Running Head: COMPUTER AIDED PRONUNCIATION TRAINING

Literature Review:
Computer Aided Pronunciation Training

Jonan P. Donaldson
ED 633/Dr. Rachel Harrington
Western Oregon University
July 29, 2009
Abstract

This literature review provides an overview of what computer-based technologies are used in the teaching of pronunciation of English to speakers of other languages, the specific pronunciation goals of those technologies, and the effectiveness of such technologies. Teachers of pronunciation to English language learners need assistance in selection of technologies for use in their classrooms. Issues of intelligibility and the goals of Computer Aided Pronunciation Training (CAPT) are central to this literature review. The effectiveness of Automated Speech Recognition (ASR) tools in CAPT, as well as non-ASR-enabled CAPT systems will be reviewed.
Literature Review: Computer Aided Pronunciation Training

The use of technology in education has grown exponentially over the last few decades, from the limited use of audio/visual equipment such as overhead projectors and video players to the use of clickers, smart boards, PowerPoint presentations, online classes, video recording, and statistical analysis applications in classes of all kinds and levels. The use of technology in the teaching of English to speakers of other languages has also increased dramatically, including in the teaching of pronunciation in English language classes.

Teachers of pronunciation of English to speakers of other languages have a wide array of technologies available. However, as emerging technology, there is relatively little research available to help teachers select the technologies that will be of greatest benefit. Further complications arise from the lack of clarity in the articulation of the goals of teaching pronunciation due to a recent shift from the definite goal native-like pronunciation to the indefinite goal of intelligibility.

This literature review will attempt to provide a clear picture of what computer-based technologies have been used in the teaching of pronunciation, the specific goals of those technologies, and their effectiveness. The primary focus will be to assist teachers of pronunciation in their selection of technologies for classroom use. The secondary focus will be to point to the holes and limitations in available research, which may stimulate ideas for future research in this field. First we will consider the issues of intelligibility and the goals of Computer Aided Pronunciation Training (CAPT). Then we will look at the use and effectiveness of Automated Speech Recognition (ASR) tools for teaching pronunciation. Finally we will look at the effectiveness of other CAPT applications.

One of the issues concerning CAPT which teachers and developers must first address is goals. Many pronunciation teachers adopt the goal of intelligibility. Others adopt a goal of near-native pronunciation. These two goals are very different and require different approaches to teaching. Intelligibility is an indefinite goal somewhere along a continuum from complete unintelligibility to native pronunciation. Using CAPT for near-native pronunciation would focus on supra-segmental aspects of pronunciation such as linking, assimilation, and intonation. Using CAPT for intelligibility would focus on segmental aspects of pronunciation such as vowels, consonants, and phoneme clusters.
One problem with intelligibility as a goal for pronunciation training is the difficulty in defining intelligibility. This literature review will start by introducing a study by Isaacs (2008) which highlights the difficulties raised by setting a goal of intelligibility. Although intelligibility is the most common goal for pronunciation training, there is no common definition of intelligibility, nor is there a commonly accepted measure thereof. Almost no empirical evidence exists to help teachers know which pronunciation features are necessary for intelligibility. Intelligibility is “an evasive concept that we know little about” (Isaacs, 2008, p. 556). It has been defined in many ways, including pronunciation free of aspects that interfere with communication, pronunciation that listeners can comfortably understand, pronunciation that does not distract listeners, and pronunciation that does not irritate listeners (Isaacs, 2008).

The research goals in this study were to ascertain if intelligibility is a “sufficient goal and adequate assessment criterion” for evaluating pronunciation and if so, what the minimum acceptable threshold level is, and if none can be identified, what criterion would be better (Isaacs, 2008, p. 561).

Non-native-English-speaking graduate students were rated by native English speakers in terms of what percentage of the speaker’s speech they understood, ease of understanding, and on “speech clarity, rate of speech, pitch, sentence rhythm, word stress, and individual consonant or vowel sounds” (Isaacs, 2008, p. 563). Each rater chose the two best speakers and the two worst speakers (Isaacs, 2008).

Isaacs found that the speakers were rated with varying degrees of intelligibility and comprehensibility. The ratings agreed at the extremes, but were not consistent in the middle. The raters were all given the same instructions and materials, but they all interpreted intelligibility differently. Raters reported that many of the non-native speakers did not have pronunciation good enough to teach classes even though they had high ratings (Isaacs, 2008).

