ReCALL **13** (1): 67–78. Printed in the United Kingdom

© 2001 Cambridge University Press

Interactive pronunciation training

WOLFGANG MENZEL
Fachbereich Informatik, Universität Hamburg, Vogt-Koelln-Strasse 30, D-22527 Hamburg, Germany
(*e-mail: menzel@informatik.uni-hamburg.de*)

DANIEL HERRON, RACHEL MORTON
Microsoft SDN UK, Compass House, 80-82 Newmarket Road, Cambridge CB5 8DZ, UK
(*e-mail: dherron@microsoft.com; rmorton@microsoft.com*)

DARIO PEZZOTTA
Dipartimento di Informatica, Sistemistica e Comunicazione, Università di Milano Bicocca, Viale Sarca 202, I-20126 Milano, Italy
(*e-mail: pezzotta@disco.unimib.it*)

PATRIZIA BONAVENTURA
Conversa, 15375 NE 90th St., Redmond, WA 98052, USA
(*e-mail: Patrizia.Bonaventura@conversa.com*)

PETER HOWARTH
Language Centre, University of Leeds, Leeds, Yorkshire LS2 9JT, UK
(e-mail: p.a.howarth@leeds.ac.uk)

Abstract

Improving the feedback quality of a computer-based system for pronunciation training requires rather detailed and precise knowledge about the place and the nature of actual mispronunciations in a student's utterance. To be able to provide this kind of information, components for the automatic localisation and correction of pronunciation errors have been developed. This work was part of a project aimed at integrating state-of-the-art speech recognition technology into a pronunciation training environment for adult, intermediate level learners. Although the technologies described here are in principle valid for any language pairs, the current system focuses on Italian and German learners of English.

1 Introduction

Computer-based solutions for pronunciation training are becoming increasingly commonplace for foreign language learning purposes (Dalby, Kewley-Port & Sillings, 1998; Herron *et al.*, 1999; Eskenazi & Hansma, 1998; Witt & Young 1997, 1998). From a

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

pedagogical point of view two major design goals seem to be of particular importance:

1. Pronunciation should be treated as much as possible as part of a complex productive skill rather than as a purely reproductive activity. Thus it needs to be integrated into a general setting that first of all invites the student to pursue his or her communicative goals and in which all aspects of pronunciation and other form-related considerations are primarily treated as particular means to successfully achieve these goals.
2. The training system should provide the student with sufficient feedback to allow him or her to identify possible deficiencies and should clearly point out the way to improve.

To establish and realise a communicative goal is possible only when at least a minimal degree of choice is offered in the student's response. Thus speech recognition capabilities are required to determine which option the student has actually chosen. Helpful feedback, on the other hand, can only be generated if sophisticated analytical tools are available that are able to extract the necessary diagnostic information from the speech signal. It is obviously in this area that currently available solutions for computer-assisted pronunciation training offer considerable room for improvement.

In most current tutoring systems, a simple playback facility along with a global scoring mechanism and a visual presentation of the signal form or some derived parameters like pitch are provided. All these tools, however, leave the task of identifying the place and the nature of the problem entirely to the student. Even if word-level scoring allows the student to directly pinpoint the mispronounced word, this might not be sufficient to give the student a helpful hint on how to improve the response. Only by discovering the precise nature of a student's mistake, at the phone level, is the system in a position to provide detailed error explanations, problem-specific speech stimuli, or individualised suggestions for improvement. Furthermore, such a low-level diagnosis makes it possible to guide the student into specifically tailored opportunities for practice.

In order to demonstrate the potential of contemporary speech recognition technology for a range of additional services, components for the automatic localisation and diagnosis of pronunciation errors have been developed and integrated into a multimedia-based training environment. Although the system is targeted at intermediate Italian and German learners of English, the underlying solutions are mostly generic and could therefore be ported to arbitrary pairs of languages. The system offers the student a wide range of communicatively relevant exercises (such as question answering) in which the answer needs to be constructed from small sets of pre-specified building blocks. It leaves it up to the student to decide which kind of feedback information to make use of.

This paper gives an overview of the pronunciation training system and its most important speech processing components. In addition, results of an off-line evaluation using a phonetically annotated corpus of non-native speech as well as some preliminary findings from a user trial with learners and teachers of ESL are presented.

