Integrating Spoken Dialogue with Adaptive E-Learning: A Framework for Remediation

J. Baca, D. Brown, G. Lazarou, and J. Picone

***Abstract*—**E-learning is playing an integral role in training non-traditional adult learners to fill the global need for a more technically skilled labor force. The lack of human contact in e-learning environments presents challenges for many categories of learners, particularly remedial students, who are often deficient in basic literacy skills, and heightens the need for an automated adaptation to the learner. Spoken language dialogue systems have been shown to be effective at reinforcing learning through their ability to adapt to individual learning styles in a natural and transparent manner. However, unsupervised adaptation to the learner requires fundamental advances in many disciplines such as speech recognition, natural language processing, and human computer interaction. This paper reviews these two research areas and presents a framework to integrate these two technologies in a way that enhances learning.

***Index Terms*—Speech recognition, dialogue systems, e-learning, human computer interaction.**

This work was supported in part by the National Science Foundation under Grant No. IIS-0095940.

J. Baca is with the Center for Advanced Vehicular Systems, Box 5405, Mississippi State, Mississippi 39762-5405 (tel: 662-325-5442; fax: 662-325-5433; email: baca@cavs.msstate.edu).

D. Brown is with the Center for Advanced Vehicular Systems, Box 5405, Mississippi State, Mississippi 39762-5405 (email: brown@cavs.msstate.edu).

G. Lazarou is with the Center for Advanced Vehicular Systems, Box 5405, Mississippi State, Mississippi 39762-5405 (email: glaz@cavs.msstate.edu).

J. Picone is with the Center for Advanced Vehicular Systems, Box 5405, Mississippi State, Mississippi 39762-5405 (email: picone@cavs.msstate.edu).

1. Introduction

E-learning is playing an integral role in training non-traditional adult learners to fill the global need for a more technically skilled labor force, a need further exacerbated by continual declines in basic literacy skills [1]. Indeed, many economically disadvantaged states have defined adult workforce training as their top legislative priority for the next several years. Despite these changing legislative priorities, however, local community colleges still do not have the economic resources to train an adequately skilled labor force using human instruction. A critical need exists for a technological solution.

The research presented in this paper was motivated by the goal to bolster workforce training in disadvantaged areas through the development of an online training system that can adapt to the needs of the adult remedial learner. The lack of human contact in e-learning environments heightens the need for an automated understanding of and adaptation to this type of learner who is often deficient in basic literacy skills. The cornerstone of our approach is the inclusion of a spoken dialogue capability that can monitor and adapt to user learning styles and stress levels.

Spoken language tutorial systems have been evaluated and found effective for learning, in areas ranging from literacy skill acquisition by children [2,3,4] to adult workforce training [5]. Potential enhancements to learning include: the ability of the learner to vocalize concepts naturally [6], the ability of the system to assess the learner's mental state from speech [7], and to provide auditory corrective and motivational feedback during the interaction [8]. These issues are of critical importance for the remedial student in an e-learning environment.

This paper reviews research in the areas of spoken dialogue tutorials and adaptive e-learning and presents a framework to integrate these two technologies in a way that enhances learning with emphasis on the remedial student. Adaptation in e-learning is reviewed first using examples from a case study of an online Pre-Algebra course, designed using a tool developed in our laboratory [9]. This is followed by review of spoken dialogue tutorial research. A framework for integrating the two technologies to optimally support the remedial learner is presented through example interactions with the case study using a spoken dialogue interface developed in our laboratory [10]. The paper concludes with a summary of this framework and the role of this work in a larger ongoing effort to support workforce training in economically disadvantages states.

2. Adaptation in E-Learning

2.1. Overview

A majority of e-learning systems adapt to learning styles using one of two popular pedagogical approaches: instructor-based [11] and learner-based [12]. A comparative analysis of these two approaches is given in [13]. To summarize, the instructor-based approach uses an instructor's specification of possible sequences through a course to dynamically determine and adapt the course content presented, in response to the learner's performance. While the learner's performance is considered, the instructor determines the sequence of content, not the learner, and the mode of content delivery remains the same for all learners. In contrast, the learner-based approach records significant attributes of learner behavior, e.g., navigation or media preferences, and adapts both the mode and content of information displayed, based strictly on these observations and no information specified by the instructor.

Both of these approaches offer benefits to the learner. The former provides the benefit of incorporating the instructor's pedagogical and subject matter expertise. The latter provides the ability to customize the mode of presentation to the learner. While excess conformance to individual learner preferences may impede learning, the benefits of individual adaptation in certain circumstances merit consideration.

IMS Simple Sequencing (IMS SS) embodies the instructor-based approach [11]. IMS SS has been adopted by the Sharable Courseware Object Reference Model (SCORM). SCORM recommends a set of technical standards and specifications allowing online learning systems to import, share, reuse, and export learning objects in a standardized way [14]. SCORM 2004 supports IMS SS, which allows specifying and guiding a sequence of learning activities, based on an instructor's knowledge of the course material. This sequence changes dynamically according to a learner's performance.

