

domain of learning how to interact with civilians in other cultures. For example, the *Tactical Language Training System (TLTS)* employed theory of mind intelligent agents in an educational environment focused on users practicing interactions with civilians in another culture using their native language (Johnson, Marsella, & Vilhjalmsson, 2004). An intelligent tutor would attempt to address motivational concerns based on techniques observed from real human language tutors. The *ELECT BiLAT* prototype (Hill et al., 2006) trains individuals on how to properly engage in bilateral negotiations within a cultural context. This system also focuses on virtual human non-verbal responses and dialogue. A virtual coach runs in the background and assess user actions, determining if the selected action contributed to a learning goal, whether or not to give explicit feedback, and gives targeted feedback based on a rudimentary student model. While these examples do not focus on health communication, they do provide useful examples of the current approaches to pedagogical conversational agents, whether they be agent you speak to directly or assistive agents that help via tutoring during conversations with agents.

Architecture Overview

This section provides an overview of the system design and details regarding each of the individual components. Throughout this section we provide justification for our design choices considering the requirements and constraints.

Figure 1 gives a pictorial view of the architecture of our system. As visible in the figure, the system is designed using one of the more common Web Application frameworks, the Model-View-Controller (MVC) framework. The justification for having a web framework was to make it available to users on the run (e.g. medical residents) without the need for an installation of any kind.

The high-level flow for the system is as follows. The web page presents the user with contextual dialog choices that represent the steps in the BNI. The user selects one of these choices and that selection is forwarded to the servlet. The servlet calls the bean with this data, which processes this dialog and forwards the responses to the servlet that is eventually presented to the user. Below we describe the details of the view, the model and the database.

The View

The *view* is the face of the application, and has a dual purpose like in any MVC. It is responsible for taking the input from the user and passing it on to the controller and must also get the output from the controller and present the contents to the user. One of the constraints while designing for the view was the instructional nature of the software and the length of the BNI. Taking these into account, it was really crucial to keep the user engaged and motivated. For this reason we decided to have a life-like

patient and coach, to be designed in Flash. The Flex framework was an obvious fit considering the need for a Flash-based front-end and its infrastructure for support of Rich Internet applications.

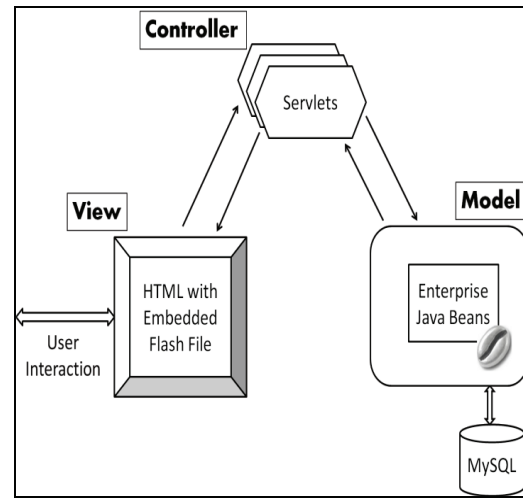


Figure 1. Overview of the Virtual BNI Trainer architecture.

The Model

The *model* is where the main cognition for the virtual coach resides. The model has been implemented as an Enterprise Java Bean. This bean implements both the coach and the patient model for this system. The reason for having an Enterprise Java Bean for a model was to be able to store user data across multiple requests. Although, we could have accomplished this using web sessions, we chose to use the model to store the session data. This choice was contingent upon the fact that we were dealing with a lot of session data that was needed to implement the user (Student) model described below.

The *controller* forms the input and calls the appropriate bean method with this input. The bean in turn uses the student, coach and patient models to retrieve the relevant information from the database and passes it on to the controller which then formats the output and sends it to the view. Details regarding the student, coach and patient models will be discussed in the next main section.

The Database

Implemented using MySQL, the database stores the entire authored dialog for the system along with user related data. While authoring the dialog, it was necessary for us to move away from simply taking the user through the BNI steps and introduce some complex dialog choices to fully gauge the understanding of the user. The database is responsible for storing all this in a structured format so that it can be retrieved quickly as response time is also of vital importance to the system.

