Foreign Language Tutoring in Oral Conversations Using Spoken Dialog Systems

Sungjin LEE^{†a)}, Hyungjong NOH[†], Jonghoon LEE[†], Kyusong LEE[†], *Nonmembers,*
and Gary Geunbae LEE[†], *Member*

SUMMARY Although there have been enormous investments into English education all around the world, not many differences have been made to change the English instruction style. Considering the shortcomings for the current teaching-learning methodology, we have been investigating advanced computer-assisted language learning (CALL) systems. This paper aims at summarizing a set of POSTECH approaches including theories, technologies, systems, and field studies and providing relevant pointers. On top of the state-of-the-art technologies of spoken dialog system, a variety of adaptations have been applied to overcome some problems caused by numerous errors and variations naturally produced by non-native speakers. Furthermore, a number of methods have been developed for generating educational feedback that help learners develop to be proficient. Integrating these efforts resulted in intelligent educational robots — Mero and Engkey — and virtual 3D language learning games, Pomy. To verify the effects of our approaches on students' communicative abilities, we have conducted a field study at an elementary school in Korea. The results showed that our CALL approaches can be enjoyable and fruitful activities for students. Although the results of this study bring us a step closer to understanding computer-based education, more studies are needed to consolidate the findings.

key words: intelligent computer-assisted language learning, robot-assisted language learning, language learning game, grammatical error detection, corrective feedback

1. Introduction

It is a fact that the private English education fee in Korea, reaching up to 16 trillion won annually, adds a great burden to Korean economy, resulting in countless articles overflowing in the media on strengthening the public education system that focuses on enhancing students' speaking ability to straighten out their hunchbacked English ability compared with the excessive grammar knowledge. This shows clear evidence for the necessity for changing our current foreign language education system in public schools which mainly focuses on vocabulary memorization and grammar-translation methodology. Although there have been enormous investments into English education all around the world, not many differences have been made to change the rote learning style in English instruction. In addition, computer-based English learning is in the center of interest, however, this method also fails to provide the opportunity for free conversation and stays at the level of simple repetition of the given text. These teaching-learning methods can-

not provide persistent motivation for learners to reach the high proficiency levels in foreign language learning. Considering the shortcomings for the current teaching-learning methodology, we have been investigating English learning systems using natural language processing technology in immersion context based on the assumptions of second language acquisition theory and practice. Through the systems, foreign language learners practice English conversation in natural contexts and are provided with corrective feedback based on the error correction procedures. POSTECH and KIST's Center for Intelligent Robotics have been cooperating in developing robots as educational assistants, called Mero and Engkey. Another system, Pomy (POstech iMmersive English studY), presents a virtual reality immersion environment, where learners experience the visual, aural and tactual senses to help them develop into independent learners and increase their memory and concentration abilities to a greatest extent (Fig. 1).

The remainder of this paper is structured as follows. Section 2 describes related studies. Section 3 introduces the speech and language technologies used in our approaches. Section 4 presents a detailed description of our preliminary field study and the results and discussion. Finally, Sect. 5 gives our conclusion.

2. Related Work

2.1 Second Language Acquisition Theory

Since the advent of Second Language Acquisition (SLA), a number of crucial factors have been revealed for improving students' productive conversational skills: 1) comprehensible input [1], 2) comprehensible output [2], 3) corrective feedback [3], and 4) motivation and attitude [4].

In relation to oral understanding, accumulated work on the process of listening suggests that comprehension can only occur when the listener places what she or he hears in context. While comprehensible input is invaluable to the acquisition process, it is not sufficient for students to fully develop their L2 proficiency. The output hypothesis claims that production makes the learner move from 'semantic processing' prevalent in comprehension to more 'syntactic processing' that is necessary for improving accuracy in their interlanguage [2]. Specifically, producing output is one way of testing one's hypotheses about the L2. Learners can judge the comprehensibility and linguistic well-formedness

Manuscript received April 28, 2011.

Manuscript revised August 20, 2011.

[†]The authors are with POSTECH, Pohang, 790-784 South Korea.

a) E-mail: junion.sjlee@gmail.com

DOI: 10.1587/transinf.E95.D.1216

Fig. 1 Mero, Engkey, and Pomy.

of their interlanguage utterances against feedback obtained from their interlocutors, leading them to recognize what they do not know, or know only partially.

On the other hand, it has been argued that corrective feedback plays a beneficial role in facilitating the acquisition of certain L2 forms which may be difficult to learn through input alone. For example, some L2 forms are rare, low in perceptual salience, semantically redundant, do not typically lead to communication breakdown, or lack a clear form-meaning relationship.

Motivation and attitude is another crucial factor in L2 achievement [4]. For this reason it is important to identify both the types and combinations of motivation that assist in the successful acquisition of a foreign language. In order to make the language learning process a more motivating experience, researchers need to put a great deal of thought into developing programs which maintain students' interest and have obtainable short term goals. The use of an interesting computer-based method can help to increase the motivation level of students, and computer-based learning has an advantage over human-based learning in that it seems to give a more relaxed atmosphere for language learning [5]–[7].

There have been few serious attempts to provide students with natural contexts that embody most of the aforementioned attributes.

2.2 Related Research Projects

Many research projects have tested the idea of providing pronunciation training using a speech recognizer in a forced recognition mode [8], [9], but a few systems exist that allow the user to engage in some form of meaningful dialog.

DEAL, developed at KTH, is a roleplay dialog system for second language learners, using a spoken dialog system [10]. It is intended as a multidisciplinary research platform, particularly in the areas of human-like utterance generation, game dialog, and language learning. The domain is the trade domain, specifically flea market situation. DEAL provides hints about things the user might try to say if he or she is having difficulties remembering how things are called, or if the conversation has stalled for other reasons.

Another system is the Spoken Electronic Language Learning (SPELL) system [11]. It provides opportunities for learning languages in functional situations such as going to

a restaurant, expressing (dis-)likes, etc. Recast feedback is provided if the learner's response is semantically correct but has some grammatical errors. This system combines semantic interpretation and error checking in the speech recognition process. Thus, it uses a special speech recognition grammar to cover both normal speech and erroneous speech.

Spoken Conversational Interaction for Language Learning (SCILL) was developed based on the spoken dialog system of MIT [12]. This system covers the topics of weather information and hotel booking. They implemented the simulated user to produce example dialogs to expose language learners to language use and to expand the training corpus for the system. It decides stochastically what to say on the basis of the system's previous reply [13].

