International Letters of Social and Humanistic Sciences
ISSN: 2300-2697, Vol. 8, pp 24-33
doi:10.18052/www.scipress.com/ILSHS.8.24

Online: 2013-09-26

Text-To-Speech Software: a New Perspective in Learning and Teaching Word Stress, Word Intonation, Pitch Contour, and Fluency of English Reading

Hussein Meihami

Department of English Teaching, Ghorveh Branch, Islamic Azad University, Ghorveh, Iran

E-mail address: hussein.meihami@yahoo.com

ABSTRACT

When mentioning technology in the language classroom, the first impulse is to think computer technology, almost because computers have so pervade our daily home, workplace, and society contexts. The aim of this experimental research is to investigate the effect of using Text-To-Speech Software (TTS), one of Computer Assisted Language Learning (CALL) resources in teaching reading, in different aspects of reading fluency. In this study, we investigated teaching and learning of Word Stress, Word intonation, Pitch Contour, and Fluency of English reading through TTS. It should be stated that comprehension had been worked in the program but wasn't investigated in the study. The study indicated that word stress; word intonation, pitch contour, and fluency had significantly improved by using TTS software.

***Keywords*:** computer assisted language learning; reading fluency; speech to talk software; teaching reading; text-to-speech

1. INTRODUCTION

There is a controversial question on when technology entering the language classroom. Some like to think that technology first entered language classroom in the 1950s and 1960s in the form of language laboratory (Brown, 2007). Institutions hastened to dedicate rooms to the installation of multiple tape-deck-equipped booths where students gathered to listen to native speakers modeling the drills of the current day's lesson (chapelle, 2005). Often users of language labs were able to record their own voice and latter on repeated it to see its problems and to consulate with their instructors about it. The advent of language labs brought promises of great breakthroughs in language teaching: technology would come to rescue the less effective methods (Brown, 2007). When the personal computers came on the scene in the 1980s, some pioneers in language teaching thought of it as a salvation for the situation of less effective methods. Once again, this time with a more confidence, pioneers of language teaching thought of the state of art method which was emerged by the computer technology can be used as a relief to the existing complexes in the field of language teaching. Over the changing in century, the computer technology factories started to develop language learning software due to the demand of new language learning importance and new language learning

preference. The former as Fromkin et al. stated in 2009 is all about speaking English as the glob lingua franca. Due to this fact and spreading English as the world communication more people tended to learn English language. The later, new language preference is all about the change in the way of learning new language. In the world that new technology can be seen in all aspects of human lives then you can not deprive language learning and language teaching of it so computer technology easily and may be by its own self entered the world of language learning. The recent advances in education applications of computer hardware and software have provided a rapidly growing store of resources for language classes.

The computer assisted learning (CALL) are flourishing with such a high speed that it is almost impossible for language instructors to keep up with them and CALLs are always a step ahead. According to Jamieson & Chapelle 2010 CALL materials are intended to b[illegible]ive and beneficial for learners, and publishers tend to claim that their materials succeed in achieving those goals. Other like Brown 2007 stated that instructors shouldn't [illegible]he allure of computer-based technology fool them into thinking that computers will m[illegible]ically make their students happy and successful. There are also other terms in accordance and some [illegible]s in conflict with the term CALL. Among them we can first men[illegible] computer-mediate communication (CMC) which was being more widely accepted w[illegible]n practitioner integrating computer technology into language learning methodology (E[illegible]rt, 2005). [illegible]e second most accepted term after CMC is technology-mediated languag[illegible] lea[illegible]g (TMLL). As the term speaks for itself, computer is not all technology we can [illegible]se in our [illegible]ssrooms and there are lots of other technologies that can be used (Kern, 200[illegible], p.185). In ver[illegible] simple terms, speech synthesis is the process of making the computer talk (Handly, 2008).