2 System architecture

The system consists of a series of modules that are arranged on three levels (please see

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

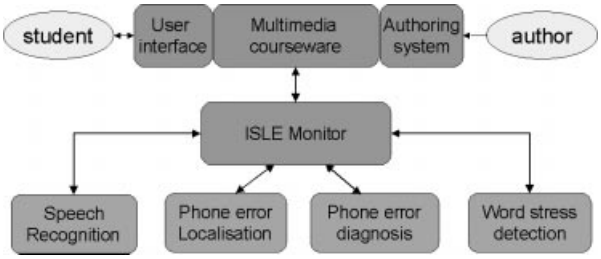


Fig. 1. Architecture of the system.

Figure 1). On the top level a standard authoring system is used to provide the communication interfaces with the student, on the one hand (see Figures 3, 4 and 5 for examples), and the courseware author, on the other. It facilitates the presentation of the learning material as well as control and synchronisation of the multimodal interaction with the student: input and output of spoken language, visualisation of feedback information, and the mouse-based selection among different training and feedback options.

The four components on the lower level of the system carry out the analysis of the speech signal. A central monitor component serves as a kind of switchboard that facilitates the communication between the upper and the lower level. It invokes the necessary analysis modules according to the student's input and reports back the diagnostic findings.

Generally, the speech signal, as produced by the student, is first subjected to the speech recognition component, which attempts to determine what the student has actually said. Thus, within the other modules for speech quality assessment and diagnosis it is always assumed that the spoken utterance is already known. In the case of a successful recognition the system tries to identify those positions in the signal that might contain a pronunciation error. This so-called localisation of errors is carried out using the confidence scores provided by the speech recogniser for each word and phone. On request, the error is further analysed in terms of phone substitutions, insertions, or omissions and checked for misplaced word stress patterns as well.

3 The speech processing components

3.1 Speech recognition

Despite considerable progress in the area of speech technology, the recognition of highly accented learner speech remains a largely unsolved issue. At the moment, it can be carried out with an acceptable degree of reliability only if the recognition task itself is kept very simple. This can be achieved by restricting the set of allowed utterances to a relatively small one, from which the student can select the correct or desired response. Additionally, the lexical inventory should be chosen in such a way that phonetically similar words are avoided at the same sentence position. Usually, the difficulty of a recognition task is described by means of a measure called perplexity, which roughly correlates with the size of the space from which the allowed utterances can be chosen.

The recognition system used in this system is based on a state-of-the-art Hidden

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

Markov Model (HMM) recogniser for continuous speech (Morton, Whitehouse & Ollason 1999), which allows the system to accept fluently spoken utterances during pronunciation training. After being adapted to the current speaker, in a short enrolment session, it achieves 100% accuracy on very simple recognition tasks, such as an oral selection from a menu of three to four short phrases (see Figure 4). The perplexity of such a task usually remains below 1.5. An example of a slightly more ambitious recognition task is given in Figure 3. For tasks of this type, with a perplexity between 2.5 and 4.3, recognition rates above 96% are still achieved. Especially difficult are recognition tasks with a high phonetic confusability such as a minimal pair exercise, as shown in Figure 5. Here, the system reached an accuracy of 94% on a task with a perplexity of 4.0.

3.2 Error localisation

Localisation is the process of identifying the areas of an utterance that are likely to contain pronunciation errors. This is necessary for two reasons. The first is one of perception; because diagnosis may take a long time, and because the student may often wish to know where errors were made but not how, it is desirable to have the system (relatively) quickly pinpoint errors. The second reason, which is more important for performance, is that diagnosis is a difficult task that is much simplified if the speech signal has been presorted into correct and incorrect regions.

A pronunciation error can be defined as some deviation from a target or model pronunciation. Non-native speech may contain many kinds of pronunciation errors. These may vary in their origin and their degree of deviation from the target, and may be more or less serious in the extent to which they hinder communication. For example, some errors may merely signal a marked non-native accent while the speech remains intelligible. It would be infeasible and possibly damaging to provide feedback for every possible deviation, as almost every student would quickly become discouraged, even if the system never made a false-alarm type error. Thus a sensible alternative is to select the most severe errors.

