Task-Based Second Language Learning VR System

Tetsuya Saito
University of Tsukuba
1-1-1 Tennodai, Tsukuba Ibaraki Japan
stetsuya@graphic.esys.tsukuba.ac.jp

Junichi Hoshino
University of Tsukuba
1-1-1 Tennodai, Tsukuba Ibaraki Japan
jhoshino@graphic.esys.tsukuba.ac.jp

Abstract

In traditional English learning as a second language, learners rarely have the opportunity to practice oral communication, so the acquisition of oral proficiency is a slow process. In this paper, we propose a task-Based second language learning VR system. This method enables learners to obtain communicative skills through the practice of particular "missions" using voice and gesture communications with life-size 3D virtual human. To construct communicative virtual human for learning, more complex modeling of real-world conversation scenes are required. We have developed dynamic locomotion control technique and negotiation of meaning model to construct realistic conversation scenes.

1. Introduction

The task-based learning method enables learners to obtain communicative skills through the "missions". In a 3D visual environment, learners can practice their communication skills, where they must converse with the local people, who in turn will help them accomplish missions. In the past, the Tactical Language Training System [1] has provided rapid training in foreign language and culture through task-oriented spoken language instruction and intelligent tutoring. However, conversation scenarios are fixed, and negotiation of meaning is not employed.

In this paper, we propose a task-based second language learning VR system. The task-based learning method enables learners to obtain communicative skills through the practice of particular missions using voice and gesture communications with life-size 3D virtual human.

First, we model the conversation process with error correction mechanism. In the process of acquiring language skill through task-based learning, it is important for learners to try to convey information to one another and reach mutual comprehension through restating, clarifying, and confirming information via the process of communication. Therefore, we constructed a model of negotiation of meaning for the virtual human to clear a stumbling of meaning during conversation with learner.

Then, we model the dynamic spatial locomotion of the virtual human. We construct a locomotion network for a virtual environment. The optimal walking path is calculated by a multi-pass searching algorithm which uses node activation from the locations and conversation units. The character also locally adjusts its position not to occlude the referenced object from user's sight.

2. System Design Model

Figure 1 shows a snapshot of using second language learning VR system. The use environment of the system is a public classroom or a personal study room. The learners are assumed to be people who studied English in junior high school and high school. The system interface consists of a large screen, a camera for the user behavior analysis, and voice recognition. The virtual human recognize user's voice and return conversation sentences based on the negotiation of meaning model.

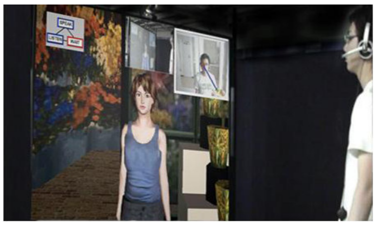

Figure 1: A snapshot of second language learning VR system

The learner experiences task-based language training in various daily situations such as shopping, cooking, and job. This learning method enables learners to obtain communicative skills through the practice with life-size 3D virtual human. Moreover, the focus is applied to the acquisition of necessary English knowledge and skill for use in daily life. The system architecture is designed to support several important internal requirements, which contains the Task Model and Conversation model (Figure 2).The Task Model provides a task based on learner, and initializes the scenario and position of virtual human. The Conversation Model includes the Negotiation of Meaning. When a virtual human answers a question, it uses a speech recognizer and natural language parsers that can annotate phrases based on structural information and refer to relevant grammatical explanations. And the virtual human can understand input and request the learner for confirmation of the incomprehensible phrase.