It was important to establish metrics to understand the progress of the student to implement an intelligent and

adaptive student model. It is also crucial to be able to use past metrics data to provide the user with the most appropriate challenges that would allow him to improve his learning of the BNI. The VBT keeps track of the user progress and thus this information also persists in the database to provide this desired functionality

The Virtual Coach, Dr. Vicky

Our current approach to computer assisted BNI training centers on the use of a virtual coach, Dr. Vicky (see), who updates a model of student performance over time, provides feedback and remediation during training, and informs an after action review for users. This section discusses the decisions made in student model design,

Student Model

Having an adaptable student model is a common key component of intelligent tutoring systems. As explained in (Zhou & Evens, 1999), there exist various types of student model to address different issues. We are targeting student models for designing a customized solution path, responding with content appropriate feedback and also to assess the student level of mastery.

Several factors that influence student performance, such as ambiguous errors, skill deficiencies, cognitive slips, and random guesses can be taken into consideration while

constructing a student model (Katz, Lesgold, Eggen, & Gordin, 1994). Model-tracing methods are one such strategy, which includes an internal mapping of production rules students are expected to master (Koedinger, Anderson, Hadley, & Mark, 1997). By dividing the concepts into a series of topics in which student responses are monitored, it is possible to accurately access the student's current skill level when compared to the ideal system model. However, the rigidity of this system eliminates the ability to learn by "trial and error" (Katz et al., 1994). Other, less precise methods evaluate students based on their problem-solving strategies instead of specific responses. Our student model employs fuzzy set theory, which provides more flexibility in item categorization. Instead of categorizing skills into exclusive categories, it recognizes overlapping skills and captures the degree of overlap. We apply this fuzzy set theory to distinct knowledge variables using the methods used in the seminal work on SHERLOCK (Katz et al., 1994). The value in the set is increased or decreased depending on the level of evidence supporting the action. These variables, called *fuzzy variables*, are a set of attributes that represent a student's level of competence in a particular skill. We track five knowledge states for each of these variables, which include: no knowledge, limited knowledge, unautomated knowledge, partially automated knowledge, and fully developed knowledge.

The distinct knowledge variables that we track for each user stem from the requirement of having the users