Isaacs said that intelligibility is a useful measure, being “a necessary but not sufficient condition to be a TA in an undergraduate course” (Isaacs, 2008, p. 571). This study could not find a minimum threshold level of intelligibility. One problem is that “intelligibility presupposes the existence of both a speaker and a listener” (Isaacs, 2008, p. 572). Accented speech, although intelligible, may negatively impact communication because “listener attitudes have the potential not only to adversely impact their interactions with NNSs, but also to bias their assessments of non-native speech . . . Ratings are, by nature, subjective” (Isaacs, 2008, p. 573).
Intelligibility is difficult for native English speakers to rate. Not only is intelligibility impossible to define, any attempt to measure it will by nature be highly subjective. It is therefore impossible for computers to measure intelligibility. Therefore, CAPT systems will always be limited if the goal of pronunciation teaching is intelligibility because computers, by nature, require clearly defined goals. One approach is to define goals in terms of feedback.

A study conducted by Engwall and Balter (2007) addressed the issue of what students and teachers want from CAPT. For CAPT to work, the feedback should be the right kind of feedback, based on what pronunciation teachers and students report as the necessary kind of feedback. The majority of pronunciation specialists now believe that not every error should be corrected, and the correction that is given should not be given immediately so as to not affect the self-confidence of the student or their self-monitoring practices. However, the majority of students think that most errors should be corrected (Engwall & Bälter, 2007).

In this study looked at six teacher interviews, five student interviews, and three classroom observations. A one-hour focus-group interview with teachers, a one-hour focus group with students, individual interviews with two teachers, and individual interviews with two students provided additional in-depth data (Engwall & Bälter, 2007).

The teachers reported that they give almost no pronunciation feedback because of time limitations or because they did not want to interfere with communication. Both teachers and students felt that feedback should not interrupt communication. The students reported wanting feedback as soon as possible without interrupting communication. The teachers believed that pronunciation feedback should only be given if what the student says cannot be understood, if the same mistake occurs repeatedly, if the listener could get a bad impression of the speaker, or if the error is one made by many students. Students wanted individual sounds explained in addition to being modeled. In classroom observations, the teachers did not give immediate feedback unless a student was struggling and only pronunciation errors which were the focus of study, or errors which interfered with communication, were corrected (Engwall & Bälter, 2007).

This study showed that very little pronunciation feedback is given in classrooms, but CAPT systems could provide more feedback. The students in this study wanted CAPT to focus on one pronunciation feature at a time, rather than giving feedback on every error. The teachers wanted CAPT to adapt to the student. The students and teachers in this study said CAPT should allow students to decide the kind and amount of feedback given (Engwall & Bälter, 2007).
Students should have the ability to select what kind of, and how much, pronunciation feedback is given by CAPT. It should be designed to provide information “without user request” and students should be “made aware of the additional feedback that may be received by request. . . . Focus should be paid to how classroom feedback may be transferred and adapted to individualized pronunciation training with computers” (Engwall & Bälter, 2007, p. 260).

When CAPT systems are being chosen by teachers or being designed, the goal should be feedback in the way and amount that professional pronunciation teachers and students suggest. The amount of feedback should be a balance between too little and too much. It should focus on one aspect at a time and should give useful advice and practice for making improvements.

A study by Mitra, Inamdar, and Dixond (2003) looked into the effects of students using CAPT systems on their own without the guidance of a teacher. The authors believed that automated speech recognition (ASR) could be used to determine the quality of pronunciation and that children can learn to use computers for self-instruction without the presence of a teacher. Pronunciation patterns heavily influenced by the native language are often incomprehensible. Most English teachers are not native English speakers, so CAPT is important. If ASR-enabled CAPT software is available, students can improve their pronunciation on their own (Mitra, Inamdar, & Dixond, 2003).

Sixteen students in a high-poverty area of India, after individually watching up to four hours of movies on the computer, and after using vocabulary and grammar language software for four months, read passages and the ASR would make transcriptions which were compared with the original passages. The data from the transcripts produced by ASR was compared with the data from the scoring by human judges who rated each recording (Mitra et al., 2003).