The Let's Go system [14] is a spoken dialog system that provides bus schedule for the area around Pittsburgh, PA, U.S.A. This system is an extension of a previously developed system [15]. Raux and Eskenazi adapted non-native speakers' data for the use of language learning. They modified the grammar for the native speaker. Modifications include the addition of new words, new constructs and the relaxation of some syntactic constraints to accept ungrammatical sentences.

In Japan, the educational use of robots has been studied, mostly with Robovie in elementary schools, focusing on English language learning. Robovie has behavior episodes with some English dialogs. To identify the effects of a robot in English language learning, the researchers placed a robot in an elementary school, and compared the frequency of students' interaction with the English test score. The students who showed a lot of interest at the starting phase had a significantly elevated English score. This implies that robot-aided English learning can be effective for students' motivation [16].

IROBI was recently introduced by Yujin Robotics in Korea. IROBI is both an educational and home robot, containing many features. IROBI was used in [17] to compare the effects of non-computer-based media and web-based instruction with the effects of robot-assisted learning for children. Robot-assisted learning is thought to improve children's concentration, interest, and academic achievement. It is also thought to be more user-friendly than other types of instructional media.

Studies on dialog-based computer-assisted language learning (DB-CALL) are still relatively new and most are

in the early stages in a starting phase. Therefore, many attempts need to be made to investigate the effects of their use.

The following section gives an account of the speech and language technologies which have been used in our systems.

3. Speech and Language Technology

We have constructed DB-CALL systems, including speech recognition, language understanding, dialog management and corrective feedback generation modules, which can perceive the utterances of learners, especially Korean learners of English, to provide effective feedback and the opportunities for practicing free conversation.

3.1 Automatic Speech Recognition

Speech recognition is performed by the DARE recognizer [18], a speaker independent real-time speech recognizer. Since data is costly for a fully trained acoustic model for a specific accent, we have used a small amount of transcribed Korean children's speech (17 hours) to adapt acoustic models that were originally trained on the Wall Street Journal corpus using standard adaptation techniques, both of maximum likelihood linear regression (MLLR) [19] and maximum a posteriori (MAP) adaptation [20]. The occurrence of pronunciation variants was detected with a speech recognizer in forced-alignment using a lexicon expanded according to all the possible substitutions between confusable phonemes. Korean speakers tend to replace the following consonants with the correspondingly similar consonants and the following eight pronunciation variants of vowels are common to Korean speakers (Table 1). By virtue of adaptation of acoustic model and pronunciation dictionary, and use of restrictive grammars, the average word error rate was about 22.8% at the vocabulary size of 1250.

3.2 Language Understanding

Since language learners commit numerous and diverse errors, a system should be able to understand language learners' utterances in spite of these obstacles. To accomplish this purpose, rule-based systems usually anticipate error

types and hand-craft a large number of error rules, but this approach makes these methods fragile to unexpected errors and diverse error combinations [11], [14], [21].

Therefore we statistically infer the actual learners' intention by taking not only the utterance itself but also the dialog context into consideration, as human tutors do. The ways to incorporate dialog context perhaps just combine all features both from utterances itself and dialog contexts into one feature set which is then used to train classification models. For DB-CALL, however, such approaches can be problematic, because distinct handling for each of the proficiency level is important in a language learning setting. Given a dialog scenario, the dialog-context model is relatively invariant; thus we prefer a hybrid model that combines the utterance model and the dialog-context model in a factored form (Fig. 2). This approach allows us to adjust the hybrid model to a required proficiency level by replacing only the utterance model.

The hybrid model merges n-best hypotheses from the utterance model with n-best hypotheses from the dialog-context model to find the best user's intention. In the language production process, user's intentions are first derived from the dialog context; subsequently the user intentions determine utterances [22]. By using this dependency and probabilistic inference rules, the most likely expected user's intention $I(U, D)$ given the utterance U and the dialog context D can be stated as follows:

$$I(U, D) = \underset{I}{\operatorname{argmax}} \frac{P(I|U)P(I|D)}{P(I)}$$

In this formula, $P(I|U)$ represents the utterance model and $P(I|D)$ represents the dialog-context model.

To predict the user intention from the utterance itself, we use the maximum entropy model [23] trained on linguistically-motivated features:

- **Lexical word features:** Lexical word features consist of lexical tri-grams using current, previous, and next lexical words. They are important features, but the lexical words appearing in training data are limited, so data sparseness problems can arise.
- **POS tag features:** POS tag features also include POS tag tri-grams matching the lexical features. POS tag

Table 1 List of possible substitutions.

Consonant	Vowel
CH → T	IH → IY
DH → D	OY → IY
TH → T	ER → R
TH → S	UH → OW
ZH → JH	EH → AE
F → P	AA → AO
R → L	AO → OW
V → B	AH → AA

Fig. 2 Hybrid model of language understanding.

Table 2 Representation of dialog context and an example for the shopping domain.

PREV_SYS_INT	Previous system intention Ex) request(quantity)
PREV_USR_INT	Previous user intention Ex) inform(item)
SYS_INT	Current system intention Ex) confirm(quantity)
INFO_EX_STAT	A list of exchanged information states which is essential to successful task completion; (c) denotes <i>confirmed</i> , (u) <i>unconfirmed</i> Ex) [item='apple'(c), quantity=3(u)]
DB_RES_NUM	Number of database query results Ex) 0

features provide generalization power over the lexical features.

On the other hand, determining user's intention from dialog state can be solved by finding similar dialog states within a dialog-state space. Each dialog segment is represented as one dialog state (Table 2). A dialog-state space is built by first collecting a dialog corpus. Semantic tags (e.g., current user/system intention and named entity) are then manually annotated to utterances. A hand-crafted automatic system is also used to extract discourse contextual features (e.g., previous intentions and exchanged information status) by keeping track of the dialog states for each point in the dialog. Then, the possible user's intentions can be selected from similar dialog states to the current dialog state. The best user's intention is obtained from the dialog state that maximizes the similarity. This idea naturally can be formulated as the k-nearest neighbors (KNN) problem [24] which provides high controllability for incrementally tuning the model during operation, which is practically very desirable property. The similarity function is defined as the following equation:

$$\text{Similarity}(D, D') = \sum_{k=1}^K \lambda_k f_k(D, D')$$

where D and D' are dialog states, K is the number of features, f_k denotes the feature functions, λ_k the weighted parameters for features. Our feature functions first include the simplest tests, whether a feature is shared or not, for each feature of a dialog context (Table 2). In addition, we include a number of feature functions based on general discourse and world knowledge. For example, if the system's intention is "inform(list_items)", the number of database query results becomes an important feature. If the number of results is greater than one, the most likely expected user's intention would be "declare(select_item)". If the number of results equals one, "declare(buy_item)" would be the most probable intention. The detailed algorithm is described in [25].

