COMEC 2019

Online Subjective Assessment of the Speech of Deaf and Hard of Hearing Children

Varga, A. K.[1]., Illés, B[2]., Czap, L.[3]

1- associate professor, Institute of Automation and Infocommunication, University of Miskolc, Hungary, varga.attila@uni-miskolc.hu
2- professor, Institute of Logistics, University of Miskolc, Hungary, altilles@uni-miskolc.hu
3- associate professor, Institute of Automation and Infocommunication, University of Miskolc, Hungary, czap@uni-miskolc.hu

Abstract. The aim of this paper is to present the results of a two-year speech production training of hearing impaired children with the help of a Speech Assistant system developed as part of a research that has been carried out as a cooperation between the University of Debrecen and the University of Miskolc in the framework of the project called 'Basic and Applied Research for Internet-based Speech Development of Deaf and Hard of Hearing People and for Objective Measurement of Their Progress'. The project is aimed at solving basic and applied research tasks to develop an application for supporting the improvement of speech production of deaf and hard of hearing people more effectively than the methods known already. The idea of the Speech Assistant has come from an audio-visual transcoder for sound visualization developed at the University of Debrecen, and a three-dimensional head model for articulation presentation, called 'talking head' developed at the University of Miskolc. The most important aim of the research is to create a complex system to assist the speech production improvement of hearing impaired children by the visualization of speech

sound and articulation. However, it is important to note that personal contribution of specialized teachers cannot be replaced. The module performing the audio-visual transcoding required is language-independent, the talking head and the automatic assessment of utterance can be adapted to other languages. While in 2015 an earlier state of the research has been introduced in the International Scientific Conference on Advances in Mechanical Engineering, this paper is presenting new results and conclusions as well.

Keywords:speech processing, deaf and hard of hearing, subjective assessment, development of online system

1. Introduction

Children who are deaf and hard of hearing need special attention in speech teaching, thus the development of info-communication based systems that can assist them to improve their cognitive skills has great significance. Our research project has been inspired by this concept, in the implementation of which researchers from different disciplines - working in areas such as medicine, education and IT - have been involved. The theoretical basis of our research is that studies have shown that the integration of acoustic and visual modality in the human brain is optimal for producing maximum clarity [1]. The combination of acoustic signal and visual modality is proved to help speech recognition [2]. If one-modal detection is difficult, perception strongly relies on the other one [3]. Since speech production is based on perception, an obvious assumption is that visual modality will promote the beneficial effect of acquiring correct productions. Image information processing of hard of hearing people is smoother and even more experienced than that of the normal hearing people. The human brain integrates acoustic and visual signals both in hearing impaired and normal hearing

people. The poorer the quality of the acoustic signal, the more we rely on visual modality [4].

To our best knowledge there has not been any other speech enhancement project with such a long period and as large vocabulary. Similar contributions have about a couple of weeks long training period: 10 weeks [5], 21 weeks [6]. Our training period was two years long with 60 children. 30 of them have practised with visual support. The other 30 formed a control group and they've got extra lessons without the Speech assistant. Both groups had 25-minute lessons two times a week. Other speech trainings that have been published involved from a couple of dozen, up to some hundreds words vocabulary: 24 items [7], 100 words [8] 104 test words [6], 480 key words [9], 270 short Swedish everyday sentences, 138 symmetric VCV and VCC{C}V words and 41 asymmetric C1VC2 words [10]. Our application worked with 3031 words, 307 oppositional word pairs and 593 other items (see 4.1).

First of all, the speech samples of children recorded and stored in the server of the University of Miskolc had to be subjectively rated by the pedagogues and naive students involved in the project. Since the available evaluation systems had been considered unsuitable for our goals (e.g. integration with our own database having specific structure, data archiving and long-term storage), we had to develop our own client-server based application to provide user specific features, customized functions and tools for data management, as well as to allow for creating unique reports, statistics and trends for further research aims.

