

# Computer Assisted Conversation Training for Second Language Learners

Preben Wik<sup>1</sup>, Anna Hjalmarsson<sup>1</sup>, Jenny Brusk<sup>2</sup>

<sup>1</sup>Department of Speech, Music and Hearing KTH, Sweden

<sup>2</sup>Dept. of Game Design, Narrative and Time-Based Media, Gotland University, Sweden

## **Abstract**

*This paper describes work in progress on DEAL, a spoken dialogue system under development at KTH. It is intended as a platform for exploring the challenges and potential benefits of combining elements from computer games, dialogue systems and language learning.*

## **1 Introduction**

DEAL is a dialogue system under development, built in order to explore the challenges and potential benefits of combining elements from computer games, dialogue systems and language learning. Our objective is similar to that of the Tactical Language Training System (TLTS) (Johnson et al., 2005), in the utilitarian sense that both systems are simulation games for acquisition of language and cultural skills. But where TLTS places focus on realism (teaching US military appropriate manners and phrases to be used on foreign ground), we wish to focus more on creating something entertaining. In that respect our objective is closer to *Façade*, a one act interactive drama where the player's interaction affects the outcome of the drama, and where the goal of the interaction is to create a good story (Mateas & Stern, 2003).

### **1.1 Serious games**

There is a growing trend among educational researchers to look at games and game design in order to make education more effective. Game designers have in their strive for success developed several effective design strategies both to get and to keep players engaged and motivated throughout a game. The same design principles are starting to find their way into other fields as well. Serious games is an initiative focusing on using game design principles for purposes other than solely to entertain, - e.g. training, advertising, simulation, or education (Iuppa & Borst, 2007). Good gameplay adds to any existing motivation to learn if there is one, and may create a motive by itself if there isn't.

### **1.2 Second language learning**

Our aim with this work is to allow language learners to practice conversational skills in a fun and challenging context. Language learning can be modelled as a series of developmental steps going from declarative to procedural knowledge. First an item is noticed in a meaningful contrastive situation, then it occurs repeatedly in meaningful input, is practised in communication until it is first somewhat internalised, and finally automatised (Ellis, 2006). To automatise language we need a meaningful situation where conversational skills can be practised repeatedly. Creating CALL that makes it possible to practice conversational interaction is therefore highly desirable.

## **2 DEAL**

DEAL is intended as a multidisciplinary research platform, particularly in the areas of human-like utterance generation, game dialogue, and language learning.

### **2.1 Ville, the framework for DEAL**

DEAL is a free-standing part of Ville, a framework for language learning developed at KTH (Wik, 2004). Ville is an embodied conversational agent (ECA). He is a virtual language tutor designed to help students improve their listening and pronunciation skills in a new language. Ville is designed to present language-specific distinctive features in a meaningful contrastive situation. He can detect and give feedback on pronunciation errors, and has many challenging exercises that are used in order to raise the student's awareness of particular perceptual differences between their L1 and L2, or to teach new vocabulary.

Ville has exercises on phone level, syllable level, word level, and sentence level. DEAL adds the possibility to also practice conversational skills. Ville is the teacher, giving corrections and feedback on a student's pronunciation and language use. The ECA in DEAL is not a teacher, but should act as a conversational partner with the objective of keeping an interesting conversation using the limited shared vocabulary between him and a student. Ville will teach the students the skills they need in order to interact in DEAL (and later interact with real people in real life)



Figure1. Ville, the virtual language tutor teaching aspects of Swedish

## 2.2 Domain in DEAL

Our first choice of domain for this work is the trade domain. More specifically we have placed the scene at a flea market. Our decision to look at this domain is based on several factors.

- A trading situation is a fairly restricted and universally well-known domain. It is something everyone is conceptually familiar with, regardless of cultural and linguistic background.
- It a very useful domain to master in the new language
- The flea market allows for, almost invites characters that are eccentric or otherwise out-of-the-ordinary in an interesting way.
- A flea market is a place where it is ok to negotiate about the price and to trade items. This type of negotiation is a complex process which includes both rational and emotional non-rational elements.
- The shop can include almost any type of item. In a larger framework (sect 2.4) vocabulary just learned can easily become items in the shop.
- Second hand items may have rich interesting characteristics such as a personal history or affectional value for the shopkeeper - or they may be defect and thus invite another type of conversation

These factors combined, a domain in which the user can engage in a dialogue that is well known but still includes elements of surprise, social commitment and competition (i.e. getting a good price).

