A tool for automatic feedback on phonemic transcriptions

Martin Cooke¹, Maria Luisa Garcia Lecumberri² and John Maidment³

¹Department of Computer Science, University of Sheffield (m.cooke@dcs.shef.ac.uk)
²Department of English Philology, University of the Basque Country (fipgalem@vc.ehu.es)
³Department of Phonetics & Linguistics, University College London (johnm@phonetics.ucl.ac.uk)

Abstract

A tool which provides relevant feedback on learners’ attempts at phonemic transcription is described. The tool aims to complement courses in transcription which are currently taught in both linguistics and language learning settings. A variety of types of feedback are provided. These can be staged by a tutor in order to support customization for different groups of learners and course levels. The tool consists of two similar standalone applications (for tutors and learners). The system performs an optimal alignment of student versus model transcriptions using a dynamic programming algorithm, modified to handle optional and alternative pronunciations. As a result, it computes a summary of errors and their locations within the attempt. Portability and internationalization are key design goals, supported in practice through the use of Java and XML. The tool is currently being tested in a controlled experiment which will provide considerable information on its actual usefulness and necessary refinements.

1. Introduction

Phonemic transcription is a component of most phonetics and phonology courses, and of some university-level linguistics courses. Transcription of one’s native language is a useful mechanism for raising awareness of the sound system and its behaviour in connected speech, making speakers’ competence explicit [1]. For language learners, transcription has the additional benefit of highlighting pronunciation errors. It can also be used diagnostically by tutors, and is introduced in language courses for non-linguistics students [2].

Certain languages have a complex relationship between orthography and phonemic forms. It is well known that English orthography, for example, is particularly unhelpful as a representation of the phonemic shape of words. In such languages, phonemic transcription practice is of great importance [3]. By contrast, the sound-symbol correspondence of Spanish, for instance, is more regular and predictable, and it is not clear that much can be gained by Spanish transcription practice beyond acquiring basic familiarity with phoneme symbols.

Transcription is not merely a reflection of speakers’ pronunciation, but is a skill requiring repeated practice [4]. As in any learning situation, specific, relevant and rapid feedback on errors is of immense importance. Unfortunately, resource limitations, both of time and skilled tutors, mean that these feedback ideals are rarely realized. Consequently, student self-monitoring is a necessity. Autonomous learning can sometimes be a frustrating experience for the student. Although resources such as pronunciation dictionaries [5] and phonemic descriptions of the language are available, they are unable to reflect the numerous variations which words may experience in context. Students must apply general contextual rules to specific examples, but will not be able to check whether their deductions are correct. Transcription books which include comments and explanations on connected speech phenomena [4,6] help fill the gap between citation form consultation and personalised feedback from instructors.

These considerations led to the current project, which aims to provide for automated feedback on phonemic transcription attempts. Such a tool should be seen as an adjunct to traditional approaches to teaching and learning rather than a replacement. The primary motivation is to provide rapid, relevant and incremental feedback in a way which, it is hoped, leads to improvements in basic transcription skills without requiring intervention from a tutor.

Section 2 enumerates our design goals for the transcription feedback tool. The initial version is described in section 3, while section 4 details its main implementation features. An initial evaluation of the tool’s ability to provide accurate and relevant feedback has been made, and an extensive controlled experiment with over 90 linguistics students at the University of the Basque Country is underway. These evaluations are summarized in section 5. Finally, future plans for the tool, including internationalization and the provision of audio feedback for dictation practice, are outlined in section 6.

An earlier description of the project, dealing primarily with pedagogical issues, can be found in [7]. Here, we focus on design and implementation features.

2. Design goals

Apart from the requirements to produce an easy-to-use, modern and intuitive interface, our design was based on the following criteria:

- **Support for typical transcription exercises**: the tool should support transcription exercises ranging from words and sentences to multi-paragraph tasks such as those found in [4]. Longer tasks (> 300 symbols) are particularly valuable since they enable regular confusions and erroneous symbol usage to be detected and fed back to the learner. It should be noted that the tool is not designed for large-scale manual transcription of corpora in the style of Transcriber [8].

- **Accuracy and relevance**: any responses to the learner should be correct and not result from computational artifacts. While this criterion sounds obvious, it is not necessarily easy to provide in practice. Consider, by analogy, the feedback provided by a compiler in the form
of error messages, some of which are the result of a chain-reaction of earlier errors. Likewise, the tool should go beyond a simple “correct” versus “incorrect” response and instead provide detailed feedback on the location and nature of errors.

