Mobile Adaptive CALL (MAC). A lightweight speech-based intervention for mobile language learners

Maria Uther[1], James Uther[2], Panos Athanasopoulos[3], Pushpendra Singh[4] & Reiko Akahane-Yamada[5]

[1] Centre for Cognition and Neuroimaging, School of Social Sciences, Brunel University, London, United Kingdom.
[2] School of Information Technology, University of Sydney, NSW 2006 Australia
[3] Department of Language and Linguistics, University of Essex, Essex, U.K.
[4] Culture Lab, University of Newcastle Upon Tyne, Newcastle Upon Tyne, United Kingdom.
[5] ATR-Promotions Inc./ ATR Cognitive Information Science Laboratories, Kyoto, Japan

maria.uther@brunel.ac.uk, james.uther@gmail.com,pathan54@hotmail.com, pushpendra.singh@ncl.ac.uk, yamada@atr.jp

10.21437/Interspeech.2007-631

Abstract

A computer-assisted language learning software for mobile devices (MAC) is presented, that was aimed to helping speakers acquire speech contrasts not native to their own language. The software is based on the high variability phonetic training (HVPT) technique. An overview of the software is given, followed by results from an efficacy study. Two groups using slightly different variations of the MAC software. One group received software training that was equivalent to an ordinary HVPT training, but delivered on mobile devices. The second group received a version that used an adaptive algorithm to determine which words and speakers that the learner had most difficulty with, and gave them more practice on those. Results showed that both groups showed significant but equivalent improvements. These results also showed that the magnitude of improvement using mobile phones was similar to those obtained using fixed PCs.

Index Terms: language learning, training, phonetic discrimination, mobile devices

1. Introduction

This study sought to use advances in the field of mobile devices to be applied to the area of second language learning. We chose one specific aspect of second language learning for this initial effort: the perception of non-native speech sounds (/r/ and /l/) for Japanese speakers of English. This application was thought to be ideally suited for delivery on mobile phones for a variety of reasons: it requires only simple interfaces, can be done on a purely 'ad hoc' or casual basis (e.g. while waiting for a bus), does not require much processing power to deliver the application but in particular its requirement for audio interaction suits the affordances of mobile phone technology well.

We specifically focussed on the case of Japanese /r/-/l/ discrimination, primarily because a large body of work has shown that PC-based training can very reliably induce changes in perception. Work by Bradlow and colleagues has for example shown that training individuals to perceptually discriminate /r/ and /l/ contrasts results in improved pronunciation [1,2]. Their approach has shown to be highly effective [2,3] resulting in improvements in discrimination to around 82 percent accuracy from 65 percent and lasting 3 months [2] from a training period of 15-22 hours over 3 weeks. The specific method used in these /r/-/l/ training studies was to train on as wide a phonetic variability (often termed 'high variability') as possible. This involves providing the learner with speech samples across a full range of linguistic variability that s/he would be expected to encounter in everyday language.

This 'high-variability' training provides presentations of different samples from different speakers and using different types of words. In terms of word-type variability, the target contrast can occur at the beginning or elsewhere in the word as in 'rake' vs. 'lake', 'blew' vs.'brew', etc. This linguistic variability was used by Logan et al. [3] and later also in other studies [2,4] with effective and long-lasting outcomes. On the basis of the success of these studies, word type and talker type variability would appear to be highly important factors in successful training. Indeed these results has also been validated by many subsequent studies demonstrating successful and long-lasting training effects [2,5]. Variability is thought to be a way of ensuring that the learner sees the perceptual constancy across a wide range of environments.