Fig. 3 Comparison between the hybrid model and the utterance model on user intention classification task.

To evaluate the proposed model, instead of involving real language learners, we simulated them by injecting grammatical errors into clear utterances generated using the user simulation method described in [26]. In the first step of the error generation procedure, we set the Grammatical Error Rate (GER) between 0% ~ 100% and determined error counts to be produced based on the GER. Then, we distributed the errors among categories and error types according to the percentages in the error types list. To verify the effectiveness of the dialog state-awareness, we compared the hybrid model with the utterance model. The utterance model just omits the dialog-context model from the hybrid model. We conducted 200 dialogs for each model per 10% GER intervals. The hybrid model significantly outperformed the utterance only model for overall range of GER. As the GER increased, the performance of the utterance model decreased dramatically, whereas the performance of the hybrid model decreased smoothly (Fig. 3). It verifies the effectiveness of dialog state-awareness through our hybrid approach.

3.3 Dialog Management

The dialog manager generates system responses according to the learner's intention. Our approach is implemented based on example-based dialog management (EBDM) framework, a data-driven dialog modeling, which was inspired by example-based machine translation (EBMT) [27], a translation system in which the source sentence can be translated using similar example fragments within a large parallel corpus, without knowledge of the language's structure. The idea of EBMT can be extended to determine the next system actions by finding similar dialog examples within an annotated dialog corpus. A dialog example is defined as a set of tuples that have the same semantic and discourse features. Each turn pair (one user turn and the corresponding system turns) in the dialog corpus is represented as one dialog example (Fig. 4) and each dialog example is mapped to semantic records in a dialog example database

Fig. 4 Indexing scheme for dialog example database.

(DEDB) because structured query languages (SQLs) can be easily manipulated to find and relax the dialog examples with some features. The index constraints represent the state variables which are domain-independent attribute. Our basic constraints consist of general features to define the dialog state such as the current user intention (dialog act and main goal), slot flags, discourse history vector, and lexico-semantic string of the current utterance. The DEDB is automatically built by first collecting a human–human dialog corpus related to pre-defined scenarios in each task. Then semantic tags (e.g., dialog act, main goal, and slot entity) are manually annotated to the user utterances, and system action tags to the system utterance. A hand-crafted automatic system is also used to extract discourse contextual features (e.g., previous intention and slot-filling status) by keeping track of the dialog states for each point in the dialog. To determine the next system action, the EBDM framework uses the three following processes:

- Query generation: DM generates an SQL statement using discourse history and the current dialog frame.
- Example search: DM searches for semantically similar dialog examples in the DEDB given the current dialog state. If no example is retrieved, some features can be ignored by relaxing particular features according to the level of importance given the dialog's domain.
- Example selection: DM selects the best example to maximize the example score based on lexico-semantic similarity and discourse history similarity.

The EBDM framework is a simple and powerful approach to rapidly develop spoken dialog systems for multi-domain dialog processing [28]. However, this framework must solve three problems for practical dialog systems for domain-specific tasks: (1) Keeping track of the dialog state to ensure steady progress towards task completion, (2) Supporting n-best recognition hypotheses to improve the robustness of dialog manager, and (3) Enabling error handling to recover automatic speech recognition (ASR) and spoken language understanding (SLU) errors. Consequently, we sought to solve these problems by integrating the agenda

Fig. 5 Global error handling procedure.

graph as prior knowledge to reflect the natural hierarchy and order of subtasks needed to complete the task. The graph is used to both keep track of the dialog state and to select the best system action using multiple recognition hypotheses for augmenting the previous EBDM framework. Dynamic help generation was also adopted as an error recovery strategy that provides immediate help messages using the agenda graph and dialog examples. Our error recovery strategies can use the discourse information to provide an intelligent guidance based on the agenda graph, and the help delivered may reflect what the user was trying to achieve at the current turn. The detailed algorithm is described in [29].

3.4 Corrective Feedback Generation

While free conversation is invaluable to the acquisition process, it is not sufficient for learners to fully develop their L2 proficiency. Corrective feedback to learners' grammatical errors is necessary for improving accuracy in their interlanguage. For this purpose, we have developed a component that handles learners' errors and helps learners to use more appropriate words and expressions during the conversation. When a learner produces ungrammatical utterances, the system provides both implicit and explicit feedback in a form of recast or elicitation, which was manifested as effective ways in the second language acquisition processes. When error is found, we judge whether the error that the student is making is global error (totally unintelligible) or local error (intelligible but partly incorrect). If it is global error, the system then stops the learner and gives correct utterances. If it is local error, it would be wiser to let the student keep talking while providing him with implicit corrective feedback.

3.4.1 Global Error Handling

When it is desirable to offer corrective feedback to global errors, the dialog manager provides fluent utterances which realize the learner's dialog act and main goal which can be regarded as the learner's intention. Corrective feedback generation takes two steps (Fig. 5): 1) Example Search: the dialog manager retrieves example expressions by querying Example Expression Database (EED) using the learner's in-

Fig. 6 The relation between dialog completion rate and the performance of the hybrid model and the average dialog length.

tention as the search key. EED is a database that stores native-like English expressions composed by English tutors corresponding to possible user intentions. 2) Example Selection: the dialog manager selects the best example which maximizes the similarity to the learner's utterance based on lexico-semantic pattern matching using Levenshtein distance. If the example expression is distant from the learner's utterance more than a pre-defined threshold, the dialog manager suggests the example as recast feedback and conducts a clarification request to induce learners to modify their utterance. Sometimes, students have no idea about what to say and they cannot continue the dialog. In such a case, timeout occurs and the utterance model does not generate hypotheses. Hence, we search EED with only the result of the dialog-context model and suggest the retrieved expression so that students can use it to continue a conversation [25].

