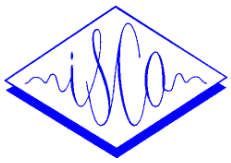

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

AN INTERACTIVE SYSTEM FOR AUTOMATED PRONUNCIATION IMPROVEMENT *

Jean-Paul LEFEVRE(1), Mervyn JACK(2), Claudio Maggio(3), Mario REFICE(4), Fabio GABRIELI(5), Michelina SAVINO(6), Luigi SANTANGELO(6)

(1) Agora Conseil, 185 Hameau du Château, 38360 Sassenage, France
(2) CSTR, University of Edinburgh, 80 South Bridge, Edinburgh EH1 1HN, Scotland
(3) Alcatel Face, Research Center, Via Generale Clark 21, 84000 Salerno, Italy
(4) Dpt. of Electrical and Electronics Engineer., Polytechnic of Bari, Via Orabona 4, 70125 Bari, Italy
(5) Tecnopolis CSATA Novus Ortus, P.O. Box 775, 70010 Valenzano (Bari), Italy
(6) Dpt. of Informatics, University of Bari, Via Amendola 173, 70126 Bari, Italy

Abstract

In this presentation, we will introduce a project whose final objective is the development of a workstation designed to improve the pronunciation of a given language by non-native speakers. The SPELL (Interactive System for Spoken European Language Training) workstation is intended to be a self-instructional device aimed at intermediate ability foreign language learners. Three languages are involved including English, French and Italian.

This study is based primarily on the analysis of the speech characteristics of non-native speakers and on the development of tools to be used in the automated assessment and improvement of non-native language pronunciation. The output of the first phase of the project will be an initial demonstrator system able to process the speech of non-native speakers in order to identify and correct pronunciation errors. This will involve the analysis of the phonetic characteristics of the speaker's mother tongue, the identification and measurement of his/her pronunciation errors, as well as the use of aids helping him/her to understand his/her errors and showing how he/she can correctly pronounce words and short phrases using both audio and visual feedbacks. In a second phase, refinements will be introduced in order to achieve a marketable product.

The technical objectives of the project are to develop methods for analyzing the characteristics of speech produced by non-native speakers (both at macro and micro levels), to develop metrics for identifying differences between a non-native speaker's pronunciation and a model offered by the system, and to provide user friendly feedback which will help to improve pronunciation. The main technical innovation behind SPELL is the application of well founded phonetic and phonological principles for teaching selected aspects of the pronunciation of the language under consideration. This paper begins with a general description of the framework of the SPELL system. This general discussion will set out some basic assumptions about the system and the phonetic issues which it addresses. Then, some technical issues about the user interface software, which takes advantage of an object oriented approach in a powerful windowing environment, are also discussed.

I. INTRODUCTION

With the consolidation of the European Community and the advent of 1992, the single European market and the opening of borders throughout Europe, a substantial demand exists for more and more people to study foreign languages, while such a necessity will probably become one of the most exciting social challenges.

However, limited human resources for the provision of foreign language teaching assistants represent a crucial barrier. Therefore, it is evident that use of some technological approach may be a key solution to satisfy the requirements of this potential exploding market, the efficient instruction of so many human learners by conventional means being logistically impossible to consider.

The SPELL (Interactive System for Spoken European Language Training) workstation is intended to be a self-instruction device designed to improve the pronunciation of a given language by non-native speakers.

The SPELL project, partly funded by the Commission of the European Communities through the ESPRIT programme, has been split in two two-years phases. The output of the first one, now near completion, is an initial demonstrator system, involving French, English and Italian, able to process the speech of non-native speakers in order to identify and correct pronunciation errors. This feasibility study involves the analysis of the phonetic characteristics of the speaker's mother tongue, the identification and measurement of his/her pronunciation errors, as well as the use of aids helping him/her to understand his/her errors and showing how he/she can correctly pronounce words and short phrases, using both audio and visual feedbacks. In the second phase of the project, planned to be started in the coming months, the on-going research will be developed towards a marketable product by refining the hardware and software through new studies focused on phonetic analysis, particularly of consonants and nasalization, user interface design and user evaluation, and by carrying out market surveys and detailed user requirements definitions in order to derive a range of products from the project.

