Acoustic Output of the Railway Information Systems for Visually Impaired Passengers

Milan Rusko, Marián Trnka, Sakhia Darjaa
Department of Speech Analysis and Synthesis
Institute of Informatics of the Slovak Academy of Sciences (II SAS)
Dúbravská cesta 9, 845 07 Bratislava, Slovakia

Ladislav Schichman
ELEN s.r.o.
Ľubochnianska 16, 080 06 Ľubotice, Slovakia

Abstract. *The Decree of the Ministry of Environment of the Slovak Republic, No.532 / 2002, art. 2.5.2, laying down details of the general technical requirements for constructions and structures used by persons with reduced mobility, states that every basic information system must be complemented by an alternative solution for providing the blind and visually impaired persons with information (for example, an informant, acoustic or tactile system, or telephone information service) and an optical system for the hearing impaired. The installation of several new information boards for the Slovak Railways was a good opportunity to introduce an automatic remotely controlled information audio output providing the same information as displayed on the information boards, to the visually impaired passengers. The architecture of the information system is presented in the paper. Several types of speech synthesizers are introduced that were candidates for the speech generation. It is explained which of them was used in the final solution and why. Potential issues of the system are pointed out and the future solution of railway information systems is discussed. The system was installed as a part of six new information tables at* Spišská Nová Ves *railway station, it is being tested and at the time this article is published, it should be already in regular operation.*

1 Introduction

1.1 Motivation

In the last years efforts are intensifying to make information more accessible to blind people and to make their orientation in the urban environment easier. The most natural solution seems to be the use of audio signals and speech announcements to supply the visually impaired with the information that is provided in a visual form to the rest of population.

1.2 Legal Status in the Czech Republic and Slovakia

Thanks to the activities of the Union of the Blind and thanks to the understanding of the responsible authorities in the Czech Republic (CZ), as well as support equipment manufacturers, the solution to making information available to the blind and partially sighted people has advanced considerably further than in Slovakia (SK). It has even been incorporated into the legislation and has been included in the Building Act. This has probably become an inspiration for implementing similar systems in SK.

The rules and obligations of building information systems with regard to the needs of blind and partially sighted citizens are in the CZ given by the Methodology of the Ministry for Regional Development of the CZ for implementing Decree of the Building Act 398/2009 Coll., "On general technical requirements ensuring barrier-free use of buildings."

The guidance systems for visually impaired can be generally devided in two main groups - information on the construction and operation of facilities in the construction site, such as lifts, and a group of information systems for the visually impaired, providing them with information on the operation of transport systems, such as departures, train arrivals, etc. The infotables must have acoustic output, as according to the Decree the basic information for public orientation must be both visual and, if it is possible, acoustic and tactile.

These systems are generally activated remotely; the information is given in a form of audio signals and voice messages. The remote control shall in CZ be provided by means of an electronic coded command receiver emitted from a distance of at least 40 m. The radio command signal frequency is 86,790 MHz for CZ (it is 87,100 MHz for SK).

In SK, the rules for information systems, with regard to the needs of the blind, are addressed only in the Decree of the Ministry of Environment of the Slovak Republic laying down details of general technical requirements for construction and general technical requirements for structures used by persons with reduced mobility (Decree no. 532/2002 Coll., Art. 2.5.2).

It states that the basic information device must be complemented by an alternative solution for providing information to a blind person (for example, an informant, acoustic or tactile system, telephone information service) and an optical system for the hearing impaired.

2 Devices for information system for the blind and visually impaired

Elements and devices for the information system for the visually impaired are designed as a complement to the existing hardware and software elements of electronic information systems for the public.

People have certainly long noticed the use of acoustic beacons for blind people at traffic lights, however, many other functions are also fulfilled today by sound beacons. The following subsections describe the most common devices used in SK and CZ.

2.1 Acoustic beacon

An acoustic (or sound) beacon is an electronic acoustic device that has two main functions for the visually impaired: orientation/navigation and information. The beacon is controlled by a signal from a command transmitter operated by the visually impaired person.

