

Introducing Voice Recognition into Higher Education

Dittrich, Toby^a and Star, Sequoia^b

^aDepartment of Physical Science, Portland Community College, USA

^bSequoia Star Consulting, USA

Abstract

Voice Recognition (VR) software has now evolved to be fast and accurate enough to be useful in many educational settings. This paper describes two new uses for VR technology, both protected by patents, which can effectively address the lack of universal oral training in education today. The first use is Instant Note Capture (INC) which can be employed in live computer presentations and in an online software add-on tool called Incredible Classroom (IC) to place and store voice to text records in educational activities. The second is a new assessment tool called Virtual Oral Recitation Examination System (VORE) which enables oral discourse to be automatically and instantaneously identifies the necessity for and demonstrates the uses of voice recognition systems in education.

Keywords: *Voice recognition; speech recognition; educational technology.*

1. Introduction

Computer advances in memory and speed coupled with continuous efforts by many research teams have advanced voice recognition technology (VRT) sufficiently that it has made its way to the consumer market. The 1990s saw the first introduction of commercially successful VRT. In the early 2000s, speech recognition was still dominated by traditional approaches such as Hidden Markov Model (HMM) (Baum & Petrie, 1966) combined with feed forward artificial neural networks. Today, however, many aspects of speech recognition have been taken over by a deep learning method called long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997), a recurrent neural network. VRT has reached maturity based on these deep learning techniques and is delivered by Google, Microsoft, IBM, Apple, Amazon, and Nuance. Even before the turn of the century, the vocabulary of the average commercial speech recognition system was larger than the average human vocabulary (Huang, Baker, et al., 2015). Currently, VRT is in common use in many applications and disciplines; however it has not yet appeared in the field of education.

The *Instant Note Capture* (INC) technology listed herein will provide the first live, dynamic, interactive classroom experience in online education. Additionally, the component introduces oratory elements into online classrooms, which has never before been available.

The *Virtual Oral Recitation Examination System* (VORE) initially sprung from the teaching of a conceptual physics course. The professor, after having read a similar written paragraph from dozens of students, thought there must be a better way. His concept was using a computer to assess the verbal answer to questions by using key words from a rubric used to assess points to the written answer. With this concept, the idea of automated assessment of oral discourses was born. The door is now open to implement technologies that allow all students to answer content questions orally. These discoveries are of great importance to the field of educational as it has been shown, repeatedly, that the introduction of oral components to learning processes improves students' academic, personal, and professional success (Beers & Nagy, 2008; Berninger & Abbott, 2010; Semel, Wiig, et al., 2003; Berninger, Abbot, et al., 2010; Nelson, 2011).

Toward the end of introducing VRT to education, efforts are underway by Northwest Educational Software Inc. of Vancouver, Washington USA, to pioneer the herein delineated VR technologies into educational tools and applications. This effort is strengthened by the company's intellectual property (IP) portfolio, which includes three US Patents (Dittrich, 2009a, 2009b, 2009c) and two similar Canadian patents.

2. Applications for the technology

Another recent and applicable educational development is called Active Learning. Active learning is an approach to instruction in which students engage the material they study through reading, writing, talking, listening, and reflecting (Carr, Palmer, et al., 2015). Active learning stands in contrast to other standard models of instruction in which teachers do most of the talking and students remain passive. Perhaps the easiest and most effective way to engage students in active learning is by getting them to talk with others and explain, reason, and debate curriculum. This invaluable process of engaging students in talking while learning is at the heart of both the technologies introduced above and described below. When students transition into the workforce, depending upon the employment position, 75% to 85% of a typical work day is spent in oral exchange (Blalock, 2006). In management level positions the percentage is as high as 90%. These workplace communication requirements stand in stark contrast to oral skill capacity garnered by students (Peter D. Hart Research Associates, 2005). The INC and VORE technologies are aimed directly at this deficiency.