Though termed "simple", IMS SS allows specifying highly complex possible sequences of activities through a course. Such sequencing can be critical for courses requiring the level of remediation needed for the non-traditional learner. Remedial courses typically require learners to repeat educational modules in complex, learner-specific patterns to complete the remediation process. A teacher knowledgeable in the subject area and appropriate pedagogy can best enumerate these sequences and guide the direction of the learner's progress; however, presenting the content in a style that supports the remedial learner is also important. Thus, this category of learner requires aspects of both the instructor-based and learner-based approaches to e-learning.

An Adaptive Hypermedia System (AHS) exemplifies the learner-based approach. It employs a model of the learner and the learning domain to determine both the mode and content of information displayed to the learner. The user model is typically initiated by pretests given to the learner and is continually updated by tracking the user's browsing behavior, e.g., pages visited, links, etc. [15, 16]. These systems offer distinct advantages by incorporating the learner's preferences in determining the mode of information displayed. However, they do not incorporate the instructor's pedagogical knowledge of the student's needs in mastering the material. Instructors require the flexibility to implement courses using the teaching philosophies they deem relevant. The adaptive learning style often used in remediation is based in the pedagogical literature [17, 18]. The following subsection describes in greater detail an instructor-based adaptive approach to an example e-learning Pre-Algebra course.

2.2. Instructor-Based Adaptive E-Learning

SCORM 2004 includes a specification for complex sequencing of information, based on the notion of a hierarchical activity tree. An example of an activity tree developed for a Pre-Algebra remediation course [10] is shown in Figure 1a. The Pre-Algebra box represents the highest level in the hierarchical course abstraction. Modules in the second tier of the tree represent the next lower level in the course hierarchy. Instructors can add, modify, or remove modules from each tier to decompose the course structure. One main advantage of the activity tree approach is its support for adaptive instruction through the application of sequencing rules to modules defined within the tree. This allows an instructor to adapt how a student will progress through a course based on objective criteria tracked for that individual. This is illustrated in the expansion of the "Comparing Numbers" Module from the activity tree, shown in Figure 1b.

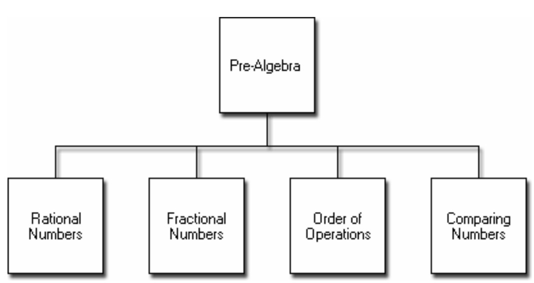

Figure 1a. Simple Activity Tree for Pre-Algebra Course

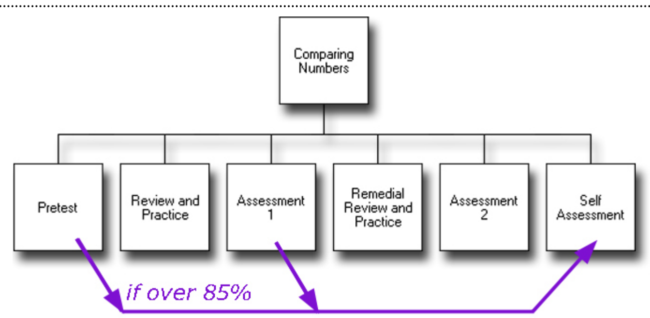

Figure 1b. Expansion of Comparing Numbers

The combination of the activity tree modules and the sequencing rules defined for each module provides the expressive capability for remediation. In this example, all students exploring the “Comparing Numbers” module are first presented the “Pretest” module. An overall score of 85% or above will enable a student to “skip” subsequent modules and go directly to final module, “Self-Assessment”. However, a student who does not achieve an overall score of 85% of “Pretest” will be presented with the sequenced material in “Review and Practice”, followed by “Assessment 1”. If the score on “Assessment 1” is less than 85%, remediation of a student’s specific deficiencies begins in the module, “Remedial Review and Practice”. In this module, a finer-grained presentation of the material specific to the learner is presented, based on answers given in the “Assessment 1” module.

This adaptive presentation is possible because instructors can specify the rules regarding the sequencing of course content presentation according to a set of skills and objectives. For example, several lower level skills and objectives are required for comparing numbers, including the ability to compare fractions to other fractions, fractions to percentages, and percentages to decimal numbers. The skill of comparing fractions to percentages requires further lower level skills of a) converting a fraction to decimal notation and then b) converting decimal notation to percentage. Sequencing rules can specify checking the specific skills for which a student exhibited deficiencies in the “Assessment 1” module and then specify the course material to present in order to remediate those deficiencies. Assume comparing percentages to fractions is *skill_1*, converting fractions to decimal is *skill_1a* and converting decimal to percentage is *skill_1b*. A rule to test these skills might be:

```
if (skill_1_score < 85)and (skill_1a_score < 85)
then show skill_1a_content
```

where *skill_1a_content* is a file with exercises to review converting fractions to decimals.