The navigation device emits a periodic sound signal that allows a person with limited vision to locate the object (for example, entrance to the building) or to receive other important information, such as the receive information about the current status of the traffic light, or the length of time remaining for safe crossing the road.

The use of beacons and information systems is already widely used in some European countries. For instance the SNCF (French National Railway Company) chose the NAVIGUEO+HIFI navigation and information devices for its train and subway stations. Okeenea won the RATP, (Parisian Public Transportation Company) tender for audio beacons and will be the audio beacons supplier for the entire Parisian metro network. More than 1,500 of these devices are going to guide visually imparaired people by the end of 2019 [1].

In his article from 2015 Harušťák informs: “in the CZ, there are 12 cities with sound-beacon equipped urban public transport. In the SK, the past two years have moved. The first swallow was the Transport Company of the City of Žilina, which in early 2014 equipped the first 39 vehicles (buses and trolleybuses) with acoustic beacons and all newly purchased vehicles will have beacons.” [2]

2.2 The TYFLOSET® System by APEX

In SK, TYFLOSET® System devices by APEX are used to control the functions of audio information systems for the blind (see Fig. 1.).

The handheld transmitter has six buttons. Buttons 1 and 2 are used in the SK for orientation on streets, in public buildings, moving staircases, sidewalks; for activating sound information terminals. Buttons 3 and 4 are reserved for public transport vehicles, buttons 5 and 6 are reserved for other applications. The 4-button command transmitter can be built directly into the white stick. After pushing the button a control signal is transmitted and the beacon answers by playing the pre-recorded sound or speech information according to the desired function.

The beacon can give a short beep that shows the visually impaired person the direction towards the beacon marked location. In buildings, the system can guide the blind to enter the building and find their way or the contact person. In public transport and railways the system can provide a line number, driving direction, and driver announcements on boarding or stops.

The sound-equipped information boards should provide voice information about the current timetable. Furthermore, it is possible to equip acoustic beacons to ensure the orientation and safe movement of visually impaired in the areas of stations and terminals [2].

The TYFLOSET® transmitter/receiver system produced bu APEX Ltd. company consists of a set of portable, mobile and static devices that serve both for acoustic and voice information and easier orientation of blind people.

It is designed as a unified system for all types of acoustic information and orientation in the Slovak Republic. This means that one command transmitter operates the blind, activating all acoustic and voice information and orientation systems. Transmitter frequency is 86.790 MHz for the CZ and 87,100 MHz for SK [3].

a) b)

Fig. 1. The TYFLOSET® System: a) pushbutton transmitter VPN 02, b) command transmitter in folding stick VPN 403. (Published by courtesy of APEX s.r.o.)

2.3 The STARMON information systems with LED boards by ELEN s.r.o.

ELEN Ltd. was established in 1991 in Prešov, Slovakia, by developers who have been especially involved in microprocessor applications, automation and robotics applications. Since its establishment, it has focused on the development and production of electronic information panels and displays.

Their most important applications range from railway station information systems providing passengers with the information on departures/arrivals of trains, information displays on tram and bus stops, special displays for hospitals, Metro (underground) information boards in Prague, through exchange-rate boards for banks, to a large-scale information board showing the state of the environment in Budapest [4].

Passenger information systems are an essential part of any modern station. Their use significantly affects the comfort and safety of passengers in public passenger transport. Most of the ELEN information boards are currently equipped with powerful LEDs as display elements.

ELEN Ltd. has been cooperating for many years with STARMON company from Czech republic, which has been designing information systems for passengers in trains, buses, and other types of public transport [5].

2.4 Automatic railway information systems with voice output

The idea of using automatic speech processing in railway information system is straightforward, as the communication via speech is the most natural one for humans. To give an example, PHILIPS has introduced their sophisticated train timetable information over the telephone that provided accurate connections between 1200 German cities, using speech recognition, speech understanding, dialogue management and voice output based on pre-recorded utterances in 1995 [6].

Similar voice-controlled system was developed in Slovakia by four academic institutions in the years 2002 to 2006 [7].

3 The architecture and hardware components of the designed system

When it was decided that the newly installed information LED boards by ELEN Ltd. Slovak Railways should be equipped with an on-demand voice-information feature for the blind, STARMON delivered a hardware solution and the Institute of Informatics of the Slovak Academy of Sciences (II SAS) designed a synthetic voice for this new feature.

