

Achieving interactivity in online learning of phonetic skills

Michael Ashby, Kayoko Yanagisawa, Young Shin Kim,
John Maidment & Joanna Przedlacka, UCL
{m.ashby, k.yanagisawa, young.s.kim, j.maidment, j.przedlacka}@ucl.ac.uk

1 Introduction This paper gives an account of research which aims at enhancing interactivity in the online teaching of phonetics. We report on a project¹ with the following broad goals: (i) to define and quantify the need for interactive feedback in phonetic skill acquisition; (ii) to develop, extend and evaluate the full potential for interactivity in a previously developed and successful online introductory course; (iii) to survey and evaluate possible systems for adding parallel functionality to existing VLEs (audio and video tools, VoIP, conferencing, etc) for more advanced courses.

2 Background and previous work Previous reports (Ashby *et al* 2007a, 2007b) describe a project which set out to replicate within a Virtual Learning Environment (VLE) most of the elements of an on-campus introductory training in phonetics. The resulting course already merits the description ‘interactive’ in that it provides communication forums for students and tutors, practice exercises with answers, a virtual “Listening Centre” (self-access listening practice), weekly assignments of increasing difficulty, graded and returned with individual feedback, and a final assessment, including a listening test. However, it arguably falls short of the spontaneous, improvised and equal interaction of a live phonetics class. The desirability of greater interactivity was already commented on by participants in the trials, and the need for interactive exchange among all participants would become increasingly important in more advanced courses.

3 Interactivity Though it is widely regarded as something desirable, there is no general agreement on the definition of interactivity in education (Yacci, 2000). One useful concept is that of a closed-loop system (Figure 1). For true interactivity, a participant must have the impression that input which he/she provides to the system potentially modifies what is fed back to them. Our own proposal is to distinguish levels of interactivity in a hierarchy. A lecture, a recording, or online resources that can merely be browsed are at Level 0 (unresponsive to student input). Level 1 is reached if a student can give **commands** to select activities, and regulate pace and repetition. The present project has aimed at pushing interactivity to Level 2, where students can effectively pose **questions** to the system (*I understand x, but tell me more about y*). Beyond this, at Level 3, we can envisage adaptive, intelligent systems which would infer a student’s **needs** from his/her patterns of behaviour and response.

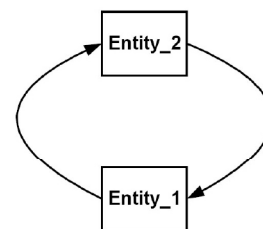


Figure 1. A closed loop (Yacci, 2000)

4 Interactive ear-training Ear-training aims to enhance skill in auditory identification and discrimination by systematic exposure to a range of stimuli which differ in controlled respects, accompanied by appropriate feedback. Plainly, a completely pre-determined programme of ear-training (whether on physical media or a webpage) may lose value through familiarity on re-use, and must lead all students through identical steps regardless of their needs and progress. Advantages of computer-based ear-training are thus the possibility of unpredictable re-permutation of stimuli drawn from a database (producing ear-training exercises which are re-usable indefinitely) and the potential to tailor an individual's path through training materials in response to progress and

¹ Funded by a Teaching and Research Award from CDE (Centre for Distance Education, University of London), whose support is gratefully acknowledged.

difficulties. Stand-alone demonstrations of these principles are straightforward to engineer; more challenging are the integration of such exercises within a VLE (the native exercise paradigms of which may be ill-suited), separation of the content-authoring and programming tasks, and development of a significant body of worthwhile materials making a measurable impact within a teaching programme.

5 Practical Phonetics Moodle course A Practical Phonetics Moodle course was developed in early 2008, available to students across a number of Speech Sciences and Linguistics degree programmes, but aimed particularly at those on first-year modules. At the most basic level, it is an online reference resource in which a defined repertoire of vowels and consonant sounds can be heard alone or in simple combinations such as CV or CVC. As an illustration, Figure 2 shows a page presenting Cardinal Vowels in [dVd] syllables, and a page introducing nasal consonants in [aCa] context. Similar pages are provided for all categories of segments.

