LIBRARY AUTOMATION DESIGN FOR VISUALLY IMPAIRED PEOPLE

Nilüfer YURTAY[1,a], Yücel Bicil[2,b], Sait ÇELEBİ[1,c], Gülüzar ÇİT[1,d] and Deniz DURAL[1,e]
[1]Computer Engineering, Sakarya University, Esentepe Campus, 54187, Sakarya, Turkiye
[2]Tubitak-UEKAE Mulltimedia Teknolojies Lab., Gebze, Kocaeli, Turkiye
[a] + (90) 264 295 5898, nyurtay@sakarya.edu.tr,
[b] + (90) 262 648 1430, bicil@euake.tubitak.gov.tr,
[c]celebisait@gmail.com,
[d] + (90) 264 295 5908, gulizar@sakarya.edu.tr,
[e] + (90) 264 295 5895, ddural@sakarya.edu.tr,

ABSTRACT
Speech synthesis is a technology used in many different areas in computer science. This technology can bring a solution to reading activity of visually impaired people due to its text to speech conversion. Based on this problem, in this study, a system is designed needed for a visually impaired person to make use of all the library facilities in Sakarya University. Certain number of books in the library is transferred in digital media via a scanner and they are transferred into their own server. A visually impaired person can use the system with the help of sound orientation of the program and the keyboard commands of the user. System will be developed according to new requests. So, the purpose of this study is to bring a solution to the social problem of visually impaired people.

INTRODUCTION

Speech is one of the methods for providing communication between people. The production process of human speech by an external computer or device according to the phonetic expansion of text or message is called synthezing (DUTOIT, 1997). Speech synthesis can be done by adding audio tracks to each other which is stored in the audio recording database. Phonemes r systems that use audio tracks as phoneme binaries have the opportunity to synthezing all kinds of words in a small amount of record using Lego logic (DUTOIT, 1996). However, these kinds of synthesis systems are very poor for intelligibility and naturalness. In this regard, the unit selection systems that use longer pieces of audio are used more widely today (KOMINEK, 2003) (J.ZHANG, 2004).

Turkish studies are still limited even though very large numbers of systems are developed for western languages. MBROLA (DUTOIT, 1996), FESTIVAL (DUBUISSON, 2009), MULTEXT (VERONIS, 1994), GENGLISH (DUTOIT, 2005), HTS (YAMAGISHI, 2007) have been developed for synthesizing more than one language. From these systems MBROLA is adapted for Turkish and a working system is developed (BOZKURT, B., 2001).

In this study, triple sounds which is the most frequently used in Turkish and an additive synthesis system which is developed by using double voices that were not covered by triple sounds is planned to use (YURTAY, 2010). This system is a simple system that works by taking string data in digital media and adding sound pieces in sound database as Lego and it is developed in Turkish-based.

In the study of TUBITAK, frequently mentioned 3000 triple voices are determined and it is seen that these voices represent Turkish 90%. By adding 383 double voices to the list which do not exist in the triple voices, a database is developed that is formed totally from 3383 number of pieces. (BICIL, 2010).

As is known, today visually impaired people cannot go to libraries to read books and they are deprived from this social activity except books read by very little number of volunteers or Braille books. In this study, solutions are developed with the help of technology to overcome these shortcomings.

Many studies have been done to facilitate the social lives of visually impaired people, provide their training and ensure their happiness. Chen, C. And Lin, S.Y. (2011) evaluated the effects of rope jump exercise on the visually impaired students and determined a difference in the flexibility and aerobic capacity for them. Vervaart, E., Janssen, N.M., Vervloed, M.P.J. (2005) have worked on a procedure is called in-sight that has been developed to screen higher levels of visual functioning related to educational process for around twelve years old visually impaired children. Fetton,E.A., Blenkhorn, P. (1986) have mentioned technological development about educational implication of communication and necessary to meet communication needs of visual impaired in different environments. Simsek, O. Altun, E., Ates, A. (2010) have talked about the difficulties experienced by visually impaired learner during developing information and communication technologies skills and they have suggested regulation for these people to develop their skills. Sacks,S Gaylord-Ross, R. (1989) have studied about comparison of peer-mediated and teacher-directed training packages for upgrading aspects of variety of social behaviours for visual impaired students. Bayir, S., Keser, H., Numanoglu, G.(2010) have researched that through the computer literacy trainings, freedom is provided for visual impaired in Turkey. In the study by Lisi, F. (2005), some

methodologies used to encourage visually impaired at social integration in the world of work. Through these methods, they are seen more successful integrated. Owsley,C., McGwin, G., Phillps, J.M., McBeal, S.F., Stalvey, B.T. (2004) have studied over an educational program that allows to reduce rates of accidents caused by older drivers who have visual acuity deficit or slowed visual processing speed or both of them after a certain age.