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

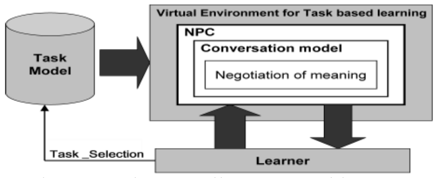

Figure 2: The overall system architecture

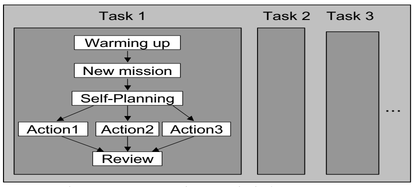

Figure 3: Task model for TBLT

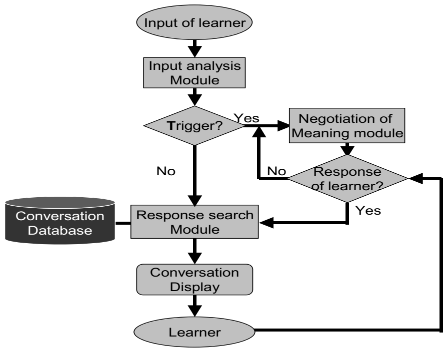

Figure 4: Flow chart of conversation

3. Task-Based Learning System

3.1. About The Task

There are various definitions of the word 'task', such as Nunan's definition [2], who stated that a task can be defined as "a piece of classroom work which involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is primarily focused on meaning rather than form." In this paper, the task provides a goal. To achieve the goal, learners have to talk to virtual human.

3.2. Task Model

Each task is composed of "Input-oriented", "Output -oriented" and "Review", and this flow is made based on the "Computational model of L2 acquisition" [3]. We constructed a task model based on it (Figure 3).

In the "Warming up" stage (Input-oriented), the learner studies the necessary vocabulary and grammar to accomplish the task. The learner starts to "play" the task when "Warming up" is completed, and the system moves to the "Self-planning" stage (Output -oriented). Here, the actions of the learner decide how they will accomplish the task. There are several ways to reach the goal, and the learner will move on to an "Action" phase after making their choice. Finally, when the learner reaches the goal, the system changes to the "Review" phase, where feedback with a focus on learning is provided.

4. Conversational Virtual Human

4.1. Conversation Database

To construct a task accomplishment type conversation database, we collected the conversation task that occurred between two interlocutors in a specific task, and these conversation data were classified into various subsets. The basic conversation templates are stored in the conversation database, with about 100 subsets, such as "Greeting", "Position", "Means", and "Person".

4.2. Mechanism of Conversational Processing

Figure 4 shows the internal conversation processing of a virtual human and the relation of conversation between the learner and virtual human.

The "*Trigger*" is the point where comprehension of the learner's input is confirmed or rejected by the virtual human. If the input of the learner were understood, the response of the virtual human would be to search the "*Conversation Database*" using the "*Response Search Module*". At the same time, if the input of the learner weren't understood, the negotiation of meaning of the virtual human in the "*Negotiation of Meaning Module*" would activate. Consequently, based on the "*Response of Learner*", the virtual human's response would be retrieved from the "*Conversation Database*" if the learner understands. Otherwise, the negotiation of meaning will continue until the learner understands.

4.2.1Input Analysis Module

When the learner inputs dialogue, the virtual human analyzes the input sentence and acquires the following information in order to judge the correct intention of the learner. To determine the component part of speech of each word, we used the morphological analysis tool "*TreeTagger*" [4]. And declarative sentences and interrogative sentences are classified according to the use of interrogatives and interrogation marks in the input sentence. Question type was classified into four kinds (How, What, Where and Others) of interrogative.

The Input Analysis Module finds the word that corresponds to a question in the conversation data, and the Response Search Module searches for the appropriate answer template base based on that word. In addition, it is necessary to complement the conversation template to generate the final reply of the virtual human based on the basic response template like that outlined in Table 1.

Table 1: Examples of response templates

Question type	**Response example**
What	"I mean" *&char_explain*
How	"You can" *&avenue* "from here."
Where	*&position_name* "is" *&address*
Others	Sorry, I can't understand what you said

4.2.2Negotiation of meaning module

There is a tremendous possibility that a problem such as a misunderstanding will arise when the learner has a conversation with a virtual human. Kotter [5] provided a method of negotiating meaning when the meaning of conversation can't be understood. Based on Table 2, our system uses the *Word repair function* and the *Synonym conversion function* to solve problems in conversation.