To evaluate the appropriateness of the feedback, we conducted 200 dialogs per 10% GER intervals from 10% to 90%, and observed the Dialog Completion Rate (DCR) as the GER increased. As the GER increased, the performance of accuracy of the language understanding module (hybrid model) decreased, whereas the DCR decreased very slightly (Fig. 6). Because of the clarification sub-dialogs, the average dialog length increased as the GER increased. Based on this result, we can conclude that our method is suitable to produce appropriate feedback even when the inferred intention is not the same as the actual one. This is because the dialog context model effectively confines candidate intentions within the given context.

3.4.2 Local Error Handling

To provide corrective feedback to local grammatical errors, we have developed a method which consists of two sub-models: the grammaticality checking model and the error type classification model (Fig. 7).

The grammaticality checking task takes the recognized hypothesis in the form of a confusion network (CN) [30] and

Grammatical Error Detection	I	am	here	at	business
1) Grammaticality Checking	0	0	0	1	0
2) Error Type Classification	None	None	None	PRP_LXC	None

Fig. 7 The grammatical error detection model consists of two sub-models.

Table 3 Description of features extracted from each error pattern to train the grammaticality checking model.

Feature	Description
S1	Confidence score of the word hypothesis matching the first word in the error pattern
S2	Confidence score for the second word in the error pattern
S3	Confidence score for the third word in the error pattern
S4	Confidence score for the fourth word in the error pattern
S5	Confidence score for the fifth word in the error pattern
TS	Total score of S1, S2, S3, S4, and S5
SD	Indicator of structural error type: 1 for Deletion and 0 for Substitution

determines the grammaticality at each word position in sequence. Even without error type information, the grammaticality checking function may be very useful for some applications, e.g., categorizing learners' proficiency level and generating implicit corrective feedback such as repetition, elicitation, and recast feedback. To judge the grammaticality, we first extract error patterns from the ungrammatical responses. The error pattern is a tuple of the erroneous word and the two left and right neighbor words. For example, the error pattern for the proposition error at 'at' for the utterance 'I am here at business' will be a tuple ('am', 'here', 'at', 'business', '-'). The error pattern is also tagged with the error type and structural deviation (e.g., deletion or substitution) for the error type classification task. When a speech is recognized, at each position in the CN, we extract a feature vector by comparing the error patterns with the segment of the CN, consisting of the target position and the two left and right neighboring positions. We extracted seven features (Table 3) for each error pattern. The idea is that the higher matching scores an error pattern has, the more likely the recognized result has the relevant error in it. As the number of error patterns is very large but most of them are not informative, only the features extracted from top 10 error patterns ranked by the TS feature are used. In addition, we perform a similar feature extraction process at the parts-of-speech (POS) level. Figure 8 depicts the aforementioned feature extraction process for more understanding. We use the LIBSVM [31] Support Vector Machine (SVM) classifier with a radial basis function (RBF) as kernel to produce a model by a grid-search using 5-fold cross-validation.

To provide meta-linguistic feedback (i.e., detailed explanations about the grammatical error), we need to identify the error type. Identifying the error type is also beneficial to construct the learner model. Thus, we perform error type

Fig.8 An illustration of feature extraction process. PRP_LXC and AT denote proposition lexical error and article error. SUB and DEL mean substitution and deletion. DT, IN, NN, RB, VBP, and VBZ denote determiner, preposition, noun, adverb, non-3rd person singular present verb, and 3rd person singular present verb, respectively.

classification for the words that are determined as ungrammatical by the grammaticality checking model. We classify the error type by choosing the error type associated with the top ranked error pattern which is reordered by weighting errors that occur more frequently in a learner corpus. The detailed algorithm is described in [32].

To evaluate the proposed method, we took 100 utterances from the NICT JLE corpus [33] and expanded it to form the pool of 5000 utterances using the grammatical error simulator (see Sect. 3.5). For the training data, we randomly chose 250 utterances from the utterance pool and ten male Korean speakers read 50 utterances for each, resulting in 500 recordings. For the test data, we randomly chose 50 utterances from the utterance pool and ten male Korean speakers read the utterances for each, resulting in 500 recordings. The result showed that the grammaticality checking method achieved 75.3% in F-score (91.82% in precision and 63.82% in recall) while keeping the false positive rate very low (0.46%). The error type classification model exhibited a very high performance, 99.6% in accuracy. Because the high precision and low false positive rate are important criteria for the language tutoring setting, the proposed method will be helpful for intelligent DB-CALL systems.

3.5 Grammatical Error Simulation

One of the key elements to the development of DB-CALL systems for morphology and syntax training is to expand the recognition grammar to include not only grammatical responses but also ungrammatical responses. In previous studies, as each new scenario is developed, it is essential to create ungrammatical responses by hand. However, using human experts to anticipate various types of grammatical errors and list all possible realizations of the errors is too laborious and costly. Thus, automatic generation of realistic

grammatical errors to create recognition grammars is crucial to the development of such systems.

We developed a grammatical error simulator that generates errors which Korean learners of English usually make. To generate realistic errors, expert knowledge of language learners' error characteristics was imported into a statistical modeling system that uses Markov logic [34]. A Markov logic network (MLN) can be seen as a first-order knowledge base with weights attached to each of the formulas. For this study, a total of 119 MLN formulas were written. For example, English learners often commit pluralization errors with irregular nouns. These errors result because they over-generalize the pluralization rule, i.e., attaching 's/es' to the end of a singular noun, so that they apply the rule even to irregular nouns such as 'mice' and 'feet'. This characteristic is captured by the simple formula:

$$\text{IrregularPluralNoun}(s, i) \wedge \text{PosTag}(s, i, \text{NNS}) \Rightarrow \text{ErrorType}(s, i, \text{N_NUM}),$$

where $\text{IrregularPluralNoun}(s, i)$ is true if and only if the i -th word of the sentence s is an irregular plural, NNS stands for plural noun, and N_NUM is the abbreviation for noun number error.

The overall grammatical error generation procedure involves four steps: 1) Generating probabilities of error types for each word in the well-formed input sentence through MLN inference; 2) Determining an error type by sampling the generated probability for each word; and 3) Determining how errors structurally deviate from correct usage (omission or replacement) by sampling according to the ratio of the number of omission errors to the number of replacement errors in the NICT JLE corpus; 4) An ill-formed output sentence is created by realizing the chosen error types (Fig. 9). More detailed algorithm is described in [35].