The judges in this study were consistent in their ratings and reported an improvement in pronunciation. The ASR provided an average score for the study group of 72%, and for the control group of 30%. The ASR, when normalized to human judging, provided reliable judgment of pronunciation improvement, reporting study group improvement and no improvement for the control group. This study showed that although ASR is unreliable in giving feedback on specific pronunciation problems, it can give reliable feedback on pronunciation improvement over time. Furthermore, using computers for video and other English-exposure activities helped students improve their pronunciation without any explicit training (Mitra et al., 2003).
A more recent study by Kim (2006) looked into the reliability of ASR scoring of pronunciation. The purpose of this study was to determine the correlation between ASR and human scoring (Kim, 2006). Reliable machine transcription of intelligible spoken language at the time of the study it was less than 90% and was influenced by accent and environment. The most popular type of ASR compares utterances with phonemes in a database of recordings of native speakers. This technology lets students receive a pronunciation score or it highlights words when the pronunciation is not understood (Kim, 2006).

Using CAPT software, students listened to sentences spoken by native English speakers as many times as they wanted, and then made recordings. The CAPT-produced word scores and intonation scores for each student, on a scale of 1 to 100, were recorded. The recordings were rated by three native English speakers on a four-point scale (Kim, 2006).

The ratings given by the scorers were not consistent. The correlation between the pronunciation scores given by the software and those given by the raters was weak. The correlation between the intonation scores given by the software and the pronunciation scores given by the raters was extremely weak. The reliability of the rating this software gave was interpreted as being moderate (Kim, 2006).

Although ASR scoring has low correlation with human scoring, it can be a valuable tool in teaching pronunciation, especially when used alongside classroom instruction. The present level of this technology is “far below the desired level of accuracy” (Kim, 2006, p. 330). The author concluded that although the reliability of ASR rating was not good, it could still be used as a valuable tool (Kim, 2006).

Two articles were found reporting studies led by Neri. The first was a study by Neri, Mich, Gerosa, and Giuliani (2008) in which the authors tried to ascertain whether children using CAPT systems can improve their pronunciation to the same degree as children studying in a traditional classroom setting with a teacher. ASR technology is advancing and now automatic feedback “can vary from rejecting poorly pronounced utterances and accepting ‘good’ ones to pinpointing specific errors either in phonemic quality or sentence accent” (Neri, Mich, Gerosa, & Giuliani, 2008, p. 394). ASR feedback can help students become aware of pronunciation problems, which helps prevent habits of incorrect pronunciation. Unfortunately, most studies about ASR focus on technology advances rather than pedagogy (Neri, Mich, Gerosa, & Giuliani, 2008).
This study compared an ASR-enabled CAPT system with traditional classroom teaching in terms of pronunciation improvement. The children in both the study group and the control group worked with repetition and vocabulary games, so the treatment for both groups was the same except one was with a teacher and the other was with CAPT. Recordings were scored by three native English-speaking teachers on a ten-point scale (Neri, Mich, Gerosa, & Giuliani, 2008).

Inter-rater reliability in this study was high. Each student got two sets of scores – one by averaging the student scores given by the raters, and the other by averaging the individual ASR word scores for each student. There was high correlation between the two scores for each student. The scores for both groups were similar in the pre-treatment sample and both improved in the post-treatment scores (Neri, Mich, Gerosa, & Giuliani, 2008).

Children who trained with CAPT improved just as much as children with a teacher. The results of this study “might be explained by the fact that the children using [CAPT] enjoyed the computer’s ‘undivided attention’ for all 30 minutes of training, while the children training with the teacher could seldom practice and receive feedback individually during the 60-minute lesson” (Neri, Mich, Gerosa, & Giuliani, 2008, p. 404). The authors concluded that: “CAPT systems could be used to integrate traditional instruction, for instance to alleviate typical problems due to time constraints or to particularly unfavourable teacher/student ratios. In this way, children could benefit from more intensive exposure to oral examples in the FL, and from more intensive individualized practice and feedback on pronunciation in the FL. This would free up time for the teacher, which could be employed to provide individual guidance on how to remedy specific pronunciation problems – something computers are not yet capable of doing in a reliable way” (Neri, Mich, Gerosa, & Giuliani, 2008, p. 405).

This study showed that CAPT can be a valuable tool for teaching pronunciation to children. Such systems can, because of their individualized nature, help students with pronunciation just as reliably as traditional classroom instruction.