Our Mobile Adaptive CALL (MAC) software builds directly on the work done using a high variability training approach and also tackles the problem of Japanese /r/-/l/ discrimination which is notoriously difficult for Japanese language learners of English [6]. We advance on current high-variability approaches in CALL for /r/ vs /l/ training by first exploiting the dimensions used as part of that approach to tailor the training for the learners. We chose a tailored approach because research has shown that more personalised tutoring approach can be a particularly successful strategy to train learners [7]. Indeed, other work [9] has also successfully used adaptive algorithms for /r/-/l/ training. Also, there is evidence to suggest that although Japanese speakers will, at a general level tend to be worse at certain word positions, at an individual level there is a large degree of variation in the discrimination of /r/-/l/ for different word and talker types [7]. Developing a model of individual differences would appear to be a way of more effectively delivering this training. This adaptation can tap into the learner's possible weaknesses with respect to a talker's voice, e.g. male or female talker, or a characteristic of the word, e.g. positioning of /r/ or /l/ within the word. Based on these premises, MAC aims to adapt to individual learners' needs in order to focus training in areas that are most difficult for the individual whilst at the same time, maintaining some variability.

Another important and novel aspect of our approach is to use mobile technology to deliver the training. Previous studies using /r/-/l/ training (at least to our knowledge) have only been delivered on traditional PC-based platforms. This has significant drawbacks for the learner: s/he would need to attend training sessions at a fixed location (usually by means of attending a lab every day for a fixed period) and would not be able to practice in their own time. Even if implemented as an Internet application, this generally still makes the application less accessible and potentially costly. Learners with unlimited web access still have to boot up and use a PC. Those using PDAs/phones with Internet connectivity would still have to pay for Internet access. In contrast, a large percentage of the population see their phone as a personal, trusted device that they always carry. MAC provides the opportunity of downloading the application to a Java-enabled phone that the learner may already own. If learners were able to practice in their own time, whenever and wherever they wanted to (e.g waiting for or on public transport, sitting at home), then it would likely provide much quicker and more effective training. Indeed, recent examples of mobile learning applications [e.g. 10, 11, 12] have pointed to these advantages. Another often overlooked consideration (particularly for mobile phones) is that there is a natural affordance for phones to speak into and listen from (i.e. audio interaction). In this way, applications that involve audio interaction may actually be better suited for mobile phones than for the PC environment. Modern handsets are often used for games, push-to-talk, and even listening to music. For these types of mobile applications, the audio is often played using a speaker functionality. MAC also follows this interaction style (but also could be used in conjunction with a headset).

The MAC software adapts (using a Boltzmann-like algorithm) according to the learner's responses and presents to the learner a speech contrast of the type in which they will most need further practice. The software was designed for J2ME -compliant mobile phones and programmed using Java MIDP. We have published the software under a General Public Licence on the web at http://code.google.com/p/macall/ and provide a detailed description of the algorithm and the software in our previous work [13].

2. Study of MAC efficacy

To assess the efficacy of the intervention, we loaned MAC to 11 Japanese adult learners of English for two weeks. We implemented a pre- and post-test design, assessing perception and production of minimal pair word tokens contrasting /r/ and /l/. To specifically assess the utility of *adaptability* in remedial speech training using the high variability approach, we also designed a non-adaptive version of MAC (which equated to an 'ordinary' High Variability Phonetic Training (HVPT) program for mobile phones), which we loaned to a control group of 11 Japanese adult learners of English. The aim was two-fold: 1. to test whether a lightweight device such as mobile phone can provide equivalent benefits in training to the kind of results yielded by PC interventions and 2. To determine whether an *adaptive* version of HVPT could provide improved outcomes compared to 'ordinary' HVPT. We chose not to include another control group of PC-only or no-intervention learners as there was already an abundance of studies demonstrating that (a) ordinary PC-based HVPT yields between 15-18% improvement and (b) that there is no spontaneous improvement from the simple act of pre- and post- testing [1-5], at least not for /r/-/l/ perception in Japanese speakers of English.

