Regarding fine-grained annotation of prosodic errors, we are considering to further support the annotator by presenting measures of learner's speech deviations from teacher's speech, in terms of pitch and duration. This can be done by automatically comparing learner's and teacher's pitch contour and durations at phoneme level.

5. Acknowledgements

This research was partially supported by the German Federal Ministry of Education and Research (BMBF) through the project Sprinter (contract 01IS12006A) and the project Deependance (contract 01IW11003).

6. References

Ai, R., Charfuelan, M., Kasper, W., Klüwer, T., Uszkoreit, H., Xu, F., Gasber, S., and Gienandt, P. (2014). Sprinter: Language technologies for interactive and multimedia language learning. In *LREC*, Reykjavik, Iceland.

Bigi, B. and Hirst, D. (2012). SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody. In *6th International conference on Speech Prosody*, Shangai China.

Brognaux, S., Roekhaut, S., Drugman, T., and Beaufort, R. (2012). Train&Align: A new online tool for automatic phonetic alignment. In *IEEE Spoken Language Technology Workshop (SLT)*, pages 416–421, Miami, FL, USA.

Eskenazi, Maxine. (2009). An overview of spoken language technology for education. *Speech Communication*, 51(10):832 – 844.

Goldman, J. (2011). EasyAlign: an automatic phonetic alignment tool undert Praat. In *Interspeech*, Florence, Italy.

Handley, Zöe. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication*, 51(10):906 – 919.

Menzel, W., Atwell, E., Bonaventura, P., Herron, D., Howarth, P., Morton, R., and Souter, C. (2000). The ISLE corpus of non-native spoken English. In *LREC*, Athens, Greece.

Schröder, M., Charfuelan, M., Pammi, S., and Steiner, I. (2011). Open source voice creation toolkit for the MARY TTS Platform. In *Interspeech*, Florence, Italy.

Strik, Helmer, Truong, Khiet, de Wet, Febe, and Cucchiarini, Catia. (2009). Comparing different approaches for automatic pronunciation error detection. *Speech Communication*, 51(10):845 – 852.

Witt, S. M. (2012). Automatic error detection in pronunciation training: where we are and where we need to go. In *International Symposium on Automatic Detection of Errors in Pronunciation Training*, Stockholm, Sweden.