Unlike other methods of providing the com[illegible]ter with a voice, such as the digital recording of human speakers , text -to-speech (TT[illegible] synth[illegible]sis systems, which generate speech from text input, have the unique [illegible] to generate speech models , which can be exploited for the provision of talking text fac[illegible]ities [illegible]mel, 2003a), the automated generation of exercises with spoken language support (d[illegible]Pijper, [illegible]997), and the generation of feedback (Sherwood, 1981) and conversati[illegible] turns (Egan and LaRocca, 2000) on demand to unanticipated learner interacti[illegible]ns. Y[illegible], the u[illegible] of TTS synthesis in Computer-Assisted Language Learning (CALL) [illegible]not [illegible]idely accepted (Egan and LaRocca, 2000; Sobkowiak, 1998) and t[illegible]e numbe[illegible]f commercial applications which put together TTS is quite restricted (Handley [illegible]d Hamel, [illegible]5). One possible reason for this is that the suitability and advantages of t[illegible]e use [illegible]f TTS synthesis in CALL have not been proven. One way in which this can [illegible]e achieved [illegible]through evaluation. In this study we have just worked on phonological aspects of reading and comprehension wasn't assessed. we investigated the use of TTS in tea[illegible]ing Word Stress: the stressed syllables in every content words (Fromkin, et al,2009) Word I[illegible]nation: the intensity of producing a word, Pitch Contour: The intonation of a sente[illegible] (From[illegible]et al,2009), and Total Fluency: the relative easiness in reading.

1. 1. L[illegible]rature Review

CAL[illegible]pplications integrating speech technology have emerged from the general need in language learning and teaching for ‘‘self-paced interactive learning environments” which provide ‘‘controlled interactive speaking practice outside the classroom” (Ehsani and Knodt, 1998, p. 45). Though little heard of in CALL until recently, it was identified that TTS synthesis could play a role in responding to this need over twenty five years ago (Sherwood, 1981). Specifically, Sherwood made the observation that typing /editing text is easier than recording voice and that navigating through a textual database is easier than retrieving recorded samples from an audiotape. He also observed that TTS synthesis has the capacity to generate speech models on demand, and that this capacity could be exploited in CALL to

provide learners with personalized feedback. A decade or so later, the same advantages were again put forward, this time by the technology specialists themselves (Dutoit, 1997; Keller and Zellner-Keller,2000). They saw TTS synthesis as an ‘‘indefatigable substitute native speaker” (Keller and Zeller-Keller, 2000, p.111), which because it is not human is perceived as non-judgmental. It has been suggested that the advantages of TTS synthesis presented above could be exploited in a number of different CALL applications.

Regarding the evaluation of TTS synthesis for use in CALL, different operational contexts often impose different requirements and therefore require different methods of evaluation (Sparck, Jones and Galliers, 1996). Applications in which TTS synthesis assumes the role of a reading machine include: talking dictionaries, talking texts and dictations.

A talking dictionary is an electronic dictionary which integrates either digital recordings of human speakers or speech synthesis for the oral presentation of dictionary entries. The experimental pronunciation tutor SAFexo , a module of the CA LL system SAFRA (System d’Apprentissage du FRANcais; Hamel, 1998, 2003a), focuses on this kind of practice. An example of a CA LL application that uses TTS synthesis in the teaching of prosody is Mercier et al.’s (2000) prosodic tutor for Breton. Examples of spoken dialogue systems which integrate TTS synthesis that are currently being developed for use in language learning include the Let’s Go Spoken Dialogue System (SDS) (Raux and Eskenazi, 2004) and SCILL (Spoken Conversational Interaction for Language Learning) system (Seneff et al., 2004).

Our review of the literature reveals that very few ‘‘formal” evaluations of TTS synthesis for the specific purposes of CA LL have been con ducted (Stratil et al., 1987a; Stratil et al., 1987b; Cohen, 1993 ; Santiago- Oriola, 1999; Hincks , 2002). Moreover, general purpose tools for the evaluation of speech synthesis systems such as the ITU-T Overall Quality Test (Schmidt-Nielsen, 1995; van Bezooijen and van Heuven , 199 7) which is exploited in the Blizzard Challenge (Bennett, 2005; Black and Tokuda, 2005), a speech synthesis comparative evaluation campaign do not address some of the criteria which are believed to be important for language learning applications, such as naturalness, expressiveness and register. Regarding evaluations of TTS synthesis for the specific purposes of CALL, identification of the potential benefits TTS could bring to CALL could be considered to fulfill the function of basic research evaluation. However, regarding the next stage of evaluation recommended by Handley and Hamel (2005), namely technology evaluation, only one report of an evaluation of the adequacy of TTS for use in CALL was found in the literature. In this study, we are going to investigate different effect TTS on reading fluency. Among the most important aspects we can mention Word Stress, Word Intonation, Pitch Contour, and Total Fluency.

2. METHOD

2.1. Research Hypothesis

Does using TTS Software for intermediate EFL students improve their reading quality such as, Word Stress, Word Intonation, Pitch Contour, and Fluency?