3. Spoken Dialogue Tutorial Systems

Spoken language dialogue systems seek to provide users a flexible, natural interaction with a computer system. Rather than querying using fixed commands and phrases, users may speak in a natural conversational style to request information and receive that information in spoken natural language, which may be combined with other modalities. At a minimum, a spoken dialogue system contains a capability for speech recognition and synthesis, natural language understanding and generation, and dialogue management. The latter capability differentiates a spoken dialogue system from simply a spoken language interface. Dialogue management enables the system to recall all utterances spoken over during a user's session or "dialogue" with the computer and refer to that contextual history in responding to the user. Spoken dialogue systems have been deployed successfully in a variety of domains [19, 20]. Their use in tutorial systems, however, presents unique issues for consideration.

Note in the Pre-Algebra course examples that the choice of information displayed is determined by sequencing specified by the instructor, in response to the learner's performance. Note also that any student who performs such that these rules are invoked will view this material in the same mode, with no adaptation of the presentation delivery specific to the individual. Spoken language dialogue systems implicitly adapt to the learner by allowing natural, freely formed spoken language input and adjusting spoken output according to the conversational context with the learner. They can also support more explicit forms of adaptation [7, 8]. The adaptive methods, presented in [17, 18] and often used in remediation, assumes learners can better correct misconceptions with immediate contextual feedback.

Although spoken dialogue offers potential benefits, it must be chosen carefully for the learning application and the learner. The following section reviews research issues and results regarding when and how to use spoken language dialogue tutorial interfaces most effectively. Spoken dialogue research spans many different areas, including speech recognition, natural language processing, and dialogue management. A comprehensive review of each of these areas is beyond the scope of this paper. The reader is referred to [21, 22] for an introductory overview of these areas. This paper focuses on a select set of issues relevant to spoken dialogue tutorial systems: advantages of full spoken language interaction, challenges in speech recognition, and issues in dialogue management.

3.1. Spoken Language Dialogue Interaction

One question to consider is whether textual, e.g., typed natural language (NL) dialogue, or spoken dialogue systems are more effective or appropriate since a text-based interface may be less complex to implement than a spoken one. Several benefits of spoken interaction have been hypothesized and empirically studied, including increased tendency towards self-explanation [6], access to the learner's mental and emotional state via verbal and non-verbal cues in speech [7], and contextual feedback to the learner [8]. The issue of textual versus spoken dialogue has not been conclusively decided and probably cannot be for all possible applications and learners.

Both NL and spoken dialogue systems have demonstrated enhancements to learning [23, 5] but many factors should be considered in a comparison and selection. Further no studies of this specific issue, text versus speech, have considered either the remedial learner with marginal literacy skills or the e-learning environment, where additional interaction with a human instructor is minimal and the benefits of spoken interaction hold greater significance.

Several textual NL tutoring systems have been developed, including AutoTutor and Why/AutoTutor [23-25], which tutor students learning about computer literacy and physics through generating explanations of their reasoning during problem-solving, and CIRCSIM [26] which assists medical students to learn the human circulatory system Each of these tutors uses textual, rather than spoken natural language input, though some use spoken or multimodal output. AutoTutor has been evaluated and shown to enhance learning in computer literacy and physics [23]. This system also incorporates an animated agent with spoken output, which raises a question as to whether it is the NL dialogue or the spoken agent that increases learning.

Evidence is offered in [27] that it is the NL dialogue content that affects learning, whereas the agent affects learner's subjective feelings about it. Again, however, neither AutoTutor nor the other systems referenced use spoken language input. While typed NL dialogue offers value, a fully spoken interaction, i.e., input and output, offers a means of adaptation to the learner beyond textual. The potential benefits of this adaptation for remedial learners are significant.

One advantage of spoken input concerns the increased tendency of the learner to self-explain reasoning in a fully spoken interaction. This was observed in a study of typed versus spoken tutorial interaction [6], and has been correlated with increases in learning [28]. Another comparison of typed versus spoken human tutorials also found increased numbers of words per student to teacher ratio [29]. However, in a study comparing spoken and typed human and computer dialogue tutors, the difference in number of words spoken and the gains in learning in the spoken condition were only significant when the tutor was human, not when the spoken tutor was a computer [30]. Several possible limiting factors were cited in the study, including the quality of the dialogue system, which simply "speech-enabled" an existing NL dialogue tutor without considering the differences in parsing natural versus spontaneously spoken speech [30]. Further, evaluations of a spoken dialogue system using reflective reasoning [5], i.e., tutorial sessions which occur *after* the actual problem-solving session using contextual feedback, designed to increase learner self-explanation, found significant gains in learning, attributed to the tutorial dialogue.

The potential of spoken input to increase self-explanation and the accompanying gains in learning represent a form of adaptation specific to the learner that could complement the remediation process automated, through teacher instruction, using the IMS SS. The IMS SS allows repetition of complex sequences needed to remediate, but does not innately elicit learner's reasoning patterns, explain why the remediation is presented, or set it in context. These issues are significant for the remedial learner. With the minimal human contact provided in an e-learning environment, this type of learner is more likely to repeat mistakes based on misconceptions and lose motivation. As an example, refer to Pre-Algebra course in Fig. 1b. If the student cannot pass 'Assessment 2' after one iteration of 'Remedial Review and Practice', rather than simply repeating this sequence with new values for the same fundamental exercises, a spoken dialogue tutor could engage the learner in a reflective contextual discussion of problems and elicit self-explanation to correct misconceptions:

Tutor: "Your scores showed problems with fraction conversion."
"How did you convert five halves to a decimal value?"
Learner: "Five halves is five over two, so I … um…multipled two times five and …divided by one hundred."
Tutor: "That's not quite right. Is five halves a compound fraction or an improper fraction?"
Learner: "uh..I think ..a compound fraction has …um…"

This example illustrates another advantage of spoken interaction concerning learner speech: it contains many features used by human tutors to assess the learner's grasp of the material and level of emotional stress to adapt responses to the learner. These features include hedging (e.g., "I think, I guess"), disfluencies (e.g., "um"), and prosodics (e.g., pitch, intonation, number of pauses, and speaking rate) [7]. A spoken dialogue interface can detect such cues [31-32] and adapt the system responses accordingly [7], just as a human tutor does. An example interaction from the Pre-Algebra fraction conversion task, showing hedging and disfluency from the student and an adaptive response from the Tutor is shown below.

Learner: "I …um…I guess I…would like to convert decimals to fractions."
Tutor: "Should we work one more example converting fractions to decimal?"

Results of research investigating the feasibility of recognizing student emotional states using acoustic and prosodic features have been promising [7, 33-34]. This research entailed studying and manually labeling human tutorial dialogue to determine predictive features and then training a system to recognize these features, categorized as positive (e.g., fast tempo, louder speech), neutral (e.g., moderate tempo and energy), and negative (e.g., pauses combined with disfluencies, "um"). To date, results have shown that predictors of emotional states correlated with increased learning can be identified in human-computer dialogue [34]. Although this research has not emphasized the remedial learner, a spoken language tutoring system for remedial reading has investigated similar issues. Though not a true dialogue system (it remembers only the most recent sentence spoken), The Reading Tutor, developed to enhance literacy acquisition, listens to children read aloud and responds with corrective feedback if necessary [35]. A recent study of this system has shown that integrating human-provided emotional scaffolding, e.g. “You’re doing fine” with an automated reading tutor significantly enhances student persistence [36], i.e., time on tasks, which leads to improvements in learning.

Despite the many possible advantages for spoken dialogue, it presents specific challenges which must be considered, perhaps the most important of which concerns misrecognition errors. Approaches to handling this issue are discussed in the following section.

3.2. Automatic Speech Recognition (ASR) Challenges

Speech recognition errors present perhaps the greatest potential barrier to acceptance of speech interfaces by users. It is thus encouraging that studies of the effects of misrecognition errors in spoken dialogue tutorials on learners have shown no negative correlation with learning [5, 37]. However, errors were negatively correlated with student desire to use the system again [5]. This finding confirmed what is generally understood for any interface using spoken input, i.e., misrecognition errors frustrate and confuse the user. This is significant in the e-learning environment to the remedial learner for whom motivation is critical. However, a combination of methods can be employed to reduce recognition errors and handle them in ways that are least intrusive for the learner when they occur.

First, confirmation strategies are important in error-handling and entail asking questions to confirm a user’s request to prevent errors. As an example, the student may say “I’d like to review subtraction” and the system then attempts to confirm: “Did you want to review fractions?” Strategies range from minimal, i.e., infrequent, only for severest consequences, which disrupts the user less often to confirm but may allow more errors, to more frequent confirmation [38]. The latter approach prevents more errors, but disrupts the user more often. Confidence scoring is a technique often used to address this issue [39, 40]. It involves assigning confidence measures to hypotheses from which low-scoring and thus, poorly understood segments of a user’s spoken utterance may be identified. This allows the system to clarify just those segments and hence, require users to re-enter only specific parts of the input. Also, the system may revert to a more explicit confirmation strategy if necessary.

The approach of the Reading Tutor [41] employs a computed certainty level to determine its confirmation strategy. It never says the student is right or wrong and thus confirms minimally. Instead it responds by speaking the correct word the student should have read and saying “mmmm?” if it believes, based on its confidence score, the student’s response was incorrect. The non-judgmental tactic offers particular value for the remedial learner and can reduce user frustration and its possible effects on the user’s vocal and verbal performance, as discussed next.

User frustration caused by cognitive load or emotional tension, feelings experienced by the remedial learner, also tend to increase vocal stress, which can then result in a downward spiral of recognition errors and increased user stress. Detecting vocal stresses through prosodic features of the voice has shown promise in addressing this issue [31, 32]. By detecting such stressors early in the interaction, the system can choose strategies or provide responses to decrease user stress and thus prevent a downward spiral of recognition errors. This is similar to the line of reasoning behind research to use prosodic features as predictors of student understanding and stress levels [7] and adapt tutor responses to the learner. Again, studies of recognition errors in spoken dialogue tutorials have not been correlated with decreases in learning [5, 37]. However, motivation and persistence present significant challenges for the remedial learner especially in e-learning, thereby increasing the value of a study analyzing the effects of these approaches on recognition errors and the interaction effects with learner satisfaction.