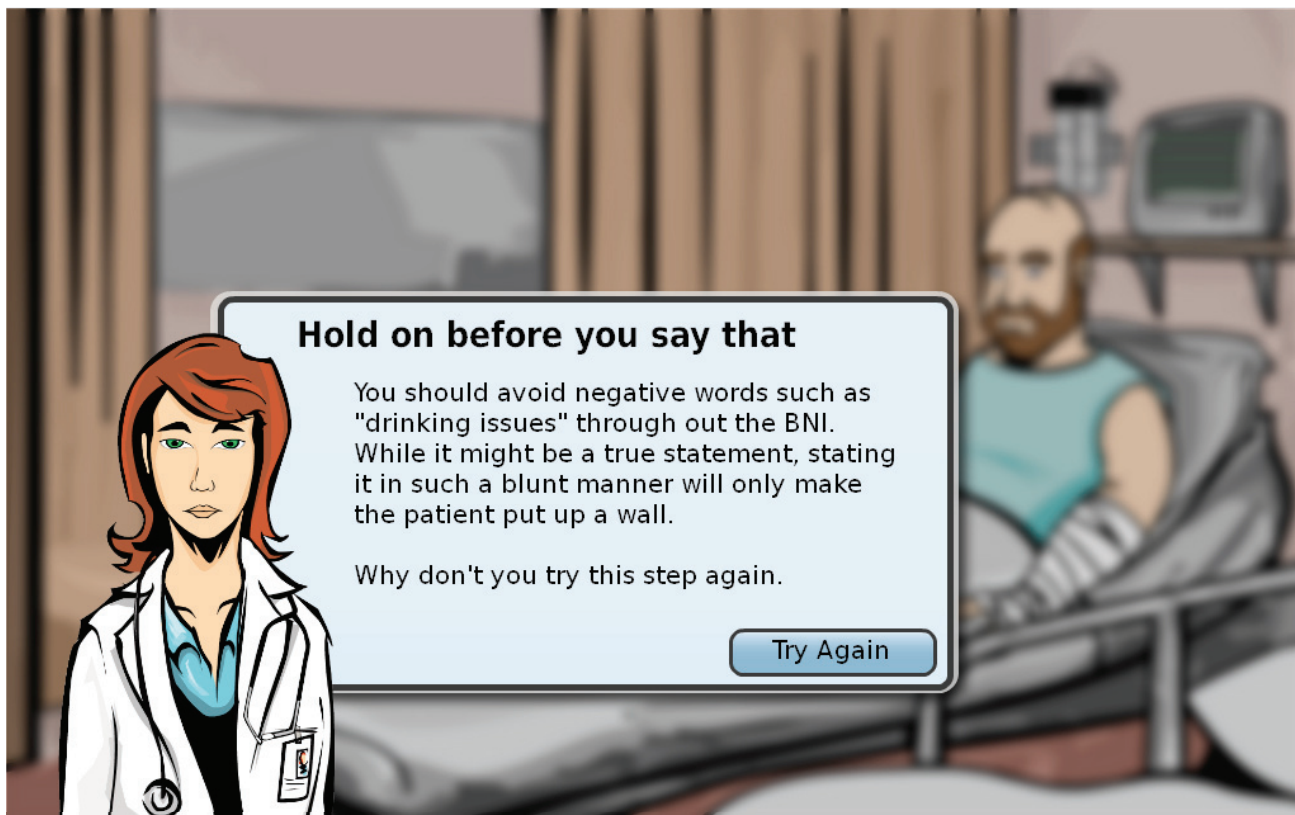


Figure 2. Screenshot of Dr. Vicky offering guidance to a user of the Virtual BNI Trainer (VBT).

perform the BNI in the correct order and being sensitive and responsive to the patient. We have two broad levels of competencies called *order competencies* and *soft competencies*. The order competencies evaluate the user on how well they perform in choosing the various steps in the BNI in the correct order. The soft competencies test how well the user performs the step. The soft competencies for the BNI domain include things like how well the student is establishing rapport with the patient or how well he is able to motivate the patient and other similar things. The soft competencies are further decomposed using the fuzzy set logic described above. We establish a system of local and global competencies, where global competencies are the ones described above. Each of these global competencies is an aggregation of the relevant local competencies. The local competencies for the global competency of *establishing rapport* might include things like *avoiding fatal words* like alcoholic while interviewing and *avoiding arguments* with the patient.

We use these concepts of order competencies and soft competencies to present the student with the most appropriate coach feedback based on the user's experience level. We also apply this concept to promote the student difficulty level. Using these concepts and the difficulty level, we also plan different dialog paths for the users and vary the amount of flexibility allowed to the user. These are discussed in more detail in the next sections.

Patient as level of difficulty

One of our key design goals is to have the VBT adapt to suit the student's expertise level (i.e. maximizing the learning potential). One way to accomplish this customization would be to have the virtual patient dynamically adjust its behavior within a conversation to reflect the user's knowledge. However this dynamic approach would have to handle the issue of how to change the virtual patient during the conversation and would greatly increase the complexity of the design. The VBT design has multiple virtual patients each with a difficulty rating; users are given a patient that is appropriate to their skill level (similar to the application of *knowledge tracing* in intelligent tutors (Koedinger et al., 1997)). A beginning student will get an easy patient that is quick to admit they have a problem and will not be resistant to change. The virtual patients will become more resistant and require more complex negotiation techniques to get them to commit to a change as the student advances. The multi-patient system allows for easy dialog authoring and management while still fulfilling the requirement of having the VBT adjust to suit the student's expertise level. It also allows the virtual patients to have varying personalities and scenarios (one deals with alcohol abuse while another with harder substances). Our current prototype has the beginning patient developed, with an established architecture (described earlier) and authoring criteria for both Dr. Vicky's behaviors as well as patient dialogue.