3.1 The architecture of the system

A block diagram of the information system is presented in Fig. 2. The text information for the boards is sent from the STARMON Information Server via the RS 485 bus to the RS 485 interfaces both in the LED board motherboard, and the Raspberry-based mini PC. When the blind person presses the button on the TYFLOSET® VPN 02 handheld transmitter or the command transmitter in folding stick VPN 403, the control 87,1 MHz is emitted. Once this signal is recived by the TYFLOSET® receiver, the control unit of the Raspberry PC prepares message based on the actual text displayed on the board, launches the text preprocessing program and the speech synthesis program itself. The text preprocessing turns abbreviations and numeralls into full text, corrects diacritics and pronunciation. The Text to Speech Synthesizer (TTS) transforms text into a synthetic human-speech-like audio signal. This is then amplified and played via a loudspeaker.

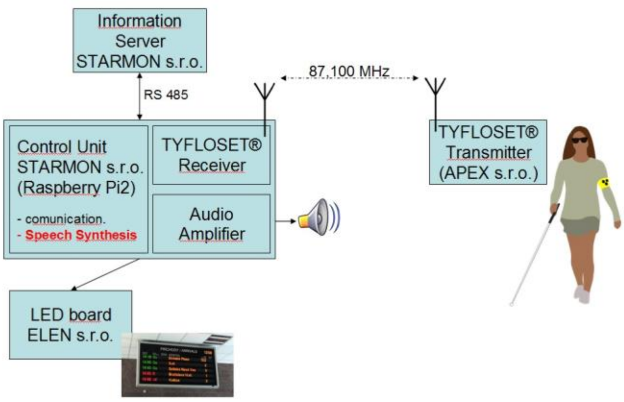

Fig. 2. The schematic diagram of the railway information system with visual (LED board) and speech (loudspeaker) outputs

3.2 The mini computer and audio hardware

The heart of the sound generating system is a Mini PC Raspberry Pi2 computer equipped with RS 485 bus communication circuits (See Fig. 3.). Raspberry Pi is a small single-boot computer whose primary operating system is Raspbian Linux.

It is equipped with Broadcom Quad-Core CPU BCM2836 with 900MHz clock, 1GB RAM, four USB 2.0 interface connectors, HDMI interface and microSD slot.

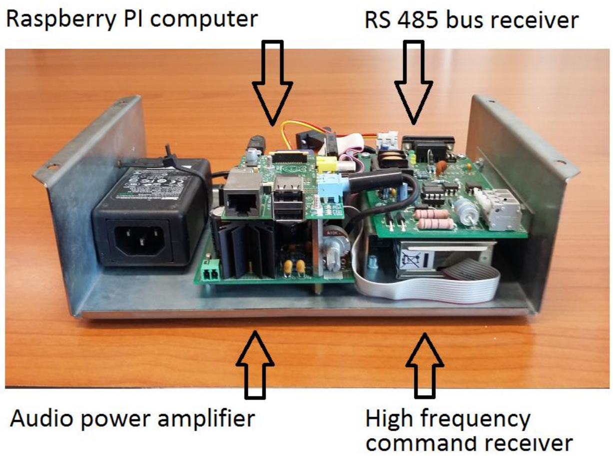

Fig. 3. Typical hardware configuration of the control unit by STARMON

The output power sufficient for driving the loudspeaker system is provided by a miniature audio amplifier mounted to the PC board.

The used AMC VIVA 4IP loudspeaker is suitable for outdoor installation in wet conditions (IP55). The IP55 category means almost complete ingress protection from particles and a good level of protection against water. The dual band speaker system provides maximum power of 20 W and sound pressure level 89 dB at 1 W of power and 1 m distance. Maximum sound pressure level (SPL) at 1m distance (at 3254 Hz) is 102 dB, and the frequency response is 90 Hz - 20 kHz. The radiation angle at 1000 Hz is 150° (horizontal) and 120° (vertical), which gives the system a good spatial coverage. Low tone speaker is 4" and the high tone has 1" in diameter.