## 2.3 User interface



Figure2. The DEAL interface with a stingy shopkeeper trying to sell a teddy-bear with a missing ear, for an outrageous price.

The top part of the user interface contains the shopkeeper, an ECA - also known as an animated talking head. Our ECAs are developed at KTH (Beskow, 2003), and can use either synthetic or natural, pre-recorded speech. The head is capable of producing lip-synchronized speech. The model can also convey extra linguistic signs such as frowning, nodding, and eyebrow movements. It has been shown that subjects listening to a foreign language often incorporate visual information to a greater extent than do subjects listening to their own language (Granström & House, 2005). The use of an ECA will hopefully also help to entice a more human-like interaction between the student and the dialogue system.

## 2.4 Dialogue system in DEAL

The dialogue system component in DEAL is based on Higgins, a spoken dialogue system built at KTH (Skantze, 2005). Higgins includes modules for semantic interpretation and analysis. Pickering, a modified chart parser, supports continuous and incremental input from a probabilistic speech recognizer. Speech is unpredictable and chunking a string of words into utterances is difficult since pauses and hesitations

will likely be incorrectly interpreted as end of utterance markers. This will be even more evident for second language learners whose conversational skills are not yet automatised and whose language contains disfluencies such as hesitations and false starts. Pickering uses context-free grammars (CFG) and builds deep semantic tree structures. Grammar rules are automatically relaxed to handle unexpected, ungrammatical and misrecognized input robustly. The discourse modeler, Galatea, interprets utterances in context and keeps a list of the communicative acts (CA) in chronological order. Galatea resolves ellipses, anaphora and has a representation of grounding status which includes information about who added a concept, in which turn a concept was introduced and the concept's ASR confidence score.

## 2.5 The task in DEAL

First Ville will teach the rudimentary vocabulary and grammar that is associated with the trade domain. The student is then given a mission to go to the nearby flea-market and use his newly acquired vocabulary in order to buy a given set of items from the shopkeeper in DEAL. The student is given a certain amount of money, but the money may not be enough to buy all the items on the student's list, unless he is creative. The stingy shopkeeper in the flea-market will try to get as much as possible for his goods. This scene can then unfold in different ways depending on what the student says, in combination with how some parameters are set in the ECA (personality). The willingness of the ECA to reduce the price of an item for example, may be affected by how the user gives praise or criticize an item of interest.

U1: I'm interested in buying a toy.  
 S1: Oh, let me see. Here is a doll.(a doll is displayed)  
 U2: Do you have a teddy-bear?  
 S2: Oh, yeah. Here is a teddy-bear.  
 (a teddy-bear is displayed, see Figure 2)  
 U3: How much is it?  
 S3: You can have it for 180 SEK  
 U4: I give you 1 SEK.  
 (shopkeeper is offended, willingness decrease)  
 S4: No way! That is less than what I paid for it.  
 U5: Ok how about 100?  
 S5: Can't you see how nice it is?  
 U6: But one ear is missing. (willingness increase)  
 S6: Ok, how about 150?  
 U7: 130?  
 S7: Ok, it is a deal!

## 3 Discussion

Our objective is similar to that of the Nice project (Gustafson et al., 2004) in that we wish to create a game where spoken dialogue is not just an add-on, but is used as the primary means for game progression. So far, the scenario, rules and possible actions in DEAL are fairly limited and much can be added in the long run. We feel, however, that the combination of game and CALL is a fortunate one, opening up for novel ways to use traditional techniques from speech technology. This paper describes the beginnings of a research platform that can produce a number of new and interesting questions on how to design dialogue systems and CALL applications.

### 3.1 ASR limitations as part of gameplay?

Can the well known difficulties in using automatic speech recognition (ASR) in combination with foreign accent be reduced to an obstacle within the gameplay? Rather than trying to adapt the ASR to be able to handle the strong accent, can it become a part of the gameplay to encourage the students to adapt their own pronunciation? The limitations of the ASR can be interpreted as a measure of the student's communicative skills, where the challenge is to be able to communicate with the ECA.

### 3.2 A dialogue system as a game

In a similar fashion, the criteria a non-native speaker (NNS) has for judging a dialogue system are different compared to a native speaker (NS). When a misunderstanding between a user and a spoken dialogue system occurs, a NS knows he has done nothing wrong, and will ascribe the misunderstanding to a weakness in the system. A NNS on the other hand will often be critical about his own ability in the new language, and might instead ascribe the misunderstanding to his own pronunciation, or incorrect use of grammar. A NNS will in a similar way be able to reason that if he or she is able to communicate with the system without errors, it can be seen as a confirmation of his or her abilities to communicate in the new language. Getting an ECA to understand you in a new language can be a very satisfying event, and can be achieved if the dialogue system is played more like a musical instrument in rehearsed cooperation rather than put on the line, and tested for its strengths

or weaknesses (as may be the case when a NS is using the same system).