2.1. *The Incredible Class Room and Instant Note Capture*

The primary educational tool under development, using the INC software, is the Incredible Classroom (IC). This tool is designed to maximize verbal exercise and face-to-face communication in the on-line educational setting. Current learning management systems (Blackboard, Desire2Learn, Canvas) have negligible or nonexistent built-in conferencing systems, or they contain imbedded links to an external conferencing program, which are neither tied to, nor associated with, the student's online learning platform. Instant Note Capture, within the IC platform, textually stores all verbal discourses in class sessions, office hours, tutoring, and student-to-student conferencing. This enables the teacher to assess students, to review lecture content, tutoring sessions, and student conferences. The text can easily be converted into thirty languages (currently available through Google). Additionally, and of invaluable importance, disabled students, those unable to type, can join the conversation by voice.

More specifically, the INC offers a synchronously interactive component to online learning not previously available. The software's unique features will address several discreet concerns that current Learning Management System (LMS) platforms exhibit. One major LMS issue resolved by IC is the use of the current *raise your hand* feature used by students when asking questions during online class time. Currently, entering a question is done by typing into a provided text box. This process is cumbersome, and often leads to awkward delays in the flow of an online class, as others must wait while one student types. With IC's INC technology, the voice audio signal and the output of voice recognition software will

both record and instantly display the spoken question and the instructor's answer on the wall of the online classroom. Additionally, questions, answers, and lectures are stored both in writing and recorded format for future review or replay, a feature useful to student and instructor alike. Microsoft Powerpoint™ is currently testing a system for introduction in the Office™ suite next year called Presentation Translator, which is essentially the application of INC.

2.1.2 VORE Oral Discourse Automated Assessment Technology

VORE is a patented software programming code for learning and assessment which enables a teacher to assign oral work questions and give oral examinations. VORE tasks are performed by the student with the aid of a computer and software. The graded results of these oral tasks are available immediately, both to the teacher and the student. The essence of the VORE technology is the use of computer code to connect voice recognition software output in text document format with essay or written document evaluation software code. This combination enables the computer to evaluate, correct, and assess or grade verbal answers and presentations. Potential applications for this technology include but are not limited to: synchronous remote online class time, office hours, and study and tutoring sessions. To further example the implementation of the intervention, an application scenario has been created and a figure of the technology provided:

First, a teacher or a textbook publisher creates an oral question to be answered by the student user. In creating the question and answer key, certain phrases, concepts, and linked concepts, which will be used for grading, are selected using indicia on the keyboard. When the instructor creates the answer for the answer key, it is spoken into a microphone for conversion to text with VR software. Due to the differences in the syntax for written text and the spoken word, the answer would be created verbally and input with a microphone (hence preventing comparing the teacher's written answer to students' oral answers). The question is then stored in a VORE answer key file of the VORE Program Test application of the administrator's computer.

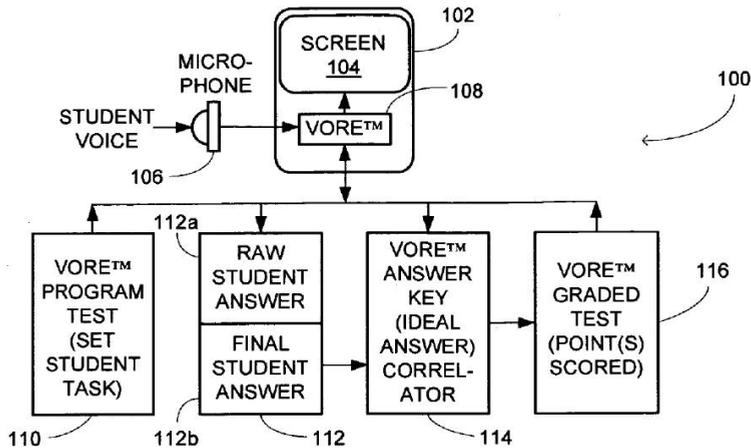


Figure 1. Vore Technology. Dittrich (2009a)

The student subsequently initiates the VORE task and a question appears on the screen of the user's computer. The student orally responds to the question using their computer's microphone, (students can also be videotaped and digitally recorded). Their verbal answer is digitized into text by VR software. Editing software can be used to correct punctuation, grammar, word usage, and sentence structure. The number of editing corrections is noted. The final student answer is assessed by automated assessment software and the grade is displayed on the screen. If the user is not accurate in their response, after further research and the formulation of a new answer, they would be allowed to delete the session and start a new VORE task. The number of times the student attempts an answer is recorded for the teacher. When the user is satisfied with their final score they deliver the score, and the text document (and audio and video recording, if one is made), to the teacher's computer (grade book file). The above delineated process possesses an additional very desirable advantage, security in the process of course administration.