2. 2. Participants

For the purpose of this study, Azad University of Ghorveh was chosen as the context due to this university facility in computer assisted language learning (CALL). This university has equipped with a big language laboratory with 40 computers. On these computers there have been different CALL material installed.

A total of 83 students of accounting all male and ranging in age from 22 to 25 were recruited to participate in an English Reading Program (ERP) as summer free credit course. Prior to course start, a placement test was conducted to rank students. It should be stated that scoring procedure was done discretely in which scoring was done based on the scales that pertain to Word Stress (WS), Word Intonation (WI), Pitch Contour (PC), and Total Fluency.

More information would be prepared in material section about this scoring procedure. Based on the achieved result 46 students were ranked as intermediate, 23 students were ranked as high intermediate, and 14 students were ranked as low intermediate in English reading based on the scales have been mentioned. A questionnaire then was held to find out about students current English program and English exposure.

The results showed that 10 out of 46 students were studying English in some English institutes. For the purpose of removing intervening factors, these 10 students were put away. Finally 36 students of accounting were entered the CALL English reading.

2. 3. Materials

Fig. 1 shows the IVONA soft ware.

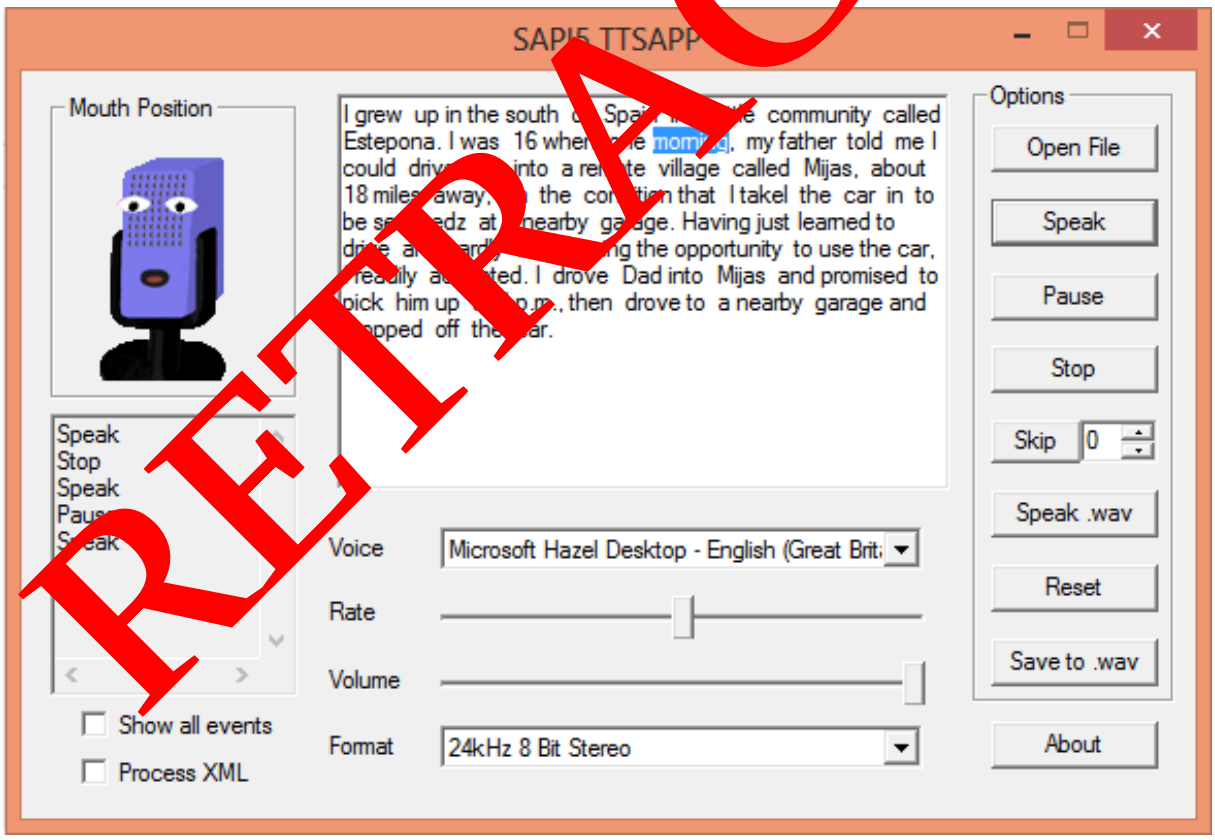

Fig. 1. IVONA Software.