Such selection strategies are performed in the system by an error localisation component that assigns and sorts confidence scores for each speech segment of interest, which might be a phone, word, or entire utterance. In simple terms, low confidence scores represent increased certainty that the utterance was mispronounced and high confidence scores represent certainty that the utterance was correctly pronounced. Therefore areas of high confidence can be filtered out, leaving segments that have a high probability of containing an error. These segments can be ranked in order of ascending confidence and passed to the diagnosis module so that the most serious errors can be attended to first.

Typically error localisation is directly based on the acoustic scores obtained by the recogniser. In order to obtain a more reliable estimation of acoustic qualities we use a more indirect approach, where confidence scores are computed via Gaussian classifiers that use several predictors taken for each frame of speech in the input signal: (1) the acoustic likelihood of the recognised path; (2) the output probability of the most likely ('best') state in the model set; and (3) the acoustic likelihood of a background model. It is assumed that the best state scores provide the reference ceiling and the background model score provides the floor and that the acoustic likelihood should thus lie somewhere in this range. The distance between the acoustic likelihood and the best state

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

defines how close the hypothesised path is to what the person actually said, or how it was said. The component was then trained over a set of predictor values for the correct and for the incorrect data. For each test observation, the likelihood that the observation vector came from the error distribution or from the correct distribution was calculated; the ratio of these two calculations was the basis for the Gaussian confidence score. Separate classifiers were trained for different classes of phones (e.g. vowels, fricatives, liquids, etc.) in order to increase the modeling accuracy.

3.3 Phone error diagnosis

Diagnosis proper follows as a second stage. It tries to discover how each localised word varied from the canonical pronunciation. For that purpose, a large number of mispronunciation hypotheses is checked against the speech signal for plausibility, again using the speech recogniser. A set of mother-tongue specific letter-to-phone and phone-to-phone rules is applied to the target word (in its orthographic and phonemic forms, respectively) in order to produce the candidate mispronunciations. Each of these is temporarily added to the recognition dictionary, and the recogniser is then re-run over the same speech signal in forced-alignment mode (at the word level). The most plausible hypothesis is taken to generate specific feedback messages (for example, "you inserted a sound like 'House' before that word", to indicate the insertion of an /h/ sound.)

Error diagnosis is based on general and expert knowledge as well as on data collected by the ISLE project from potential users (teachers and students). It was assumed that intermediate level, non-native learners of English will make various types of errors, all of which are manifest as 'pronunciation' errors. These types can be generally described as:

- articulatory difficulties producing particular sounds or clusters of sounds (e.g., the notoriously difficult /th/ sound in English)
- receptive difficulties, because of which the student is unable to perceive and therefore to reliably produce the distinction between two sounds (e.g. /ih/ and /iy/ for Italian speakers)
- orthographic carry-over from the mother tongue; because so much of language use and learning is written, peculiarities of the student's native orthographic system may interfere with pronunciation of English (e.g. the sequence 'ie' is pronounced as /iy/ in German, but can be pronounced in many ways in English.)
- orthographic difficulties of English; because of the high degree of ambiguity in mapping written to spoken English, the student may be expected to mis-apply or mis-generalise 'rules' of English pronunciation.

Since it can be assumed that the student's mother tongue is known, and because there are presumed to be regularities in the way that those speakers make mistakes, it is sensible to try to predict what errors a student might make based on his or her mother tongue. This is accomplished by means of string transformation rules, which allow one to rather precisely tailor the space of error hypotheses to be considered. Even a rather complete set of rules will generate a candidate set that is small in comparison to what an unrestricted open phone-loop recogniser is effectively using, making this a computationally attractive constraint.

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

Two types of error rules are used by the system: letter-to-phone rules, which are designed mainly to capture orthographic errors, and phone-to-phone rules, which attempt to model articulatory or receptive difficulties.

The first step in producing candidate mispronunciations for a given word is to generate a mapping between the letters in the word and the phones in the correct pronunciation. For example, the word *folk*, pronounced /f ow k/ (with the phone symbols of the recogniser used in this system), is mapped as:

orthographic string	F	O	L	K
canonical pronunciation	f	ow		k

using a set of letter-to-phones rules designed to generate the correct pronunciation for a large number of English words.