4 Speech synthesis on the railways

Playing the acoustic signals, voice messages and announcements is the most widely used way of informing passengers on railway stations and trains (e.g. [8]).

The traditional solution was to use recorded prompts with "slot filling" with various sections of speech utterances to produce the required voice-message.

This approach is however very inflexible. It is unable to interpret new messages with unforeseen utterance structure or out-of-vocabulary words. On the other hand some years ago, the intelligibility and naturalness of the synthesized speech utterances could hardly compete with the voice messages obtained by concatenation of pre-recorded words and phrases.

4.1 The “true” TTS on the foreign railways

The alternative represented by the use of “true” speech synthesis has been currently used more and more. For the sake of simplicity, let us define the “true” speech synthesis as a system that is able to interpret – i.e. read aloud – *any* (even unknown) text in a given language with sufficient naturalness and intelligibility. Let us introduce some examples.

In the 2010 it was decided that the Swedish Transport Administration, Trafikverket will use a text-to-speech public announcement system to relay passenger information to travelers at train stations across the country. The text to speech synthetic voice was created by Acapela Group [9], [10].

The TextSpeak company reports, that their “TTS-EM modules have been integrated in 2017 to announce passenger information across New York City (DOT) and Los Angeles (LA Metro) for audio and ADA compliance in new smart bus shelters. Additional deployments in 2017 include 1000s of information displays across Europe including France, Germany and Scandinavia.” [11]

Hungarian developers are probably the farthest in the deployment of synthesizers in station reporting systems from nearby countries. Their system has been in operation at the largest passenger railway station of Hungary since June 2014 and has been installed for more than 60 other stations and stops [12].

4.2 TTS - The candidate systems for voice generation

There have been many technical approaches in the history of modern speech synthesis, that were successfully tried out. And most of them we have tried in our systems too. (Sorry, we skip the historical experiments, like Volfgang von Kempelen’s speaking machine or even Homer Dudley's vocoder.)

4.2.1 Formant synthesis - eSpeak

One of them was based on the role of the speech spectrum resonances – formants in the human ability to identify various phonemes. Using two simple types of excitation signals and formant filters, Klatt was able to design a formant synthesizer [13] creating an intelligible speech.

The STARMON company has delivered their sound hardware equipped with eSpeak Slovak voice.

eSpeak is a compact open source software speech synthesizer which uses a "formant synthesis" method. This allows the TTS to be provided in a small size. The speech is relatively clear and needs only a short time to produce utterances, but is not as natural or smooth as the synthesizers based on newer approaches of concatenation of the elements of pre-recorded human speech [14]. As the Department of Speech Synthesis and Analysis of the II SAS have been working on speech synthesis in Slovak since 1989 and have developed several synthesizers of much higher quality, they soon eSpeak excluded from the list of potential candidates because of unnaturalness of the produced speech.

4.2.2 Concatenative synthesis - Kempelen 2.0

The expression “Concatenative synthesis” designates in general any synthesis method using concatenation of pre-recorded speech segments (e.g sentences, phrases, words, syllables, diphones, phonemes, or their parts). One of our concatenative synthesizers, Kempelen 2.0 [15], had been used in the services of the Slovak telephone operators in their SMS-to-Voice services for about fifteen years. It was a diphone synthesizer using an ovelap-add method similar to PSOLA for pitch manipulation following the CART-trees based F0 and Duration models. This synthesizer was called by the Slovak Telekom Robo-teta (Robo-aunt) for the robotic character of its female voice. This was mainly due to the small number of implementations of diphones (synthesis elements), which led to spectral monotony. The second weak point was the imperfect modeling of intonation and speech rhythm, leading to repetitive prosodic patterns sounding mechanically.

4.2.3 Unit Selection synthesis (UniSel) - Kempelen 3.0

The Unit Selection synthesis [16] is probably the most successful and most used method among the approaches using waveform concatenation algorithms.

The synthesis elements can be of different length (triphones, diphones. phones, subphones etc.). These are chosen from multiple candidates contained in a large speech database according to their phonetic, word, sentence context and to their F0, and duration.