2. Concept of the Speech Assistant System

The research served the purpose of creating a new aid for the deaf and the hard of hearing in learning to speak, which is called Speech Assistant System. The foundation of the research is represented by the 'talking head' developed at the University of Miskolc and the audio-visual transcoder developed at the University of Debrecen. The

objective of the research is to create a complex system which provides the audio-visual representation of the speech process, by the visual representation of the sound images of speech on the one hand, and of the articulation on the other, set in a training framework system. The 3D head model with its transparent face can visualize the tongue motion better than a natural speaker. In addition, the system includes a number of functions (visualization of prosody, automatic assessment and implementation of the knowledge-based system) that facilitate individual practice not only on the computer, but also on a mobile device. The module of the technology developed performing the audio-visual transcoding is language independent, while the talking head and the automatic assessment can be adapted to other languages besides Hungarian. An example for the visualization of speech sounds and the talking head can be seen in Figure 1.

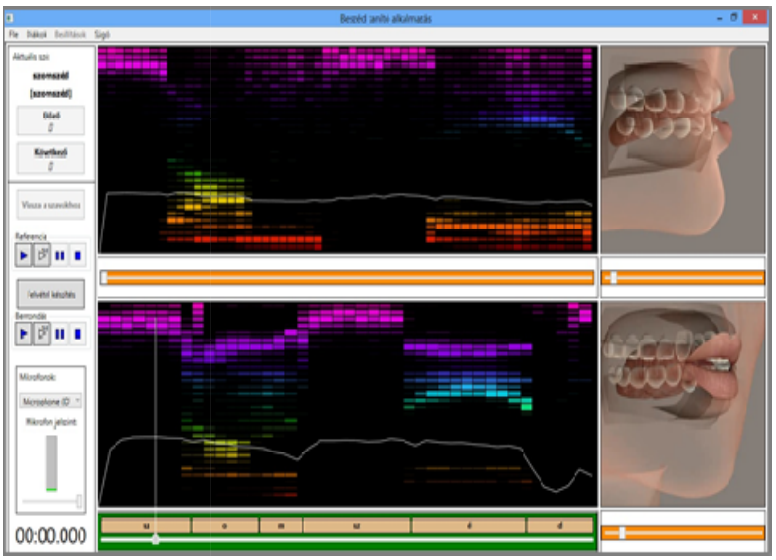

Figure 1. Practicing user interface of Speech Assistant. 45-degree-angle and 90-degree-angle views of the talking head (right), segmented visual representation of the reference speech (bottom left) and the representation of the speech sound recorded during practicing (top left).

The development of the Speech Assistant system [11] began in September 2013, with the participation of 13 teachers specialised for dealing with deaf and hard of hearing children. 60 pupils of different ages and stages of speech production were selected for the training.

Before the start of the development work the children had to tell various words and sentences that we had recorded and stored in a high capacity storage server. The same words had to be recorded again after the first and second academic years of the development work [12].

In the following, the paper is going to introduce the online system developed for evaluating the words and sentences recorded and stored on the central server of the Department of Automation and Infocommunication, University of Miskolc.

3. Development of an Online Evaluation System

In the course of learning the pronunciation of the reference, which is presented by the server or the teacher, the pupil tries to imitate it with his/her current utterance. By the analysis of different feature extraction and distance calculation methods, a similarity measure can be created in accordance with the subjective assessment method [11]. This is the basis for progress assessment and feedback generation. The evaluation can apparently be formed by comparing the earlier results, since the same pronunciation can be one pupil's success or the other's failure as well. Verification of automatic evaluation can be performed by investigating clarity, for which purpose a client-server based online system had to be developed for storing the ratings of speech samples.

The Concept of the Evaluation System

As regards the implementation of the system, a primary goal was to ensure accessibility. Thus the most obvious solution is a web-based application that provides a 24-hour 'anywhere and anytime' access (see Figure 2).

The system has been developed using a combination of PHP and MySQL that provides support for submitting and storing data in a database. The hardware and software infrastructure necessary for the operation has been provided by a central server at the University of Miskolc as follows:

- PHP module: for running PHP based program codes on server side,
- MySQL module: for centralized data storage and for performing filtering and searching operations.

PHP (Hypertext Preprocessor) is an open-source computer scripting language, which can run on any server-side operating systems in cooperation with most server programs. Its main application field is the creation of dynamic websites. [13]

Figure 2. Concept of the online evaluation system

MySQL is a very popular database management system, which is famous for its simplicity and effectiveness. Data are stored according to the relational model and standardized SQL (Structured Query Language) commands can be used for data management. [14]

Various features have been implemented through PHP functions. Users can be managed entirely by an administrator having supervisor rights. Supervisors can perform operations like: creating new user account, disabling users, deleting user account, managing user rights, changing user profile.