The letter-to-phone rules are then applied in a left-to-right fashion across the letters in the word and each of the phones might be substituted by others, generated according to a letter-to-phone rule from the student's mother tongue. This can result in alternative segmentations of the phonetic string. For example, a student might be expected, based on his mother tongue, to explicitly pronounce the letter *L* in the word 'folk', resulting in the alternative transcription:

orthographic string	F	O	L	K
alternative pronunciation	f	oh	l	k

Finally, all the phonetic transcriptions generated so far are modified according to the phone-to-phone rules. For example, a word-final consonant may create an intrusive schwa:

orthographic string	F	O	L		K
old hypothesis	f	oh	l		k
additional hypothesis	f	oh	l	k	ax

All the mappings between the original orthographic string and the newly created phonetic transcriptions are kept, for use in generating error explanations later on.

The result of this transformation procedure is a set of possible pronunciations, one of which is the correct (canonical) pronunciation, and the rest of which are variant mispronunciations created by the application of one or more error rules, either or both letter- or phone-to-phone (for example, the pronunciation / f ow l k ax/ above contains two errors: the pronunciation of the normally-silent L and the insertion of a schwa). Because the size of this set grows greatly as the number of errors per (mis-) pronunciation increases, the actual demonstration system was limited to produce only those candidate pronunciations that contain one error. The offline results described below allowed up to three errors per candidate pronunciation, because speed was not a consideration. It should be pointed out that less than 1% of the errors made by the target speakers contain more than three errors; most contain just one.

This approach allows one to diagnose a large number of pronunciation errors typically made by learners of a foreign language. However, the system is constrained to operate with the phone set of the underlying recogniser, which in this case has been trained on

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

native speech for the target language alone. Thus, error explanations can only be given in terms of the 41 phones from the British phone set. A proper diagnosis of those phonetic variants that are highly influenced by the mother tongue of the student and therefore most frequent at the beginner level would require one to train a recognition system with an extended bilingual or even non-native model set.

3.4 Word stress detection

The last in the series of speech processing steps was developed for the purpose of detecting lexical-stress errors (i.e., putting the stress on the incorrect syllable of a polysyllabic word.) It is also based on a Gaussian classifier using phone segment length, energy and some spectral features of the vocalic region as input. For all polysyllabic words the determined position of the main word stress is compared to the canonical stress position as indicated in the pronunciation dictionary and each deviation is reported as a possible diagnostic result to the error explanation component.

3.5 Performance

The speech processing components of the system have been evaluated by comparing the diagnostic results of the system to those of trained linguists who annotated a speech corpus of non-native intermediate-level English (Menzel *et al*., 2000; Bonaventura, Howarth & Menzel, 2000). This corpus contains approximately 20 minutes of speech from each of 23 German and 23 Italian speakers, recorded in a typical language learning scenario. The speech of each subject was annotated by one of six trained annotators at the word, phone, and stress level. Phones, in particular, were scored as either correct, changed (a different phone was substituted), deleted or inserted.

The performance of the system can then be described in terms of the following measures:

- Hit rate: Percentage of phones marked as incorrect by the human annotator that have also been localised or diagnosed by the system.
- False alarm (FA) rate: Percentage of phones considered correct by the annotator, but flagged as erroneous by the system.

Because there were multiple annotators operating independently, there was an imperfect agreement among them. By introducing a certain amount of redundancy into the annotation procedure the degree of disagreement can be measured, in order to calculate both how likely any two given annotators were to find the same error (essentially, an inter-annotator hit-rate), and how likely one annotator was to 'incorrectly' find an error that was not found by another annotator (who is considered as the 'perfect' annotator). By performing all such pair-wise comparisons between annotators, and averaging the results, a rough expectation can be drawn for upper limits on performance.

Figure 2 shows a comparison between human and system performance for the three tasks: error localisation, error diagnosis, and word stress detection. Several findings are immediately apparent. Overall performance in terms of hit rates is seemingly poor, but humans do not perform much better. There is one notable exception, however. Stress detection, which is performed with a high degree of reliability by humans, seems to be

