To what extent can Automatic Speech Recognition (ASR) support the effective development of the intelligibility of careful speech for English for Academic Purposes (EAP) students preparing to give presentations on short pre-sessional courses in the U.K.?

Introduction

Pre-sessional EAP courses in a Higher Education (HE) setting in the UK are typically short intensive courses of ten to twelve weeks, with students from a range of linguistic backgrounds, often with a predominance of speakers of Arabic and Asian languages, at a level of 5.0 to 6.0 according to the International English Language Testing System (IELTS), needing to pass a series of assessments in order to progress onto their degree courses. One common element of these courses is an academic presentation from the students' subject area, usually of ten to fifteen minutes in duration, given to a group of peers and instructors, using Power Point, delivered in English from minimal notes, using high level academic and subject specific vocabulary that may present a major challenge to their existing pronunciation and speaking skills (Hincks, 2005). One of the assessment criteria is typically the clarity of pronunciation (BALEAP, 2013; DE Chazal, 2014), in other words, the students' intelligibility, which can be defined as the extent to which the listener is able to understand the speaker (Derwing and Munro, 2005). EAP courses need to provide the necessary course elements in order to help students improve their intelligibility. However, beyond the basics required to pronounce the language adequately at the beginning of their learning, pronunciation seems to often be neglected in the teaching of English as a Foreign Language (EFL) (Gilakjani, Ahmadi, and Ahmadi, 2011). If anything, this could be said to be magnified in an EAP classroom, where time constraints often mean it is possible to allocate only a few sessions to pronunciation. Therefore, there is a learner need for a structured approach to improve students' pronunciation. Given the constraints mentioned above, this will probably need to facilitate pronunciation development outside the classroom (Gilakjani, Ahmadi, and Ahmadi, 2011). I will therefore explore the extent to which existing technology can be implemented to help students in this process. I will outline the capabilities of automated speech ASR software, give an overview of the elements present in a scaffold for pronunciation teaching for presentations,

before exploring the affordances of ASR in terms of helping students to improve their careful speech for presentations.

ASR Technology

ASR is often mentioned in research papers and articles in terms of its potential for developing pronunciation (Derwing et al., 2000; Deng and Trainin, 2015), or its shortcomings, such as ASR's limited ability to provide useful feedback, especially to non-native speakers (NNS) (Derwing, Munro, and Carbonaro, 2000). However, ASR has developed rapidly over recent years. Liakin (2015) states that the available literature suggests that integrating ASR technology into pronunciation teaching may have positive effects on pronunciation. There is a wide range of applications (APPS) and programmes that either provide ASR functions, for example, *Nuance Dragon Dictation* and *Praat*, or incorporate ASR into their software, for example, *Duolingo*. According to Levis and Suvorov (2014) and McCrocklin (2016), ASR is software that transcribes speech from input received through a microphone, analyzed using algorithms and probability with a database to judge similarity before producing textual, visual or audio output. The type of input varies from 'speaker dependent', which is focused on one person's voice, 'speaker independent' that works with any speech falling within a given range, and 'speaker adaptable' that adapts to the speaker's voice (Rosen and Yampolsky, 2000; Young and Mihailidis, 2010; Liakin, 2015). ASR also differentiates between 'discrete word recognition', 'connected word recognition' and 'continuous speech recognition' (Rosen and Yampolsky, 2000). Negative assessments of ASR have tended to focus on practical and technical aspects. Chen (2011) found it too expensive and lacking accessibility; Meisam and Tavakoli (2015) deemed it too sophisticated, and Liakin (2015) stated that high levels of volume or the lack of a strong enough internet connection make it difficult to use. However, most systems such as *Praat* and *Dragon Dictation* are available for free and easily downloadable on most platforms, can be used offline once downloaded, and require little instruction (Meisam and Tavakoli, 2015); nearly all research uses microphones to improve accuracy (Kawahara et al., 2010), and *Dragon Dictation* recommends using a headset with a microphone (Nuance, 1991); research is also being done into noise-reducing algorithms (Luan et al., 2012). In sum, provided that learners have access to a good internet connection when required, the practical and

technical problems should be limited. But the question remains how effective ASR can be in terms of helping students develop their intelligibility through a scaffolded approach to presentations.