3.3. Dialogue Management

As noted previously, a dialogue system must go beyond understanding based on single utterances. It must seek to understand the user within the entire context of a given human-computer conversation or dialogue. Dialogue management is necessary to mediate this conversation between the human and computer, identifying and remembering the user’s goals when generating responses, and managing turns for the human and computer to speak. As an example:

Tutor (t1): "Convert ten percent to a proper fraction."
Student (s1): "A proper fraction."
Student (s2): "Ten percent is one over ten."
Tutor(t2): "You got this right when you converted from a proper fraction to percent."
Student (s3): "That was twelve over one hundred… twelve percent."
Tutor (t3): "Yes."

This simple example illustrates how the dialogue manager monitors the learner's goals in the conversation as well as her previous responses and current progress or "moves" toward those goals. In this way it determines that for response **s1**, the student is not moving toward the goal yet, so the tutor does not respond. After response **s2**, however, it recognizes a wrong move toward the goal, but also remembers a previous correct answer given during the dialogue. A comprehensive review of all approaches taken to dialogue management is beyond the scope of this paper, but most have involved the use of assumptions regarding discourse theory. A thorough and recent review of this topic can be found in [42].

With respect to dialogue tutorial systems, certain dialogue management issues are common to both typed and spoken interaction. In fact, the approaches of AutoTutor [23-25], the textual dialogue system referenced previously, and that of ScoT, A Spoken Conversational Tutor [5], are among two of perhaps greatest relevance for the remedial learner. AutoTutor employs a strategy that is based on observations of how human tutors actually perform, using techniques less sophisticated than many commonly cited in the pedagogical literature, relying more heavily on a strategy to increase learner self-explanation, which as previously noted, has been significantly correlated with increases in learning [28]. This strategy attempts to elicit self-explanation to trace the learner's reasoning using a set of anticipated correct answers and misconceptions [25]. The learner's responses are continually compared against this set of expectations codified in the system using a finite state machine architecture. This enables the dialogue manager to track learner progress and adapt tutorial responses accordingly, using techniques such as "pumps", "prompts", and "hints" to elicit self-explanation of answers. As the learner explains in response to these techniques, she exposes false assumptions or misconceptions which the tutor can correct and at the end of a subdialogue, summarize the correct line of reasoning. As an example:

Tutor: **"**The item you are purchasing is two dollars and you have only two dollars and fifty cents. What will be the total cost of the item with sales tax?**"**
Learner: "Fourteen cents.
Tutor: "Can you explain how fourteen cents and the purchase price are related?" (PUMP)
Learner: "Seven percent is 7/100. That is .07 times two dollars. And that is fourteen cents"
Tutor: "So is the sales tax is ___?" (PROMPT)
Learner: "Oh yeah… the sales tax is 14 cents. if you add that to two dollars, the amount is two dollars and fourteen cents."
Tutor: "Yes! To compute the total cost of an item with sales tax, you must first calculate ….then add…"

One salient feature of this system is its demonstration of the computational feasibility of building tutorial dialogues using finite state machine architectures. The nodes in the networks represent knowledge goal states for the learner, while the arcs represent tutorial dialog moves, such as pumps or prompts. In addition, the conversational smoothness and quality of dialogue moves have been evaluated positively by experts [23]. Integrating such a strategy could complement an existing adaptive e-learning tool which already maintains scores on student skills and knowledge.

ScoT, a spoken dialogue system, also seeks to elicit explanation from the learner, but by providing reflective tutoring after a problem-solving session [5, 43-44]. Used to train workers in on-board ship control, it requires real-time problem-solving performance. Therefore, ScoT, provides reflective sessions *after* the problem-solving activity to correct misconceptions and wrong answers. Research has shown that in reflective sessions with human tutors, students are more likely to ask questions and explain their reasoning [45]. However, such dialogues must carefully construct the original context in which the student performed the activity. ScoT addresses this by presenting a contextualized dialogue of the student's original answers, and a model for the correct answers to elicit explanation from the student to produce the correct answer. Its dialogue manager also employs a model of tutorial goals, but separates this from a model of student knowledge and domain knowledge to more easily contextualize information presented to the learner.

4. Recommended Research Framework and Directions

To summarize the issues presented in Section 3, there are at least three potentially significant advantages of integrating a spoken tutorial dialogue capability with e-learning: increased learner self-explanation and hence learning, an increased ability of the system to infer the user's cognitive and emotional state and adjust tutorial responses; both of these advantages lead to the third, which is the increased ability of the system to provide corrective and motivational feedback to the user. The main potential disadvantage concerns the issue of speech recognition errors and their effects on learning and motivation.

To reiterate, each of the advantages hold particular promise for the adult remedial learner, though none of the research cited pertaining to these issues focused on this user population. The Reading Tutor has been extensively studied, evaluated, and shown beneficial to grammar school children who are remedial readers; nonetheless it is focused exclusively on reading, is not a true dialogue system and is used by children. The latter issue is important; while some of its findings may prove applicable to adults and should be investigated, the significant differences in the speech of children and adults mean that the transfer of knowledge gained from this work requires careful testing on adults. A study of the integration of a speech synthesis capability with e-learning for adult literacy reported in [46] showed promise, but it used speech synthesis only, no spoken input, and furthermore, was not a dialogue system. Several other international e-learning sites devoted to adult literacy reported in [47] (www.thestudyplace.org, www.cyberstep.org, www.bbc.co.uk/skillswise) have integrated limited speech synthesis also, but again with no spoken input or dialogue capability. Nonetheless, these efforts demonstrate the broad attempts to use e-learning to reach the remedial learner.