Structure of the Reference Database

For automatic evaluation, first of all, the speeches of hearing-impaired children have been recorded as reference data. The initial main database has been systematized by the members of a research team at the University of Miskolc using several criteria: classification of speech; topic classification; number of syllables; number of voices; vowel-consonant formula, etc.

The current database consists of exactly 2,355 words (some words occur multiple times, but the announcers and their intelligibility are different), which have been evaluated by 13 teachers and 23 naive students. The basis of the five-scale rating for evaluation is determined by the teachers. All teachers have rated only the students of other schools to avoid bias resulting from the recognition of the speakers. Results have been recorded via the web application developed for this purpose.

ratings

datafield type	datafield name
id	int(5)
speech_sample_id	varchar(5)
user_id	varchar(5)
score	varchar(10)
date	varchar(15)
remarks	varchar(1000)

indexes

key	type	unique	packed	column	cardinality	join	NULL
PRIMARY	BTREE	YES	NO	ID	21273	A	NO

Figure 3. Data table for storing evaluations

Given that the speech samples have been recorded in 3 different schools (Budapest, Eger, Debrecen) educating deaf and hard of hearing children, it was reasonable to store them in three different directories on the server (see Figure 3).

The structure and scheme of the tables within the database can be described by data fields, in which the following additional parameters have to be defined:

- type of data field: number, char, string, Boolean, date, etc.,
- length of data field: necessary number of characters for digital storing of data,
- integrity constraints bound to data fields: an inner rule system, defining the accuracy of the information stored (for example, whether it can be left blank in the data under recording, or whether it is a primary key, etc.).

The definition of data fields has been done by entering a name and selecting a type for them. Since structures do not clearly identify tables, as more tables may exist with the same structure, and on the other hand defining structures may take a long time, thus a unique identifying name has been assigned to each table within the database. This name is used to clearly identify a table during the operations. Therefore, the name of the tables within the database, and that of the data fields within a table must be unique.

Five-scale Evaluation

One of the Speech Assistant system's features is the automatic rating and feedback for the hearing-impaired pupils to practice the sample words. In the course of learning each pupil's pronunciation is compared to the reference produced by the server. For verification of clarity, the following scale has been defined by the teachers:

- Unintelligible (1): articulation is completely distorted; the vowels and consonants are unrecognizable; the reproduction of the syllable number is not adequate or discernible; breathing and management of breath is faulty; tempo and rhythm are incorrect; the utterance is unmelodious, non-dynamic or too tense.
- Difficult to understand (2): grave distortions, omission of speech sounds, speech sound replacement; only some of the vowels can be discerned; distortions due to insufficient breathing, e.g. too breathy or choked; characterized by irregular, disturbing tonality, rhythm and tempo.
- Moderately intelligible (3): the articulation of vowels is correct, the number of syllables is appropriate; serious speech defects may occur, e.g. dyslalia (the speech impediment in which certain vowels are incompletely formed), nasality, head voice, prosodic inadequacies.
- Easy to understand (4): slight speech defects; slight prosodic inadequacies.
- Understandable at the same level as the speech of the hearing (5): at most 1-2 speech sound defects may occur.

Before the evaluators submitted their scores, they could listen to the speech samples repeatedly, eliminating this way the loading problems caused by the loss of network or narrow bandwidth Internet connection. Moreover, private notes about the samples can be put into a textbox to indicate any types of problems.

Using the Evaluating System Developed

After submitting the user name and password of the account created by the supervisor on the login interface (see Figure 4), user data are loaded. When the supervisor creates a new user account, he/she specifies which speech samples have to be evaluated by the user.

The system performs a check and after successful authentication it loads the profile on the basis of the user accounts defined in the user profiles. After a successful login, the current speech sample for evaluation is loaded. The evaluation process can be interrupted at any time, and be continued until a specified deadline has expired.When the user has scored each of the samples, the system indicates it to the user, and closes his/her account. When listening to a sample, the word or sentence which is actually concerned is displayed in a textbox.