The elements of intelligibility

There are different elements to consider in adopting a scaffolded approach to developing students' intelligibility for academic presentations. A typical approach on a pre-sessional course would include the following: individual sounds, word stress, sentence stress (chunking), intonation, volume and pace (Clementson, 2017). This approach is supported by much available research. When an aspect of pronunciation hinders intelligibility, communication will be impeded (Setter and Jenkins, 2005). However, if teachers understand the effect on intelligibility of the different aspects of learner pronunciation, they can help students improve (Derwing & Munro, 1997). The main aim is that students can be understood, for which they need to have good pronunciation but not a perfect accent (Wang and Young, 2014).

Volume and pace

There seems to be little research into volume and pace in English Language Teaching (ELT) literature. Most references seem to come either from speech therapy or practical manuals on how to improve presentation skills. In a typical approach taken in such manuals, Levin and Topping (2006, p.17) relate the 'small voice' effect to tension and nerves or cultural inhibitions, and recommend a range of physical and articulatory exercises to help overcome it. In terms of academic presentations on EAP courses, the importance of volume and pace needs further research.

Segmentals and suprasegmentals

Most research seems to broadly agree on the importance of segmentals and supra-segmentals in improving learners' pronunciation, although the relative degree of focus is often controversial. Lambacher (2001) finds that segmentals and not just supra-segmentals are important for intelligibility. In addition to individual sound

disparities between L1 and L2, it is also important to focus on syllables and where they occur in a word, for example, the final consonant problem in languages such as Thai and Chinese (Setter and Jenkins, 2005; Gilakjani, Ahmadi, and Ahmadi, 2011), and the context in which the sounds occur (Gilakjani, Ahmadi, and Ahmadi, 2011). This seems to indicate that it is not simply the articulation of the sounds that poses problems, but that conceptual and cognitive problems seem to lie behind the failure to produce correct sounds. Also, most recent research stresses the importance of prosody to intelligibility (Anderson-Hsieh et al.,1994; Magen,1998; Munro and Derwing, 1999; Gilakjani, Ahmadi, and Ahmadi, 2011: Meisam and Tavakoli, 2015; Lima, 2016), and focusing on supra-segmentals can bring improvement in a short time (Derwing, Munro and Wiebe, 1998) and training in prosody in preparation for presentations was helpful to students' academic success (Lima, 2016). However, Setter and Jenkins (2005) state that supra-segmental features pose problems because they do not take place at a conscious level, and that misplacing syllables and vowel reduction can lead to misunderstanding and suggests that stressed syllables serve as a way to access lexical meaning. This is supported by Cutler (1984), who puts forward the idea that the stress pattern of a word is part of the information stored in the mental lexicon, and if this is not adhered to, intelligibility will suffer. Intonation is a controversial area where there is some disagreement about how it functions and how to teach it (Setter and Jenkins, 2005). Correct speech rhythms also seem important in terms of intelligibility (Setter and Jenkins, 2005) and problems are likely to be exacerbated if the learners' first language is syllable-timed rather than stress-timed (Gilakjani, Ahmadi, and Ahmadi,). There is some controversy over the stress-syllable timed distinction. Cauldwell and Hewings (2002) describe English speech as 'irrythmical', and Cauldwell (1996) stated that differentiating in this way hampers progress in understanding spontaneous speech. However, students making presentations on EAP courses are not expected to produce spontaneous speech, but careful, rehearsed speech. The fact that many teachers nowadays prefer to call stress-timing a tendency only (Setter and Jenkins, 2005), and one that is mostly suitable in formal speech, could be seen as a reinforcement of this approach to teaching presentations.