Our Unit Selection synthesizer, Kempelen 3.0 was completely developed at the Department of Speech Analysis and Synthesis of the II SAS [15]. The CART trees, that were used in the first versions for prosody modeling were later replaced by HMM models that generate the target values of F0 and duration for every phoneme sought in the database.

A syllable was chosen as the base unit of synthesis, which contributes to the natural rhythm of the resultant speech. Unwanted artifacts at the connection points of the syllables are rare and mostly come from imperfect automatic phoneme allignement in the database.

The minimum size of the speech database is about two hours of speech recordings. This database has to be stored on the disk or uploaded in the memory. The memory footprint of this database is big and the process of reading the candidate elements is time consuming as it is not optimized for speed in the current version.

4.2.4 Statistical Parametric synthesis with Hidden Markov modeling (HMM TTS), Kempelen 4.0

The statistical parametric speech synthesis uses statistical modeling based on Hidden Markov Models (HMMs) to create estimates of F0, duration and spectral envelope (in a form of Mel-cepstrum coefficients) to drive the vocoder and generate the synthetic speech. [17]

Our “HMM speech synthesizer” was developed in 2011 [18]. It is based on HTS Speech Synthesis Toolkit [19]. The context-dependent HMM models were trained from our Slovak speech databases, as generative models for speech synthesis process. The system was supplemented by various language-specific components, such as text preprocessing, letter-to-phoneme conversion, etc.

The original version of the synthesizer uses the Mel Log Spectrum Approximation (MLSA) Vocoder [20]. Speech parameters are generated from HMMs with dynamic features, namely multi-space probability distribution HMMs (MSD-HMMs). The MLSA filter is excited using a simple impulse – random noise excitation.

Experiments and comparisons were done with HMM synthesizers using more sophisticated vocoders [21], however these were not public domain and would increase the price of the system.

4.2.5 Statistical Parametric synthesis with DNN modeling (DNN TTS), Kempelen 5.0

Recent massive increase in available computing power and memory capacity, the use of parallel computing and the use of graphics processors has led to the possibility of using different types of neural networks to model models for statistical parametric synthesis [22].

Our “Deep Neural Network (DNN) synthesizer“ was designed using the Merlin toolkit for building DNN models for statistical parametric speech synthesis [23]. It was used in combination with a front-end text processor designed at II SAS and a WORLD vocoder [24].

We found out that the amount of training speech data necessary for getting satisfactory quality of the resulting voice was highly speaker-dependedent. While it was enough to use about two and a half hours of speech of our male speaker Milan to get a reasonable naturalness and intelligibility, about ten hours was needed to create our female voice Dagmar. Further increasing the volume of training data should lead to an increase in quality, but one has to make sure that the recordings are consistent in style, recording channel, etc.

The quality of DNN voices is generally very high especially in terms of natural intonation and rhythm, and timbre of voice. However the artifacts of vocoding are still audible in a form of a slight buzz.

5 Results and discussion

As mentioned in the description of the legal status, the Decree of the Ministry of Environment of the Slovak Republic no. 532/2002 Coll., Art. 2.5.2., introduces an obligation to provide information to a blind person in an appropriate way (for example, an informant, acoustic or tactile system, telephone information service) and an optical system for the hearing impaired. This offers several alternative possibilities to the voice messages.

Providing information by a human informant is a relatively expensive solution taking into account that the number of railway stations with so called “Comprehensive services for passengers” in SK, which should provide this service is more than 60.

Tactile displays are both rare and expensive (see e.g. TeslaTouch [25]). Vidal-Verdu and co-authors present an up-to-date survey of graphical tactile displays which could be used for the visually impaired people. However most of them are research prototypes and the expenses to produce them commercially would be currently too high. Thus the goal of an efficient low-cost tactile display for visually-impaired people has not yet been reached [26].

An information system equipped with a speech output using a speech synthesizer thus proves to be one of the most appropriate solutions at present.

The authors considered the properties, possibilities and hardware requirements of five types of synthesizers – one public domain formant synthesizers and four synthesizers produced by II SAS.

The eSpeak was excluded from the list of potential candidates because of unnaturalness of the produced speech.