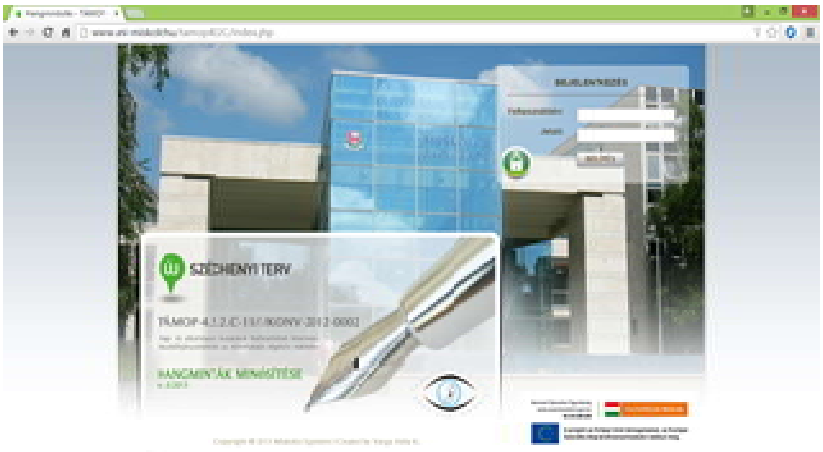

Figure 4. Login page of the evaluating system developed

The system indicates to the user how many samples he or she has already rated. Teachers in Budapest have evaluated 1,441 speech samples, teachers in Eger have

evaluated 2,043 samples and teachers in Debrecen have evaluated 1,226 samples. Lay students have had to score all the samples, which means 2,355 evaluations for a student.

The evaluation consists of the following steps:

- By clicking on the play button the current speech file loaded for rating can be listened. Each sample can be played multiple times.
- The rating of the audio file can be performed by clicking on a radio button next to the scores.
- If the evaluator wants to write notes to the speech sample currently scored, the textbox can be used for this purpose.
- The evaluation of a speech sample is completed by clicking on the 'Submission of scores', after which the next audio file isloaded for rating.

We have used the online evaluation system for two purposes:

- For automatic assessment of speech samples produced during the lessons a speech quality scale was needed as a reference. Subjective evaluators (specialized teachers and naive students) listened to 2,355 prerecorded speech samples of different level of utterance and scored them according to the five-grade scale.

For the assessment of the speech production progress during the two-year training the evaluators listened to the records of the same word or sentence recorded before and after the training period and scored the progress according to the five-grade scale.

4. Results of the Improvement of Speech Production

Some amendments to the application described in Section 3 also allow for measuring the results of the two-year improvement of speech production. There is one difference in the course of evaluation compared to the previous one. Namely, the evaluator listened to the recordings produced both at the beginning (in September 2013) and at the end (in

May 2015) of the research giving scores to the improvement according to the evaluation scale.

Introducing Speech Assistant into Education

The testing of the system started in September 2013 with 14 participating pedagogues. Four of them got methodological and organizational tasks. In the research period, each of the 10 speech therapist pedagogues worked with 6 children: 3 of them were taught with traditional methods while the other 3 used the Speech Assistant software. The pupils selected for training are of different ages and stages of cognitive and speech development, having hearing impairment to varying extent. The 60 pupils involved were distributed among the group working with the talking head and the control group evenly, considering the extent of their hearing impairment, their capabilities and occurrent joint handicaps.

Before using the system, registration is required, during which only some basic information must be entered to get a full access to use the Speech Assistant. After logging in, teachers have the opportunity on the home page to select who they want to deal with and what words the selected child has to practice (see Figure 5).

On the initial page potential new students and important notes can be registered with unique identifiers, previously registered pupils can be deleted, and saved workspaces can also be loaded. Patterns recorded during the exercise will be automatically uploaded to the server dedicated to the pupils' accounts for the purpose of carrying out further investigations and research. In case a pupil is deleted, his/her samples won't be automatically deleted from the server.

The system contains a training word database of 3031 words. These were selected by the participating pedagogues based on their previous experiences considering several aspects:

- part of speech (verb, noun, adjective, numeral, article, pronoun, conjunction, adverb);
- topic (family, school, entertainment, etc.);
- number of syllables;
- number of speech sounds;
- vowel-consonant formula.

Words with the following features got unique notation:

- there are consonant obstructions in the middle that possibly need more practice;
- there are consonant obstructions at the end;
- pronunciation significantly differs from the written form;
- having two totally different meanings (ambiguous words).