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

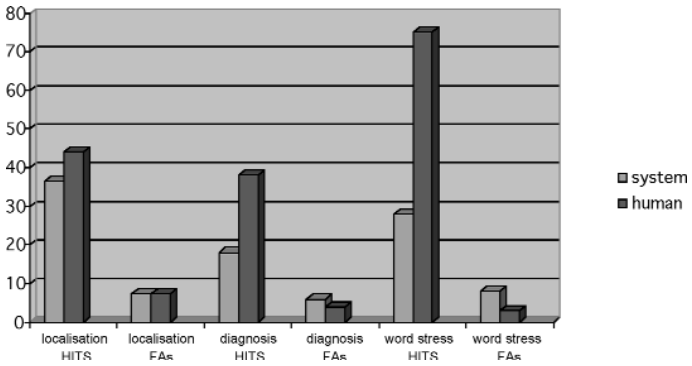


Fig. 2. Performance of the diagnostic components compared to human performance in terms of hit and false alarm rates.

particularly difficult to detect automatically. In general a low hit rate can easily be tolerated, though, because enough errors probably remain to be reported. More worrying, however, is the relatively high false alarm rate, since false alarms might confuse the student considerably. In particular, for utterances with a rather low prior probability for errors (e.g., those from advanced students) it will result in an unacceptably high proportion of misleading system responses. Still, a number of measures can be taken to reduce the perceived effect of such high false alarm rates. For example:

- exercises should be difficult enough for the target user group in order to produce enough pronunciation errors,
- error explanations should focus on errors with a high prior probability, and
- those phone errors that can be diagnosed with higher degrees of reliability can be targeted.

Finally, a much tighter integration of the diagnostic components can be attempted. Because, for instance, a shift in the word stress position is almost always accompanied by a change in vowel quality, word stress errors should be reported only if they are consistent with the phone error findings. Thus, heuristics like “a schwa is never stressed” can be included to improve reliability.

4 Spoken language exercises

All pronunciation related activities have been embedded as far as possible into a communicatively motivated environment. After listening to a sample dialog the student might enter into different types of exercises that range from reading aloud and listen-and repeat tasks to question-answering, in which the student can generate his or her response by combining appropriate items from multiple selection menus. Using the built-in speech recogniser, the system first determines whether the selection made by the student was appropriate and, if so, subjects the utterance to further inspection by the diagnostic components. In this manner, even simple dialog situations can be simulated (see Figure 3).

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

5 Feedback for errors

Feedback is provided to the student through a multimedia-based interface, which offers various amounts of detail about the place and the nature of possible mispronunciations to be provided. First, words that might include pronunciation errors are flagged and the student can listen to them again either as produced by his or her own voice or by a model speaker. If requested, the system points out the erroneous phones as well and provides an error explanation, as shown in Figure 4.

By clicking on the corresponding words or letters in the error explanation the student can again listen to the target phone and the substituted one, both in (a different) context and in isolation (for example, if the correct phone was /p/, the student can click on the example word 'pear' to hear it pronounced or on the p in that word to head the /p/ sound alone). Note that all the interaction is carried out using only the orthographic representations. This avoids any reference to a phonetic symbol system, with which the student might not be familiar. In the example of Figure 4, the student has mispronounced the word 'cheaper' (by substituting the vowel /iy/ with /eh/). The location of the error in the word is indicated to the student by colouring red the letters 'ea', and the student is told that 'that should sound like media, not like else'. The e is coloured blue in the word 'media' to demonstrate the correct vowel, and the e in 'else' is red, to indicate the sound the student (incorrectly) made. This type of feedback becomes more complicated with more extreme errors (e.g., substituting one sound with two different sounds), so in some cases the system 'drops back' to simply telling the student where the error was, or what the correct sound was, without providing an example of what sound the student made.

Finally the student is invited to engage in specifically tailored follow-up exercises (via the 'improve' button), which require him or her, for example, to repeat the problem word in isolation, to practise the target phone in a variety of different phonetic contexts, or to contrast it with the substituted phone in a minimal pair exercise (see Figure 5). Moreover, the system keeps track of recent student performance and offers an opportu-