The above description opens the door to a wide variety of educational applications that exist for the implementation of VORE technology. These include:

- Providing an opportunity to utilize oral examinations for students with poor writing skills or disabled students not capable of writing or typing;
- A tool for underperforming students, with test scores that are lower than average. VORE can be implemented to allow for additional study and retakes of exams orally;

- Enabling teachers to provide a method to create closure to a daily class lesson, in a quick and easy manner, with immediate feedback identifying low performers in the class;
- Providing positive identity verification for testing in online courses, since the task could be completed while being videotaped (Voiceprint identification software can also be used for positive identification of the test-taker as well);
- Provide a method of automated assessment of the oral expression CCSS standards in schools and districts nationwide; and
- Increase the capacity of student's oral skills.

Several specific educational tools have been designed and are available for development. They include the following.

VORE in Reading Comprehension: Students were given a reading assignment and after reading a portion of the assignment, the computer would present them with a content related question. The reader would be required to answer the question verbally and their answer would be graded instantly with VORE. If the assessment was successful, the reader would be presented with the next portion of the assignment. If not, the reader would be directed to reexamine the original portion and answer another related question. This process continues until the reader successfully answers the question and can proceed. At the end of the assignment a series of VORE tasks are presented for completion of the reading assignment. This creates a research opportunity to compare groups of students who have used VORE and those who have not with regard to their overall reading comprehension abilities.

VORE in Mathematics: The ORAL Equation Solver: The following is an example of a VORE Mathematic task titled Oral Equation Solver (OES). When the student boots up the OES software they would be able to choose from mathematical tasks such as: 1) Add and subtract fractions; 2) Multiplying polynomials; 3) Find linear equation from (x_1, y_1) and (x_2, y_2) ; 4) Solve two linear equations in x and y for (x, y) ; 5) Factor polynomials of second degree; 6) Determine zeros or roots of polynomials; 7) Determine the derivative of a function at a point given its algebraic form; 8) Determine the derivative of a function at a point given its graph; 9) Describe the end behavior of functions (rational, linear, polynomial); and 10) Find the determinant of a 2×2 matrix.

To demonstrate VORE's basic working architecture, adding fractions has been chosen. Upon selection of this math task, the student would see the following:

$$\frac{2}{3} + \frac{1}{5} = \frac{\boxed{}}{\boxed{}}$$

Figure 2. Sample Mathematics Problem

Initially the student would be asked to make a general statement about this task, e.g., "Fractions may be multiplied and divided easily but cannot be added or subtracted without a common denominator" (Note: the task's key words are underlined). Then an instructor's voice or text on the screen would instruct the student to go through the individual steps of the mathematical task, in this case, to add this fraction. Each step would contain interactive boxes and specific computer generated voice instructions, for example: "Please enter the proper common denominator and explain the process by which you arrived at your answer." After entering 15 into the denominator box, the student's response would be something to the effect, "The common denominator is fifteen because it is the product of three and five." If the VORE assessment was passed by the student statement, the software would then accept the 15 in the common denominator box and go to the next mathematical operation with a voice prompt. This process would repeat until the final answer was stated, in this case 13/15. Finally, a voice prompt would ask for, a final closing statement, such as: "Explain why this task is complete?" The student would respond, "Because the fraction or answer - thirteen fifteenths - cannot be simplified or further reduced to a simpler fraction." The VORE assessment score would then be stored and presented to the student.

VORE Interactive Biology- Skeleton/Human Body: Software would be developed utilizing the patented VORE technology to construct an interactive skeleton for the purpose of training Anatomy/Physiology and General Biology students on the constituent parts of anatomy. This software could be used both for training and for assessment. The fact that this tool is much like an educational game enables it to increase content knowledge while simultaneously providing students with oral training.

When the program is booted up in the training mode, a colored skeleton showing bones is seen. The cursor arrow randomly points to a bone or joint and the student is then required to name that component and provide a short statement as to its function. The VORE technology would then score the answer and if the answer score is above the passing level, the arrow would move to another bone. This process would be repeated a certain number of times and a cumulative score would be presented. As an example, this repetitive process could occur twenty times at five points each time, for a total possible score of 100 points. Each repetition would need a minimum score of three points to proceed. If the score was

less than three points, a tutoring session would automatically start on that component, but the user could not immediately redo that repetition.