Guidelines for feedback authoring

The VBT adjusts the virtual coach's feedback based on the student's experience. This feedback must also reflect the correct and incorrect aspects of the specific dialog the student selects. For example if the student selects dialog that contains phrases with negative connotations (e.g. "drinking problem", "alcoholic", or "drunk") the virtual coach must provide feedback that takes into account: a) the negative phrase, b) when in the conversation it was selected, and c) the student's current experience level. This presents a huge challenge for authoring the dialog, as every combination of time, dialog, and experience level must be accounted for. The problem becomes even more complex if the dialog tests multiple competencies from the student model and the user has varying experience ratings for each one. For example, if a selected dialog is related to competency A and B there must be feedback authored for a student that is: an expert in A but a novice in B, the reverse (expert in B and novice in A), and every combination in between. The VBT's solution to this problem of complex feedback authoring is to designate four categories of feedback: Heavy, medium(1), medium(2), and light.

Using the previous example of competencies A and B here is how feedback would be adjusted.

- If A & B are both novice level: Heavy feedback
- If A is novice & B is expert: Medium(1) feedback with a focus towards competency A
- If A is expert & B is novice: Medium(2) feedback with a focus towards competency B
- If A & B are both expert: Light Feedback

This setup is tailored to each dialog option to maximize the relevance. Another reason to reduce feedback to only 3 to 4 response levels is to increase the difference between responses. If there are six feedback responses per dialog there is less of a pedagogically significant difference between responses 5 and 6. Reducing the total responses to 4 helps the problem and makes each response more significant. The guidelines used for authoring each feedback category are:

Heavy feedback: Reinforces what the student did well and why it is important. Then points out the student's important.

Medium feedback: Reinforces what the student did well. Then points out the student's error and what s/he can do to correct it.

Light feedback: Points out the student's error and gives a brief reminder on how to correct it.

The feedback guidelines reflect the idea of incorporating scaffolding into the virtual coach. When the student is starting the training program, a novice, the virtual coach has a heavy presence and guidance. As the student gains knowledge the presence and feedback from the virtual coach eases off. When the student reaches an expert experience level the feedback is mainly just reminders of previous lessons and feedback.

User Errors

Dialogue is annotated with a categorization of correct, semi-correct, or fatal. Dialogue marked as correct contains no errors and is representative of how a BNI should be conducted. Semi-correct dialogue is mostly correct but contains an error or two. Selecting a semi-correct dialogue will allow the student to proceed but will make the virtual coach appear giving feedback on the error. Fatal dialogue is completely wrong or contains fatal words (e.g. “drinking problem”, “alcoholic”, or “drunk”) and will prompt the student to repeat the step. Every dialogue choice affects the student models; raising the competencies that were done correctly and lowering the competencies that were not.

interface needed to be inviting and easy to use both for initial users and long-term users. A quality backend architecture can be wasted if the frontend interface is not designed to match the user's needs. The VBT places a large focus on the environment, characters, and art design in order to provide an engaging experience. The menu structure is placed within the environment, floating in the screen, and not in the typical location of being docked to a corner or side.

A floating menu structure has the feel of fitting the created environment as opposed to being something added outside of the environment. Not only does this give a more modern feel, it also helps direct focus towards the avatars and give more weight to the environment they live in.



Figure 3. An example of the floating menu design used in the Virtual BNI Trainer.

Students can also make an ordering error by trying to execute a step in the BNI at the incorrect time. An example of an ordering error would be trying to negotiate a goal before establishing a problem. If the student selects dialogue in the wrong order the virtual coach will appear providing feedback and direct the user to try the step again. The student model will then be updated by deducing the order competency of the step the student is currently on.

User Interface Design

The goal of the user interface design was to provide a fresh modern look that promoted clear information display. The

The user has approximately nine dialog choices to choose from at any given time in the conversation. This could be confusing and time consuming to read if the dialog were all displayed at once so the VBT groups dialog choices into categories called buckets. When the student starts a new round of dialog in a conversation they are presented with approximately three buckets. Selecting a bucket expands it and displays approximately three dialog choices related to that bucket's category (three buckets times three dialog choices with in each equals nine total dialog choices). The exactly number of dialog and bucket choices depends on the difficulty level and location within the BNI process. Examples of bucket categories are "Greet", "Use Reflective

Listening" and "Discuss Low Risk Amounts." This also serves as another form of scaffolding as an experienced user might know the correct bucket, or step, for that phase of the BNI and thus only have to review three dialog options; while a novice user might have to explore as three buckets before deciding on the correct dialog choice.