2.1. Method

2.1.1. Participants

We recruited a total of 23 adult native speakers of Japanese (4 males, 19 females), ranging in age from 18 to 33 years. They were all recruited from language schools in London in the UK. These students had just arrived in the UK to complete a short course in English as a foreign language, and they were selected from basic English courses. They also self-reported that their level of speaking and listening skills in English were poor to basic (from a scale of 'poor-basic intermediate-advanced'). None had lived in an English-speaking country before and were in England no more than 1 month at the time of testing. Participants were divided into 2 groups – MAC intervention and a control mobile HVPT. Four participants had to be excluded due near ceiling level performance in the pre-training task, leaving 10 in the MAC group and 9 in the Mobile HVPT group.

2.1.2. Stimuli

A large database of digitally recorded spoken words for the perception tests was provided by the Department of Phonetics and Linguistics at University College London. These stimuli were identical to those used a previous study [14] and the recording procedures were described in detail in that paper. The pre-training and post-training *test* stimuli consisted of 99 minimal pairs that contrasted /l/ and /r/ in three phonetic environments: initial singleton, initial cluster and intervocalic, with 33 minimal pairs from each category. The *training* stimuli consisted of a corpus of 360 different minimal pairs also contrasting /l/ and /r/ in the three phonetic environments mentioned above, and recorded by 4 speakers of Standard British English.

2.1.3. Procedure

The perceptual training program followed the procedures used by the previous high-variability studies [1-4]. Participants first underwent a minimal pair perceptual identification test. For this test, two minimal pair /r/-/l/ words (e.g. 'rake' and 'lake') appeared on a PC, and a spoken word was presented. The participant's task was to identify using a button press which written word corresponded to the spoken word. Their accuracy and reaction times were recorded by the PC. Following the initial pre-training test, participants then underwent a two week training phase. In contrast to previous studies, the training phase did not require participants to attend laboratory sessions. Instead, participants were given a Nokia 6630 mobile phone with the MAC software. They were instructed that they must use the intervention at least for 30 minutes everyday for two weeks, and to make a written record of the date and duration of each time they use it. The program itself also recorded the participants' usage data, allowing the administrator to see when and how many times individual participants used the program, although sometimes software bugs in early implementation prevented this data from being completely accurate or indeed available at all in some cases.

After the training phase, the participants were tested again using the same minimal pair perceptual identification task as in the pre-training test. Production samples of /r/-/l/ words were also recorded to determine whether there were any effects of perceptual training on production. The MAC training software itself involved (as with other HVPT interventions) minimal pair identification. However, in the training phase, feedback and 'correction trials' were given (whereas in the test phases they were not). The feedback was in the form of a chime signaling a correct response or a buzz signaling an incorrect response. For the 'correction trials', this involved giving an additional trial for the participant when they responded incorrectly. After the incorrect response buzz, the test pair was repeated (with a random order of presentation on the screen) until the participant gave a correct response. Control subjects were given exactly the same pre and post-test tasks, but during the training they were given an alternate, non-adaptive version of the software (Mobile HVPT). This non-adaptive version was equivalent to an ordinary high-variability phonetic training, but delivered on a mobile phone. All testing was carried out in a quiet room at either the participants' language school in London or in a quiet room at UCL Phonetics Department, London.

2.2. Results

Our results showed that there was a clear improvement in participants' performance for the minimal perceptual identification task from the pre-training to the post-training tests. The accuracy increase was around 15% across both groups, which is comparable to other HVPT studies and also was statistically significant ($F_{1,17}$=39.184, p<0.01), see Figure 1.

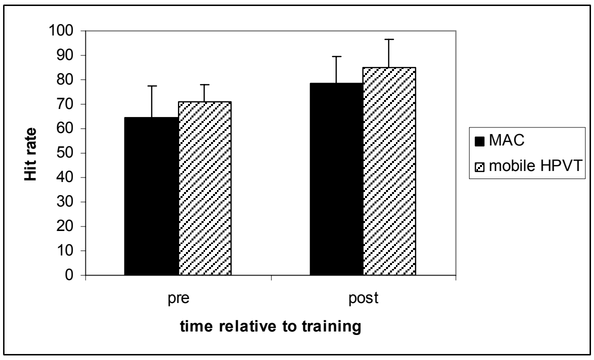

Figure 1. Improvements in performance for MAC and mobile HPVT. Error bars show standard deviation.