The main technical innovation behind SPELL is the application of well-founded phonetic and phonological principles for teaching selected aspects of the pronunciation of the language under consideration.

* This project is supported by the European Community's ESPRIT programme, under contracts N° 5192 and 7153.

II. THE APPLICATION DOMAIN

The SPELL workstation is intended to be a self-instruction device for intermediate ability foreign language speakers. This assumption avoids the need for the SPELL system to teach the basics of language use as well as pronunciation. It needs to be determined what is meant by "intermediate ability" and pronunciation tests must be devised accordingly for the three SPELL languages.

Since it is assumed that improvement in intelligibility is the most common objective in the successful use of a foreign language, this parameter is used to gauge the user pronunciation improvement. In SPELL, this is done by a suitable comparison of the user performance against a model offered by the system, according to some similarity criteria.

As far as the pronunciation is concerned, two basic aspects have been taken into account : the sentential (macro) and the segmental (micro) levels. At the first stage of the project, it has been decided to consider them as separated; i.e. the user can exercise on one aspect at a time, being made no attempt at the moment to analyze the possible interactions between them.

This is one of the main basic assumption which characterize the SPELL system with respect to the more traditional "global" approach in learning a foreign language pronunciation. However, these two approaches may be integrated in the future in such a way to take full advantage of the possibilities offered by speech technology in the learning process domain.

The reference offered by the SPELL system as a model obviously reflects the characteristics of a "typical" native speaker : this abstract concept is made concrete in SPELL by means of two basic mechanisms. For each language, a set of possible characteristics which reflects all the correct realizations of any specific feature is built up; statistically based variance is allowed in the similarity criteria to set up the thresholds used to judge the user performance. For example, in the intonation exercise the user is asked to make his/her pitch contour fit in a computed band (tunnel) rather than adhere strictly to the offered model. In the vowel quality exercise, instead, the pronunciation of the user selected vowel is considered correct if it fits in a precomputed target area. These possibilities are another key point which makes the SPELL workstation a more flexible tool when compared to the mimicry technique most used by foreign language teachers in pronunciation learning tasks. In fact, there are very few foreign language learners who genuinely need to produce fully native versions of pronunciation. For the majority of students, improvement in intelligibility is a more practical objective in the successful use of a foreign language. It should be noted that mimicry is required to achieve success at pronouncing unfamiliar sounds but that functional ability to communicate with foreign listeners is of more interest than "speaking like a native" (see, for example, [3], [4]).

Finally, since the system is intended to run autonomously, that is without any human assistance, provision is made for a comprehensive set of feedbacks which let the user know exactly what to do next and for what purpose in any interactive situation.

III. FUNCTIONAL DESIGN OF THE WORKSTATION

3.1 SPELL Architecture

As summarized on figure 1 hereafter, an unified architecture can be stressed out for the whole system.

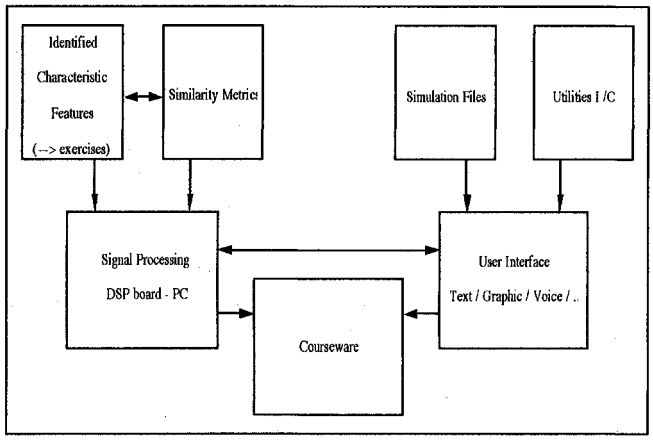

Fig. 1 - The SPELL System

Firstly, distinctive features of interest which characterize the pronunciation are identified. All these features belong to three families related to the intonation, stress and rhythm at the macro (sentential) level and to the vowels and consonants at the micro (segmental) level. In each family, one or more distinctive features have been addressed.

Accurate feature analysis depends on the location of phonetic segments within an utterance produced by a student. The analysis of vowel quality (micro) features certainly requires the prior location of the vowel segments. Proper prosodic analysis too, cannot be completed without first locating these phonetic segments which have had durational and intonational features overlaid on their structures. Therefore, a segmentation program has been implemented for application to the speech signal prior to feature extraction and analysis.