The next study in which Neri was involved was a study by Neri, Cucchiarini, and Strik (2008) in which phoneme-specific pronunciation training using an ASR-enabled CAPT system was analyzed. The authors started with the hypothesis that ASR can provide automatic feedback of an individualized nature, so they studied whether ASR-based feedback on individual phonemes known to cause problems can improve pronunciation (Neri, Cucchiarini, & Strik, 2008).
A CAPT system was specifically built for this study. This system provided “an overt and clear indication that an error occurred” (Neri, Cucchiarini, & Strik, 2008, p. 227). The phonemes chosen for attention in this study were selected because they were frequent, common for students with various native languages, persistent, possible to cause communication difficulty, and were easy for an ASR to detect. The software indicated which phoneme was mispronounced by the student, after which the student could listen to their own utterance again or listen to the model.

Each student in the study group and two control groups was pre- and post-tested by recording sets of sentences that included all the phonemes of the language. Six experts rated the recordings, focusing only on segmental quality. The participants in the study group completed anonymous questionnaires to obtain their opinions about the programs (Neri, Cucchiarini, & Strik, 2008).

The ratings in this study were found to be highly reliable. Data analysis showed that the pronunciation of targeted phonemes improved for the experimental group: “It would seem that training . . . benefited these learners by accelerating their development” (Neri, Cucchiarini, & Strik, 2008, p. 239). Participants indicated positive reactions, saying that the feedback function was necessary (Neri, Cucchiarini, & Strik, 2008).

The authors concluded that ASR-based feedback helps students improve their pronunciation of difficult phonemes, but does not have an effect on overall pronunciation quality of non-targeted phonemes. ASR-based feedback on target phonemes may help students develop skills in distinguishing between specific segmental errors and global errors. They argue that an ASR system which provides automatic feedback on specific targeted phonemes “can be a useful pedagogical tool to supplement regular teacher-fronted classes” (Neri, Cucchiarini, & Strik, 2008, p. 241).

Not all CAPT systems rely on or integrate use of ASR, as seen in the final articles. A study by Seferoglu (2005) aimed to discover if CAPT software without ASR could help students improve their overall supra-segmental pronunciation skills (Seferoglu, 2005). The author believed that although language learners rarely acquire native-like pronunciation they can improve their pronunciation when the teaching of pronunciation is approached on both the segmental level and on the supra-segmental level (Seferoglu, 2005).

Computer technology has been used for teaching pronunciation since the 1960s, but only in the last ten years has it become commonplace. Most CAPT systems focus on accuracy, while classroom methodology usually focuses on fluency or communicative skill. Most CAPT systems
focus on segmental aspects of pronunciation. Sereroglu studied whether a CAPT system would improve segmental and supra-segmental pronunciation (Seferoglu, 2005).

A control group studied as a class, but the experimental group studied during class time on their own using the CAPT. For the pre- and post-study tests, students gave interactive presentations. The presentations were rated for pronunciation by the researcher and videotaped for later rating by another rater. Participants were rated on phonemes, diphthongs, consonant clusters, linking, word stress, sentence stress, rhythm, and intonation. At the end of the study the researcher interviewed all participants about their pronunciation (Seferoglu, 2005).

The experimental group showed improved pronunciation. The author interpreted the findings to indicate that the CAPT system is helpful in improving pronunciation, especially in situations where little native language exposure is available and should be used for structured drills in conjunction with communicative activities (Seferoglu, 2005).

Unlike the study by Seferoglu where all aspects of pronunciation training including phonemes, diphthongs, consonant clusters, linking, word stress, sentence stress, rhythm, and intonation were targeted, Verdigo (2006) focused on a highly-specialized CAPT system which gave students feedback on intonation alone because language learners usually use the intonation patterns of their native language, which may reduce intelligibility (Verdugo, 2006).

This study looked at the effectiveness of a CAPT system for improving both awareness of intonation and production of natural intonation patterns. It used acoustic analysis of participant’s utterances for intonation patterns, assessment of recordings of participant’s conversations by four native English speakers, observations, and questionnaires. Two issues were addressed: the effectiveness of visual pitch displays and the effectiveness of comparing native English speaker pitch displays with those of utterances by the participants (Verdugo, 2006).

A control group was given the same pre- and post-tests which were recorded. All the recordings were assessed by four native-English-speaking raters. Raters also were asked about each participant’s intelligibility and intonation quality. Raters were finally brought together for a follow-up discussion (Verdugo, 2006).