In the system design, the visually impaired person who wants to take advantages of the library services after arrival in the library is directed to the designed system by a librarian. It is aimed to ensure the new book requests, book search and reading a found book from a requested page number with voice guidance and keyboard commands done by the visually impaired person.

It can be said that, visually impaired people can easily use libraries with the help of this system. The design of the system is fully applicable and after applying the processes mentioned in Section 2 and 3, it is planned to dedicate the system automatically and with the support of very few people to visually impaired people. The processes in implementation and application of the system can be examined in two main topics: Preliminary Processes and Application Stage.

PRELIMINARY PROCESSES

Hardware

In the design of the proposed system, a server and a computer with a minimum 2.53 Ghz processor, 4GB DDR2 Ram, 200GB hard disk are needed. Using the existing server to store books in digital media in the Sakarya University library is considered. The number of computers is limited to one as the initial number and then can be increased depending on the ratio of users.

Library staff must be convinced of using the system as an active and reliable way. Furthermore the system can also be used by visually impaired users who did not before. For this reason, the need for a monitor and a mouse appeared.

During the system work, visually impaired user will direct the system with an input device. At this stage, a choice must be done between two important input devices. These devices are a keyboard and a microphone. They have advantages and disadvantages among each other. In this sense, if the keyboard is selected by visually impaired person who knows to use the keyboard, it is seen to be more efficient and reliable. In the case of visually impaired person who selects to use the microphone eliminates the requirements of using the keyboard and even without using their hands he/she can manage the system. But today's speech recognition technology efficiency, most of the library environment is not completely isolated from sound and most of visually impaired person can use the keyboard. Because of all these reasons the keyboard will be preferred in this study. In addition, a scanner is required for digitization of printed documents in the library.

Software

In addition to serve for visually impaired person, the system must have the software infrastructures that must be compatible with libraries own automation systems. Thus, the proposed system and library hardware will be used more efficiently and they can be used like other computers in the library.

Paid or free software can be selected to use in the speech synthesis module. However in this proposed system, speech synthesis module that we have developed before will be used (YUCEL, 2010).

Most important parts of the system are voice guidance and management parts of the program used by the visually impaired person. At this stage, the developed software will guide the visually impaired person vocally and then management will be provided as a result of commands taken from the keyboard. A scanner will be used during the digitization of the printed documents and books in the library. While scanning the papers of the relevant document, they are converted to image format and an OCR (Optical Character Recognition) system to translate the photos into text format is needed. There is much commercial software developed to translate printed documents into digital media like Fine Reader, ReadIris, etc. For example, if Fine Reader is preferred, we can adapt easily the program to own developed system by using APIs presented. System requirements of this software can be summarized that 128 MB, 16 GB RAM for every additional processors (in case of multi- processors system), 250 MB empty disk space for typical program installation, 100 MB empty disk space for running the program, %100 TWAIN compatible scanner, digital camera or graphics card and a graphics unit (at least 800x600 resolution) with fax modem.(http://www.abbyy.com/sdk/,2011).

PREPARATION STAGE

In this stage, converting specific printed documents which do not exist in digital media into digital media process is done. The resources in the library can be simply separated into two groups:

- Digital Resources
- Non-digital Sources

Master's and doctoral theses in the library are in digital resources group. Copies of these documents in Acrobat Reader or other formats can be found usually in the digital medium. All of them are accessible for users. So, these sources can be transferred into the used system rapidly.

Non-Digital Resources are the only the resources referred to as ink printing in the library. Stories, novels, magazines, newspapers can be considered as examples of these resources. Speech synthesis can be done to the resources by translating them into digital medium with the help of a software support. In this sense, the system can be dedicated to the visually impaired users. The problem of resources that do not exist in the digital media (non-digital resources) is necessity of digitization.