We learned the weights of first-order formulas from the

Phase	Error type	He	wants	to	go	to	a	movie	theater	
Inference	v_agr	0.000	0.371	0.000	0.000	0.000	0.000	0.000	0.000	1 step
	prp_lex	0.000	0.000	0.284	0.000	0.269	0.000	0.000	0.000	
	at	0.000	0.000	0.000	0.000	0.000	0.355	0.000	0.000	
	***	***	***	***	***	***	***	***	***	
	none	0.921	0.449	0.604	0.866	0.605	0.506	0.781	0.798	2 step
Sampling	none		v_agr	prp_lex	none	none	at	none	none	
Structural Deviation			replace	omit			omit			3 step
Realization		He	<i>want</i>		go	to		movie	theater	4 step

Fig. 9 An example process of grammatical error simulation.

Table 4 Comparison between the precision and recall rates of the proposed and baseline methods.

Model	Precision (%)	Recall (%)	F1 Score (%)
Baseline	72.00	33.33	45.33
Proposed	78.00	41.67	53.67

NICT JLE corpus. The current version of the error tagset targets morphological, grammatical, and lexical errors and can describe diverse grammatical errors. The error tagset currently includes 46 tags. Due to the space limitation, please refer to [33] for the full list of error types. To verify the quality of the simulated grammatical errors, the proposed grammatical error simulation method was compared against the real learners' errors and the baseline model that uses only the POS bigram. The baseline method is comparable to the method of [36].

We employed precision and recall rate and human evaluation to measure the quality of grammatical error simulation. There are no generally accepted criteria as to what constitutes a good grammar error simulation. A good learner model should be able to generate learner-like behavior and errors. Precision and recall are a common measure of quality in user modeling [37] and were first used in evaluating user simulation models in dialog systems by [38]. To assess whether grammar error simulation models produce learner-like errors we need to compare their output with real responses in the same contexts. For this purpose, we used the precision and recall metrics. The results showed that the precision and recall rates of the proposed model were higher than those of the baseline model by 6.00% and 8.34%, respectively (Table 4). For human evaluation, the overall quality of the simulated errors was assessed by experienced researchers in English as a foreign language. Ninety sentences (30 sentences for each proficiency level) were randomly selected from the erroneous sentences in the NICT JLE corpus. For each erroneous sentence, we took the corresponding correct sentence from the corpus and then generated two sets of three sentences, one set using the baseline model and one set using the proposed model. The evaluators were given a correct sentence and a set of seven corresponding erroneous sentences (one real error and six simulated errors) for each of the 90 sentences. They were required to give a

Table 5 Mean, standard deviation, and ANOVA results of human evaluation.

Model	N	Mean	SD	F	Pr > F
Human	270	4.644	0.920	39.95	< 0.0001
Proposed	810	4.285	1.245		
Baseline	810	3.874	1.524		

Table 6 Results of Scheffé post-hoc analysis.

Model	Human	Proposed	Baseline
Human	--	0.359*	0.770*
Proposed	-0.359*	--	0.411*
Baseline	-0.770*	-0.411*	--

* Significant difference ($p < 0.05$)

score on the five-point Likert scale, which uses a scale of 1–5 to indicate “strongly disagree” to “strongly agree” with the statement “The error sentence seems to be generated by real language learners, not randomly generated by machine”. Descriptive statistics, analysis of variance (ANOVA), and multiple post-hoc comparisons were used to find the effects of different error sources on the human evaluation. The result showed that the human-generated errors recorded the highest score, and the proposed model obtained a higher score than the baseline model. The descriptive statistics and ANOVA results indicate that significant differences exist between the error sources (Table 5). To identify the differences in detail, a Scheffé post-hoc analysis was performed. Significant mean differences were found between all pairs of models (Table 6).

4. Field Study

We performed a field study at a Korean elementary school to investigate the educational effects of our approaches using the educational robots, Mero and Engkey. The following subsections describe the method of the study in more detail.

4.1 Setting and Participants

A total of 24 elementary students were enrolled in English lessons two days a week for a total of about two hours per day and had chant and dance time on Wednesdays for eight

Fig. 10 Students interacting with Mero and Engkey.

weeks during the winter vacation. However, three students left the study, resulting in a total of 21 students. The students in this study were recruited by the teachers of the school and divided into beginner-level and intermediate-level groups, according to the pre-test scores. They ranged from second to sixth grade; in general, there are six grades in a Korean elementary school. All of them were South Korean, spoke Korean as their first language and were learners of English as a foreign language. None of the participants had stayed in an English-speaking country, such as the United States and United Kingdom, for more than three months, which may indicate that this group had limited English proficiency. Figure 10 shows the layout of the classroom: 1) PC room where students took lessons by watching digital contents, 2) Pronunciation training room where the Mero robot performed automatic scoring of pronunciation quality for students' speech and provided feedback, 3) Fruit and vegetable store, and 4) Stationery store where the Engkey robots acted as sales clerks and the students as customers.

4.2 Material and Treatment

The researcher produced training materials including a total of 68 lessons, with 17 lessons for each combination of the level, beginner and intermediate, and the theme, fruit and vegetable store and stationery store. Among other things, the course involves small talks, homework checking, purchases, exchanges, refunds, etc. Participants in this course should become thoroughly trained in various shopping situations. With this aim in mind, when dealing with task assignment, the instructors proceeded in subtle gradations, moving from the simple to the complex. Throughout the course of the study, each student was asked to enter the four rooms in the order of PC room, Pronunciation training room, Fruit and vegetable store, and Stationery store so that students were gradually exposed to more active oral linguistic activities.

4.3 Data Collection and Analysis

In order to measure the cognitive effects, i.e., improvement of listening and speaking skills, different pre/post-tests were developed with the same level of difficulty. All students took a pre-test at the beginning of the study and a post-test at the end. For the listening skill test, 15 multiple-choice questions were used which were developed by experts in evaluation of educational programs. The items in the test were mainly selected from the content taught during the course. The test was used as the assessment tool in both the pre-test and the post-test phases of the study. The speaking skill test consisted of 10 1-on-1 interview items. The topics of the interviews were selected from the content taught. The evaluation rubric measured speaking proficiency on a five point scale in four categories: pronunciation, vocabulary, grammar, and communicative ability. A paired t-test was performed using the mean scores and standard deviations to determine if any significant differences occurred.