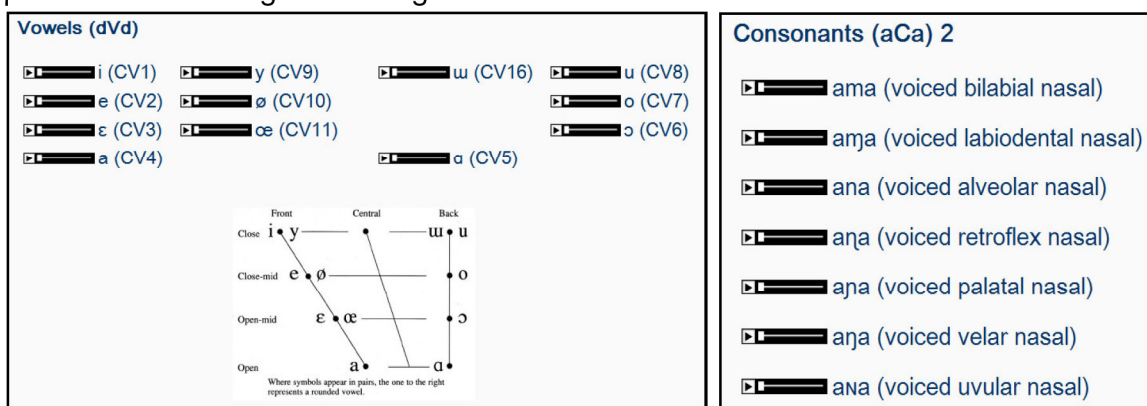
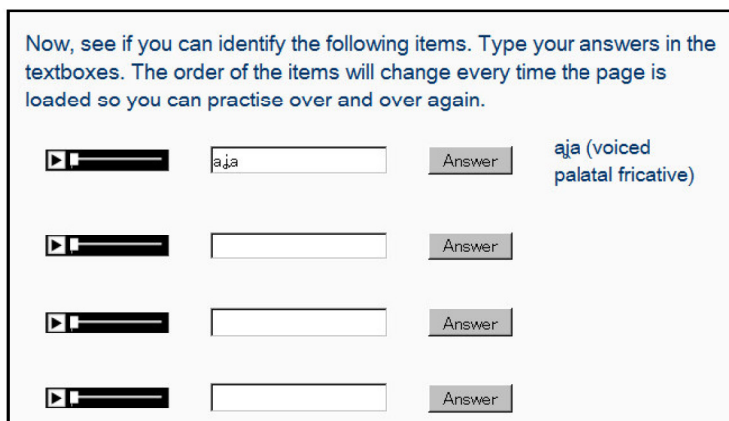


Figure 2. Screenshots of course pages introducing vowels (left) and nasal consonants.

Even for the single speaker used so far, it was necessary to create some 680 individual mp3 files. A semi-automated recording workflow, with cueing of the speaker linked to segmentation and titling of the sound files, was followed by manual editing to crop silences, and remove extraneous noises. Then a script-assisted audio workflow resampled each file at 22.05kHz, normalised for amplitude, and encoded in mp3 format at 160kbps, a value we had selected after trials.

The resulting resources resemble those available in clickable IPA charts already available, but differ in providing a wider sampling of contexts, more explanatory material,



and systematic ear-training (see Figure 3) in which the selection and order of items varies with each re-loading of the page. JavaScript was used to produce randomisation and selection on the ear-training pages.

Figure 3. An ear training page devoted to fricatives in VCV structures. The user has transcribed the first item and revealed the answer.

Along with consonant and vowel identification, and simple nonsense words such as [ʒɔʏ, ɥetʰ], other ear-training tasks presented this way include **substitutions**, in which labels are required for successive changes made to one segment within an English word – for example, *arrow* with intervocalic /r/ replaced in turn by [v, r, ʁ, ʁ, ...] – and **intonation**, in which a short English sentence such as *The meeting's scheduled for Tuesday* is presented in a variety of realizations differing in tonicity and tone choices which must be analysed and described. The tasks were chosen to correspond with those in the practical and oral examinations which the target audience was approaching.

6 Interactive response to questions

Certain pages are provided with pop-up windows (Figure 4) which offer a range of contrasting alternatives when a particular symbol in a listening item is queried by clicking. The alternatives are based on classroom experience of likely difficulties and confusions. A student who experiences no difficulty at that point can pass straight on without branching into the additional practice window.

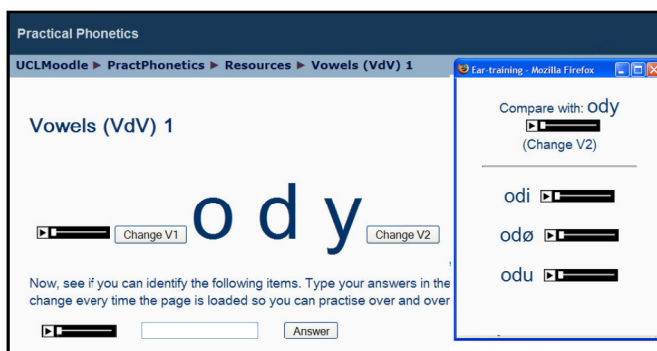
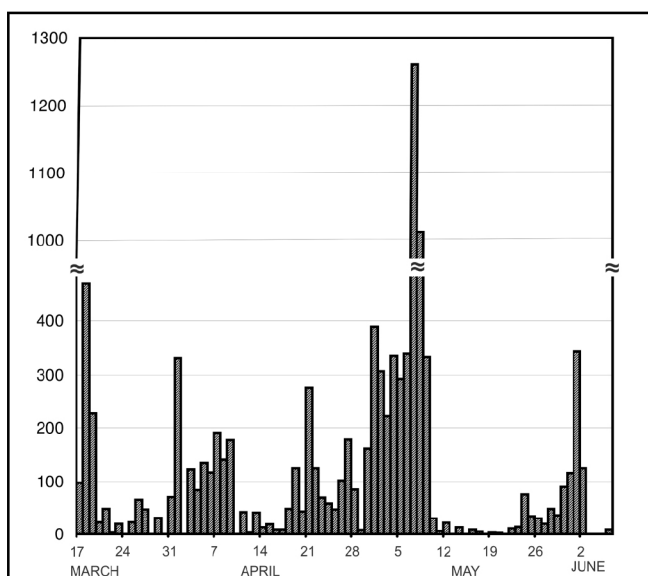