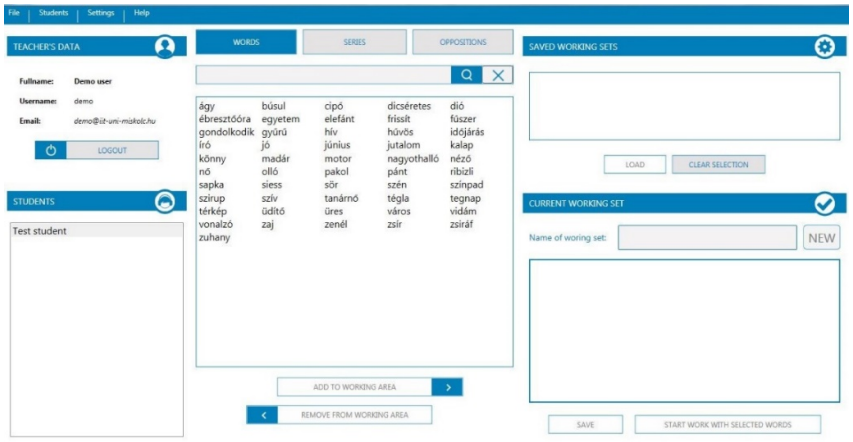

Figure 5. Word selection in the Speech Assistant system

By the end of the second year, the set of word samples are augmented with 307 oppositional word pairs (e.g. "bab-pap", "bont-pont", "lombos-lompos") and 593 fixing word sequences (e.g. "bu-be-bi-...", "ub-eb-ib-...", "ubu-ebe-ibi-...") [15].

The Course of Development

After having assessed the speech status and pronunciation of the pupils, the pedagogues personally assigned the speech elements to be improved in the next areas.

- Correction, fixing and automation of speechsounds.
- Long-shortsounds, voiced-unvoicedsoundpairs, correctpronunciation of soundsinconsonantobstructions.
- Awareness of prosodicelements.
- Improvement of speechhearing, differentiatingafterhearing.

Initially, the methodology of how the system should be used was not uniform. Pedagogues could freely decide when and how to apply the Speech Assistant system during the lessons in order to collect experiences and provide feedback for further development, as well as to work out a methodological recommendation to its use. By the end of the second year, the uniform course of a speech development lesson, i.e. the methodological steps of speech development have been specified as follows.

- According to the development plan, the words containing the sound to be practiced are determined depending on its position within the word.
- The selected word is read, listened, interpreted and embedded in a sentence to make sure that it is properly understood.
- Pupils are asked to observe the articulation pattern – that is analyzed by the pedagogue - on the talking head in both degree-angle views and also in slow motion.
- In case the sound to be practiced is incorrectly pronounced, pupils are asked to observe the correct utterance and the specific motion of the tongue and lips step by step.
- The pupil's speech is recorded.
- The recorded speech is listened and compared to the reference using the bar

charts. The mistakes are analyzed and prosody is also discussed. If necessary, the speech recording should be repeated and the changes should be examined.

- At the end the performance of the pupil is evaluated.

The Results of the Two-year Speech Development

From the 60 pupils selected initially for the speech development extra lessons 5 quitted the training, thus only the progress of the other 55 pupils could be measured. 28 of them were girls and 27 were boys. 28 of them used the talking head and 27 of them did not (control group) in the development lesson (see Table 1 and Figure 6) that lasted half an hour twice a week.

For measuring progress, speech samples of all participating children were recorded (60 words and 30 sentences) at the beginning and at the end of the two-year development. The samples were compared and evaluated by the pedagogues from other schools. When evaluating the results, we realized that the teachers have different score-ranges. Since it is a subjective test, it can be considered normal. They gave scores above the average more generally than below, therefore the outstanding values caused an increase in the averages. Consequently, the median (the middle value in the ordered list of scores) was selected for the evaluating progress. This is lower than the average, but falls closer to most of the scores. (Concerning all pupils, the average of the average scores is 1.12, while the average of the median scores is 0.86.). Figure 7 shows the progress of each pupil for the two-year development period.

Median	>1.5	1-1.5	0.5-1	<0.5
Talking head	4	9	11	4
Control group	0	9	9	9

Median	>1.2	0.8-1.2	0.4-0.8	<0.4
Talking head	11	5	9	3
Control group	2	10	10	5

Table 1. Distribution of children using talking head and those in the control group according to the rate of progress.