Designed system aims to do this job as static firstly and then dynamically. At first stage, most popular books in the library by selecting the first 500 of them are planned for digitization with the help of hardware and software support. Transferred sources will be stored in PDF format. However, one or more people are needed to select books and then transfer in the designed system.

In this study, Microsoft MS-Project program is used to define project's activity, distribution of resource-task. Project is analyzed under the heading of scope, analysis/hardware/software requirements determination, design, development, test stage, documentation, application, dissemination and last revisions. Project's scope determination takes 3,5 days, analysis/hardware/software requirements determination takes 12,5 days, design of suitable and functional environment for the library and obtaining permits takes 7,5 days, development part includes supplying the using software and integrating this to the system takes 30 days, testing of the system takes 4 days, training process takes 6 days, preparing the help documentation takes 18 days, application takes 7 days, dissemination process takes 3 days and last revisions take 3 days. As a result, the estimated time opening the system to use was calculated and found 94,5 days approximately. Project management designed by using MS-Project program is shown in Figure 1.

Many people work in the designed project. They are management part for determining the scope, project manager for resource assignments, choice of software/hardware and following the project, analyst for the design of suitable and functional environment for the library and obtaining permits, developments for the software, tester for testing the system, trainer, technical service for the documentation process, and distribution team for the user opinions work. Resource assignments are made using MS-Project program is shown in Figure 2.

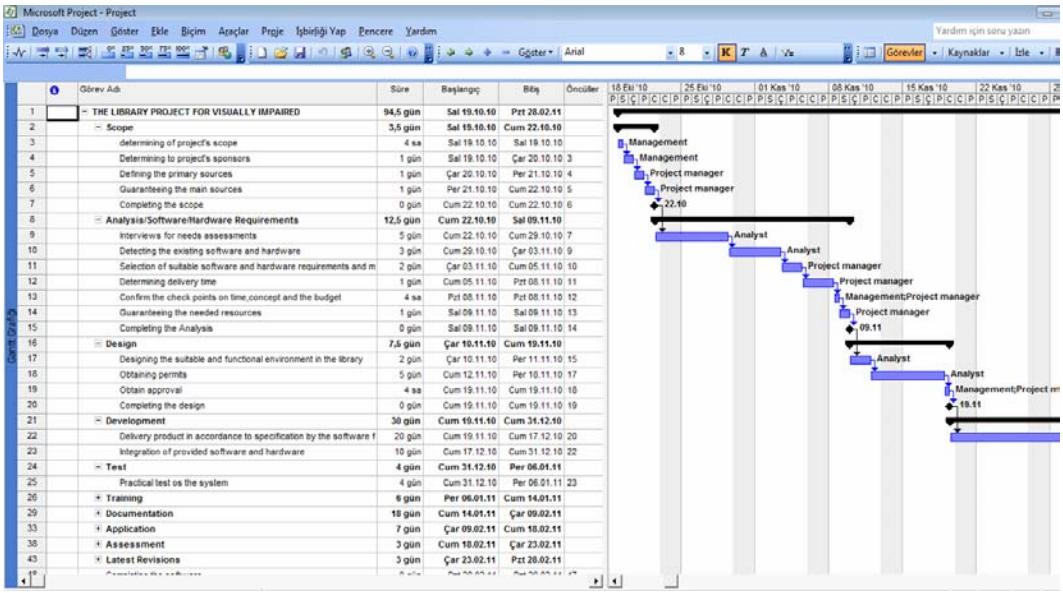
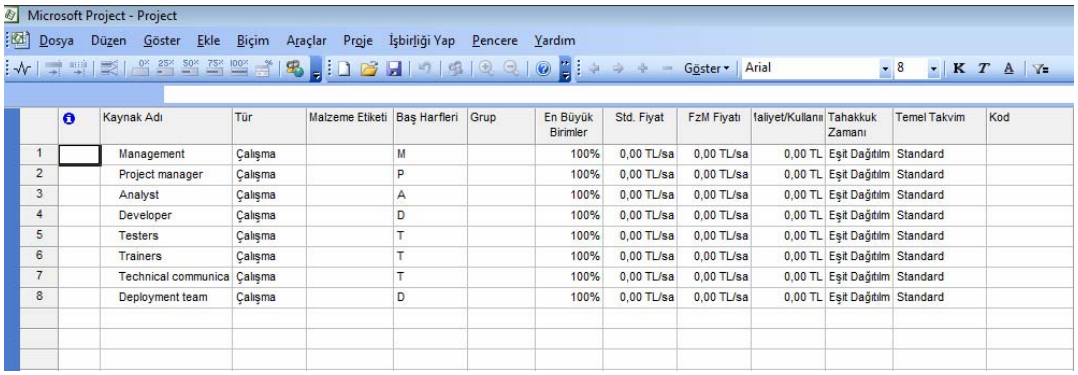