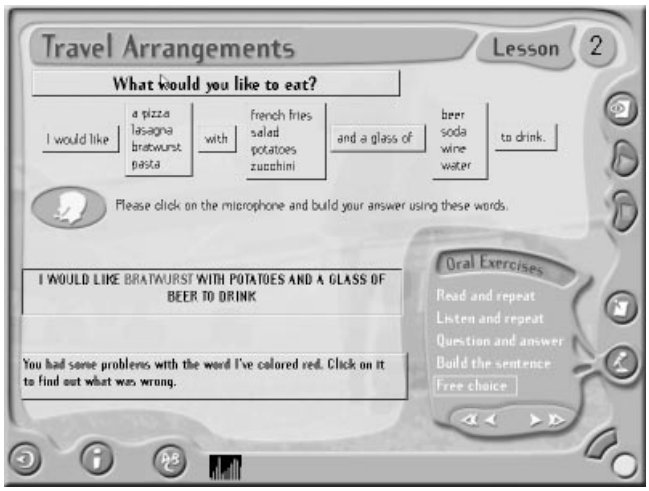


Fig. 3. Example of an exercise, that allows the student to construct her response from a set of menus.

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

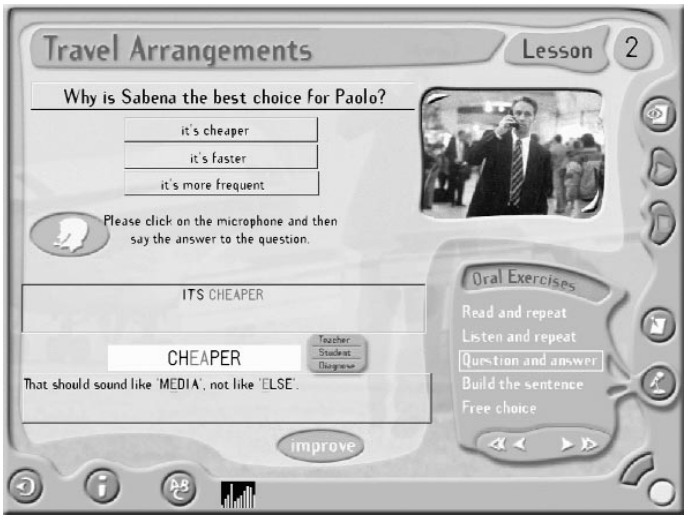


Fig. 4. A screenshot from the demonstration system, showing the highlighting and explanation of errors (the error-containing word is located in the white box, and the example correct and incorrect words in the phrase beneath).

nity to practise the (individually) most difficult phones again. Thus, based on the diagnostic findings a great variety of individualised error explanations and exercise opportunities are created for the student, in order to allow him or her to overcome even highly persistent kinds of pronunciation mistakes.

6 Evaluation

The complete demonstration system has been evaluated in a field trial involving fifteen learners (among them six Italians and nine Germans) and thirteen teachers where eight spoke German and five English as their first language. All of them carried out a supervised trial session of approximately one hour in length and filled in a questionnaire in which they were asked to quantify the system performance on a scale between one (very good) and five (very poor). Table 1 summarises their judgments for the three most important questions.

With the exception of the accuracy judgement made by English-speaking teachers these results show a remarkably high degree of acceptance across different user groups. Since in our sample of test users the Italian students had a much lower level of proficiency than their German counterparts, the evaluation might also indicate that the degree of acceptance is particularly high among students at a lower level of achievement.

7 Conclusions

Experience with the development of the pronunciation training system described here shows that current state-of-the-art speech technology can be used to improve the quality and detail of feedback to the student considerably. Thus it is possible to point out the

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

Table 1. *Acceptability judgements of teachers and students on a scale between one (very good) and five (very poor)*

	Teachers		Students		
Was the feedback	English	German	Italian	German	Average
understandable?	2.2	2.6	1.8	2.3	2.2
accurate?	3.8	2.4	1.5	2.3	2.5
helpful?	2.6	2.0	1.7	2.0	2.1

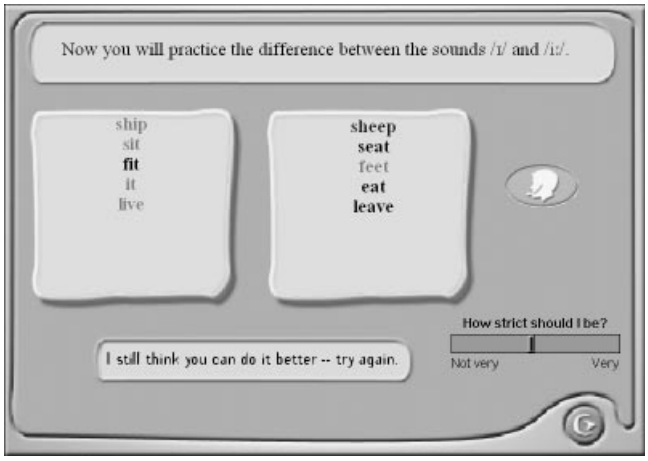


Fig. 5. Example of a minimal pair exercise, as triggered by a previously detected pronunciation difficulty.