However, even with efficiency provided by the buckets some of the steps can still be quite lengthy. Some of the dialog options can be three to four sentences long and with three dialog options per bucket that can be a lot of reading for the user. Especially considering students do not read dialog like a book, only one time through, rather they will reread dialog options comparing them to each other to determine the correct one. Cutting out words from the dialog choices is not sufficient because it would not reflect how an actual BNI is conducted. The VBT addresses this overload problem by first displaying dialog in a truncated form. This truncated provides the key points of the dialog choice while removing some of the "filler" words. The truncated form could also be thought of as a kind of summary. This allows a knowledgeable user to glance at the dialog and make quick decisions on how to proceed. A student could determine the truncated dialog has red flags and ignore it or that appears correct and warrants further investigation, at which point the user can mouse over the truncated dialog to expand it to the full dialog. An example of truncated and expanded dialog is as follows:

Truncated version

"...you drink about 2 to 3 times a week...this is very risky and irresponsible behavior. You have a serious drinking problem."

Full version (provided on mouse over)

The nurse said you drink about 2 to 3 times a week with 6 to 8 beers per occasion. As a doctor I would like to let you know this is very risky and irresponsible behavior. You have a serious drinking problem.

A knowledgeable student could quickly glance at the truncated version and understand that the dialog reviews "drinking amounts," which is correct but it takes a very aggressive tone and uses the phrase "drinking problem" which should be avoided. They would then recognize it as incorrect and then move on to the next choice while a novice user might have to expand the dialog to arrive at that conclusion. Since the BNI is a collection of key techniques, not a script to rehearse, the truncated dialog presentation method helps train students to use these key techniques.

We conducted a preliminary usability study to inform the direction of the design. Each participant ran through two different versions of the interface and provided feedback about each design. We found that participants valued the atheistic design and the interface animations were not seen as a hindrance, however we are aware this might change

with longer term use. Also while participants initially were confused with the expanding dialog, where dialog is first displayed in a truncated summary form and then expanded on mouse over, after understanding the purpose of it participants expressed that it could be a useful element. From this feedback we added a tutorial tool tip to help with the initial confusion over the expanding dialog.

Dialog Authoring

As mentioned earlier, there are typically three dialog choices within each bucket. However there are at least 10 dialog choices authored for each step that are randomly selected at run-time for those three slots. Dialog is annotated with a categorization of *correct*, *semi-correct*, or *fatal*. The VBT selects one dialog that is marked as *correct* and then randomly fills the remaining slots with dialog from that bucket category to populate a bucket with dialog choices. It is possible to have multiple correct dialog choices per bucket, which will encourage users to reflect on the dialogue that they select as they cannot assume there is always just one best answer.

This approach allows for increased repeatability without worrying about students easily identifying the right and wrong paths through rehearsal through multiple learning sessions. The overall structure of the conversation is the same but how the student gets from beginning to end is different each time. This lead to the challenge of authoring dialog that "sounds different" but all lead to the same outcome.

For example:

The user can say "tell me about the TV you mentioned buying" or "do you see any connections between your drinking habits and difficulties in your life?" Both of these options would lead to the virtual patient making the realization that cutting back on drinking will save money.

The overall effect of this design decision is that we have to write for depth (i.e. the entire possible conversation tree) as well as breadth (i.e. multiple dialogue statements when a single one would suffice). The practical experience of authoring can prove quite challenging and time consuming because of the need to keep multiple dialog paths coherent when presented in different orders (determined by the user).