Table 2: Repair Types

Repair Types	**Explain**
Confirmation check	Yes or No
Clarification request	Did you mean···?
Comprehension check	Do you understand?
Repetition	Repetition of word
Recast	Implicit error correction
Overt	Indication of understanding

5. Conversational Locomotion

In the interactive task environment, characters should be able to compose gestures and locomotion based on the series of task locations and surrounding objects. The proper location and timing of the character is influenced by various contexts such as the connection of scene locations and the current environment. The character also locally adjusts its position not to occlude the referenced object from user's sight.

5.1. Overview of the architecture

Figure 6 shows the conversational locomotion architecture. The system has a locomotion module, conversation modules, and a task manager. A task consists of a set of scene units and controls the discourse of the conversation. A scene unit has a pre-condition, scene location, and links to a collection of possible conversation modules. Proper scene units are selected using the pre-conditions, such as environment change and the user's utterance. Locomotion module dynamically plans locomotion paths and generates walking motion pattern based on key locations.

5.1.1 Key location control

To compose locomotion and conversation, we need to decide the character's location and the timing of walking during conversation. The actor's locations are influenced by where the scenes are talking place and the content of the conversations. Locomotion and conversation is composed by following three types of location constraint.

1) Scene location: The scene location corresponds to where the actions and conversations are taking place.
2) Interpersonal location: Character changes relative locations from the other actors during conversation.
3) Referent location: This is the relative locations of the character and the referenced object.

These location constraints are used as key locations in the conversational locomotion planning. The key location consists of a position in the floor coordinate system, and a standing duration. In most scenes, the proper standing position of the character has a degree of freedom.

Figure 5: Example of conversational locomotion and gestures

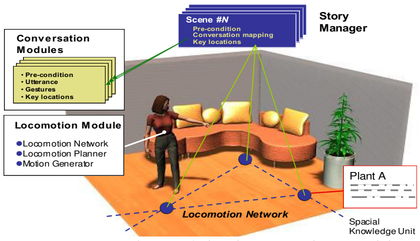

Figure 6: Overview of the locomotion architecture.

5.1.2 Conversational locomotion network

Locomotion is controlled by activating the locomotion network using task locations and conversation units. An optimal locomotion path is selected by calculating the optimal locomotion path with the maximum activation. The initial locomotion network is constructed by sampling the possible standing locations. The candidate node positions are task locations and before objects referenced in conversation units. Timing of locomotion is controlled by associating a key location at a proper clause in utterance. For example, the referent location can be associated with clauses including the referenced object. Even if the key location specification is pre-determined, the actual character motion is dynamically changed depending on the task locations and the order of the conversations.

5.2. Conversational Locomotion Planning

The locomotion path is dynamically selected by using a multi-pass searching algorithm that calculates the maximum activated path. When the status of activation is changed by conversation units, the path is re-calculated.

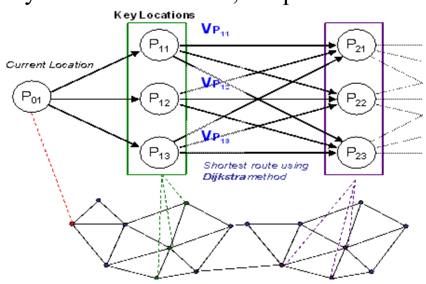

Figure 8: Key location and multiple pass searching. N-best key location is selected to search maximum activated pass.

5.2.1 Multiple pass searching

Multiple key locations are set with a different activation value. By selecting N-best key locations, the possible locomotion segments between key locations are selected. The total activation along the locomotion segments is calculated. Locomotion segments between candidate key locations are obtained by Dijkstra method [Dijkstra 1959]. We calculate candidate locomotion segments such as P00 -P01 to obtain the total activation value. (Figure 8)

5.2.2 Activation functions

In addition to the scene locations, we use the apparent object size and walking size to locally control locomotion.

1) Apparent object size (Figure 9.a):

When the apparent size of the referenced object is small, the character should move closer until it becomes large enough. The referent object is approximated by a sphere, and the view angle from user's eye position is calculated. Note that the approximated object size corresponds to the object area referred to in the conversation.