It was decided that despite its reliability and high speed of speech production, the Kempelen 2.0 diphone concatenative synthesizer is outdated and should not be used in the current public information system.

The Kempelen 3.0 Unit Selection synthesizer has a disadvantage of relatively slow speech generation caused by reading the element candidates from the memory. It was so impossible to use this synthesizer in the designed information system even though it produces a pleasant and natural voice.

The Kempelen 5.0 DNN synthesizer is four times slower than Kempelen 4.0 HTS synthesizer mainly due to higher higher volume of calculations needed by WORLD vocoder. DNN models are about 100 times larger than HMM and their memory requirements, as well as the time required to load them, are considerably higher too.

Therefore, a compromise was made between speech quality and the speed, and the Kempelen 4.0 parametric HMM statistical synthesizer was selected to be used in the current version of the information system.

Six new voice equipped information boards have been installed at Spišská Nová Ves railway station. The system is being tested and in the time of publication of this paper it should be already in regular operation.

Fig. 4. One of the new information boards in Spišská nova ves equipped with voice output for the blind (note the black loudspeaker mounted to the upper left corner of the board). The protective plastic film will be removed from the display after regular operation is started.

To conclude we have to mention several potential issues that have to be worked on.

The response time of Kempelen 4.0 is approximately 0,5 times realtime on the Rapberry PC II system. The current version processes the whole message and then reads in one block. It is planned that the following version will generate speech by sentences during playing the previous utterance. This will reduce the reaction time requirements significantly and enable the use of other types of synthesizers. It will also be possible to consider implementing the option of setting a higher emotional arousal, or voice effort, as is usual with warning messages, Lombard speech [27], or emotion cues [28]. Of course, in that case an emotional-speech database would have to be used to train the synthesizer [29].

The intelligibility of the output speech of the synthesizers should have been tested using standard methods, e.g. using phonetically balanced SUS test [30].

The range of the radio transmitter has to be set correctly to prevent multiple triggering and reading by several information systems simultaneously.

6 Conclusion

We introduced a new voice-equipped information system developed for Slovak Railways, that combines visual text information on LED information boards with reading-on-demand of the same text content using speech synthesis in Slovak.

Following the analysis and experiments, Kempelen 4.0 HMM synthesizer was implemented in the current version of the device. The authors hope that their product will help the blind and partially sighted passengers to obtain the needed information more comfortably.

Acknowledgment

The authors would like to express thanks to Mr. Milan Slanina, the Head of Service and Design of the STARMON Ltd., who provided the authors with all the necessary information and participated in the development of this product.

This work was supported by the Slovak Scientific Grant Agency VEGA, grant No. 2/0161/18.

References

[1] Okeenea, http://www.okeenea.com/navigueo-hifi-audio-beacon/

[2] I. Harušťák, ÚNSS, Akustické informačné systémy s diaľkovým ovládaním pre nevidiacich, in: *Mosty inklúzie 7/2015*. http://www.nrozp-mosty.sk/temy-cisla-7-2015/item/1690-akusticke-informacne-systemy-s-dialkovym-ovladanim-pre-nevidiacich.html

[3] TYFLOSET® electronic orientation and information system for the visually impaired persons, http://www.apex-jesenice.cz/tyfloset.php?lang=en

[4] ELEN s.r.o., https://www.elen.sk/

[5] STARMON s.r.o., http://www.starmon.cz/

[6] H. Aust, M. Oerder, F..Seide,and V.Steinbiss, "The Philips automatic train timetable informationsystem", *Speech Communication*, Vol. 17, 1995, pp. 249-262.

[7] J. Juhar, S. Ondas, A. Cizmar, M. Rusko, G. Rozinaj and R. Jarina, "Development of Slovak GALAXY/voiceXML based spoken language dialogue system to retrieve information from the internet." *Proceedings of INTERSPEECH* (2006).

[8] E. Klabbers, "High-quality speech output generation through advanced phrase concatenation", *Proceedings of the COST Workshop on Speech Technology in the Public Telephone Network.*, Rhodes, 1997, pp. 85- 88.