Implementing a scaffolded approach

In view of the elements described above, it seems reasonable that in order to help learners approach intelligibility, teachers should teach prosodic elements as well as individual sounds (Gilakjani, Ahmadi, and Ahmadi, 2011). In terms of segmentals, a common approach is to diagnose which sounds students have difficulties with before the students are instructed to work on their own (Gilakjani, Ahmadi, and Ahmadi, 2011). This is often done by identifying particular language groups, as can be seen from books such as *Pronunciation in Use* (Hancock, 2003), which, typically of materials used for this purpose, has detailed lists of which sounds to focus on for speakers of specific languages. In terms of helping students with prosodic elements, Marks (1999) is in favour of using rhymes to give students a scaffold. This approach can be seen clearly in many materials seeking to develop the presentation skills of students, in which one of the most important steps of preparing students for presentations is to practice chunking, rhythm and stress patterns and intonation. *Presenting in English* (Powell,1996), exemplifies the type of exercises widely used to help students with these issues. On the basis of this scaffold, students will develop, script, rehearse and improve their presentations, putting into practice the concept put forward by Bruce (2011) that a process of practice and rehearsing speaking tasks makes possible the internalization of these tasks. As already stated, on pre-sessional EAP courses, classroom time for this process, and therefore direct access to peer and teacher feedback, is limited. Students therefore need a way of developing their capabilities and analysing their strengths and weaknesses in order to improve. Audio-recording speech to identify errors is often suggested (Celce-Murcia et al., 2010) but with little effect (McCrocklin, 2016). This highlights the question as to whether students have the requisite knowledge or skills to successfully analyse their own speech and performance in order to improve. According to McCrocklin (2016), feedback is essential for the development of pronunciation, and integrating ASR into pronunciation training has potential in this respect. Gilakjani, Ahmadi, and Ahmadi (2011) identify the need for software that will identify and give feedback on problems thereby helping students deal with them autonomously, saving valuable class time by giving students a platform to work on their pronunciation.

ASR: Four key issues

There seem to be four main issues that need to be explored in terms of ASR's usefulness: accuracy, the model language used by the ASR, accuracy recognition and the feedback provided.

Accuracy

According to Kim (2006), the aim of ASR is to achieve 100% accuracy of intelligible sounds, regardless of speed, content or background noise. However, Kim (2006) reports that accuracy rating at below 90%, with some systems achieving only 60 %, depending on accent, background noise and what is being said in terms of subject and quantity. In a study using FluSpeak, Kim (2006) found that the software was not as accurate as human analysis and that the ASR system was of little use in terms of analyzing intonation.

The model language

This brings into question the model language used in ASR systems. Most ASR software differentiates between non-native-speaker (NNS) and native-speaker (NS) accents. Kim, Oh, and Yoon (2007) found that since ASR software is based on NS speech it performs well with NS but less so with NNS. There are claims that this has been improving (Neri et al., 2003). In contrast, McCrocklin, (2016) Dragsted et al. (2011) and Esshassah (2016) still noted recognition problems, especially with NNS.

Accuracy recognition

Many ASR misrecognitions can happen with NS just as easily as with NNS. These can be caused by system error, homophones, word boundary problems, and hesitations (Dragsted et al., 2011), and also by contextual and other phonetic clues (Esshassah, 2016). However, Dragsted et al. (2011) found that NNS students' mispronunciations were the largest group. These resulted largely from specific speech sounds being mispronounced, particularly in function words, resulting from speakers' failure to stress and unstress vowel sounds correctly, and the software being unable to distinguish between voiced and unvoiced consonants in final position. This bears strong similarities to the features of pronunciation listed earlier that account for problems of intelligibility. However, Liakin (2015) states that

specially designed ASR systems that accept mispronunciations do better with NNS. This can be done by either basing ASR systems on NNS speech, or enhancing existing systems to accommodate NNS speech by adjusting the recognition accuracy rate (Kim, Oh, and Yoon, 2007). This clearly has pedagogical implications as to whether it would be acceptable or desirable to teach NNS pronunciation. On the other hand, it would be possible to adjust the recognition accuracy rate to be more tolerant of beginners' accents and less so of more advanced students (Kim, 2006). This would tie in with the inter-language model described by Setter and Jenkins (2005); furthermore, this would work well with the English as an international language (EIL) communication models between NNS and NNS put forward by Setter and Jenkins (2005).