Figure 1: MS-Project work breakdown structure



	Kaynak Adı	Tür	Malzeme Etiketi	Baş Harfleri	Grup	En Büyük Birimler	Std. Fiyat	FzM Fiyatı	faliyet/Kullanı	Tahakkuk Zamanı	Temel Takvim	Kod
1	Management	Çalışma		M		100%	0,00 TL/sa	0,00 TL/sa	0,00 TL	Eşit Dağıtılm	Standard	
2	Project manager	Çalışma		P		100%	0,00 TL/sa	0,00 TL/sa	0,00 TL	Eşit Dağıtılm	Standard	
3	Analyst	Çalışma		A		100%	0,00 TL/sa	0,00 TL/sa	0,00 TL	Eşit Dağıtılm	Standard	
4	Developer	Çalışma		D		100%	0,00 TL/sa	0,00 TL/sa	0,00 TL	Eşit Dağıtılm	Standard	
5	Testers	Çalışma		T		100%	0,00 TL/sa	0,00 TL/sa	0,00 TL	Eşit Dağıtılm	Standard	
6	Trainers	Çalışma		T		100%	0,00 TL/sa	0,00 TL/sa	0,00 TL	Eşit Dağıtılm	Standard	
7	Technical communica	Çalışma		T		100%	0,00 TL/sa	0,00 TL/sa	0,00 TL	Eşit Dağıtılm	Standard	
8	Deployment team	Çalışma		D		100%	0,00 TL/sa	0,00 TL/sa	0,00 TL	Eşit Dağıtılm	Standard	

Figure 2: MS-Project resource assignments

APPLICATION STAGE

After preparation stage, while the designed system is continuing to serve to visual impaired people; it will continue to evolve according to their wishes. In this stage, system will not need an active care like in preparation stage. Only the new requested ink printing books will continue to transfer to system. In this way, the system will run more efficiently.

PERFORMANCE OF THE SYSTEM

The synthesis duration of TTS module in performed with hardware requirements is shown in Table 1 and Table 2 below.

Table 1: Test data for our TTS Module

Text No	Text
1	Istanbul
2	Cehalet erdemdir.
3	Yaratıcılığın yüzde doksanı terlemektir.
4	Sanatsız kalmış bir milletin hayat damarlarından biri kopmuş demektir.
5	Metinden Konuşma Sentezi çoklu-katman işlemleri içerir. Önemli bir ön katman metinin tamamen harflerine ayrılarak "normalize" edilmesidir; kısaltmaların tam metin karşılıklarıyla değiştirilmesi, tire ve belirsiz noktalamanın temizlenmesi, sayıların harflere dökülmesi ve aksanların uygun sembollerle değiştirilmesini içerir. Bu ön işlem dil bağımlıdır ve her dil için gramerine, yazılışına ve sözlüğüne dayanan özelleşmiş kurallar gerekir. MTRD bu amaçla Türkçe için bir ön-işlemci geliştirmiştir.
6	Toplumumuzun yaşam kalitesinin artmasına ve ülkemizin sürdürülebilir gelişmesine hizmet eden, bilim ve teknoloji alanlarında yenilikçi, yönlendirici, katılımcı ve paylaşımcı bir kurum olma vizyonunu benimseyen TÜBİTAK, akademik ve endüstriyel araştırma geliştirme çalışmalarını ve yenilikleri desteklemek, ulusal öncelikler doğrultusunda Araştırma-Teknoloji-Geliştirme çalışması yürüten Ar-Ge enstitülerini işletme işlevlerinin yanı sıra, ülkemizin Bilim ve Teknoloji politikalarını belirlemekte ve toplumun her kesiminde bu farkındalığı artırmak üzere kitaplar ve dergiler yay-ınlamaktadır.
7	Özgürlük ve bağımsızlık benim karakterimdir. Ben milletimin en büyük ve ecdadımın en değerli mirası olan bağımsızlık aşkı ile dolu bir adamım. Çocukluğumdan bugüne kadar ailevî, hususî ve resmî hayatımın her safhasını yakından bilenler bu aşkım malumdur. Bence bir millete şerefin, haysi-yetin, namusun ve insanlığın vücut ve beka bulabilmesi mutlaka o milletin özgürlük ve bağımsızlığına sahip olmasıyla kaimdir. Ben şahsen bu saydığım vasıflara, çok ehemmiyet veririm. Ve bu vasıfların kendimde mevcut olduğunu iddia edebilmek için milletimin de aynı vasıfları taşımasını esas şart bilirim. Ben yaşabilmek için mutlaka bağımsız bir milletin evladı kalmalıyım. Bu sebeple milli bağımsızlık bence bir hayat meselesidir. Millet ve memleketin menfaatleri icap ettirirse, insanlığı teşkil eden milletlerden her biriyle medeniyet icabı olan dostluk ve siyaset münasebetlerini büyük bir hassasiyetle takdir ederim. Ancak, benim milletimi esir etmek isteyen herhangi bir milletin, bu ar-zusundan vazgeçinceye kadar, amansız düşmanıyım.