In order to investigate the effects on affective factors such as satisfaction in using robots, interest in learning English, confidence with English, and motivation for learning English, a questionnaire was designed by 10 teachers and experts in evaluation of educational programs. It consisted of some personal information and 52 statements in accordance with four-point Likert scale, which had a sliding answer scale of 1–4, ranging from “strongly disagree” to “strongly agree”, without a neutral option. Mean and standard deviation were used to evaluate the effect on students' satisfaction, whereas a pre-test/post-test method was used for other factors.

The design of the field study, however, makes the precise role of DB-CALL approaches in facilitating L2 development less than clear. This is due to the lack of a control group. In spite of the absence of control groups, we hope the results of this study can show the trend of effectiveness of DB-CALL approaches so that promotes further investigations.

4.4 Results and Discussion

According to the findings, there were large improvements of the speaking skills in the beginner-level participants' achievement on the post-test. The score in the post-test is significantly better than that of the pre-test. The listening skill, however, showed no significant difference. Significant differences of the speaking skill were also found in the result of the intermediate group and the effect sizes are also large, whereas the listening skill showed a significantly negative effect. The combined results of both groups (Table 7) showed no significant differences in the listening skill. This finding can be explained by a number of factors such as the unsatisfactory quality of the text-to-speech component and hindrance of robots' various sound effects. In addition, there might be learners' tendency to focus on their speaking. Because they are very familiar with a shopping situation,

Table 7 Cognitive effects on oral skills for overall students.

Category	N	Pre-test		Post-test		Mean difference	t	df	Effect size
		Mean	SD ^a	Mean	SD ^a				
Listening	21	10.95	3.2	10.67	1.91	-0.29	-0.55	20	0.12
Pronunciation	21	32.14	8.86	45.62	4.28	13.48	9.48*	20	0.90
Vocabulary	21	32.95	8.21	42.38	5.31	10.43	8.00*	20	0.87
Grammar	21	31.62	7.96	40.62	4.43	9.00	7.59*	20	0.86
Communicative ability	21	33.57	9.83	47.48	3.06	13.91	7.60*	20	0.86
Total	21	123.13	34.13	176.1	16.53	46.81	8.48*	20	0.88

* $p < .01$, SD^a = Standard Deviation **Fig. 11** The effects on affective factors.

they can buy items even though they cannot listen perfectly. The large improvement of speaking skill in the overall results agrees with the findings of previous studies in general. Specifically, the gain in the vocabulary area indicates that the authentic context facilitated form-meaning mapping and the vocabulary acquisition process. The improved accuracy of pronunciation and grammar supports the output hypothesis and the effects of corrective feedback. Learners had feedback at any related point which made them reflect on their erroneous utterances. The increase of communicative ability shows that learners were getting accustomed to speaking English. It also can be attributed to the fact that when using robot-assisted learning the student gained confidence in a relaxed atmosphere. A lack of confidence and a feeling of discomfort were more related to students' participation in face-to-face traditional discussions, and less to participation in computer-based learning.

The survey results on affective effects (Fig. 11) showed that the students were highly satisfied in using robots for language learning (Table 8). But, some questions showed the need to develop a more anthropomorphic appearance and a natural voice. The students' responses to the questions about interest in learning English on pre- and post-test showed a large improvement of interest with significance level of 0.01 (Table 9). But the lower score of the question about increase of familiarity with English might reflect that engaging in studying English for only two months is not enough to get familiar with listening and speaking English. A significantly large increase of confidence was found in the responses to the questions about confidence in En-

Table 8 Students' satisfaction in using robots.

Question	N	Mean	SD [†]
The robot looks smart	21	3.24	0.70
The robot can watch you	21	3.33	0.66
The robot can listen to your song and speech	20	2.95	0.83
The robot can come to you	21	2.90	0.94
The robot's appearance looks comfortable for learning	21	2.90	1.00
The robot's body looks comfortable for moving around in a classroom	21	2.62	0.80
The robot's facial expression looks comfortable to you	21	2.67	0.86
The robot's compliment is pleasing to you	21	3.38	0.74
You like the robot's voice	21	2.67	0.97
The robot seems secure	21	3.14	0.65
Total		2.98	0.44

SD[†] = Standard Deviation

glish on pre- and post-test with significance level of 0.01 (Table 10). This can be attributed to the fact that using robot-assisted learning allowed the students to make academic achievement and get confidence through repeated exercises in a relaxed atmosphere. However, relatively low scores were given to the questions related to individual level of fear or anxiety associated with either real or anticipated communication with another person. The responses to the questions about motivation for learning English presented a large enhancement of motivation, with significance level of 0.01 (Table 11). The low score of the questions related to preparing to study English may illustrate that traditional education doesn't work for the new generation of children. The popularity of e-Learning in Korea is promoting the increasing disengagement of the "Net Generation" or "Digital Natives" from traditional instruction. Please refer to [39] for more detailed information of the educational effects.

Unfortunately, we could not gather performance data of technical components during the field study. But we found that the educational robots generally worked well although there were occasional communication breakdowns because students' initial proficiency had many pronunciation and grammatical errors as well as non-correspondence. Also the educational robot consists of numerous components that are possibly vulnerable to system errors such as a wireless net-

Table 9 Students' interest in learning English.

Question	N	Mean		t
		Pre	Post	
English class is interesting	18	2.83	3.33	1.58
Listening and speaking English is interesting	18	2.65	3.00	0.96
Singing a song, chanting, and other games are interesting	18	2.83	3.59	3.00**
You want to talk with others in English	18	2.17	2.67	1.84
You enjoy learning English	18	2.72	3.17	2.20*
You want to use the expressions learned	18	2.83	3.50	3.37**
You inquire about an unknown word to a dictionary or others	18	2.94	3.17	1.46
You want to participate in English class with passion	18	3.06	3.11	0.33
You want more English classes in school	18	2.89	3.12	0.82
You are looking forward to English class	18	2.89	3.22	2.92**
You want to study English more in the future	18	2.83	3.11	1.76
You pay attention to what you are going to learn in English class	18	2.61	3.17	3.83**
You try to remember what you have heard	18	2.89	3.33	2.68*
You think English is easier and more familiar than before	18	2.67	2.89	2.83*
You are curious about what you are going to learn in your next English class	18	2.89	3.22	1.46
You are interested in English	18	2.78	3.39	3.05**
Total		2.78	3.21	3.21**

* $p < .05$, ** $p < .01$ **Table 10** Students' confidence with English.