There was no significant difference between the two groups (MAC vs. HVPT) and no significant group by training interaction. This suggests that there were equivalent gains for the Mobile HVPT control group (who also received the 'simple' version of the program without any adaptation in the training phase) and the 'MAC' (adaptive mobile HVPT) group.

The participants' usage data were also analysed. The data from the mobile phones was only available for a few participants due to technical difficulties. However, the self-report data was available for all participants (bar one, which did not record their usage due to non-compliance). Of those that did have phone usage data (n=10), there was a significant correlation between the mobile device usage data and the self-report data (pearsons coefficient=0.681, p<0.05), suggesting hat the self-report data was fairly reliable. We also analysed whether there was any relationship between performance improvement (i.e. hit rate post-training minus hit rate pre-training) and found that there was no significant correlation between improvement level and usage level. However, a closer inspection of the data would suggest that this may be due to the fact that there was very little variability in usage time, with most participants using the software for just under the recommended 30 mins (mean=20 mins) and only 3 participants used it for less than this amount of time (shortest time was 5 mins). Not a single participant exceeded the recommended usage time of 30 mins.

3. Discussion

Our results would suggest that it was feasible to deliver training using mobile phones. However, it appears that there were no gains to be made from using adaptation in the training. However, this result may support the result of other studies using PC-based technology and variants of HVPT which suggest that HVPT is 'as good as' other, more sophisticated methods [14]. Of course, other studies may be needed in order to determine whether it still might be possible to get greater improvements with different adaptive algorithms to the one we used.

Notwithstanding the finding that adaptive algorithms did not provide superior training, we of course still had the positive finding that speech training could be adequately achieved with simple mobile phones. In this sense, the MAC intervention was a success, as one concern was the memory storage available within mobile phones and the fact that audio codecs that one needs to use in order to minimize file storage would not be sufficient to achieve training. Admittedly, in previous pilots, some of the audio codecs that we trialled (e.g. AMR) were not of sufficient quality in user studies, so we had to opt for the superior mp3 format [15]. Given that most mobile devices these days also support mp3 format, this was not an issue. What was not clear though, was whether one could make significant gains using the mp3 format compared to using native .wav files that one uses on fixed PCs. Our results clearly show that training with these lossy codecs is indeed possible and practical.

It was interesting to note that there was no relationship between usage time and performance improvement. However, as we discussed in the results, this is likely to be due to the fact that there was quite a lot of homogeneity in the usage data. Nonetheless, one limitation of the study was that the usage data from the phones was not always so reliable due to the device programming limitations. In future iterations, we would hope to improve this. As we were able to simultaneously collect self-report diary data, this was not a huge problem. However, it would be less onerous on the participants if all usage data collection were automatically and reliably collected by the device.

In terms of future work, we also collected speech production samples in order to determine whether there were any effects of perceptual training on production, however, this data has not yet been analysed, but would be interesting to look at to see if there are also production improvements in the mobile HVPT as has been shown in the PC-based studies (e.g.[1]). Also, we are looking to extend the work to see if equivalent gains can be made in other speech training contrasts (e.g. vowel perception). This latter work will of course be particularly important in the light of data showing that the HVPT training method has been used to successfully train speech contrasts other than /r/-/l/ (e.g. [16]).

4. Conclusions

In summary, we have shown that speech training useful for successfully acquiring a second language is possible on a lightweight device such as a mobile phone. Given the potentially compromised sound quality of such devices compared to PCs, it appears that equivalent gains can be made, and within a short period of time. This innovation therefore presents a useful tool for those learning second languages as it provides a much more portable and convenient method of learning than a fixed PC tutor.