A significant benefit of e-learning is its ability to deliver learning experiences remotely to those who might not otherwise have access. Ironically, the reduction in human contact that makes this delivery possible presents its major potential drawback, particularly for the learner who is already underserved and disadvantaged. A spoken language tutorial dialogue that can adapt to these learners significantly increases the value of any e-learning experience, though this paper has focused on the instructor-based method of adaptive delivery. To address the needs of these learners, the following research goals and directions are recommended:

- Target this population as a research priority. This will enable collecting sufficiently large corpora of speech and learning interactions to seriously study potential advantages and disadvantages of spoken dialogue.
- Standardize on a learning/training task or application so that results can be analyzed and compared against a baseline. This approach has lead to significant advances in speech recognition research over the past two decades.
- Standardize on an e-learning approach with which to integrate a spoken dialogue tutor for investigations, again for baseline comparisons. We have presented an instructor-based approach and contend that it offers unique advantages to the remedial learner, though others may be viable.

Establishing these research goals and directions will be a major challenge for the speech and e-learning communities. Though this population has become a more recent focus of attention, considerable commitment of resources will be required. Assuming these can be established, however, it then becomes possible to investigate a framework for integration of spoken dialogue with e-learning for remedial learners, wherein the following questions can be examined:

1. Learner self-explanation: Is this increased with spoken dialogue? Does this increase lead to gains in learning? The answer to the first question is clearly yes compared to an e-learning tool with no dialogue capability. However, the underlying dialogue management architecture is a factor that must be considered in both questions. Architectures which seek to elicit this behavior, such as those of AutoTutor or ScoT, should be evaluated and differences between typed and spoken interaction on the same architecture compared and analyzed.
2. Inferring learner mental states: Does spoken dialogue increase the ability of the system to infer learner emotional and cognitive states? More importantly, does this enhance learning and how? For the latter question, results need not be comparable to those with human tutors. Though desirable, the benchmark should be whether learning gains are significant compared to those in a standard e-learning environment with no spoken dialogue. Again, this will require collecting

corpora of spoken interactions for these learners, since little of this data currently exists. For both 1 and 2, scores on the same learning activities for those using e-learning systems with and without spoken dialogue must be compared and analyzed.

3. Dialogue management: how the dialogue manager responds to the learner, based on information about the learner provided in 1 and 2 presents another variable. Are misconceptions identified through learner self-explanation corrected by the dialogue manager in the appropriate way and time? Are the responses to user cognitive load inferred from user speech supportive to the learner? Finally, how can the knowledge about learner performance stored in the e-learning system be integrated efficiently with tutorial knowledge used by the dialogue manager.
4. Speech recognition errors: First, do these occur more frequently for this population and second, regardless of frequency, what are their effects on learning and motivation? Third can the vocal stress-detection techniques used to infer learner mental states also be used to reduce speech recognition errors? If not, can other prosodic detection techniques be incorporated to reduce learner vocal stress and the cycle of misrecognition errors it can spawn?

5. Conclusions

To conclude, the integration of spoken dialogue in e-learning presents many unique and important research challenges. In developing our approach to this integration for remedial learners, we shared the goal articulated by authors of a study to deploy instructional technology for rural community development [48]:

“[understanding]…how our on-line courses might help develop a community which would like to hold on to its children and grandchildren; not of what the system does for those of us who are privileged, but what it does for those who are not.”

We plan to continue investigating the issues raised in this paper for a targeted group of remedial learners. However, significant advances in these areas will require the efforts of a community of researchers. To that end, this paper has reviewed critical issues in spoken dialogue tutorial research and presented a framework for investigating the integration of a spoken dialogue tutorial capability with adaptive e-learning to support remedial learners. Selected example interactions were used to illustrate the intersection of critical issues in this framework. These interactions were taken from a case study of a Pre-Algebra e-learning course delivered using a spoken dialogue interface. This study is part of a larger ongoing effort with local community college instructors and learners to develop technology in support of workforce training in our state and surrounding region.

References

[1] B.V. Street, R.L. Venezky, D.A.Wagner (Eds.), *Literacy: An International Handbook* (Westview Press: Boulder, CO, 1999).

[2] J. Mostow, J. Beck, J. Bey, A. Cuneo, J. Sison, B. Tobin & J. Valeri, Using automated questions to assess reading comprehension, vocabulary, and effects of tutorial interventions, *Technology, Instruction, Cognition and Learning, 2*, 2004, 97-134

[3] J.E. Beck, P. Jia & J. Mostow, J., Automatically assessing oral reading fluency in a computer tutor that listens. *Technology, Instruction, Cognition and Learning, 2*, 2004, 61-81.

[4] J. Mostow, G. Aist, P. Burkhead, A. Corbett, A. Cuneo, S. Eitelman, et al., Evaluation of an automated Reading Tutor that listens: Comparison to human tutoring and classroom instruction, *Journal of Educational Computing Research*, 29(1), 2003, 61-117.