2) Walking distance (Figure 9.b):

When the walking distance from the current location of the character is longer, the character tries to avoid this longer path. The total activation values are calculated along locomotion segments.

$$V(P_{0,0}, P_{1,n_1}, \ldots, P_{s,n_s})$$
$$= \sum_{t=1}^{s} \left\{ w(t) \cdot \left[\alpha A(P_{t,n_t}) + (1-\alpha) D(P_{t-1,n_{t-1}}, P_{t,n_t}) \right] \right\}$$

Where w(t) is a weighting value. w(t) is used to control the number of key locations that the character should consider. Another type of activation function is easily integrated in this framework. For example by setting the negative activation value to the specific locations, the character will avoid entering that place.

5.2.3 Local position adjustment

The actor's position is locally adjusted not to obscure the user's sight of the referenced object. The viewing area of the user is calculated from the 3D location of user's eye and the referent object sphere. When the character approaches the object, it stops at the intersection of the view area and edges of locomotion networks. (Figure10)

6. RESULTS

We ran an experiment to demonstrate the effectiveness of the technique proposed in this paper. The learners were three graduate students with an average TOEIC score of 550. The content experiment was to learn the language necessary in the daily life of Japanese living in America. The virtual environment was set in town, and the task "Go to the restaurant" was assigned. The task was successful if the learner arrived at the restaurant. The conversation was about seven minutes.

Learning content (*T* shows a trigger of problem in the conversation.)
L (Leaner): Excuse me.
N (Virtual Human): Yes, what can I do for you?
L: Could you tell me how to get to Chinese restaurand? **[T]**
N: Do you mean restaurant?
L: Yes
L: Ok, It's across the street from the department store. Restaurant is in plain sight.
L: What is plain sight? **[T]**
N: I mean obvious place.
L: I see, thank you very much.
N: You're welcome.

In this experiment, a decreasing trend in spelling errors could be achieved via the *Word repair function*. In addition, the phrase that the learner didn't understand could be learned through the use of the *Synonym conversion function*. We confirmed that in the acquisition of language skill, negotiation of meaning is essential as part of the learner's conversational activity.

However, in cases where the learner's question was unclear, the virtual human could not generate an appropriate response. To solve this problem, more consideration in regards to the function of negotiation of meaning is required.

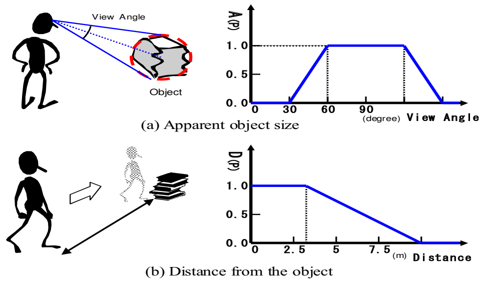

Figure 9: Scene constraints and activation value

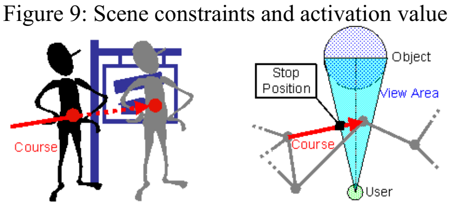

Figure 10: The local position adjustment using the user's relative position and the referenced object.

7. CONCLUSIONS

We have presented an approach to promoting communication and achieved learner's greediness for learning through the negotiation of meaning model in the task-based learning system. Moreover, the effectiveness of the virtual human's was acknowledged in the learner's evaluations of the experiment. In future, it will be necessary to evaluate a series of tasks in order to improve the conversation function of virtual human's.

References

[1] Johnson, W. L., Marsella, S., & Vilhjalmsson, H.: The DARWARS Tactical Language Training System. Proceedings of I/ITSEC, 2004.

[2] Nunan, D.: Understanding Language Classrooms. A guide for teacher-initiated action. Hemel, Hempstead: Prentice Hall International (UK) Ltd, 1989.

[3] Ellis, R.: Teaching and Research: Options in Grammar Teaching. TESOL QUARTERLY; 32: 39-60, 1998.

[4] http://www.ims.uni-tuttgart.de/projekte/corplex/TreeTagger/ last accessed Feb.8, 2007.

[5] Kötter, M.: Negotiation of Meaning and Code switching in Online Tandems. Language Learning & Technology, pp.145-172, 2003.