5. Acknowledgements

This study was supported by the Engineering and Physical Sciences Research Council (EPSRC), U.K. (Grant GR/S55095/01). The authors would also like to thank the Drs. Paul Iverson and Valerie Hazan of UCL for the use of their speech samples. We also thank the key language schools in London for their assistance and co-operation in helping us recruit participants: especially LSI International and also King's London (Beckenham). Thanks to Kota Hattori for assistance in collecting a portion of the data. Thanks also to Shenene Kara and Rashmi Punja for their assistance in analyzing a portion of these data.

6. References

[1] Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R.. and Tokhura., Y. (1997) Training japanese listeners to identify english /r/ and /l/: IV. some effects of perceptual learning on speech production. Journal of the Acoustical Society of America, 101(10):2299–2310.

[2] Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R.. and Tokhura., Y. (1999) Training japanese listeners to identify english /r/ and /l/: Long term retention of learning in perception and production. Percept Psychophys., 61(5):977–985.

[3] Logan, J. S., Lively, S. E. and Pisoni., D. B. (1991) Training japanese listeners to identify /r/ and /l/: A first report. J. Acoust. Soc. Amer., 89(2):874–886.

[4] Lively, S. E., Logan, J.S. and Pisoni, D.B. (1993) Training japanese listeners to identify english /r/ and /l/: II. the role of phonetic environment and talker variability in learning new perceptual categories. J. Acoust. Soc. Amer., 94(3):1242–1255.

[5] Bradlow, A.R., Pisoni, D. B., Akahane-Yamada, R., Tokhura, Y., Lively, E. and Yamada, T. (1994) Training japanese listeners to identify english /r/ and /l/ iii. Longterm retention of new phonetic categories. J. Acoust. Soc. Amer., 96(4):2076–2087.

[6] Ingram, J.C. and Park, S-G. (1997) Cross-language vowel perception and production by japanese and Korean learners of english. Journal of Phonetics, 25:343–370.

[7] Bloom, B. S. (1984) The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13:3–16.

[8] Takagi, N. (2002) The limits of training japanese listeners to identify english /r/ and /l/: Eight case studies. J. Acoust. Soc. Amer., 111(6):2887–2896.

[9] McClelland, J.L., Fiez, J.L. and McCandliss, B.D. (2002) Teaching the /r/ -/l/ discrimination to Japanese adults: behavioral and neural aspects. Physiology and Behavior, 77:657–662.

[10] Chen, Y.S., Kao, T.C. and Sheu, J.P.(2003) A mobile learning system for scaffolding bird watching learning. Journal of Computer Assisted Learning, 19(3):347–359.

[11] Davis. S. (2002) Research to industry. four years of observations in classrooms. Proceedings of IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE), 99:31–38, Aug, 2002.

[12] Naismith, L., Sharples, M., & Ting, J. (2005) Evaluation of CAERUS: a context aware mobile guide. In H. van der Merwe & T. Brown, Mobile Technology: The Future of Learning in Your Hands, mLearn 2005 Book of Abstracts, 4th World Conference on mLearning, Cape Town, 25-28 October 2005. Cape Town: mLearn 2005, p.50.

[13] Uther, M., Singh, P., Zipitria, I., & Uther, J. (2005) 'MAC: An adaptive, perception-based speech remediation s/w for mobile devices'. Artificial Intelligence in Education (AIED) workshop on language tutoring, July, 2005, IOS Press.

[14] Iverson, P., Hazan, V. & Bannister, K. (2005) Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults J. Acoust. Soc. Am. 118, 3267.

[15] Uther, M., Zipitria, I., Uther, J. & Singh, P. (2005) 'Mobile Adaptive CALL (MAC): A case-study in developing a mobile learning application for speech/audio language training.', IEEE International Workshop on Wireless and Mobile Technologies in Education, 28-30 Nov. 2005; Pages:187-191.

[16] Wang, Y., Spence, M.V., Jongman, A. and Sereno, J.A. (1999) Training American listeners to perceive Mandarin tones. J. Acoust. Soc. Am. 106, 3649-3658.