- **Support for pronunciation variations**: for phonemic transcription exercises (as opposed to dictation tasks where an audio signal is available), it is necessary to cater for *alternatives* (e.g. “again” rendered as /ə/ or /eɪgən/) and *optional* phonemes (e.g. “history”: /hɪstəri/ or /hɪstɹɪ/).

- **Multiple styles of staged feedback**: to encourage active styles of learning, the tool should be equipped with a number of levels of feedback which can be employed in a staged fashion, under the control of a tutor. This enables the tool to be customized for different groups of learners and tuned to particular stages of an accompanying course.

In addition to design goals related to transcription itself, the tool was built with portability in mind, leading to the following desirable features:

- **Internationalisation**: while the initial development is for British English (RP dialect), the tool should allow ease of localization to other languages, such as French and Swedish, for which transcription is a useful task.

- **Portability**: to promote the use of the tool in a wide variety of educational settings, the tool’s design should enable it to be employed on all common platforms, and, ultimately, on the web.

- **Customisable**: as part of the localization process, such things as key-bindings, feedback messages and tooltips should be external to the tool itself, preferably in a well-defined format such as XML [9].

Finally, the tool should enable detailed evaluation. In particular, it should record usage and results in order to provide a base of data to support research questions in, for example, L1 versus L2 transcription competences for a range of languages.

### 3. The tool in use

The initial version of the transcription tool consists of two standalone applications, one for the tutor to enter model orthography/transcription pairs, the other for the learner to enter and receive feedback on transcriptions. Since the two interfaces are very similar, we restrict our description to the learner’s view.

Figure 1 depicts the main tool window. When the tool is launched, the learner supplies a username and is prompted to choose from a list of available transcriptions (if there is more than one). The system then retrieves their last attempt at that transcription and displays both the orthography (upper panel) and transcription attempt (middle panel). If this is their first attempt, the transcription window is empty. The learner then enters IPA symbols using the keyboard (lower panel) or customizable key-bindings. When sufficient symbols have been input, the user has the option of submitting the attempt. The system performs an optimal alignment against the model transcription supplied by the tutor, taking into account optional symbols and alternatives.

As a result of the alignment, a summary of errors and their locations within the transcription attempt is available. Errors are of three types: symbol confusions (e.g. /h/ used instead of /z/), missing symbols (e.g. dropped /h/ where one is required) and extra symbols (e.g. using an /h/ where none is required).

The feedback available to the user after submission of their attempt is controlled by the tutor by means of a feedback schedule accompanying the model transcription. The aim is to promote active learning by providing a limited amount of help initially, and increasing the degree of detail on subsequent submissions. Specifically, 4 different types of feedback are available, and each may be enabled at a point specified in terms of the number of submissions (e.g. “do not identify the locations of errors for the first few attempts”; “do not show the correct transcription until at least 10 attempts have been made”). The four feedback options are:

- **Error listing**: this summarizes the errors ordered by frequency of occurrence. The idea is to focus on regular, repeated errors that might result from confusion about what sound a given symbol represents, for instance. The error listing can also supply detailed advice on specific errors.

- **Words containing errors**: this option highlights those words containing errors (but not the precise location of the error) within the transcription display itself.

- **Symbols in error**: here, symbols that are confused or extra are highlighted. However, no indication of the location of missing symbols is given.

- **Correct (model) transcription**: a correct solution is displayed in the transcription window. Learners can return to their attempt at any stage but may not edit or submit the correct version directly.

![Figure 1: Orthography (top), transcription attempt (middle) and IPA keyboard (bottom). Symbols in error are shown in red (visible on the CDROM version of the paper).](image-url)
Figure 2: Error listing for the transcription attempt shown in figure 1. The list is ordered by frequency of occurrence and includes confusions, missing and extra (inserted) symbols. Any specific help is displayed in the right-hand column.

4. Implementation

4.1. Alignment
The primary computational component of the tool is a string aligner based on a dynamic programming algorithm, modified to handle optional and alternative pronunciations. The task of the aligner is to find the sequence of substitutions, insertions and deletions which results in the optimal correspondence between the learner’s transcription attempt and the model transcription. It is then straightforward to compute a confusion matrix and determine common confusions, missing and extra symbols. The algorithm is a modified version of that commonly used in tasks such as the evaluation of automatic speech recognition systems.