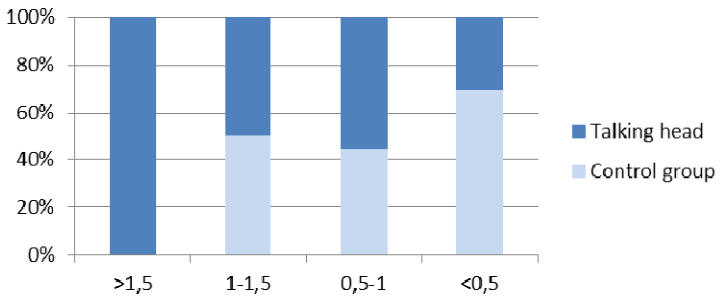

Figure 6.Distribution of children using talking head and those in the control group in ranges of progress

In the group where development was supported by the talking head the average of personal medians indicating the progress of children is 0.97, while it is 0.74 in the control group. This means that after the two-year training with the help of the Speech Assistant hearing-impaired children got one grade higher on average than at the beginning of the rating scale.

5. Conclusions

The Speech Assistant, a web-based pronunciation improving program that is still under construction proved to be a beneficial tool in the individual speech therapy of hearing impaired. Perception supported with multiple sensors improves the efficiency of teaching pronunciation in case it is optimized through the methodological steps provided by the software tool the application of which is expedient. We can conclude that the group of children achieving the most significant improvement contains mainly those who were developed with the use of the talking head. Most of the children performing the least progress were taught in the control group. According to the written assessment of the teachers the motivation and attitude of the pupils were different, however the conclusion can be drawn since the groups were composed unbiased.

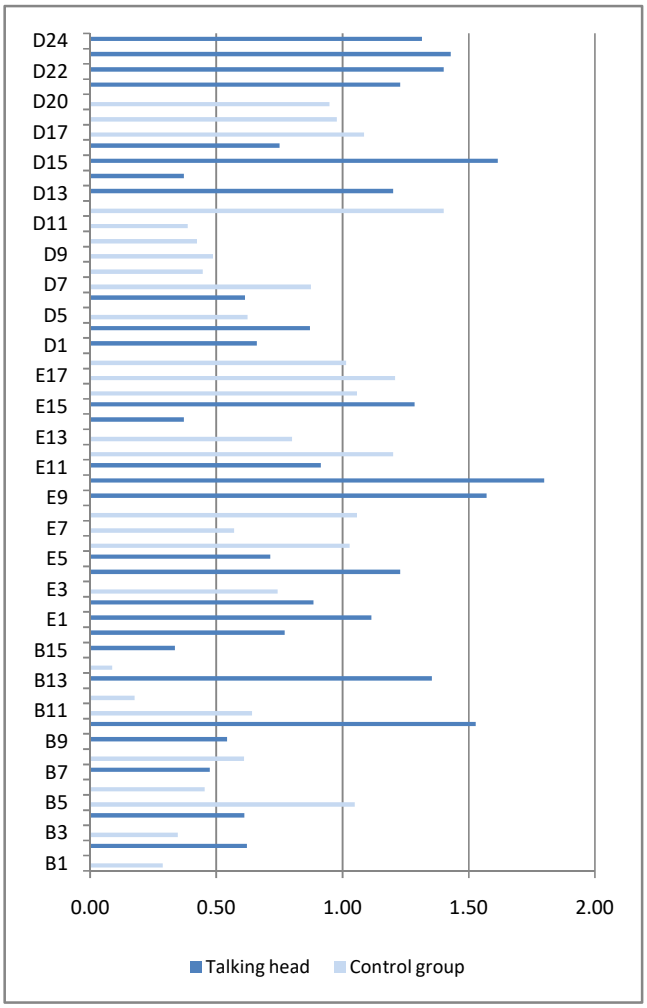

Figure 7. Detailed results of the progress of children

The motivating power of the web-based speech development program can play a significant role in drawing attention and in increasing endurance of focusing, especially because it can be supplemented with games. Children today willingly accept computer aid which makes the tiring lessons more interesting and vivid for them.