Interacting with the system in a roleplay fashion also changes other criteria for how the system is judged. A traditional spoken dialogue system in an info-seeking application will be judged by factors such as efficiency in reaching task completion. A good system, in this domain, will try to minimize the number of turns needed. For the type of dialogue system described in this paper we aim for the opposite. The longer the conversation takes and the more turns between the user and the system, the better! If the interaction between the agent and the user is successful, and takes the form of a roleplay, user satisfaction will depend on other things, like for example the story, response time, social competence, character, and error handling.

### 3.3 How will utterance generation affect the player's suspension of disbelief?

Suspension of disbelief refers to the willingness of an audience to accept the premises of a work of fiction, even if they are fantastic or impossible. It also refers to the willingness of the audience to overlook the limitations of a medium, so that these do not interfere with the illusion. To maintain the suspension of disbelief when interacting with an ECA, such as the talking heads in DEAL, we carefully need to consider how the ECA is perceived. For language learning students who engage in this type of dialogues the ECA is not to act as a teacher but as a conversational partner with whom they can train their conversational skills. In the DEAL domain this does not mean to be pleasant, or to do what the user asks for. The agent can actually be rude and try to avoid the user's request as long as it is done in a way that does not destroy the user's suspension of disbelief. From this perspective it is important that the agent behaves in a human-like way (Hjalmarsson, submitted). How we choose to generate utterances is crucial for how the system will be perceived and for how the user will behave. How can the system motivate the user to talk a lot and not only in short command-like utterances?

### 3.4 Auto-adjust gameplay and CALL

If the task is too easy the player gets bored, and if the task is too difficult the player gets frustrated. In both cases the result is that the player will lose interest in the game. A well known technique to counter this in a game is to increase

the difficulty of the task as the player improves. The more skilled the player gets, the harder the task becomes. How should an interactive story be modeled structurally to support the auto-adjustment of gameplay in a CALL application such as DEAL? These are some of the things we wish to study with DEAL.

## 4 Acknowledgements

This research was carried out at Centre for Speech Technology, KTH. The research is also supported by the Graduate School for Language Technology (GSLT). We wish to thank Jens Edlund, Gabriel Skantze, Rolf Carlson Björn Granström, Olov Engwall, Philip Shaw, and Becky Hincks.

## 5 References

- Beskow, J. (2003). Talking Heads - Models and Applications for Multimodal Speech Synthesis. Doctoral dissertation, KTH.
- Ellis, N. (2006). Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Saliency, Interference, Overshadowing, Blocking, and Perceptual Learning. *Applied Linguistics*, 27, 164-194.
- Granström, B., & House, D. (2005). Effective Interaction with Talking Animated Agents in Dialogue Systems. In van Kuppevelt, J., Dybkjaer, L., & Bernsen, N. O. (Eds.), *Advances in Natural Multimodal Dialogue Systems* (pp. 215-243). Springer, Dordrecht, The Netherlands.
- Gustafson, J., Bell, L., Boye, J., Lindström, A., & Wirén, M. (2004). The NICE fairy-tale game system. In *Proceedings of SIGdial*. Boston.
- Hjalmarsson, A. (submitted). The effects of human-like variability in utterance generation. Submitted to *Proceedings of Interspeech 2007*. Antwerp, Belgium.
- Iuppa, N., & Borst, T. (2007). *Story and simulations for serious games : tales from the trenches*. Focal Press.
- Johnson, W., Vilhjalmsson, H., & Marsella, S. (2005). Serious games for language learning: How much game, how much AI?. In *12: th International Conference on Artificial Intelligence in Education*. Amsterdam.
- Mateas, M., & Stern, A. (2003). Façade: An experiment in building a fully-realized interactive drama. In *Game Developer's Conference: Game Design Track*. San Jose, California, US.
- Skantze, G. (2005). Galatea: a discourse modeller supporting concept-level error handling in spoken dialogue systems. In *Proceedings of SigDial* (pp. 178-189). Lisbon, Portugal.
- Wik, P. (2004). Designing a virtual language tutor. In *Proc of The XVIIth Swedish Phonetics Conference, Fonetik 2004* (pp. 136-139). Stockholm University.