Feedback

Perhaps the most immediate and clearest feedback provided by ASR is speech to text feedback, such as *Nuance Dragon Dictation* (Liakin, 2015), which enables students to check their oral pronouncements by reading the transcript. The transcript can also be integrated into other APPS for further work (Deng and Trainin, 2015). Despite limitations in terms of accuracy, this opens up a great deal of potential for work on preparing presentations. However, textual feedback cannot show prosodic features. Visual feedback seems to have the most potential for prosody training. Hincks (2003) highlights the potential of ASR to provide student-friendly feedback featuring a visual representation that allows users to compare intonation, a rating of the learners' pronunciation accuracy, the highlighting of words that are not pronounced accurately. It seems that systems that integrate a combination of feedback types are most effective. Kim (2006) and Wang and Young (2014) both found that systems using different modalities of feedback demonstrated the potential to improve learners' pronunciation. The *Say it* APP (Dance, 2016) takes a similar approach, though there is as yet no empirical data to show its effectiveness. Unfortunately, many attempts to use ASR for prosodic features have struggled to achieve meaningful success. A common criticism seems to be that feedback is too sophisticated to be understood, or not meaningful enough to be of use (Anderson-Hsieh, 1994; Neri, et al.,2002; Kim, 2006; Gilakjani, Ahmadi, and Ahmadi, 2011; Wang and Young, 2014). A further possibility would be to use ASR in combination

with classroom instruction and teacher feedback. This can be done online or in class to help learners develop oral skills (Kim, 2006; Cauldwell, 2013; Liakin, 2015). Cauldwell (2013) has done this with academic presentations, and has put forward the idea of ASR software used interactively in combination with the teacher to improve monotonous delivery. In terms of presentations, the central question remains the extent to which the integration of ASR into pronunciation training can improve intelligibility.

The effectiveness of ASR

There seems to be quite strong evidence to support the idea that ASR can support the improvement of pronunciation at segmental level, whereas at supra-segmental level there is less. Liakin (2015) suggests that current off the shelf dictation software has advanced in the field of recognition of non-native speech. Both Chen (2006) using *My English Tutor*, Liakin (2015) using *Nuance Dragon Dictation* and Gorjian, Hayati, and Pourkhoni, (2013), Olson, (2014), Meisam and Tavakoli (2015) and Esshassah (2016) using *Praat,* reported improvements in segmentals using ASR. Whereas Liakin (2015) speculated that this could be transferred to other segmentals, (Meisam and Tavakoli, 2015) stated that this is generalizable to other vowels in other languages. Liakin (2015) suggests targeting specific sounds with a high functional load likely to affect intelligibility, and puts forward other phonetic features such as prosody as subjects for further research. But as yet, results of studies on supra-segmentals seem inconclusive. Chen (2011) notes insufficient feedback on stress and intonation training. (González, 2012) finds that the *Dragon Dictation* APP and similar seem to have great potential but lack of research is hindering their use in terms of teaching supra-segmentals. Setter and Jenkins (2005) note that materials to deal with supra-segmentals are in development, but at best these types of materials can be used to complement classroom teaching rather than replace the teacher. However, (Gorjian, Hayati, and Pourkhoni, 2013) noted that students who learnt prosodic features through CALL did better than those who used a traditional approach such as repetition or phonetic symbols. In addition, Le and Brook (2011) reported promising results from using *Praat* to teach students how to improve intonation. Research is also being done in the field of using ASR in speech therapy. For example, Nolan et al. (2012) discusses the use of visual feedback on volume

levels in patients with Parkinson's disease. There are also APPS available, such as *Bla Bla Bla*, suggested uses of which include helping develop volume levels, pace and sentence stress (Gurevitch, 2012), though many of these are recommended on professional websites and there seems to be little research or evidence to support their efficacy as yet. Evidence to support ASR development in terms of prosody then is mixed, but there are other factors to consider.

