BRINGING TTS SOFTWARE INTO THE CLASSROOM: THE EFFECT OF USING TEXT TO SPEECH SOFTWARE IN TEACHING READING FEATURES

by **Hussein Meihami**
Allameh Tabataba'i University,
Tehran, Iran
hussein.meihami @ yahoo.com
and **Fateme Husseini**
Sisters Campus, Kashan University,
Kashan, Iran
hosseini.fatemeh27 @ yahoo.com

Abstract

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 particular, 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 a part of the program but wasn't investigated in the study. The study indicated that word stress, word intonation, pitch contour, and fluency have significantly improved as a result of using TTS software.

Keywords: Computer-Assisted Language Learning, reading fluency, speech-to-talk software, teaching reading, text-to-speech.

1. Introduction

The use of technology in language instruction dates back, according to some researchers, to 1950s and 1960s when technology entered the classroom in the form of language laboratories (Brown, 2007). Institutions dedicated 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, 2001). Often users of language labs were able to record their own voice and later on repeated it to see its problems and consult 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 them as a salvation for the situation of less effective methods. Once again, this time with more confidence, pioneers of language teaching thought of the

state-of-the-art technology as a relief to the existing complexes in the field of language teaching. Over the 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. As Fromkin et al. (2009) stated, it "is all about speaking English as the lingua franca of the entire globe". Due to the spread of English as a means of communication more and more people tended to learn the English language.

The recent advances in educational applications of computer hardware and software have provided a rapidly growing store of resources for language classes. The area of Computer-Assisted Language Learning (CALL) is flourishing at such a high pace that it is almost impossible for language instructors to keep up with it. According to Jamieson, Chapelle and Preiss (2010), CALL materials are intended to be attractive and beneficial for learners, and publishers tend to claim that their materials succeed in achieving those goals. Other scholars, like Brown (2007), state that instructors should not let the allure of computer-based technology fool them into thinking that computers will magically make their students happy and successful.

There are also more specific areas of CALL, such as computer-mediated communication (CMC – Egbert, 2005) or technology-mediated language learning (TMLL - Keller et al. 2000, p.185). However, for the purposes of the present study, we take speech synthesis into consideration, which is the process of making the computer talk (Handley, 2006). Unlike other methods of providing the computer with a voice, such as the digital recording of human speakers, text-to-speech (TTS) synthesis systems generate speech from text input and have the unique ability to generate speech models. This can be exploited for the provision of talking text facilities (Hamel, 2003a), the automated generation of exercises with spoken language support (dePijper, 1997), and the generation of feedback (Sherwood, 1981) and conversational turns on demand to unanticipated learner interactions (Egan & LaRocca, 2000).

Yet, the use of TTS synthesis in Computer-Assisted Language Learning (CALL) is still under development (Egan and LaRocca, 2000; Sobkowiak, 1998) and the number of commercial applications which put together TTS is quite restricted (Handley & Hamel, 2005). One possible reason for this is that the suitability and advantages of the use of TTS synthesis in CALL have not been fully proven yet. One way in which this can be achieved is through evaluation. In this study we have worked on phonological aspects of reading to investigate the use of TTS in teaching word stress (Fromkin, et al., 2009): the stressed syllables in every

content words, word intonation (the intensity of producing a word), pitch contour (the intonation of a sentence) and Total Fluency (the relative easiness in reading).

2. Literature review

Research in Computer-Assisted Language Learning (CALL) has shifted from investigating if CALL is superior to non-CALL to how CALL can be used effectively in language learning (Hegelheimer, 2004). Few studies, however, have investigated how CALL software designed according to principles of second language acquisition theory can be used in authentic settings. Chapelle (2001) points out that "computer-assisted language learning research has tended to be conducted in laboratories" and that it often involved either artificial languages or languages the participants were not studying. Empirical investigations conducted within these parameters have relied more heavily on internal validity, which can be achieved more easily in laboratory settings and structured observation, than on external validity present during actual classroom use of CALL programs. Second language acquisition (SLA) theory suggests that learners need to interact with the target language to acquire it (Larsen-Freeman and Long, 1991; Pica, 1994; Chapelle, 1998). Computer programs that offer students opportunities for interaction may help learners begin to use the language effectively and draw closer to understanding how to use the language in actual environments (Harless et al., 1999). Thus, CALL offers researchers an ideal medium to investigate how students use options provided by the software in an authentic environment.

