

DEAL

A Serious Game For CALL

Practicing Conversational Skills In The Trade Domain

Preben Wik¹, Anna Hjalmarson¹, Jenny Bruski²

¹Department of Speech, Music and Hearing KTH, Sweden

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

preben, anna [at]speech.kth.se] jenny.bruski@hgo.se

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

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 automatized [1]. Automaticity is achieved when a task can be performed almost effortlessly. The learner does not have to think about individual steps and can carry out the task from start to end while thinking about other things. Dramatic changes in brain activity can be seen on fMRI scans as automaticity develops, because the brain can process information more quickly and with less effort [2]. When we speak our first language, it is to a large extent automatic. When we learn a second language it is not, and to achieve some level of fluency in a language, what we really want to achieve is automaticity.

Research and literature on second language acquisition (SLA) and Computer assisted language learning (CALL) is diverse, with no single theory or model seen as the most appropriate way to achieve such automaticity. However, there seem to be a consensus about the value of conversational interactions. The more you talk the better it is. To automatize 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.

1.1. Motivation from games and gameplay

Because of its complexity, learning a language requires a substantial effort, and the motivation for doing so varies both over time and between individuals. A wish to be like the speakers of the language (integrative motivation) is often a strong motivating factor for younger learners, whereas the utility of what is learnt (instrumental motivation) is a stronger motivator for others. Motivation can also come from the

pleasure of learning (intrinsic motivation), or from the task itself (task motivation) to mention some sources.

There is a growing trend among educational researchers to look at games and game design in order to make education more effective. Game designers focus on finding ways to keep the players engaged and have in their strive for success developed several effective design strategies both to get and to keep players engaged and motivated throughout a game. This is known as *good gameplay*. According to Prensky [3] “Gameplay is all the doing, thinking and decision making that makes a game either fun or not”. Good gameplay is what makes games addictive, and what makes millions of people spend a significant amount of their time and money on playing games. The pleasure of engagement is the motivating force to play.

The same design principles that are used by game developers 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 [4]. Good gameplay adds to any existing motivation to learn if there is one, and may create a motive by itself if there isn't. The idea to transform education and create more engaging educational material by looking at the games industry has been suggested and described by several authors, for example [5][6]

1.2. How do we get students to talk?

Our aim with this work is to allow language learners to practice conversational skills in a fun and challenging context. Our objective is similar to that of the Tactical Language Training System (TLTS) [7], 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 [8]. Our objective is also similar to that of the Nice project [9], 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. Our focus is however on language learning.

2. DEAL

DEAL is a dialogue system under development, built in order to explore the challenges and potential benefits described in the introduction. In short, it is a roleplay for second language learners, using a spoken dialogue system. It 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 [10]. 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 (e.g. Figure 1). 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)

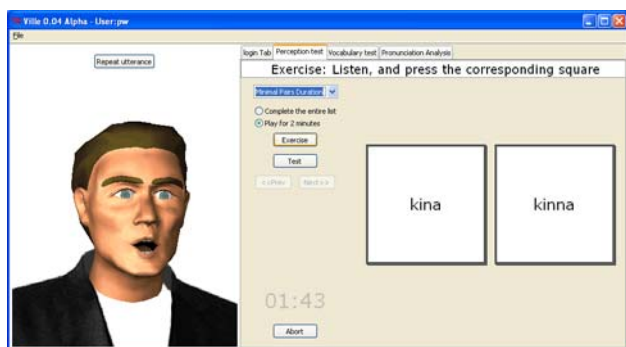


Figure 1. Ville, the virtual language tutor teaching aspects of Swedish

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 common 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

In combination it is 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

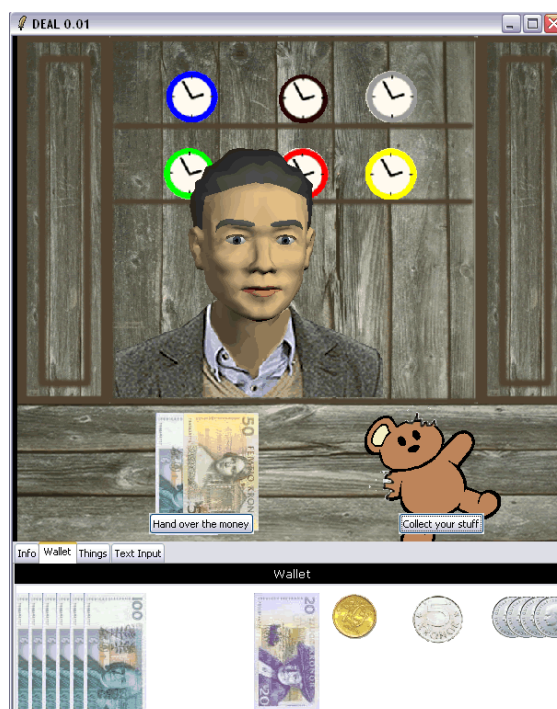


Figure 2. The DEAL interface with a stingy shopkeeper trying to sell a teddy-bear with a missing ear, for an outrageous price.