Autonomy and motivation

A great deal of the literature on ASR seems to concentrate on how accurate it is, but it is also relevant to consider the generally positive view taken of ASR by students (Chen, 2011; Dragsted, Mees, and Gorm, 2011; Wang and Young, 2014; Ahn and Lee, 2015; Liakin, 2015; Esshassah, 2016), and the contribution this technology could make to developing learner autonomy (Jenkins, 2003; Meisam and Tavakoli, 2015; McCrocklin, 2016). Motivation seems to be a major factor in improving pronunciation (McCrocklin, 2016), and ASR seems to enjoy a positive reception amongst students for several reasons. Firstly, Cauldwell (2016) noted that there seems to be a strong student perception that ASR can help improve pronunciation. Secondly, ASR enables students to practice independently in a risk free environment (Banafa, 2008; Wang and Young, 2014; Ahn and Lee, 2015; Mccrocklin, 2016; Esshassah, 2016). Furthermore, despite awareness of the limitations of the software, ASR also enables students to rehearse and practice, thereby developing their awareness of their pronunciation issues (Anderson-Hsieh, 1992; Ahn and Lee, 2015; Liakin, 2015; Esshassah, 2016; McCrocklin, 2016), freeing up valuable teaching time (Liakin 2015; Meisam and Tavakoli, 2015), improving in a shorter time frame due to the immediate one to one feedback not possible in most classes (Ahn and Lee, 2015), and adapting to different learner styles (Hsu, 2015; Liakin, 2015;). If students are encouraged to focus on how to make the message clear rather than the pronunciation of individual sounds (Setter and Jenkins, 2005), this would seem to fit well to developing intelligibility in presentations by providing a non-linguistic goal that would help to focus the students on the task by allowing them a platform on which to practice their presentations.

ASR and teaching theory

ASR technology still does not seem to be widely used by EAP teachers, despite the increasing number of research studies being carried out that seem to underline its potential and support its integration into teaching as being theoretically sound. For example, it can be seen as both behavourist, when used for recording, repetition and shadowing, and constructivist according to Vygotskii (1978) when used for practice and role-play (Ahn and Lee, 2015). Although Chen (2011) noted the limited scope for creativity because of the reliance on fixed dialogues and exercises, Liakin (2015) states that ASR can be used for explicit teaching and learning, but also makes interaction with computers possible, facilitating its mobile use. It seems clear then that it depends what the technology is used for rather than the technology itself. In order to evaluate this, the model of Substitution, Augmentation, Modification, Redefinition (SAMR) proposed by Puentedura (tycro1, 2013), can be applied. In terms of task types that it potentially makes possible, ASR can be said to have the potential to redefine pronunciation teaching. But in terms of the current situation and its capability to give meaningful feedback, it is hard to argue that it substitutes the teacher performing the same task, as the teacher is still in a position to give better, more informed, more detailed feedback. Chen (2011) sums this up when reporting that there are useful aspects but lists some of the limitations, especially limited feedback. There seems to be a lot of potential for ASR to change the balance and type of activities in the classroom and encourage autonomy, increase creativity and change the balance between individual learning and classroom activities (Kim, 2006). This would seem in many ways to be subject to the levels of tasks and also about having a more tolerant language model that is more responsive to non-native English speakers (Chen, 2011). Perhaps ASR's true current potential lies in its integration by teachers into their classroom teaching, equipping students to continue to use it outside the classroom. For example, if the approach that Cauldwell (2013) takes to presentations is observed, and even combined with *Dragon Dictation* to provide text feedback in addition to ASR chunks and audio, the task model could be said to be moving towards redefinition. This approach enables learners to listen to their own spoken utterances whilst receiving written feedback in the form of text, visual feedback in the form of chunks and also feedback from their teacher. This approach then clearly places the onus on the teachers to develop their Technological, Pedagogical and Content Knowledge (TPACK) (Koehler and Mishra, 2009) in order

to provide the best help to their students (Barrett and Liu, 2016). TPACK can be summarized as being the foundation for teaching effectively with technology (Koehler and Kishra, 2009) and means explicitly that its three parts fuse into one and are not regarded individually. In practice this means that teachers should be able to select the most appropriate technology for an individual task, integrate that technology into their classes, and train students to use that technology, thereby encouraging the normalization of technology in the classroom (Bax, 2005) through making students comfortable with the technology and encouraging them to use it themselves outside class and after the course has finished (McCrocklin, 2016).