These features are more deeply described in the two companion papers [1], [2].

The SPELL system must be able to deal with a number of different types of pronunciation errors produced by non-native speakers of a given language. They can be classified as : phonemic, phonetic and graphemic errors [3].

Phonemic errors may occur when sound to be pronounced have no correspondences in speaker's mother tongue, so that he/she tends to perceive and reproduce them as realizations of his/her native sounds closest to the foreign language articulation area. This removes the distinctive function of such non-native language phonemes, causing misunderstandings in communication.

Phonetic errors may occur when substitutions of target sounds with native ones are due to their strong articulatory and acoustic similarity, even though they are not equivalent in the strict sense. These replacements do not cause problems of intelligibility, but still show the "non-nativeness" of the speaker (foreign accent).

Graphemic errors may arise when pronunciation errors are caused by spelling interferences of the speaker's mother tongue. Finally and obviously the speaker may produce gross mistakes such as misreading, stuttering, false starts, etc.

In order to have the input waveform be segmented correctly when such errors may exist, two basic solutions have been adopted in SPELL. Firstly, the test utterances are constructed in such a way as to limit the types of error which might occur. Secondly, the SPELL segmenter has been designed using a segmental transition network which includes the more predictable types of error which may occur between two given languages.

Then, for each feature, a set of exercises is proposed according to an instructional paradigm as discussed in the following section.

Each feature is associated with both signal processing functions and similarity metrics. Signal processing is needed to compute automatically the relevant parameters corresponding to the features. According to their complexity and to other technical requirements, the signal processing functions are encoded either on a DSP extension board or directly on the host computer. Metrics are needed to evaluate the similarity of the responses of the user when compared to a reference. In some cases, these metrics also involve signal processing functions .

A set of elementary functions involving text, graphics, voice playback and potentially synthetic speech, all necessary in building the above mentioned exercises, is under control of the user interface. Associated with it is a library of utilities which handles the various I/O needed in the interaction. As input, the user interface normally accepts data computed by the signal processing functions, but obviously some simulation files can be alternatively used for other purposes such as demonstrations.

The last stage of the SPELL system is the courseware itself. For a given feature to be studied, the courseware allows the user to follow a set of exercises organized according to a chosen instructional strategy.

3.2 The SPELL Courseware

The courseware foreseen for this project is intended to provide directed instruction to the student, allowing better predictions to be made about the student's performance, and enabling a proper evaluation of the system itself as a teaching aid. The development of such a teaching courseware concentrates research efforts towards a system which touches on all levels of pronunciation teaching. Courseware exercises would be prepared in advance by a teacher using a system authoring software.

The system is designed in such a way to allow any possible instructional sequence; however, the present prototype is based on a so-called **DELTA** paradigm :

Demonstrate -- Audio demonstrations by the system of various utterances (as well as visual representations as, for example, in the case of exercises on intonation) are used to highlight the pronunciation features of interest. For example, differences in vowel quality would be demonstrated by playing out words which contain minimal pairs such as *beet* versus *bit* for the / **i** / versus / **I** / vowel distinction in English. The same pair of words might appear in sentences to demonstrate the phonological significance of the vowel distinction (e.g. "I would like a bit/beet.").

Evaluate Listening -- Small listening tests are completed by the student to evaluate his/her ability to perceive the pronunciation features of interest. For example, the student would take an ABX test: the two utterances A) "bit" and B) "beet" are played to the student, who must decide if the test utterance X is the same as A or B. If the student fails to perceive the distinction for a given pronunciation feature then he/she may be asked to go back to the Demonstrate stage for more examples.

Teach -- The actual teaching of the pronunciation features of interest takes place at this stage, with quantitative feedback for the student and directions for modifying inadequate performances.

Assess -- This stage is the formal evaluation of the student's ability to pronounce the features of interest. For example, the student is given N attempts to produce utterances containing the features of interest. If the student achieves a certain percentage of correct pronunciations then he/she should proceed onto the next features to be taught; otherwise the student should go back to the Demonstrate phase for this particular task.