IVONA UK Brain 1.4.21 was used in this research study. This TTS software is among the newest brand of converting text to speech without any limitation in the length of the texts. It is also providing students with four voice two with British accent and two with American accent. The program also has different rates of speaking production. Due to the level of the students we selected 24 KHz 16 Bit Stereo.

This TTS software gives the facility of copying any text in it to be produced in oral form. So we decided to work on Select Reading Intermediate by Linda Lee and Erik Gundersen. Students of the TTS class were given IVONA UK Brain software and Select Reading Intermediate PDF to install them on their personal computers for further activities and exercises. We also used a discrete test of reading to rank students. It is a quite comprehensive test which was rated by a quite comprehensive reading scales like word stress, word intonation, pitch contour, and total fluency. Table one illustrates these scales.

Table 1. Guideline for Assigning Reading Rating.

Scale	Point	Behavioral statement
Word Stress	6	Phonemically acceptable word stress throughout
(WS)	5	Few phonemic word stress errors but never making reading problematic
	4	Occasional phonemic word stress errors necessitate attentive reading
	3	Frequent phonemic word stress errors require demands for repetition
	2	Constant phonemic word stress errors make reading very bad
	1	Severe errors make understanding impossible
Word	6	Acceptable word intonation throughout
Intonation	5	Few word intonation errors but never making reading problematic
(WI)	4	Occasional word intonation errors necessitate attentive reading
	3	Frequent word intonation errors require demands for repetition
	2	Constant word intonation errors make reading very bad
	1	Severe errors make understanding impossible
Pitch	6	Acceptable pitch contour throughout
Contour	5	Few pitch contour errors but never making reading problematic
(PC)	4	Occasional pitch contour errors necessitate attentive reading
	3	Frequent pitch contour errors require demands for repetition
	2	Constant pitch contour errors make reading very bad
	1	Severe errors make understanding impossible
Fluency	6	Fluent and effortless speech like a native speaker
(FL)	5	Natural and continuous speech with pauses at unnatural point

4	Fluent speech with occasional problems
3	Frequent problems hinder fluency and demand greater effort
2	Slow speech, hesitant, and sometimes silent
1	Virtually unable to make connected sentences

Table 2 shows the weighting of each point. To obtain a reader total fluency score, the rating on each of the four scales-averaged for the three teachers- are transformed into values in the weighting table.

Table 2. Weighting Table.

Rating Point	1	2	3	4	5	6
Word Stress	3	5	10	15	20	25
Word intonation	3	5	10	15	20	25
Pitch Contour	3	5	10	15	20	25
Fluency	3	5	10	15	20	25
					Total	……

Suppose, for instance, that a reader receives three in word stress, four in word intonation, four in pitch contour and three in Fluency. These ratings are worth a score of 50:

Table 3. Example of Rating Procedure.

Scale	Rating	Weighted
Word Stress	3	10
Word intonation	4	15
Pitch Contour	3	10
Fluency	4	15
	Total	50

For the final thing here to mention, we used SPSS version 16 to analyze data.

2. 4. Procedure

The study was conducted in a period of three months in 2012 summer. The program was divided into two sections. A-one-month class period-July 2012- for making students familiar with (1) different aspects of English reading features such as Word Stress, Word Intonation, Pitch Contour, and Fluency; for the purpose of this students went through an intensive instruction, (2) instruction on how to use CALL materials and specifically IVONA Software. During this one month period students participated in 12 sessions. Then in a-two-month period, August and September 2012, they were starting English Reading Course.

The program was 24 sessions, three days a week. From the very beginning of the course TTS Software was used in TTS class for the purpose of teaching English reading; however, the control group went through an ordinary method of teacher reading and students' repetition. In TTS class, each and every student had a computer in front of himself. The procedure of TTS class was quite interesting in which for each unit of the Select Reading the teacher read it just for once and then the text was copied into IVONA and students were required work on it. Students in TTS class were asked to work on different aspects and features, WS, WI, PC, and FL, as they were instructed in the earlier class. On the other class, the control one, the teacher read the text and then students repeated it and then students were asked to ask about the features.