Future Work

The future of the VBT will focus on two key areas. First, the current prototype only presents a single patient that represents a low difficulty BNI situation (a middle-aged man who was drinking beer while working with power tools in his garage). As the user applies the BNI correctly, there are no major difficulties in reaching a positive conclusion. In other words, the current patient is an ideal one intended as a scaffold for novices. As users gain proficiency, they should be introduced to a more complicated BNI situation. Our future design and

development work will address the authoring of more complex situations for intermediate and advanced users, as well as the development of mechanisms for advancing the user to the next difficulty. This should also include considerations for how to design patient attitudes, either through design or modeling, to generate less than ideal responses to proper BNI actions.

The second focus of our future work is on the individual differences that users have that may have a direct relevance to training situations. Learner motivation, mindset, goal orientation, and playstyle are just a few of the dimensions that shape how receptive learners are to educational digital media learning situations (Magerko, 2008). Our future work intends to explore these dimensions and their potential relevance to how the VBT can use not only models of the user at the knowledge, but across these broader dimensions to affect how and when guidance is given to the user and content is selected.

References

- Andre, E., Herzog, G., & Rist, T. (1988). On the simultaneous interpretation of real world image sequences and their natural language description: The system soccer. In *Proc. of the 8th ECAI* (pp. 449–454).
- Bickmore, T., & Gruber, A. (2010). Relational Agents in Clinical Psychiatry. *Harvard review of psychiatry*, 18(2), 119–130.
- Bickmore, T. W., Pfeifer, L. M., & Jack, B. W. (2009). Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the 27th international conference on Human factors in computing systems*, 1265–1274.
- Bickmore, T., & Cassell, J. (2001). Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seattle, WA, 396–403.
- Cassell, J., Bickmore, T., Billingham, M., Campbell, L., Chang, K., Vilhjálmsón, H., & Yan, H. (1999). Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 520–527.
- D’Mello, S. K., Craig, S. D., Witherspoon, A., Mcdaniel, B., & Graesser, A. (2008). Automatic detection of learner’s affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1), 45–80.
- D’Onofrio, G., & Degutis, L. C. (2010). Integrating Project ASSERT: A Screening, Intervention, and Referral to Treatment Program for Unhealthy Alcohol and Drug Use Into an Urban Emergency Department. *Academic Emergency Medicine*, 17(8), 903–911.
- Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, 36(2).
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31(1), 104–137.
- Green, N., Lawton, W., & Davis, B. (2004). An Assistive Conversation Skills Training System for Caregivers of Persons with Alzheimer's Disease. In *Dialogue Systems for Health Communication, Papers from the 2004 AAAI Fall Symposium, Technical Report FS-04-04*, AAAI Press, Menlo Park, CA (pp. 36–43).
- Hill, R. W., Belanich, J., Lane, H. C., Core, M. G., Dixon, M., Forbell, E., Kim, J., et al. (2006). Pedagogically structured game-based training: Development of the ELECT BiLAT simulation. In *Proceedings of the 25th Army Science conference*, 82–95.
- Johnson, W. L., Marsella, S., & Vilhjálmsón, H. (2004). The DARWARS tactical language training system. In *Proceedings of IITSEC*. Marina del Rey, CA.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2), 75–86.
- Magerko, B. (2008). Adaptation in digital games. *IEEE Computer Magazine*, ed. van Lent, M., 41(6), 87–89.
- Marsella, S., Johnson, W. L., & LaBore, C. (2003). Interactive pedagogical drama for health interventions. In *Conference on Artificial Intelligence in Education, Sydney, Australia*.
- National Institute on Alcohol Abuse and Alcoholism. (2007). Helping Patients Who Drink Too Much: A Clinician's Guide. *Helping Patients Who Drink Too Much: A Clinician's Guide*. Retrieved October 18, 2010, from http://pubs.niaaa.nih.gov/publications/Practitioner/CliniciansGuide2005/clinicians_guide.htm
- Piwek, P., Krenn, B., Schröder, M., Grice, M., Baumann, S., & Pirker, H. (2004). RRL: A rich representation language for the description of agent behaviour in NECA. *AAMAS Workshop: Embodied Conversational agents - Let's Specify and Evaluate Them!*, Bologna, Italy.
- Zhou, Y., & Evens, M. W. (1999). A practical student model in an intelligent tutoring system. In *ICTAI*, 13–18.