[5] H. Pon-Barry, B. Clark, E.O. Bratt, K. Schultz, S. and Peters, Evaluating the effectiveness of SCoT: a spoken conversational tutor, *Proc. ITS 2004 Workshop on Dialogue-based Intelligent Tutoring Systems: State of the Art and New Research Directions*. Maceio, Brazil, 2004, 23-32.

[6] A.G. Hauptmann and A.I. Rudnicky, Talking to computers: an empirical investigation, *International Journal of Man-Machine Studies* 28(6), 1988, 583-604.

[7] D. J. Litman and K. Forbes-Riley, Predicting student emotions in computer-human tutoring dialogues. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, Barcelona, Spain, July 2004.

[8] J. Mostow, J.E. Beck, & J. Valeri, Can automated emotional scaffolding affect student persistence? A Baseline Experiment, *Workshop on "Assessing and Adapting to User Attitudes and Affect: Why, When and How?"* at the 9th International Conference on User Modeling (UM'03), June 2003, Johnstown, PA.

[9] J. Baca, F. Zheng, H. Gao, & J. Picone, J., Dialog systems for automotive environments, *Proc. of the European Conference on Speech Communication and Technology (Eurospeech 2003), Geneva, Switzerland, September 2003*, 1929-1932.

[10] J. Baca, D. Brown, T. Miskelly, & S. Calhoun, Improving usability of adaptive e-learning authoring tools, *Proc. International Conference on Education and Information Systems: Technologies and Applications, Orlando, Florida, USA, July 2004*, 352-357.

[11] IMS Global Learning Consortium. (2003). IMS SS information and behavior model. Retrieved August 16, 2004 from http://www.imsglobal.org/simplesequencing/ssv1p0/imsss_infov1p0.html.

[12] P. DeBra, Adaptive educational hypermedia on the web. *Communications of the ACM*, 45 (5), 2002, 60-61.

[13] N.A. Abdullah & H. Davis, Is simple sequencing simple adaptive hypermedia? *Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia (HT '03), Nottingham, UK, August* 2003, 172-173.

[14] Advanced Distributed Learning. (2004). SCORM Overview. Retrieved August 16, 2004 from http://www.adlnet.org.

[15] P. DeBra, A. Aerts, D. Smits, & N. Stash, AHA! Meets AHAM, *Second International Conference on AH and Adaptive Web-Based Systems, Malaga, Spain, May* 2002, 388-391.

[16] P. Brusilovsky & M.T. Maybury, From adaptive hypermedia to the adaptive web, *Communication of the ACM*, 45(5), 2002, 30-33.

[17] J.K. Burton, D.M. Moore, & S.G. Magliaro, S.G. (1996). Behaviorism and instructional technology, In D.H. Jonassen (Ed.), *Handbook of research for educational communities and technology* (NY: Simon and Schuster, 1996, 46-67).

[18] P.A. Cooper, Paradigm shifts in designed instruction: From behaviorism to cognitivism to constructivism, *Educational Technology*, 33 (5), 1993, 12-19.

[19] D. Buhler, W. Minker, J. Haubler, & S. Kruger, Flexible multimodal human-machine interaction in mobile environments, *Proc. International Conference on Spoken Language Processing (ICSLP '02), Denver, Colorado, USA, September* 2002, 169-172.

[20] A. Gorin, G. Riccardi, & J. Wright, How may I help you?, *Speech Communication*, 23, 1997, 113-127.

[21] X. Huang, A. Acero, H-W. Hon, Spoken language understanding, *Spoken Language Processing*, (New Jersey: Prentice Hall, 2001, 853-952).

[22] D. Jurafsky & J.H. Martin, Dialogue and Conversational Agents, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, (New Jersey: Prentice Hall, 2000, 719-760).

[23] A.C. Graesser, L. Shulan, G.T. Jackson, H.H., Mitchell, M Ventura, A. Olney, and M. Louwerse, AutoTutor: A tutor with dialogue in natural language, *Behavior Research Methods, Instruments, and Computers*, 2004, 36, 180-193.

[24] A.C. Graesser, K. Wiemer-Hastings, P. Wimer-Hastings, R. Kreuz, and the Tutoring Research Group, AutoTutor: a simulation of a human tutor, *Journal of Cognitive Systems Research*, 1, 1999, 35-51.

[25] A.C. Graesser, N.K. Person, D. Harter, and the Tutoring Research Group, Teaching tactics and dialogue in AutoTutor, *International Journal of Artificial Intelligence in Education*, 12, 2001, 257-279.

[26] M. Evens, S. Brandle, R. Chang, R. Freedman, et al., CIRCSIM-Tutor: An Intelligent Tutoring System Using Natural Language Dialogue, *Proc.12th Midwest AI and Cognitive Science Conference, MAICS* 2001.

[27] A.C. Graesser, K. Moreno, J. Marineau, A. Adcock, A. Olney, A., N. Person, and The Tutoring Research Group, AutoTutor improves deep learning of computer literacy: Is it the dialogue or the talking head? In U. Hoppe, F. Verdejo, and J. Kay (Eds.), *Proc. of Artificial Intelligence in Education*, *2003*, 47-54.