place and the nature of a possible mispronunciation rather precisely. This not only allows one to provide the student with guiding hints on how to improve performance but also opens up the possibility of preparing exercises that are specifically tailored to his or her current needs. This is a clear advance over other currently available solutions, which appear to be much more limited in this respect.

Obviously, the level of performance that has been achieved is far from being satisfactory. In particular, the reliability of diagnostic results needs considerable improvement. Nevertheless, the results of a preliminary user-centred evaluation suggest that acceptable solutions with a possibly restricted range of diagnostic capabilities can be devised with the currently available technological means.

The project web site, at `http://www.ec-isle.org`, has a demonstration version of the system for download.

Acknowledgments

This research was supported by the European Commission's Telematics R&D pro-

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press

gramme as Language Engineering project ISLE under the contract LE4–8353. The project consortium was formed by Entropic Cambridge Research Laboratories (Cambridge, UK), Ernst Klett Verlag (Stuttgart, Germany), Dida*El (Milano, Italy), Leeds University (UK), Università degli Studi di Milano-Bicocca (Italy) and Universität Hamburg (Germany). We are particularly grateful to Eric Atwell, Roberto Bisiani, Fabio Daneluzzi, Antonio Fabiano, Roberto Galbiati, Kamran Kordi, Julian Odell, Dave Ollason, Alberto Petitti, Luigi Roli, Clive Souter, Jürgen Schmidt, Hartmut Wick, and Steve Young for their valuable contributions to the success of this project.

References

Bonaventura, P., Howarth, P. and Menzel, W. (2000) Phonetic annotation of a non-native speech corpus. *Proceedings International Workshop on Integrating Speech Technology in the (Language) Learning and Assistive Interface, InStil 2000*, Dundee, 10–17.

Herron, D., Menzel, W., Atwell, E., Bisiani, R., Danelozzi, F., Morton, R. and Schmidt, J. A. (1999) Automatic localization and diagnosis of pronunciation errors for second-language learners of English. *Proceedings 6th European Conference on Speech Communication and Technology, Eurospeech 99*, Budapest, v. 2, 855–58.

Dalby, J., Kewley–Port, D. and Sillings, R. (1998) Language–specific pronunciation training using the HearSay system. *Proceedings International Workshop on Speech Technology in Language Learning, STiLL 98*, Marholmen, Sweden, 25–28.

Eskenazi, M. and Hansma, S. (1998) The Fluency pronunciation trainer. *Proceedings International Workshop on Speech Technology in Language Learning, STiLL 98*, Marholmen, Sweden, 77–80.

Menzel, W., Atwell, E., Bonaventura, P., Herron, D., Howarth, P., Morton, R. and Souter, C. (2000) The ISLE corpus of non-native spoken English. *Proceedings 2nd International Conference on Language Resources and Evaluation, LREC 2000*, Athens, Greece, 957–63.

Menzel, W., Herron, D., Bonaventura, P. and Morton, R. (2000) Automatic detection and correction of non-native English Pronunciation. *Proceedings International Workshop on Integrating Speech Technology in the (Language) Learning and Assistive Interface, InStil 2000*, Dundee, 49–56.

Morton, R., Whitehouse, D. and Ollason, D. (1999) *The IHAPI book*. Cambridge: Entropic Cambridge Research Laboratory, Ltd.

Witt, S. M. and Young, S. J. (1997) Language learning based on non–native speech recognition. *Proceedings 5th European Conference on Speech Communication and Technology, Eurospeech 97*, Rhodes, Greece, 633–36.

Witt, S. M. and Young, S. J. (1998) Performance measures for phone–level pronunciation teaching in CALL. *Proceedings International Workshop on Speech Technology in Language Learning, STiLL 98*, Marholmen, Sweden, 99–102.

https://doi.org/10.1017/S0958344001000714 Published online by Cambridge University Press