A group of English language learners had intonation training and a control group had regular English language classes. The study group received intonation training by listening to recordings of dialogues by native English speakers while watching pitch displays. At the top of the screen was the pitch display of the native English speaker and at the bottom of the screen was the
learner’s pitch display. After ten weeks, the study group recorded dialogues and free discussion and completed questionnaires concerning pronunciation and intonation (Verdugo, 2006).

The study group showed increased quality of intonation and higher levels of awareness of intonation. The control group showed no change in intonation. Native English-speaking raters gave higher ratings to the experimental group after the study in terms of general improvement of intonation. The study found “following training the spoken performance and degree of intelligibility were perceived to have improved significantly in the experimental group but not in the control group” (Verdugo, 2006, p. 150). The raters later commented that the experimental group had shown great improvement in overall intelligibility. The questionnaires indicated positive perceptions in the experimental group concerning their improvement in intonation. They also better understood the importance of intonation in conveying meaning (Verdugo, 2006).

Computers require goals. Therefore, CAPT systems are limited if the goal is intelligibility because intelligibility is impossible to measure, even by native English speakers (Isaacs, 2008). If near-native pronunciation is the goal, as is the case with ASR systems which compare speech with a database of recordings by native English speakers, the selection and use of CAPT systems will be easier.

When choosing, evaluating, or designing any CAPT system, it is important to do so following pedagogically sound criteria. The voices of pronunciation teachers and students should be taken into account. Teachers and students say a CAPT system should provide feedback on targeted aspects of pronunciation. It should give useful feedback rather than simple feedback such as “correct” or “incorrect.” It should give feedback about improvement over time. The amount and type of feedback should be adjustable by the user and should give neither too much nor too little feedback (Engwall & Bälter, 2007).

ASR is still less than ninety percent accurate, even with native English speakers. Therefore, ASR technology is unreliable as a measure of pronunciation skill. However, it can be used to measure improvements in pronunciation over time, especially when normalized to native-English ratings. For example, if the ASR indicates that a student is fifty percent accurate at the beginning of a time period, and a native-English-speaking rater indicates the student is seventy percent accurate, a subsequent measure by the ASR of sixty percent accurate would translate to an improvement of ten percent and an actual rate of eighty percent accurate. Correlation between
ASR rating and human rating of pronunciation is weak, but over time improvement in ratings are consistent (Kim, 2006). Thus, ASR-enabled CAPT systems can be used to measure improvement, but not to provide an accurate picture of actual pronunciation skill (Mitra, Inamdar, & Dixond, 2003).

CAPT systems help students improve pronunciation just as effectively as classroom instruction. However, they should not be seen as a replacement for teachers, but as a tool teachers can use to free up time for communicative activities and individualized pronunciation feedback (Neri, Mich, Gerosa, & Giuliani, 2008).

ASR-based feedback can help students improve their pronunciation of problematic phonemes. Rather than engaging the whole class in phoneme practice in which some students are proficient and others are not, CAPT can provide individualized feedback (Neri, Cucchiarini, & Strik, 2008).

CAPT systems are not limited to phoneme training at the segmental level. They can also provide practice and feedback of supra-segmental aspects of pronunciation such as linking, word stress, sentence stress, rhythm, and intonation. This is especially beneficial in situations where non-native teachers are expert in reading, writing, and grammar, but have heavily accented pronunciation which could lead to unintelligibility (Seferoglu, 2005).

Languages use different intonation patterns, but those patterns do not transfer to other languages. Since they are often subconscious and perhaps even sub-lingual, as indicated by the fact that babies recognize the meaning of intonation patterns before they recognize words, they are difficult to teach explicitly. However, CAPT systems can provide visual feedback by creating intonation graphs of sentences produced by native English speakers and those produced by students, therefore providing an opportunity for learning by comparison (Verdugo, 2006).

Pronunciation teachers, researchers, and CAPT developers have, over the last decade, been increasingly taking advantage of CAPT systems. However, the great potential has barely been sampled. The future holds incredible promise for the use of CAPT as an integral part of the language classroom. Increasing dialogue between students, teachers, researchers, and software developers is prompting a shift in the use of CAPT from being a technological innovation to being a pedagogical tool.
References