Table 2: Test times for our TTS Module

The initialization time of the TTS Module: 0.131243						
Text No	**Text Normalization**	**Selection of Logotoms**	**Creating the Wave File**	**Audio Player Initialization**	**Total Synthesis Time**	**Generated Audio File Time**
1	0.000001	0.000082	0.000054	0.000005	0.000141	0.850341
2	0.000001	0.000140	0.000089	0.000003	0.000232	1.290039
3	0.000002	0.000300	0.000224	0.000003	0.000529	3.228515
4	0.000002	0.000515	0.000336	0.000003	0.000855	5.57373
5	0.000017	0.003525	0.002086	0.000004	0.005632	39.115234
6	0.000010	0.004233	0.002174	0.000004	0.006421	47.38208
7	0.000015	0.007109	0.003109	0.000015	0.010248	80.46997

As shown in Table 2, with the number 7 texts has 138 words and has been synthesized approximately 0.010248 seconds totally. Based on this, if we assume that one book has approximately 500 words in each page, approximately 0.037130 seconds are needed to synthesize 500 words of one page.

Text Normalization Process and the Challenges of Mathematical Notations

In this study, one of the problem is faced during the synthesizing speech is Turkish non-text format mathematical notation and images. One of the processes to be done is text normalization before synthesizing speech, for example number of 269 as two hundred and sixty nine to read as. Normalization can be used for using of some mathematical notations easily. Such as "% = percent", "°C = Celsius degree", " a^3 = a cube", "$\sqrt{7}$ = square root of seven ".

However, normalization of longer mathematical notation is more difficult. Therefore, creation of clear, understandable and simple standard is necessary. After a standard is created, training and promotion are required for visual impaired. As a result of these, long and complex mathematical notations will be understood in sound format by visual impaired easily.

DISCUSSION AND RECOMMENDATIONS

If a server has been installed in a library and the system is integrated with this server, each computer do not have to require individual large hard disks and by this way hardware requirements can be minimized. So that, many people might be able to use the system at the same time.

Although speed and clarity of our developed speech synthesizer is enough, the need of natural speech synthesizer is great. Because the concentration and productivity of visual impaired people may fall in the face of monotonous speeches. However, studying of natural speech synthesis is still continuing and this is very difficult field of study.

If the system is adapted over the internet, visual impaired do not need to come to the library for requesting a new book. And this way there can be more effective and easier usage. Especially, in the digital medium data is suitable for this usage.

CONCLUSION

In this study, a simple and working system was designed needed for a visually impaired person to make use of all the library facilities in Sakarya University. A visually impaired person can use the system with the help of sound orientation of the program and the keyboard commands of the user. When the specified requirements are provided, a system can easily be established in the library, so visually impaired people can benefit facilities of the library. Also a standard is needed for longer mathematical notations.

ACKNOWLEDGEMENT

Thanks to TUBITAK-UEKAE Multimedia Technologies Laboratory.

REFERENCES

Bicil,Y. (2010). "*Türkçe Metinden Konuşma Sentezleme*", Yüksek Lisans Tezi, Sakarya üniversitesi Bilgisayar Mühendisliği.