[28] M. Chi, N.D. Leeuw, M. Chiu, C. Lavancher, Eliciting self-explanations improves understanding, *Cognitive Science* 18, 2001, 471-533.

[29] C.P. Rose, D. Litman, D. Bhembe, K. Forbes, S. Silliman, R. Srivastava, and K. VanLehn, A comparison on tutor and student behavior in speech versus text based tutoring, *Proc. of the HLT-NAACL 03 Workshop on Educational Applications of NLP*, 2003.

[30] D.J. Litman, C. P. Rose, K. Forbes-Riley, K. VanLehn, D. Bhembe, and S. Silliman, Spoken versus typed human and computer dialogue tutoring, *Proc.7th International Conference on Intelligent Tutoring Systems (ITS)*, Maceio, Brazil, August-September, 2004.

[31] M. Ostendorf, Linking speech recognition and language processing through prosody, *Communications and Cognition – Artificial Intelligence (CC-AI)*, 15 (3), 1998, 279-303.

[32] K.R. Scherer, D. Grandjean, T. Johnstone, G. Klasmeyer, & T. Banziger, T., Acoustic correlates of task load and stress, *Proc. of the International Conference on Spoken Language Processing (ICSLP '02), Denver, Colorado, USA*, 2002, 2017-2020.

[33] D.J. Litman & K. Forbes-Riley. Annotating student emotional states in spoken tutoring dialogues, *Proc.s of 5th SIGdial Workshop on Discourse and Dialogue (SIGdial)*, Boston, MA, April-May 2004.

[34] K. Forbes-Riley & D. Litman. Correlating student acoustic-prosodic profiles with student learning in spoken tutoring dialogues, *Proc. 9th European Conference on Speech Communication and Technology (Interspeech-2005/Eurospeech)*, Lisbon, Portugal, September 2005.

[35] G. Aist, & J. Mostow, Adapting human tutorial interventions for a reading tutor that listens: using continuous speech recognition in interactive educational multimedia., *Conference on Multimedia, Exeter, England*, September 1997.

[36] G. Aist, B. Kort, R. Reilly, J. Mostow, & R. Picard, Experimentally augmenting an intelligent tutoring system with human-supplied capabilities; adding human-provided emotional scaffolding to an automated reading tutor that listens, *Proc. Intelligent Tutoring Systems ITS 2002 Workshop on Empirical Methods for Tutorial Dialogue Systems*, San Sebastian, Spain, June 4, 2002.
.
[37] D. Litman and K. Forbes-Riley. Speech recognition performance and learning in spoken dialogue tutoring, *Proc. 9th European Conference on Speech Communication and Technology (Interspeech-2005/Eurospeech)*, Lisbon, Portugal, September 2005.

[38] C. Kamm, User interfaces for voice applications. In D. Roe & J Wilpon (Eds.), *Voice communication between humans and machines* (Washington, D.C.: National Academy Press, 1994, 422-442).

[39] S. Pradhan, & W. Ward, Estimating semantic confidence for spoken dialogue systems. *Proceedings of the International Conference on Acoustic Speech and Signal Processing (ICASSP-2002), Orlando, Florida, USA, May 2002,* 1, 2002, 233-236.

[40] Y.C. Tam, J. Mostow, J.,Beck, & S. Banerjee, S., Training a confidence measure for a reading tutor that listens, *Proceedings of the European Conference on Speech Communication and Technology (Eurospeech 2003), Geneva, Switzerland,* September 2003, 3161-3164.

[41] G. Aist & J. Mostow, A time to be silent and a time to speak: time-sensitive communicative actions in a reading tutor that listens, *AAAI Fall Symposium on Communicative Actions in Humans and Machines, Boston, Massachusetts, USA,* November 1997.

[42] A.C. Graesser, M.A. Gernsbacher & S.R. Goldman (Eds.), *Handbook of Discourse Processes*, (Mahwah, N.J.: Lawrence Erlbauum, 2003).

[43] H. Pon-Barry, B. Clark, K. Schultz, E.O. Bratt, S. Peters, & D. Haley, Contextualizing reflective dialogue in a spoken conversational tutor, *Educational Technology & Society*, 8 (4), 2005, 42-51.

[44] H. Pon-Barry, B. Clark, K. Schultz, E.O. Bratt, and S. Peters, Advantages of spoken language interaction in dialogue-based intelligent tutoring systems, *Proc. of ITS 2004, 7th International Conference on Intelligent Tutoring Systems*. Maceio, Brazil. Lecture Notes in Computer Science 3220 Springer 2004, 390-400.

[45] S. Katz, D. Allbritton & J. Connelly, Going beyond the problem given: how human tutors use post-solution discussions to support transfer, *International Journal of Artificial Intelligence and Education*, 13, 79-116.

[46] P. Walsh & J. Meade, Speech enabled e-learning for adult literacy tutoring, *Proc. Of 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'03)*, 2003.

[47] C. Holland, Designing an online Literacy interface, *BERA Conference*, University of Exeter, 2002.

[48] J.Burton & B. Lockee, Building equity through learning: instructional technology for rural community development, *International Journal of Educational Technology,* July 2000, http://www.ao.uiuc.edu/ijet/v2nl/v2nlfeature.html.