CALL applications 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 & Knodt, 1998, p. 45). Though little heard of in CALL until recently, 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 had 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 "indefatigable substitute native speaker" (Keller and Zeller-Keller, 2000, p.111), and because it is not human is perceived as non-judgmental.

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 & 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 CALL system SAFRA (Systeme d'Apprentissage du FRANcais; Hamel, 1998, 2003a), focuses on this kind of practice. An example of a CALL 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 literature reveals that very few "formal" evaluations of TTS synthesis for the specific purposes of CALL have been conducted (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, 1997) 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 the 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.

3. The study

3.1. The research question:

- Does using TTS software in intermediate EFL reading classroom can improve EFL

students' reading with regard to word stress, word intonation, pitch contour, and fluency?

3.2 Participants

For the purpose of this study, Azad University of Ghorveh was chosen as the context due to its provision of CALL facilities. This university is equipped with a big language laboratory with 40 computers, with different CALL resources installed. A total of 83 students of accounting, all male and aged from 22 to 25, were recruited to participate in an English Reading Program (ERP) as a summer free credit course. Prior to course start, a placement test was conducted to rank students. It should be stated that the 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.

Based on the achieved result 46 students were ranked as intermediate, 23 students were ranked as upper-intermediate, and 14 students were ranked as lower-intermediate in English reading based on the aforementioned scales. A questionnaire then was held to find out 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 entered the CALL English reading program.

3.3 Materials

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 also offers four voices - two with the British accent and two with the 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 enables copying any text in it to be produced in oral form. In the present study, Select Reading Intermediate by Linda Lee and Erik Gundersen were used. Students of the class were given IVONA UK Brain software and Select Reading Intermediate PDF to install them on their personal computers for further activities and exercises.