In the current version, only single-symbol alternatives are possible (i.e. one-for-one substitutions). It is planned to introduce many-to-many alternatives in a later version, since these are necessary to handle processes such as syllabic consonants. For example, “international” can be realised as /ɪntəˈnæʃənəl/ /ɪntəˈnæʃənəl/ /ɪntəˈnæʃənəl/ or /ɪntəˈnæʃənəl/.

4.2. XML
The transcription tool makes significant use of the eXtensible Markup Language [9] in order to promote portability, customization and localization. XML allows the definition of custom document formats via user-defined elements and attributes. Syntax checking is handled by a validating parser.

4.3. Java
Java was chosen for implementation since it is the most portable and web-savvy language in existence, and benefits from a well-developed collection of interface components. In addition, parsing and writing of XML documents is supported at a high level by the public-domain JDOM (Java Document Object Model) classes [10].

4.4. UNICODE
Ultimately, it is our intention to use the IPA subset of UNICODE [11] throughout the application. Unfortunately, since we plan to create a web version of the tool, patchy browser support led us to use images for IPA symbols in the current version.

5. Evaluation
An initial ‘offline’ evaluation was carried out using a sample of 15 exam scripts (each containing around 300 phonemes) from students undertaking an English Transcription course at the University of the Basque Country. Scripts were not chosen at random, but instead included two populations: very poor and good attempts. The purpose of this evaluation was to determine whether the tool produced relevant feedback and to assess the agreement with the assessment of the marker. Two principal findings of the evaluations were:

• The tool was successful in locating all errors in the exam scripts, and, in most cases, the feedback given related directly to the error.
• Certain types of error resulted in less relevant feedback due to the limited nature of the alternative pronunciations scheme. For instance, diphthongs incorrectly represented by a pair of monophthongs (e.g. use of /ɪ/ + /ə/ instead of /ɪə/ in “here”) would typically show a substitution of one monophthong for the
diphthong, plus an insertion error. Another example is the use of /nl/g/ instead of /ŋ/.

Two unanticipated findings point to further potential of the tool. First, exam scripts occasionally contained non-IPA symbols. By contrast, it is expected that users of the tool would learn relevant symbols through frequent exposure. Second, the tool showed up common confusions across the entire sample. For example, half the sample showed 2 or more /s/-/z/ confusions, and schwa was underused. The tool therefore has the potential to provide rapid ‘class-level’ feedback which could be employed with little delay to influence an accompanying formal course.

At the time of writing (March 2001), a large-scale controlled study has just started at the University of the Basque Country. The sample in the study consists of English Philology students doing a course in English Phonetics. The total number of students (over 90) is divided into experimental and control groups, taking into account their native languages (Spanish/Basque) and their knowledge and experience with English transcription. Both groups are doing the same ten transcription passages over a set period of time, but the experimental group is using the tool described here, whereas the control group members carry out traditional (hand-written) transcriptions which are corrected by their tutor. At the end of the test period both groups will be compared in order to evaluate their transcription improvement and thus, the relative value of the tool versus traditional methods. Additionally, all students will be given a questionnaire aimed at assessing variables such as motivation as well as their opinions on the tool usage and usefulness.

These results will be taken into account to modify the tool as seems pertinent. The corpus of student attempts will also be analysed to determine population-level information on such things as common symbol confusions and symbol overuse/underuse.

6. Discussion and future work

A tool which provides automatic feedback on phonemic transcription attempts has been described. The tool has been evaluated in a passive setting by comparison with marked examination scripts, and is currently being extensively employed in a controlled ‘live’ experiment.

Our intent is to make the transcription tool available as a free resource in the near future. However, certain developments are planned before such a release.

First, the tool will migrate to the web. Although we have not experienced significant problems running the software at partners’ sites, a web-based tool will be easier to maintain, and, more importantly, provide access to what we anticipate will be a large and growing database of transcriptions in a number of languages and from learners of a variety of linguistic backgrounds.

Currently a standalone Java application, the intention is to develop a distributed architecture consisting of a Java Applet running on the client, connected to a server application (probably a relational database) which serves up XML documents and records results.

Second, we intend to test the ease of localization of the tool in order that a detailed localization procedure can be defined for users wishing to adopt the tool for their target language.

Later versions of the tool will provide more extensive support for alternative pronunciations. We also anticipate developing a version of the tool geared towards dictation practice.

Acknowledgement: this project is being partly funded by Grant AFC2000-0286-IP from the Spanish Ministry of Education.

7. References