Conclusion

In terms of the benefits of integrating ASR into pre-sessional courses in UK HE institutions to help international students prepare for presentations, my conclusions are as follows: First of all, there seem to be some concrete uses of ASR. The technology is no longer prohibitively expensive or difficult to access, nor does it require a great deal of learner training. Integrating ASR into pronunciation training seems to help students improve the pronunciation of individual sounds and words. APPs such as *Say it* and *Dragon Dictation* could be integrated to develop the pronunciation of individual words and practice word stress. Given the restricted and specific nature of the vocabulary used in an academic presentation, words that present problems could be identified and rehearsed as part of the preparation. Beyond segmental level, ASR technology does not yet seem to have reached the stage of development where it can reliably give students working alone feedback on their spoken English. This area needs further research. Perhaps the strongest arguments in favour of using ASR technology are in in terms of motivation, awareness-raising and autonomy. ASR programmes such as *Dragon Dictation* and *Sonocent's Audio Note-taker* could be used separately or in conjunction, autonomously or with the teacher, in the process of scripting, rehearsing, reviewing and improving the intelligibility of students' pronunciation. Its usefulness at present would fall more strongly into the realm of process rather than product in the absence of empirical evidence to show otherwise, but students would have a platform to rehearse and practice their presentations with various options available for them to receive feedback and evaluate their own work.

References:

Anderson-Hsieh, J. (1994) 'Interpreting visual feedback on Suprasegmentals in computer assisted pronunciation instruction', *CALICO Journal*, 11(4), pp. 5–22. doi: 10.2307/24152754.

Ahn, T. youn and Lee, S.-M. (2015) 'User experience of a mobile speaking application with automatic speech recognition for EFL learning', *British Journal of Educational Technology*, , p. n/a–n/a. doi: 10.1111/bjet.12354.

BALEAP. (2013) *Can do framework for EAP syllabus design and assessment 1* Available at: https://www.baleap.org/wpcontent/uploads/2016/04/Can_Do_Framework__with_sample_activities_April_2013.pdf (Accessed: 15 December 2016).

Barrett, N.E. and Liu, G.. (2016) 'Global trends and research aims for English academic oral presentations: Changes, challenges, and opportunities for learning technology', *Review of Educational Research*, 86(4), pp. 1227–1271. doi: 10.3102/0034654316628296.

Bax, S. (2003) 'CALL—past, present and future', *System*, 31(1), pp. 13–28. doi: 10.1016/S0346-251X(02)00071-4.

Bruce, I. (2011) *Theory and concepts of English for academic purposes*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

Cauldwell, R. and Hewings, M. (1996) 'Intonation rules in ELT textbooks', *ELT Journal*, 50(4), pp. 327–334. doi: 10.1093/elt/50.4.327.

Cauldwell, R. (2002) 'The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms', *Apples – Journal of Applied Language Studies*, 2(1), pp. 1–24.

Cauldwell, R. (2013) *Phonology for listening: Teaching the stream of speech*. Birmingham: Speech In Action.

Cauldwell, R. (2016) *Mountains not Monotones peaks and valleys in oral presentations*. Available at: http://www.speechinaction.org/wp-content/uploads/2016/10/Handout-Mountains-not-Monotones.pdf (Accessed: 22 December 2016).

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). Teaching pronunciation: A course book and reference guide. Hong Kong: Cambridge University Press.

Chen, H.H.-J. (2011) 'Developing and evaluating an oral skills training website supported by automatic speech recognition technology', *RECALL*, 23(1).

Clementson, C., 2017. Presentations Week Six: Sentence Stress and Pacing, ... *Summer EAP.* University of Brighton, unpublished. (See appendix A.)

Cutler, A. (1984) *Stress and accent in language production and under- standing 1*. Available at: http://pubman.mpdl.mpg.de/pubman/item/escidoc:69903/component/escidoc:506931/Cutler_1984_Stress+and+accent.pdf (Accessed: 12 January 2017).

Dance, J. (2016) *How I developed a pronunciation app and got it published by OUP*. Available at: http://eltjam.com/oup-pronunciation-app/ (Accessed: 23 November 2016).

De Chazal, E. (2014) *English for academic purposes A comprehensive overview of Eap and how it is best taught and Learnt in a variety of academic contexts*. Oxford, United Kingdom: Oxford University Press.

Deng, Q. and Trainin, G. (2015) *Learning vocabulary with Apps: From theory to practice*. Available at: http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1033&context=nebeducator (Accessed: 30 November 2016).

Derwing, T.M., Munro, M.J. and Wiebe, G. (1998) 'Evidence in favor of a broad framework for pronunciation instruction', *Language Learning*, 48(3), pp. 393–410. doi: 10.1111/0023-8333.00047.