IV. THE SPELL USER INTERFACE

The appropriate feedback must be provided by the SPELL workstation for the user. Prosodic features appear to be relatively straightforward to be described to the student and therefore diagnostic feedback is appropriate (e.g. the workstation informs the student that he/she used a falling pitch contour at the end of the utterance but that a rising pitch would have been more appropriate). In the case of vowel quality, it is felt that quantitative feedback on its own (i.e. without diagnostic information) would be more appropriate, since it may be difficult to convey complete articulatory relationships to the linguistically unsophisticated foreign language learner.

As a matter of fact, it should be emphasized that the feedback to a SPELL user is not intended to be expressed in terms of explicit sophisticated linguistic or phonetic concepts : for example, the student needs not to be aware of the abstract phonological systems which form the bases of the training system and neither the knowledge of even a phonetic alphabet is considered a requirement.

The SPELL user is expected to interact with the workstation through a

simple pointing device, such as a mouse, to make the needed choices when asked for and with a microphone to utter the required sentences and/or words.

On the contrary, he will receive feedbacks from the system through suitable texts and/or graphics displayed in the appropriate areas of the screen and by means of relevant speech messages via an headphone.

The language used for textual feedback is normally the user native one, even though the system can be set up differently.

Accordingly, the system has been designed on a core of basic functions supported by a powerful windowing environment, in agreement with the Common User Access (CUA) guidelines provided by the System Application Architecture standard [6], while the actual implementation has been performed using mainly an object oriented rapid prototyping programming environment. CUA involves a set of rules and guidelines to be followed during user interface design, specifying a standard set of user interface components in terms of their appearance and how users can interact with them. It defines a style guide which needs to be preferably followed at each step of a given application, including rules for topics such as use of windows, dialog control actions, menus, buttons, use of colors, terminology and so on.

At the functional level the software system consists of four main sets of "modules" :

General handlers : devoted to the system initialization, control of the global system status and to interface with the DSP extension board;

Data handler : used for input, output, management and access to all the data used within the workstation;

Graphics handlers : deal with all the graphic capabilities;

Courseware handlers : used to implement the chosen instructional sequence.

According to the chosen object oriented approach, these modules are organized as "classes" into hierarchical trees to take advantage of their inheritance capabilities. For example, the class "TextDialog" has been defined which handles simultaneously either the "buttons" which the user is asked to click on in the different phases of the exercises, as well as the "texts" representing the questions posed by the system to solicit the user action as answer.

The class at the highest but the top level of the tree, called "SpellWindow", includes all the classes used to handle the windows made available to the user and therefore contains also all the "methods" needed for each of the children classes (here are included, for example, all the calls to the library functions schematically shown in the previous Fig. 1).

In the Windows environment, the SPELL application starts by a simple instantiation of the father class "SpellApp".

V. CONCLUSIONS

At the completion of the first two years of research activities, we have been presenting the main features of the SPELL workstation developed as a demonstrator of the main achievements of the related ESPRIT project.

Its unique technical characteristics are represented by the integration, in a single device, of well-founded phonetic and phonological principles with up-to-date signal processing techniques and software development methodologies and tools.

The target application domain, i.e. learning foreign languages pronunciation, is the challenging environment which will constitute, as we believe, the stimulus for more deep research activity in this field and will eventually allows the development of a successful market product.

References

[1] E. Rooney, S.M. Hiller, J. Laver, "Prosodic (Macro) Features for Automated Pronunciation Improvement in the SPELL System", ICSLP Conf, Banff, October 1992.

[2] M-G. Di Benedetto, F. Carraro, S.M. Hiller, E. Rooney, "Vowels Pronunciation Assessment in the SPELL System", ICSLP Conf, Banff, October 1992.

[3] A. D'Eugenio, "L'insegnamento della Pronuncia Inglese agli Italiani, Problematica Didattica e Suggerimenti Correttivi alla Luce Dell'analisi Contrastiva", Bologna, PATRON, 1977.

[4] W.M. Rivers, "A Practical Guide to the Teaching of French", New York: Oxford University Press, 1975.

[5] J. Harmer, "The Practice of English Language Teaching", London: Longman, 1983.

[6] R.E. Berry, "Common User Access - A Consistent and Usable Human-Computer Interface for the SAA Environment", IBM System Journal, Vol. 27, N° 3, 1988.