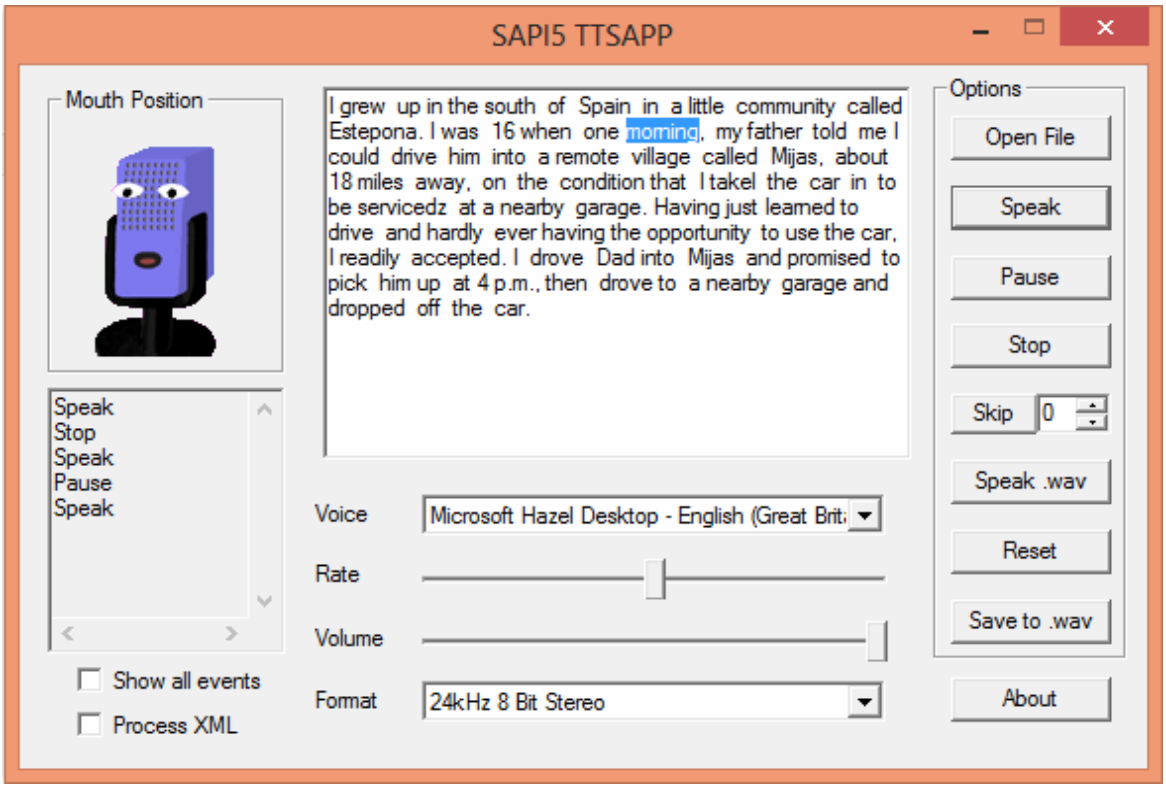

Figure 1. IVONA Software.

We also used a discrete test of reading to rank students. The test was rated by reading scales assessing word stress, word intonation, pitch contour, and total fluency, as demonstrated in Table 1.

Table 1. Scales for assessing reading.

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 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 repetition
	2	Constant word intonation errors make reading very bad
	1	Severe errors make understanding impossible
Pitch Contour	6	Acceptable pitch contour throughout

(PC)	5	Few pitch contour errors but never making reading problematic
	4	Occasional pitch contour errors necessitate attentive reading
	3	Frequent pitch contour errors require 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	**……**

3.4 Procedure

The study was conducted over a period of three months in summer 2012. The program was divided into two sections. First, a one-month class period – July 2012 – was meant to make students familiar with different aspects of English reading features such as word stress, word intonation, pitch contour, and fluency. For this purpose, students went through intensive instruction, both in how to use CALL materials and specifically the 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 the English Reading Course. The program was held in 24 sessions, three days a week. From the very beginning of the course TTS software was used for the purpose of teaching English reading; however, the control group went through an ordinary method of teacher reading and students' repetition. In the TTS class,

on the contrary, each and every student had a computer in front of himself. Each unit of the Select Reading was read once by the teacher and then the text was copied into IVONA and students worked on different parts of that text. Students in the TTS class were asked to work on different aspects and features, WS, WI, PC, and FL, as they were instructed in the earlier class. During the other class, the control one, the teacher read the text, students repeated it and then they were asked to talk about its 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 assessment by teachers. After each session, students in the TTS class were asked to work on the text at home on their personal computers. Students in the control group were also required to work on the text based on what they learned in 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 three teachers rated them based on the criteria given in Table 1. The reason for the rating by three teachers was to assure reliability of the test. Once again it should be stated that comprehension of reading was not assessed in this program.

3.5. Results

MANOVA was used to compute results. Table 4 shows the achieved result in this study. According to this table, the relationship between the variables in the TTS class is statistically significant in the two final 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 first features shows that the difference between TTS class and the control class is quite considerable.

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	1	.000
TTS Class	18.33	3.42			
Control Class	11.66	3.42			

The examination of the means indicates that the average word stress scores for the TTS class (16.11) are significantly higher than the mean score of the control class (11.83). The difference between the means in both classes for word stress is 4.28, which clearly indicates the significance of WS learning by the use of IVONA software. The same interpretation is true about word intonation learning. While the TTS class's mean average of WI is 15.16, the mean average of WI is 9.88, which shows the better WI learning through INOVA. Based on the achieved statistics the mean score for Pitch Contour is 16.66 that in comparison with the mean average of PC for Control class 11.66 is quite high. There is also 6.67 amount of difference between Fluency in TTS class (18.33) and Control class (11.66) that significantly shows the better performance in TTS class. All in all, the amount of differences in the four reading features indicate the confirmation of the hypotheses. When comparing the TTS & Control class on Word Stress p= .008, Word Intonation p=.006, Pitch Contour p=.002, and Fluency p=.000, it can be posited that using CALL material in improving reading features is prosperous and successful.