Upon completion of this session the student would see their overall score and if not satisfied they could start a new session and attempt a higher score. Potentially, these sessions could be engaged by two or more students, each student getting consecutive tries with a running score competition. The number of repetitions per session could be increased as desired, and the grading rubric for each component could be adjusted for the grade level of the course. When the student has mastered the bones, they could then click on muscles, or nerves, or circulatory system and proceed to learn all of the components of the human body. This example demonstrates how the VORE technology is similar in function to the combination of a video game, flash card training, and the memory recall game Concentration.

VORE Science Technology Engineering Math (STEM) Oral Story Problem Solver and Content Knowledge Assessment: Often in STEM, students learn to solve story problems without much regard for concepts and theoretical considerations. Students frequently pick an equation and simply substitute the variables numerically and calculate. This is termed the “plug and chug” technique. This shortcoming in STEM learning is often accentuated even more in more advanced courses, where mathematics often overshadows concept knowledge. Instead students should think, evaluate, consider alternatives, plug, and calculate. To enhance the story problem solution experience VORE can, in much the same way as the OES require verbal explanations for each step in the STEM story problem. This would require the student to understand each step and not just manipulate symbols. Many automated solutions systems, can grade the symbolic manipulation, but fail to assess the content knowledge associated with the symbolism.

3. Conclusion

In conclusion, VR is an effective software technology and can be used to accomplish multiple tasks. The troubling fact is that this technology's use in education is lagging and in fact is essentially nonexistent. Described herein are two technologies that could change this regrettable fact. Both INC and VORE are effective strategies to promote students talking and using oral communication, which is at the core of one of active learning's key components. Implementing these technologies would tackle a most glaring deficiency in education, the lack of universal oral training at every level. These platforms would expand the potential for many interesting educational activities beyond the applications described above. Once VRT is introduced into education, an avalanche of applications for it will most certainly appear.

References

- Baum, L. E., Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains, *The Annals of Mathematical Statistics*, 37(6), 1554–1563. doi:10.1214/1177699147
- Beers, S. F., & Nagy, W. (2008). The development of syntactic complexity in oral language texts: Effects of grade and genre. Proceedings from: *The International Conference on Writing Research Across Borders*. Santa Barbara, CA.
- Berninger, V. W., Abbot, R.D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Educational Psychology*, 102 (3), 635-651. doi: 10.1037/a0019319
- Blalock, M. (2006). Why good communication is good business. Retrieved from <http://www.writingforbusinessandindustry.com/uploads/2/8/7/2/2872349/goodcommunication.pdf>
- Carr, R., Palmer, S., & Hagel, P. (2015). Active learning: the importance of developing a comprehensive measure. *Active Learning in Higher Education* 16, 173-186. doi: 10.1177/1469787415589529
- Dittrich, W. A. (2009a). United States Patent No. 7,296,218 B2. New York, New York: U.S. Patent and Trademark Office.
- Dittrich, W. A. (2009b). United States Patent No. 7,562,288 B2. New York, New York: U.S. Patent and Trademark Office.
- Dittrich, W. A. (2009c). United States Patent No. 7,657,221. New York, New York: U.S. Patent and Trademark Office.
- Peter D. Hart Research Associates. (2005). *Rising to the challenge: Are high school graduates prepared for college and work?* Washington, DC: Hart, P.
- Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Huang, X., Baker, J., & Reddy, R. (2014). A historical perspective of speech recognition, *Communications of the ACM*, 57(1), 94-103. doi: 10.1145/2500887
- Nelson, M. (2011). Oral assessments: Improving retention, grades, and understanding. *PRIMUS*, 21(1), 47-61. doi: 10.1080/10511970902869176
- Rabiner, L. (1984). The acoustics, speech, and signal processing: A historical perspective, *IEEE ASSP Magazine*, 1(1), 4-10. doi: 10.1109/MASSP.1984.1162225
- Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluations of language fundamentals: Examiner's manual* (4th ed.). San Antonio, TX: Harcourt Assessment/