Then students in both classes were asked to work on some exercises that had been prepared for the purpose of examining students' knowledge on word stress, Word Intonation, Pitch Contour, and Fluency. In both classes the students' voice was recorded for weekly progress examining of students by teachers. After each session, students in TTS class were asked to work on the text at home on their personal computers. Students in control group were also required to work on the text based on what they learned in the class. At the end of the program another test was conducted based on the procedure which the placement test had been done. Students were given a text to read and the teacher rated them based on the Table 1 Guideline for Assigning Reading Rating. The reason for the rating by three teachers was to be assured of reliability of the test. Once again it should be stated that comprehension of reading wasn't assessed in this program.

3. RESULT

For obtaining the result we used MANOVA. Table 4. shows the achieved result in this study. According to it, TTS class is significant in all English reading features, word stress p = .008; word intonation p = .006; pitch contour p = .002; and fluency p = .000. this amount of p in three out of four features show that the difference between TTS class and the control class is quite big.

Table 4. Comparison between TTS Class & Control Class in WS, WI, PC, and FL.

Variable	Mean	SD	t	df	p
Word Stress			7.81	1	.008
TTS Class	16.11	5.01			
Control Class	11.83	4.11			
Word Intonation			8.59	1	.006
TTS Class	15.16	6.5			
Control Class	9.88	4.01			
Pitch Contour			11.76	1	.002
TTS Class	16.66	4.2			
Control Class	11.66	4.53			
Fluency			34		.0
TTS Class	18.33	3.42			
Control Class	11.66	3.42			

Inspection of the means indicates th average Word Stress scores for TTS class (16.11) is significantly higher than the mea sco Control class (11.83). The difference between the mean in TTS & Control class fo Word Stress is 4.28 that clearly indicates the significance of WS learning by t of IVO NA software. The same interpretation is true about Word Intonation learni Whi TTS cl ss mean average of WI is 15.16, the mean average of WI is 9.88 tha sho WI learning through INOVA. Based on the achieved statistics the mean score Pitch Contour is 16.66 that in comparison with the mean average of PC fo Control cla 11.66 is quite high. There is also 6.67 differences between Fluency in TTS c s 18.33 and Control class 11.66 that significantly shows the better performa ce in TTS cl All in all the amount of differences in the four reading features indi te the confirmation of the hypotheses. On the comparison of TTS & Control class on Word res p = .008, Word Intonation p = .006, Pitch Contour p = .002, and Fluency p = .000 it is c tal clear that using CALL material in improving reading features is prosperou nd succ ful.

4. CO LUSION

The results of the present study support the hypothesis that CALL materials has a significant effects on Fluency features like word stress, word intonation, pitch contour, and fluency. It was found that gains in knowledge of four aspects of Reading (WS, WI, PC, and FL) tended to be larger with the use of CALL in the classrooms. At on encounter, the participants demonstrated large gains in knowledge of reading fluency, indicating that for intermediate using TTS software can increased their Total Fluency. On the whole, in this study it was cleared that Total Fluency which is a combination of reading features such as word stress, word intonation, pitch contour, and fluency could be significantly improved by

using TTS software in reading classes. The very rational reason for this happening is that when students have TTS in their personal computers at home they can easily access to a powerful source for learning how to read English texts correctly, so in their spare time and very easily they use it and take advantage of it. On the other hand, in the other class students don't have such a facility and should just learn through class and the teacher is the first and final resource for them.

Overall, the results show that word stress and word intonation aspects of reading were benefited the most from using TTS in classroom, confirming Sherwood 1981results. The results also are in accordance with the previous done researches. (Seneff et al.2004; Hamel, 1998, 2003a).

It is important to note that in this study comprehension of reading was not assessed and considered. Further research examining Reading Comprehension in CALL classes should be done to find out about using CALL in Reading Comprehension improvements.

What should be considered language planners and teachers is to use CALL in teaching reading features, Word Stress, Word Intonation, Pitch Contour, and Fluency to improve total fluency of English reading and to use CALL in classrooms till to take advantage of both learning language and language and the pedagogical context of them.

References

[1] Auralog, (2002). *Talk to me: the conversation method* (French). (Version 3.5) from Auralog <http://www.auralog.fr>.

[2] Beutnagel, M., Conkie, A., Schroeter, J., Stylianou, Y., Syrdal, A., (1999). *The AT&T next-gen TTS system.* In: Proceedings of the Joint Meeting of the ASA, EAA, and DAGA. Berlin, Germany.

[3] Black, A., Lenzo, K., (2000). *Limited domain synthesis.* In: Proceedings of the ICSLP, Beijing, China.