4. Discussion

Studies of teaching reading through TTS software indicated that it is "nonjudgmental" (Keller & Zellner-Keller, 2000) due to nonhuman origin of this software. According to some researchers, they are quite unnatural and are hard to be seen as a way to improve language learning with regard to reading (Sherwood, 1981). The result of this research indicated that TTS can help improve reading features. The obtained results from mean and MANOVA indicated that there are significant differences between the experimental group in which the TTS software was the medium of teaching reading and the control group with the placebo of a traditional teaching reading method.

While the mean score for word stress in the experimental group was 16.11, the control group scored 11.83, which seems to confirm the effectiveness of TTS in word stress learning of EFL students' reading. As regards the reason, TTS software may play a part in producing and showing word stress. Using this software, students always have the facility of checking every word. In this regard, the teacher is not the only source of helping students with word stress.

Word intonation showed the same result as for the effectiveness of the TTS software. The mean score obtained by experimental group was 15.16, yet it was 9.88 for the control group. Like word stress facility in TTS software, there is a section in TTS which indicates the intonation regarding to words and sentences.

Total fluency had the biggest difference among other reading mean score features. The mean score for the experimental group was 18.33; this amount was 11.66 for the control group that accounted for a 6.67 mean score difference between these two groups. The reason for this is that due to the invention of newly developed TTS software equipped with natural delivery fluency it is very helpful for students to practice this feature. In newly developed TTS software there are different rates of fluency that make it possible for students to get more flexible practice.

As a whole, the results of this investigation showed the positive effect of using TTS software in improving reading features in EFL context of the intermediate students. The contradiction of the previous studies (Sherwood, 1981) regarding the positive effect of using TTS software can be explained by the fact that those studies were done at the very beginning of TTS software invention. At that moment, this software was very mechanical and delivery facilities were not perfect. However, it is important to conduct further research to see the effect of using TTS in different areas of language learning with the help of newly created TTS software.

5. Conclusion

The results of the present study support the hypothesis that CALL materials have a significant effect on such fluency features as 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 classroom. In this respect, the participants demonstrated large gains in knowledge of reading fluency, indicating that TTS software can increase intermediate students' Total Fluency. On the whole, in this study, 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.

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

It is important to note that in this study reading comprehension was not assessed and considered. Further research examining reading comprehension in CALL classes should be done to find out the effect of using CALL on reading performance.

What should be considered by language planners and teachers is that CALL can be used to teach reading features, word stress, word intonation, pitch contour, and fluency to improve total fluency of English reading.

References

Bennett, C. (2005). Large scale evaluation of corpus-based synthesizers: results and lessons from the blizzard challenge 2005. In: *Proceedings of the INTERSPEECH*, Lisbon, Portugal, 105–108.

Black, A., & Tokuda, K. (2005). The blizzard challenge (2005): evaluating corpus-based speech synthesis on common datasets. In: *Proceedings of the INTERSPEECH*, Lisbon, Portugal, 77–80.

Brown, H., D. (2007). *Teaching by Principles*. New York: Pearson.

Chapelle, C.A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning and Technology*, 2(1), 22–34. Retrieved 12 September, 2000 from the World Wide Web: http://llt.msu.edu/vol2num1/article1/index.html.

Chapelle, C. A. (2001). Innovative language learning: achieving the vision. *ReCALL*, 23(10), 3-14.

Cohen, R. (1993). The use of a voice synthesizer in the discovery of the written language by young children. *Computers in Educations*. 21(1/2), 25–30.

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* (pp. 575–588). London: Springer Verlag.

Dutoit, T. (1997). *An Introduction to Text-to-Speech Synthesis*. London: Kluwer Academic Publishers.

Egbert, J. (2005). Theories of CALL. *CALL Essentials: Principles and Practice in CALL Classrooms*. New York: Pearson.

Egan, B., & LaRocca, S. (2000). Speech recognition in language learning. University of Abertay Dundee, Dundee, England, pp. 4–9.

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 January31, 2005, from <http://llt.msu.edu/vol2num1/article3/> .