Derwing, T.M., Munro, M.J. and Carbonaro, M. (2000) 'Does popular speech recognition software work with ESL speech?', *TESOL Quarterly*, 34(3), p. 592. doi: 10.2307/3587748.

Derwing, T.M. and Munro, M.J. (2005) 'Second language accent and pronunciation teaching: A research-based approach', *TESOL Quarterly*, 39(3), p. 379. doi: 10.2307/3588486.

Dragsted, B., Copenhagen, Mees, I.M. and Hansen, I.G. (2011) 'Speaking your translation: Students' first encounter with speech recognition technology', *Translation & Interpreting*, 3(1), pp. 10–43. doi: 10.12807/t&i.v3i1.115.

Esshassah, F. (2016) *The effects of computer-enhanced reading aloud on the production of /p/ and /ε/ in Arabic-English Interlanguage*. MA thesis. Montreal. Available at: http://spectrum.library.concordia.ca/981239/1/Esshassah_MA_2016.pdf (Accessed: 2 December 2016).

González, J.F. (2012) 'Can apple's iPhone help to improve English pronunciation autonomously? State of the App', *CALL: Using, Learning, Knowing, EUROCALL Conference, Gothenburg, Sweden, 22-25 August 2012, Proceedings*, . doi: 10.14705/rpnet.2012.000031.

Gorjian, B., Hayati, A. and Pourkhoni, P. (2013) 'Using Praat software in teaching prosodic features to EFL learners', *Procedia - Social and Behavioral Sciences*, 84, pp. 34–40. doi: 10.1016/j.sbspro.2013.06.505.

Gurevitch, N. (2012) *IPhone/iPad app Bla | Bla | Bla: A sound Reactive App*. Available at: http://slp-tech.blogspot.co.uk/2012/03/bla-bla-bla-sound-reactive-app-for.html (Accessed: 22 January 2017).

Hancock, M. (2003) *English pronunciation in use*. Cambridge: Cambridge University Press.
(Hancock, 2003)

Hincks, R. (2003) 'Speech technologies for pronunciation feedback and evaluation', *Recall*, 15(1).

Hincks, R. (2005) 'Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism', *System*, 33(4), pp. 575–591. doi: 10.1016/j.system.2005.04.002.

Hsu, L. (2015) 'An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training', *Computer Assisted Language Learning*, 29(5), pp. 881–900. doi: 10.1080/09588221.2015.1069747.

Huensch, A. and Tremblay, A. (2015) 'Effects of perceptual phonetic training on the perception and production of second language syllable structure', *Journal of phonetics*, 52, pp. 105–120.

Kawahara, T., Wang, H., Tsubota, Y. and Dantsuji, M. (2010) *English and Japanese CALL systems developed at Kyoto university*. Available at: http://sap.ist.i.kyoto-u.ac.jp/EN/bib/intl/KAW-APSIPA10.pdf (Accessed: 15 January 2017).

Kim, I.-S. (2006) 'Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation', *Journal of Educational Technology & Society*, 9, pp. 322–334. doi: 10.2307/jeductechsoci.9.1.322.

Kim, H.K., Oh, Y.R. and Yoon, J.S. (2007) *Patent US8515753 - acoustic model adaptation methods based on pronunciation variability analysis for enhancing the recognition of voice of non-native speaker and apparatus thereof*. Available at: https://www.google.com/patents/US8515753 (Accessed: 2 December 2016).

Koehler, M.J. and Mishra, P. (2009) 'What is technological pedagogical content knowledge?', *Contemporary Issues in Technology and Teacher Education,* 9(1), pp. 60–70.

Lambacher, S. (2001) 'A brief guide to resources for developing expertise in the teaching of pronunciation', *Prospect*, 16(1), pp. 63–70.

Lan Zhang and Manwa L. Ng (2013) 'An Acoustics-Based Assessment of Error Patterns in English Sounds Produced by Cantonese ESL Speakers | Published online: 18 Jul 2013', , p. Pages 73–84.

Le, H.T. and Brook, J. (2011) *Using Praat to teach intonation to ESL students*. Available at: http://www.hpu.edu/CHSS/English/TESOL/ProfessionalDevelopment/201109_TWPspringfall11/9_1-2_Le-Brook.pdf (Accessed: 21 January 2017).