[4] Campbell, N., Black, A., (1996). *Prosody and the selection of source units for concatenative synthesis. In*: van Santen, J., Sproat, R., Olive, J., Hirschberg, J. (Eds.), Progress in Speech Synthesis. Springer-Verlag, London, pp. 279-292.

[5] Campbell, N., Hamza, W., Jo¨ge, H., Tao, J., Bailly, G., (2006). *Special section on expressive speech synthesis.* IEEE Transactions on Audio, Speech and Language Processing 14(4).

[6] Chapelle, C., (2001a). Innovative language learning: achieving the vision. *ReCALL* 13(10), 3-14.

[7] Conkie, (1999). Robust Unit Selection System for Speech Synthesis. In Procs. Joint meeting of ASA, EAA, and DAGA. Berlin. Germany.

[8] de Pijper, J., (1997). *High-quality message-to-speech generation in a practical application.* In: van Santen, J., Sproat, R., Olive, J., Hirschberg, J. (Eds.), Progress in Speech Synthesis. Springer Verlag, London, pp. 575-588.

[9] Ehsani, B.K., Knodt, E., (1998). *Speech technology in computer-aided language learning: strengths and limitations of a new CALL paradigm.* Language Learning & Technology 2 (1), 45-60, Retrieved January 31, 2005, from <http://llt.msu.edu/vol2num1/article3/> .

[10] Francis, A., Nusbaum, H., (1999). *Evaluating the quality of synthetic speech.* In: Gardner-Bonneau, D. (Ed.), Human Factors and Voice Interactive Systems. Kluwer Academic Publishers, Boston, pp. 63-67.

[11] Handley, Z., (2006). *Evaluating Text-To-Speech (TTS) Synthesis for use in Computer-Assisted Language Learning (CALL).* Unpublished Doc-toral Thesis. The University of Manchester.

[12] Henton, C., (2002). Challenges and rewards in using parametric or concatenative speech synthesis. *International Journal of Speech Technology* 5, 117-131.

[13] Huang, X., Acero, X., Hon, H.-W., (2001). *Spoken Language Processing: A G... Theory, Algorithm, and System Development.* Prentice Hall, Upper Saddle River, Ne... Jersey.

[14] Keller, E., Zellner-Keller, B., (2000). *Speech synthesis in language le...ning: c...llenges and opportunities.* Procs. InSTIL. University of Abertay Dundee, Dund..e, Engla..., pp. 109-116.

[15] Multitel, (2005). *eLite Documentation.* Retrieved from <ht...://www.m... titel.be/TTS/layout.php?page=eLite_doc>.

[16] Polkosky, M., Lewis, J., (2003). Expanding the MO... develop...t and psychometric evaluation of the MOS-R and MOS-X. *Internati...nal Journal of S...eech Technology* 6, 161-182.

[17] Raux, A., Eskenazi, M., (2004). *Using task-orie...d spoken dialogues for language learning: Potential, practical applicat...n and cha...s.* In: Delmonte, R., Delcloque, P., Tonelli, S.(Eds.), Proceedings of th... In...L/ICALL 2004 Symposium, Venice, Italy, pp. 147-150.

[18] Santiago-Oriola, C., (1999). *...l synthe...is in a computerized dictation exercise.* In: Proceedings of EUROSP...ECH'9..., Vol. 1, Budapest, pp. 191-194.

[19] Schroeter, J., Conkie, ... Sy..., A., Be...tnagel, M., Juka, M., Strom, V., Kim, M.-J., Kang, H.-G., Kapl...w, D., (20... *A perspective on the next challenges for TTS research.* In: Pr...cee...gs of the I...EE 2002 Workshop on Speech Synthesis, Santa Monica, Cali...nia, pp. ...1-214.

[20] Seneff, S., Wang, C., Zhan..., J., (2004). *Spoken conversational interaction for language learni...n:* P...ee...ings of the InSTIL/ICALL 2004 – NLP and Speech Technologies in Langua... Learnin... Systems, Venice, Italy, pp. 151-154.

[21] Spar... Jones, ... Galliers, J. R., (1996). Evaluating Natural Language Processing ...yste... An Analysis and Review. Springer, London.

[22] St...l, M., Weston, G., Burkhardt, D., (1987b). Exploration of foreign language speech synth...is. Literary and Linguistic Computing 2 (2), 116-119.

[23] TMA Associates, 2003. Nuance: US English. Retrieved from <http:// www.tmaa.com/tts/Nuance_USEng.htm>.

(Received 16 July 2013; accepted 20 July 2013)