Fromkin, V., Rodman, R., & Hyams, N. (2009). *An Introduction to Language*. Boston: Thomson/Heinle.

Hamel, M.-J. (1998). Les outils de TALN dans SAFRAN [NLP tools in SAFRAN]. *ReCALL*, 10(1), 79–85.

Hamel, M.-J. (2003). Re-using Natural Language Processing Tools in Computer Assisted Language Learning: The Experience of SAFRAN. *Unpublished Doctoral Thesis*. UMIST, Manchester. UK.

Handley, Z. (2008). Evaluating Text-To-Speech (TTS) Synthesis for use in Computer-Assisted Language Learning (CALL). *Unpublished Doctoral Thesis*. The University of Manchester.

Harless, W., Zier, A., & Duncan, R. (1999). Virtual dialogs with native speakers: The evaluation of an interactive multimedia method. *CALICO Journal*, 16(3), 313–336.

Hegelheimer, V. (2004). Using CALL in the classroom: analyzing students' interaction. *System*, 32(2), 185-205.

Hincks, R.(2002). Speech synthesis for teaching lexical stress. *TMH-QPSR*. 44, 153–165.

Jamieson, J., Chapelle, C.A., & Preiss, S. (2010). CALL Evaluation by developers, a teacher, and students. *CALICO Journal*, 23(1), 93-138.

Keller, E., Zellner-Keller, B. (2000). Speech synthesis in language learning: challenges and opportunities. *Procs. InSTIL. University of Abertay Dundee*. Dundee, 109–116.

Larsen-Freeman, D., & Long, M. (1991). *An Introduction to Second Language Acquisition Research*. LondonŁ Longman.

Mercier, G., Guyomard, M., Siroux, J., Bramoulle´ , A., Gourmelon, H., Guillou, A., & Lavannant, P. (2000). Courseware for Breton spelling pronunciation and intonation learning. *Procs. InSTIL. University of Abertay Dundee*. Dundee, 145–148.

Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes and outcomes? *Language Learning*. 44(3), 493–527.

Raux, A., Eskenazi, M.(2004). Using task-oriented spoken dialogues for language learning: Potential, practical application and challenges. In: Delmonte, R., Delcloque, P., & Tonelli, S. (Eds.). *Proceedings of the InSTIL/ICALL 2004 Symposium*. Venice, 147–150.

Santiago-Oriola, C. (1999). Vocal synthesis in a computerized dictation exercise. In: *Proceedings of EUROSPEECH* 99, 1. Budapest, 191–194.

Schmidt-Nielsen, A. (1995). Intelligibility and acceptability testing for speech technology. In: Syrdal, A., Bennett, R., Greenspan, S. (Eds.). *Applied Speech Technology* (pp. 195-231). Boca Raton: CRC.

Seneff, S., Wang, C., & Zhang, J. (2004). Spoken conversational interaction for language learning. In: *Proceedings of the InSTIL/ICALL 2004 – NLP and Speech Technologies in Language Learning Systems*. Venice, 151–154.

Sherwood, B., (1981). Speech synthesis applied to language teaching. *Studies in Language Learning*. 3, 175–181.

Sobkowiak, W. (1998). Speech in EFL CALL. In: Cameron, K. (Ed.). *Multimedia CALL: Theory and Practice.* Elm Bank, Exeter.

Sparck Jones, K., & Galliers, J.R. (1996). *Evaluating Natural Language Processing Systems: An Analysis and Review*. London: Springer.

Stratil, M., Burkhardt, D., Jarratt, P., & Yandle, J. (1987a). Computer-aided language learning with speech synthesis: user reactions. *Programmed Learning and Educational Technology*, 24(4), 309–316.

Stratil, M., Weston, G., & Burkhardt, D. (1987b). Exploration of foreign language speech synthesis. *Literary and Linguistic Computing*, 2(2), 116–119.

van Bezooijen, R., van Heuven, V. (1997). Assessment of synthesis systems. In: Gibbon, D., Moore, R., Winski, R. (Eds.). *Handbook of Standards and Resources for Spoke Language Systems* (pp. 481-563). New York: Walter de Gruyter Publishers.