Levin, P. and Topping, G. (2006) *Perfect presentations!* Maidenhead: Open University Press.

Levis, J., and Suvorov, R. (2014) Automated speech recognition C. Chapelle (Ed.), The encyclopedia of applied linguistics Retrieved from http://onlinelibrary.wiley.com.ezproxy.brighton.ac.uk/store/10.1002/9781405198431.wbeal0066/asset/wbeal0066.pdf?v=1&t=htq1z7hp&s=139a3d9f48261a7218270113d3833da39a187e74

Liakin, D. (2015) 'Learning L2 pronunciation with a mobile speech recognizer: French /y.', *CALICO Journal*, 32(1).

Lima, E.F. (2016) 'Comprehensibility and liveliness in nonnative student oral presentations before and after training: A mixed methods study', *System*, 63, pp. 121–133. doi: 10.1016/j.system.2016.10.003.

Luan, Y., Suzuki, M., Yamauchi, Y., Minematsu, N., Kato, S. and Hirose, K. (2012) 'Performance improvement of automatic pronunciation assessment in a noisy classroom', *2012 IEEE Spoken Language Technology Workshop (SLT)*, . doi: 10.1109/slt.2012.6424262.

Marks, J. (1999) 'Is stress-timing real?', *ELT Journal*, 53(3), pp. 191–199. doi: 10.1093/elt/ccm073.

McCrocklin, S.M. (2016) 'Pronunciation learner autonomy: The potential of automatic speech recognition', *System*, 57, pp. 25–42. doi: 10.1016/j.system.2015.12.013.

Meisam, R. and Tavakoli, M. (2015) 'The effectiveness of CALL in helping Persian L2 learners produce the English vowel /ɒ/', *GEMA Online® Journal of Language Studies*, 15(3).

Neri, A., Cucchiarini, C. and Strik, H. (2002) *Feedback in computer assisted pronunciation training: technology push or demand pull?* Available at: http://hstrik.ruhosting.nl/wordpress/wp-content/uploads/2013/04/a91.pdf (Accessed: 14 January 2017).

Nolan, P., Hoskins, S., Johnson, J., Powell, V., Chaudhuri, K.R. & Eglin, R. 2012, "Implicit theory manipulations affecting efficacy of a smartphone application aiding speech therapy for Parkinson's patients", *Studies in health technology and informatics,* vol. 181, pp. 138.

Nuance Inc, N.C. (1991) *Dragon NaturallySpeaking 13 installation guide and user guide*. Available at: http://www.nuance.com/ucmprod/groups/corporate/@web-enus/documents/collateral/dns13_userguide.pdf (Accessed: 15 January 2017).

Olson, D.J. (2014) 'Benefits of visual feedback on segmental production in the L2 classroom, *Language Learning & Technology*, 18(183), pp. 173–192.

Powell, M. (1996) *Presenting in English: How to give successful presentations*. Hove: Language Teaching Publ. [u.a.].

Rosen, K. and Yampolsky, S. (2000) 'Automatic speech recognition and a review of its functioning with dysarthric speech', *Augmentative and Alternative Communication*, 16(1).

Setter, J. and Jenkins, J. (2005) 'State-of-the-art review article', *Language Teaching*, 38(01), p. 1. doi: 10.1017/s026144480500251x.

Sonocent (2013) Available at: http://www.speechinaction.org/sonocent/ (Accessed: 22 December 2016).

tycro1 (2013) *The SAMR model explained by Ruben R. Puentedura*. Available at: https://www.youtube.com/watch?v=_QOsz4AaZ2k (Accessed: 9 January 2017).

Vygotskii, L.S. (1978) *Mind in society: The development of higher psychological processes*. Edited by Michael Cole. Cambridge, MA: Harvard University Press.

Wang, Y.. and Young, S.S.. (2014) 'Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system', *Journal of Computer Assisted Learning*, 31(6), pp. 493–504. doi: 10.1111/jcal.12079.

Young, V. and Mihailidis, A. (2010) 'Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: a literature review', *Assistive technology: the official journal of RESNA*, 22(2).