The user interface in DEAL is divided into six parts (Figure 2). The top part contains the shopkeeper, an ECA – also known as an animated talking head. Our ECAs are developed at KTH [11], 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. Language is multimodal, and in second language learning, visual signals may in many contexts be more important than verbal signals. 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 [12]. The use of an ECA will hopefully also help to entice a more human-like interaction between the student and the dialogue system.

The middle part of the user interface portrays the shop-counter, where any objects discussed between the user and the shopkeeper are shown, and where the money transaction takes place if the negotiation results in an agreement. The pictures also give clues about the scope of the domain, i.e., what can be talked about

In the lower part of the user interface there is a notebook with four tabs. The info-tab contains hints about things the user might try to say if he or she is having difficulties remembering what things are called, or if the conversation has stalled for other reasons. The wallet-tab (shown in Figure 2) contains the money the user has at his or her disposal (section 2.5) By clicking on a coin or a bill the money is moved to the shop-counter, where the shopkeeper can take it and give back correct change (or incorrect change as a possible dramatic action). The things-tab holds a picture of all the items the user has managed to acquire, as well as a small database of the conditions for the acquisition (the price, a log of the conversation that led to the purchase etc). Finally, the text input-tab offers a text input field as an alternative to the automatic speech recognition.

2.4. Dialogue system in DEAL

The dialogue system component in DEAL is based on Higgins, a spoken dialogue system built at KTH [13]. 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 automatized 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

The scenario for a language student that is using Ville in conjunction with DEAL will then be as follows. The first few lessons in Ville will teach the rudimentary vocabulary and grammar that is associated with the trade domain. The numbers, some colors, a few objects like a clock and a teddy-bear in addition to a few phrases like "Do you have..." "How much does that cost" and so on. 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, like for example in the dialogue below.

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

So far, the scenario, rules and possible actions in DEAL are fairly limited, and much can be added. We feel however that the combination of game and CALL in many ways is a fortunate one, opening up for novel ways to use traditional techniques from speech technology. This paper describes the humble 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 a difficulty 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. 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).

Playing a dialogue system like a roleplay also changes other criteria for how the system is to be 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 will try to minimize the number of turns needed. For the 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 role-play, 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 willing suspension of disbelief?

The willing suspension of disbelief refers to the state of accepting, for the time being, the premises of a work of fiction as though it were real. 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 [14]. 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. Acknowledgement

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, and Becky Hincks.

5. References

- [1] Ellis, N. "Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning", *Applied Linguistics*, 2006, Vol. 27: 164-194
- [2] Schneider, W. Chein, J. Controlled & automatic processing: behavior, theory, and biological mechanisms, *Cognitive Science*, 2003, Vol. 27, No. 3, Pages 525-559
- [3] Prensky, M. (2002). The Motivation of Gameplay: or, the REAL 21st century learning revolution. On *The Horizon*, Volume 10 No 1.
- [4] Iuppa, N & Borst, T. 2007, *Story and simulations for serious games: Tales from the trenches*, Focal Press
- [5] Prensky, M.(2001) *Digital game-based learning* McGraw Hill
- [6] Gee, J. *What Video Games Have to Teach Us about Learning and Literacy.* , Palgrave Macmillan, 2003
- [7] Johnson, W. L., Marsella, S., & Vilhjalmsson, H. (2004). *The DARWARS Tactical Language Training System*. Proceedings of IITSEC 2004.
- [8] 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.
- [9] Gustafson, J., Bell, L., Boye, J., Lindström, A., & Wirén, M. (2004). *The NICE Fairy-tale Game System*. In *Proceedings of SIGdial*. Boston.
- [10] Wik, P. (2004). *Designing a virtual language tutor*. In *Proc of The XVIIth Swedish Phonetics Conference, Fonetik 2004* (pp. 136-139). Stockholm University.
- [11] Beskow, J. (2003). *Talking Heads - Models and Applications for Multimodal Speech Synthesis*. Doctoral dissertation, KTH
- [12] Granström, B., & House, D. (2005). *Effective Interaction with Talking Animated Agents in Dialogue Systems*. *Advances in Natural Multimodal Dialogue Systems* (pp. 215-243). Springer, Dordrecht, The Netherlands
- [13] Skantze, G. (2005). *Galatea: a discourse modeler supporting concept-level error handling in spoken dialogue systems*. In *Proceedings of SigDial* (pp. 178-189). Lisbon, Portugal.
- [14] Hjalmarsson, A. (submitted). *The effects of human-like variability in utterance generation*. Submitted to *Proceedings of Interspeech 2007*. Antwerp, Belgium.