[9] ACAPELA report, http://nationalpainreport.com/swedens-railway-stations-get-new-text-to-speech-technology-for-public-announcements-885121.html

[10] ACAPELA news, https://www.acapela-group.com/news/public-transport-acapela-group-creates-custom-voices-for-trafikverket/

[11] TextSpeak, https://www.textspeak.com/first-case-studies-2-2-2/

[12] Zainkó, Csaba et al. "A polyglot domain optimised text-to-speech system for railway station announcements." *Interspeech* (2015).

[13] Dennis H. Klatt, "Software for a cascade/parallel formant synthesizer" *J. Acoustical Society of America*, 67(3) March 1980.

[14] eSpeak, http://espeak.sourceforge.net

[15] S. Darjaa, M. Rusko, M. Trnka: Three generations of speech synthesis systems in Slovakia, *Proceedings of XI International Conference Speech and Computer (SPECOM)*, 2006, pp. 297-302.

[16] A.J. Hunt, A.W. Black, Unit selection in a concatenative speech synthesis system using a large speech database, *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Conference Proceedings*, 1996, ISBN: 0-7803-3192-3.

[17] K. Tokuda, H. Zen and A.W. Black, "An HMM-based speech synthesis system applied to English." *Proceedings of 2002 IEEE Workshop on Speech Synthesis*, 2002. (2002): 227-230.

[18] S. Darjaa, et. al.: HMM speech synthesizer in Slovak. *In: GCCP 2011*, Bratislava, 2011, pp. 212-221.

[19] H. Zen, T. Nose, J. Yamagishi, S. Sako, T. Masuko, A.W. Black, K. Tokuda, The HMM-based speech synthesis system version 2.0, Proc. of ISCA SSW6, Bonn, Germany, Aug. 2007.

[20] S. Imai, K. Sumita, Ch. Furuichi. Mel Log Spectrum Approximation (MLSA) filter for speech synthesis, *Electronics and Communications in Japan* 66(2), 10-18, 1983

[21] M. Sulír, J. Juhár, M. Rusko, Development of the Slovak HMM-Based TTS System and Evaluation of Voices in Respect to the Used Vocoding Techniques. *Computing and Informatics*, 35, 2016, pp. 1467-1490.

[22] H. Zen, A. Senior, M. Schuster, Statistical parametric speech synthesis using deep neural networks, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 7962-7966.

[23] Z.Wu, O. Watts, S. King, "Merlin: An Open Source Neural Network Speech Synthesis System" *in Proc. 9th ISCA Speech Synthesis Workshop (SSW9)*, September 2016, Sunnyvale, CA, USA.

[24] M. Masanori, F. Yokomori and K. Ozawa, “WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications.” *IEICE Transactions* 99-D (2016): 1877-1884.
[25] O. Bau, I. Poupyrev, A. Israr, and Ch. Harrison, TeslaTouch: electrovibration for touch surfaces. *In Proceedings of the 23nd annual ACM symposium on User interface software and technology (UIST '10,*. ACM, New York, NY, USA, 2010, 283-292.
[26] F. Vidal-Verdu, M. Hafez, “Graphical tactile displays for visually-impaired people.” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2007, 15(1), 2007, pp. 119−130.
[27] J. Šimko, Š. Beňuš, M. Vainio, “Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue.” *In Journal of the Acoustical Society of America*, 2016, vol. 139, no. 1, 2016, pp. 151-162.
[28] M. Hric, M. Chmulik, I. Guoth and R. Jarina, “SVM based speaker emotion recognition in continuous scale.” Proceedings of the *25th International Conference Radioelektronika*, 2015, pp. 339-342.
[29] R. Sabo, J. Rajčáni, “Designing the Database of Speech Under Stress”, *Journal of Linguistics (Jazykovedný časopis)*, Volume 68: Issue 2, 4016, pp. 326–335.
[30] M. Sulír, J. Staš, J. Juhár, “Design of phonetically balanced SUS test for evaluation of Slovak TTS systems” / - 2014. In: Elmar-2014 : 56th International Symposium, Zadar, Croatia, University of Zagreb, 2014, pp. 35-38.