Bozkurt, B., Dutoit, T. (2011). "*An implementation and evaluation of two diphone-based synthesizers for Turkish*", Proc. 4th ISCA Tutorial and Research Workshop on Speech Synthesis, (pp.247-250), Blair Atholl, Scotland .

Chen, C. and Lin, S.Y. (2011*). "The impact of rope jumping exercise on physical fitness of visually impaired students"*, Research in Developmental Disabilities Volume 32, Issue 1, (pp 25-29) .

Dubuisson,T., Dutoit, T., Gosselin, B., Remacle, M. (2009). "*On the Use of the Correlation between Acoustic Descriptors for the Normal/Pathological Voices Discrimination*", EURASIP Journal on Advances in Signal Processing, Analysis and Signal Processing of Oesophageal and Pathological Voices.

Dutoit, T. (1997) High Quality Text- to- Speech Synthesis: An overview, Journal of Electrical & Electronics Engineering, Australia: Special Issue on Speech Recognition and Synthesis, 17, 1, (pp. 25-36).

Dutoit, T., Cernak, M. (2005). "*TTSBOX: A Matlab Toolbox For Teaching Text-To-Speech Synthesis*", Proc. ICASSP'05, Philadelphia, USA.

Dutoit, T., Gosselin, B. (1996). "*On the use of a hybrid harmonic/stochastic model for TTS synthesis by concatenation*", Speech Communication, Elsevier Publisher, 19, (pp. 119-143).

Dutoit, T., Pagel,V., Pierret, N., Bataiile, F., Van Der Vrecken, O. (1996). "*The MBROLA Project: Towards a set of high quality speech synthesizers of use for non commercial purposes"*. The 4th International Conference on Spoken Language Processing, Philadelphia, PA, USA,(pp. 1393-1396).

Fetton, E.A., Blenkhorn, P.(1986). " *Computer assisted communication in the education of the visually impaired",* Computers & Education, Volume 10, Issue 1, (pp 29-33).

http://www.abbyy.com/sdk/,2011.

Kominek, J., Black, A.W. (2003). "*CMU ARCTIC databases for speech synthesis"*. CMU-LTI-03-177. Language Technologies Institute, School of Computer Science, Carnegie Mellon University.

Sacks, S., Gaylord-Ross, R. (1989). " Peer-Mediated and teacher-directed social skills training for visually impaired students" , Behavior Therapy, Volume 20, Issue 4, (pp 619-640).

Bayir, S., Keser, H, Numanoglu G. (2010). *" General review on computer literacy of visually handicapped individuals in Turkey",* Procedia - Social and Behavioral Sciences, Volume 9, (pp 1475-1480).

Lisi, F. (2005). " *The visually impaired and their social integration in the field of work",* International Congress Series, Volume 1282, (pp 1158-1162).

Owsley, C., McGwin, G., Phillips, J.M., McNeal, S.F., Stalvey, B.T. (2004). *"Impact of an educational program on the safety of high-risk, visually impaired, older drivers"*, American Journal of Preventive Medicine, Volume 26, Issue 3, (pp 222-229).

Simsek, O., Altun, E., Ateş, A.(2010). "*Developing ICT skills of visually impaired learners" ,*Procedia - Social and Behavioral Sciences, Volume 2, Issue 2, (pp 4655-4661).

Veronis, J., Hırst, D., Esspesser, R. (1994). Ide, N., "*NL and Speech in Multext project"*, AAAI'94 Workshop on Integration of Natural Language and Speech.

Vervaart, E., Janssen, N.M., Vervloed, M.P.J. (2005). *"In–Sight: An assessment procedure for higher levels of visual functioning for visually impaired children* ", International Congress Series, Volume 1282, (pp 754-758).

Yamagishi, J., Zen, H., Toda, T., Tokuda, K. (2007). *"Speaker-Independent HMM-based Speech Synthesis System - HTS-2007 System for the Blizzard Challenge 2007*", Proc. Of Blizzard Challenge 2007.

Yurtay, N., Çelebi, S., Bicil, Y., Gündüz, A. B. (2010). "*Türçe Konuşma Sentezleme*", International Science and Technology Conference, Famagusta,(pp. 828-833).

Zhang, J. (2004). *"Language Generation and Speech Synthesis in Dialogues for Language Learning",* Master Thesis, MIT Engineering Faculty, (pp. 54-56).