Question	N	Mean		t
		Pre	Post	
You understand what you have learned in English class well	20	2.70	3.10	1.71
You answer the questions about what you have learned with confidence	19	2.58	3.16	2.63*
You are not afraid of speaking English	20	2.45	2.75	1.10
You sing songs and chant with confidence	20	2.80	3.20	1.80
You are not afraid of being questioned by a English teacher	20	2.50	2.75	1.10
You participate in English learning activities (role-play, game) actively	20	2.75	2.95	1.29
You are not afraid of English homework	20	2.55	3.15	3.04**
You think you can make a good presentation in English classes	20	2.75	3.20	3.33**
You fully understand what you have learned in English classes	20	2.95	3.35	2.99**
You feel no shame about your English mistakes	20	2.10	2.65	2.77*
You can greet foreigners with confidence	20	2.65	3.40	3.94**
You think that you can speak English better if you study harder	20	3.15	3.85	3.20**
Total		2.66	3.12	3.53**

* $p < .05$, ** $p < .01$

work interface, a robot motion controller, speech processing devices, etc. To resolve such problems, field operators had god a manual control and fixed problems when needed so that the session could proceed. We expect that we can have

Table 11 Students' motivation for learning English.

Question	N	Mean		t
		Pre	Post	
You are aware of the necessity of studying English	20	3.00	3.60	2.85**
You want to read signboards and lyrics written in English	18	2.78	3.21	3.43**
You recognize the importance of English to your present and future life	19	3.75	3.84	1.00
You want to learn English more	20	3.10	3.60	3.25**
You want to study hard in English classes	20	3.40	3.75	1.93
You have a desired level of English ability	20	3.15	3.55	2.99**
You want to buy English books and materials	20	2.40	2.95	3.58**
You make plans to study English	20	2.40	2.70	2.04
You spend more time on studying English by yourself	20	2.55	3.15	3.27**
You enjoy preparing for English classes	20	2.25	2.80	2.78*
You want to get praised for your English	20	3.55	3.90	2.67*
You want to write down your ideas in English	20	2.55	3.05	2.13*
You want to use what you have learned in English classes in everyday life	20	2.75	3.20	2.65*
You want to converse with foreigners actively	20	2.35	2.95	3.27**
Total		2.86	3.30	4.99**

* $p < .05$, ** $p < .01$

performance data of technical components on human experiments in future studies.

5. Conclusion

In this paper, we described the rationale of POSTECH approaches for CALL from a theoretical view of language learning and briefly introduced a set of technologies that we used to implement the educational assistant robots and 3D virtual language learning games. Our approaches basically apply many adaptations to the state-of-the-art technologies of spoken dialog system to overcome some problems caused by numerous errors and variations of non-native speakers. Furthermore, a number of methods have been developed for generating educational feedback. To investigate the cognitive and affective effects of our approaches, a course was designed in which students had meaningful interactions with intelligent robots in an immersive environment. The result showed no significant difference in the listening skill, but the speaking skills improved with a large effect size. Also, it showed that the systems promote and improve students' satisfaction, interest, confidence, and motivation. The results showed that our CALL approaches can be an enjoyable and fruitful activity for students. Although the results of this study bring us a step closer to understanding computer-based education, more studies are needed to consolidate/refute the findings of this study over longer periods of time using different activities with samples of learners of different ages, nationalities, and linguistic abilities.

Acknowledgments

This work was supported by Basic Science Research Pro-

gram through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0019523).