Figure 4. Interactivity achieved through optional pop-ups.

7 Uptake and utilisation

The course was monitored over a 12-week period in the Spring of 2008. From a combined cross-programme cohort of 73, the final enrolment was 67 (a takeup rate of 92%).



The course remained in almost constant use, day and night, for three months, and total page accesses exceeded 9000. Peaks in usage (see Figure 5) can be linked straightforwardly to the dates of examinations, showing response to fluctuating demand.

Figure 5. Daily page accesses over a 12-week period.

8 Audio input The discussion so far has focused on the interactive selection of speech stimuli for *output* to participants in listening tasks. But students are also learning production skills, so there is a complementary need for

speech *input*, to be judged by tutors, or shared with fellow students – and ideally for synchronous interactive audiovisual exchange among participants.

A VLE such as Moodle allows for the sharing of media files as attachments to forum postings or assessment submissions, though the process is cumbersome, especially if a single syllable or segment is the focus of interest. In this project we have evaluated the potential of Skype as a vehicle for practical phonetics teaching and assessment. Its

performance can probably be taken as representing what can be achieved over a regular internet connection by any current VoIP system – whether standalone or integrated into a VLE. Clearly, if a premium is paid for dedicated bandwidth, a multimedia conferencing tool can achieve any arbitrary level of performance.

9 Measuring performance Extreme variability in the connections established in VoIP leads us to prefer a practical measure of the overall quality of interaction rather than an assessment of the technical capabilities of the Skype codec. In our tests, participants exchange spoken productions of randomly generated nonsense syllables (Figure 5); when an error occurs, an exchange of repetitions and corrections follows until both participants are satisfied, the measure of the relative usefulness of the channel being the time taken complete a block consisting of a standard number of syllables. Audio-only and webcam-video conditions are compared. The five authors participated in various pairings over links of differing lengths, including intercontinental trials between Britain and Japan.

Although certain aspects of the speech tokens under discussion (particularly rhythm, stress, tone and intonation) are apparently preserved quite well, it is plain that many sound-types suffer significant degradation. Noted areas of difficulty include place of articulation in fricatives and nasals, the distinction between implosives and regular pulmonic plosives, distinctions of place among clicks, and certain vowel differences, such as front rounded vs back unrounded. Some of these relate to visible aspects of speech production, and under good connection conditions, use of video can halve the time taken to exchange a test block. More commonly, however, the deterioration in audio performance incurred by the demand on bandwidth for the video channel more than offsets any advantage, and in many trials the video condition had to be abandoned as unviable.

	A	B
1	p'	a
2	z	y
3	k'	u
4	ϕ	e
5	ŋ	y

Figure 6. Random syllable generator realized in Microsoft Excel.

10 Conclusions

Most of the problems in the construction of an interactive online ear-training course have been solved, and a highly successful example has been created, though development of materials remains labour-intensive.

While VoIP can provide the infrastructure of synchronous interactivity required for teaching or assessment, it requires to be supplemented by uncompressed audio and video files representing the spoken tokens under discussion, the interchange of which would require rather more than real time. A goal for development is to streamline the processes of file creation and upload, permitting a "one-click" switch into high-quality asynchronous operation at critical points.

11 References

- Ashby, M., House, J., Huckvale, M., Maidment, J. and Yanagisawa, K. (2007a). "A distance e-learning course in phonetics". In *Proceedings of the 16th ICPHS*. Saarbrücken. pp. 1713-1716.
- Ashby, M., House, J., Huckvale, M., Maidment, J. and Yanagisawa, K. (2007b). "Phonetics in a Virtual Learning Environment". In *Proceedings of PTLC2007* pp. 62 - 127.
- Yacci, M. (2000) "Interactivity Demystified: A Structural Definition for Distance Education and Intelligent CBT". <http://www.it.rit.edu/~may/interactiv8.pdf>, accessed 08 July 2009.