References

- [1] S.D. Krashen, *The input hypothesis: Issues and implications*, Laredo Pub., New York, 1985.
- [2] M. Swain, "Communicative competence: Some roles of comprehensible input and comprehensible output in its development," *Input in Second Language Acquisition*, vol.15, pp.165–179, 1985.
- [3] M.H. Long, "Focus on form in task-based language teaching," in *Language Policy and Pedagogy: Essays in Honor of A. Ronald Walton*, pp.179–192, John Benjamin, Philadelphia, 2000.
- [4] A.M. Masgoret and R.C. Gardner, "Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and Associates," *Language Learning*, vol.53, pp.167–210, 2003.
- [5] A. Liang and R.J. McQueen, "Computer assisted adult interactive learning in a multi-cultural environment," *Adult Learning*, vol.11, 1999.
- [6] J. Roed, "Language learner behaviour in a virtual environment," *Computer Assisted Language Learning*, vol.16, pp.155–172, 2003.
- [7] H. Yi and J. Majima, "The teacher-learner relationship and classroom interaction in distance learning: A case study of the Japanese language classes at an American high school," *Foreign Language Annals*, vol.26, pp.21–30, 1993.
- [8] J. Dalby, "Explicit pronunciation training using automatic speech recognition technology," in *Research in Technology and Second Language Education: Developments and Directions*, pp.379–398, Information Age Publishing, Greenwich, CT, 2005.
- [9] A. Neri, C. Cucchiari, and H. Strik, "Effective feedback on L2 pronunciation in ASR-based CALL," *Proc. Workshop on Computer Assisted Language Learning*, pp.40–48, 2001.
- [10] J. Brusk, P. Wik, and A. Hjalmarsson, "DEAL: A serious game for CALL practicing conversational skills in the trade domain," *Proc. SLATE-Workshop on Speech and Language Technology in Education*, Pennsylvania, 2007.
- [11] H. Morton and M.A. Jack, "Scenario-based spoken interaction with virtual agents," *Computer Assisted Language Learning*, vol.18, pp.171–191, 2005.
- [12] V.W. Zue and J.R. Glass, "Conversational interfaces: Advances and challenges," *Proc. IEEE*, vol.88, pp.1166–1180, 2000.
- [13] S. Seneff, C. Wang, and J. Zhang, "Spoken conversational interaction for language learning," *Proc. InSTIL/ICALL Symposium*, 2004.
- [14] A. Raux and M. Eskenazi, "Using task-oriented spoken dialogue systems for language learning: Potential, practical applications and challenges," *Proc. InSTIL/ICALL Symposium*, 2004.
- [15] A.I. Rudnicky, C. Bennett, A.W. Black, A. Chotomongcol, K. Lenzo, A. Oh, and R. Singh, "Task and domain specific modelling in the Carnegie Mellon Communicator system," *Proc. Sixth International Conference on Spoken Language Processing*, 2000.
- [16] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: A field trial," *Human-Computer Interaction*, vol.19, pp.61–84, 2004.
- [17] J. Han, M. Jo, S. Park, and S. Kim, "The educational use of home robots for children," *IEEE International Workshop on Robot and Human Interactive Communication*, pp.378–383, 2005.
- [18] D.H. Ahn and M. Chung, "One-pass semi-dynamic network decoding using a subnetwork caching model for large vocabulary continuous speech recognition," *IEICE Trans. Inf. & Syst.*, vol.E87-D, no.5, pp.1164–1174, May 2004.
- [19] C.J. Leggetter and P.C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Comput. Speech Lang.*, vol.9, pp.171–185, 1995.
- [20] G. Zavaliagos, R. Schwartz, and J. McDonough, "Maximum a posteriori adaptation for large scale HMM recognizers," *Proc. Acoustics, Speech, and Signal Processing*, pp.725–728, 1996.
- [21] D. Schneider and K.F. McCoy, "Recognizing syntactic errors in the writing of second language learners," *Proc. 17th International Conference on Computational Linguistics*, vol.2, pp.1198–1204, 1998.
- [22] D.W. Carroll, *Psychology of Language*, Wadsworth Publishing, Belmont, CA, 2003.
- [23] A. Ratnaparkhi, *Maximum entropy models for natural language ambiguity resolution*, Ph.D. Dissertation, University of Pennsylvania, 1998.
- [24] B.V. Dasarthy, *Nearest Neighbor: Pattern Classification Techniques*, IEEE Computer Society, 1990.
- [25] S. Lee, C. Lee, J. Lee, H. Noh, and G.G. Lee, "Intention-based corrective feedback generation using context-aware model," *Proc. International Conference on Computer Supported Education*, Valencia, 2010.
- [26] S. Jung, C. Lee, K. Kim, M. Jeong, and G.G. Lee, "Data-driven user simulation for automated evaluation of spoken dialog systems," *Comput. Speech Lang.*, vol.23, pp.479–509, 2009.
- [27] M. Nagao, "A framework of a mechanical translation between Japanese and English by analogy principle," *Readings in Machine Translation*, pp.351–354, 2003.
- [28] C. Lee, S. Jung, S. Kim, and G.G. Lee, "Example-based dialog modeling for practical multi-domain dialog system," *Speech Commun.*, vol.51, pp.466–484, 2009.
- [29] C. Lee, S. Jung, K. Kim, and G.G. Lee, "Hybrid approach to robust dialog management using agenda and dialog examples," *Comput. Speech Lang.*, vol.24, pp.609–631, 2010.
- [30] L. Mangu, E. Brill, and A. Stolcke, "Finding consensus in speech recognition: Word error minimization and other applications of confusion networks," *Comput. Speech Lang.*, vol.14, pp.373–400, 2000.
- [31] C. Chang and C. Lin, *LIBSVM: A library for support vector machines*, <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>, 2001.
- [32] S. Lee, H. Noh, K. Lee, and G.G. Lee, "Grammatical error detection for corrective feedback provision in oral conversations," *Proc. AAAI 2011: Twenty-Fifth Conference on Artificial Intelligence*, San Francisco, 2011.
- [33] E. Izumi, K. Uchimoto, and H. Isahara, "Error annotation for corpus of Japanese learner English," *Proc. Sixth International Workshop on Linguistically Interpreted Corpora (LINC 2005)*, pp.71–80, 2005.
- [34] M. Richardson and P. Domingos, "Markov logic networks," *Mach. Learn.*, vol.62, no.1, pp.107–136, 2006.
- [35] S. Lee, J. Lee, H. Noh, K. Lee, and G.G. Lee, "Grammatical error simulation for computer-assisted language learning," *Knowledge-Based Systems*, vol.24, pp.868–876, 2011.
- [36] J. Foster and E. Andersen, "GenERRate: Generating errors for use in grammatical error detection," *Proc. Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, pp.82–90, 2009.
- [37] J. Schatzmann, K. Georgila, and S. Young, "Quantitative evaluation of user simulation techniques for spoken dialogue systems," *Proc. SIGdial Workshop on Discourse and Dialogue*, pp.2–3, 2005.
- [38] I. Zukerman and D.W. Albrecht, "Predictive statistical models for user modeling," *User Modeling and User-Adapted Interaction*, vol.11, pp.5–18, 2001.
- [39] S. Lee, H. Noh, J. Lee, K. Lee, G.G. Lee, S. Sagong, and M. Kim, "On the effectiveness of robot-assisted language learning," *ReCALL Journal*, vol.23, no.1, pp.25–58, 2011.

Sungjin Lee received the B.S. degree in Computer Science Engineering from Pohang University of Science and Technology in 2006. He is currently a Ph.D. student of the same university. His research interests include grammatical error simulation and detection, intelligent computer-assisted language learning, spoken dialog system, machine learning.

Hyungjong Noh received the B.S. degree in Computer Science Engineering from Pohang University of Science and Technology in 2005. He is currently a Ph.D. student of the same university. His research interests include dialog management, intelligent computer-assisted language learning, spoken dialog system.

Jonghoon Lee received the B.S. degree in Computer Science Engineering from Pohang University of Science and Technology in 2005. He is currently a Ph.D. student of the same university. His research interests include pronunciation error detection, automated pronunciation assessment, statistical machine translation.

Kyusong Lee received the B.S. degree in Computer Science Engineering from Soongsil University in 2010. He is currently a Ph.D. student of Pohang University of Science and Technology. His research interests include language learning game, learner dialog corpus collection, spoken dialog system.

Gary Geunbae Lee received the B.S. and M.S. degrees in Computer Engineering from Seoul National University respectively in 1984 and 1986. He received the Ph.D. degree from in Computer Science from UCLA. He has been a professor at Computer Science and Engineering department of Pohang University of Science and Technology in Korea since 1991. He is a director of Intelligent Software Laboratory which focuses on human/natural language technology researches including spoken dialog processing, intelligent computer assisted language learning, speech synthesis and web/text mining.

processing, intelligent computer assisted language learning, speech synthesis and web/text mining.