The user-friendly interface has been of high priority during the development of the system in order to interconnect the appearance and the users' expectations. Due to this goal a comfortable, easy to use system has been resulted facilitating users' work and reducing errors in data entry and data management.

Acknowledgement

The described study was carried out as part of the EFOP-3.6.1-16-00011 "Younger and Renewing University – Innovative Knowledge City – institutional development of the University of Miskolc aiming at intelligent specialization" project implemented in the framework of the Szechenyi 2020 program. The realization of this project is supported by the European Union, co-financed by the European Social Fund.

The research has been carried out within the framework of the project TÁMOP-4.2.2.C-11/1/KONV-2012-0002 funded by the European Union, co-financed by the European Social Fund.

References

[1] A. K. Katsaggelos, S. Bahaadini, R. Molina: Audiovisual Fusion: Challenges and New Approaches, Proceedings of the IEEE | Vol. 103, No. 9, September 2015, pp. 1635-1653. DOI: 10.1109/JPROC.2015.2459017

[2] D. W. Massaro, M. M. Cohen: Integration of visual and auditory information in speech perception, Journal of Experimental Psychology Human Perception & Performance November 1983, pp. 753-771. DOI: 10.1037/0096-1523.9.5.753

[3] D. W. Massaro, M. M. Cohen: Speech perception in perceivers with hearing loss: Synergy of multiple modalities, Journal of Speech Language and Hearing Research · March 1999, pp. 21-41.

[4] Hunyadi, L., Szekrényes, I., Czap, L., Sziklai, I. "Seeing the sounds?" *ARGUMENTUM* 10: pp. 335-338. 2014.

[5] D. W. Massaro, J. Light: Improving the Vocabulary of Children with Hearing Loss, The Volta review · December 2003, pp. 141-174.

[6] D. W. Massaro, J. Light: Using Visible Speech to Train Perception and Production of Speech for Individuals With Hearing Loss, Journal of Speech Language and Hearing Research ·May 2004, pp. 304-320. DOI: 10.1044/1092-4388(2004/025)

[7] D. W. Massaro, J. Light: Improving the Vocabulary of Children with Hearing Loss, The Volta review · December 2003, pp. 141-174.

[8] S. Al Moubayed, J. Beskow, A-M. Öster, G. Salvi, B. Granström, N van Son, E. Ormel, T. Herzke: Studies on Using the SynFace Talking Head for the Hearing Impaired, Proceedings, FONETIK 2009, Dept. of Linguistics, Stockholm University

[9] C. Siciliano, G. Williams, J. Beskow, A. Faulkner: Evaluation of a Multilingual Synthetic Talking Face as a Communication Aid for the Hearing Impaired, Proceedings of the International Congress of Phonetic Sciences, 2003, pp. 131-134. ISBN 1-876346-nn-n © 2003 UAB

[10] J. Beskow, O. Engwall, B. Granström, P. Wik: Visualization of speech and audio for hearing impaired persons, Technology and Disability · August, IOS Press 2008, pp. 97-107. ISSN 1055-4181

[11] Czap, L., Illés, B., Varga, A. "Concept of a Speech Assistant System." Proceedings of the 4th Word Congress on Software Engineering WCSE 2013. Hong Kong, China, 2013.12.03-04. Hong Kong: IEEE Computer Society, pp. 207-211. DOI: 10.1109/WCSE.2013.37

[12] Vicsi, K., Szaszák, GY. "Automatic Segmentation of Continuous Speech on Word Level Based on Supra-segmental Features." *International Journal of Speech Technology*, Vol. 8, Num. 4, 2005, pp. 363–70. DOI: 10.1007/s10772-006-8534-z

[13] Nixon, R. "Learning PHP, MySQL & JavaScript: With jQuery, CSS & HTML5.", ISBN 978-1491918661, 2014.

[14] Ullman, L. "PHP and MySQL for Dynamic Web Sites: Visual QuickPro Guide.", ISBN 978-0321784070, 2011.

[15] Kovács, Sz.,Tóth, Á., Czap, L. "Fuzzy model based user adaptive framework for consonant articulation and pronunciation therapy in Hungarian hearing-impaired education." Proceedings of the 5th IEEE International Conference on Cognitive Infocommunications: CogInfoCom2014 : Proceedings. Vietrisul Mare, Italy, 2014.11.05-07. Piscataway: IEEE, 2014. pp. 361-366. DOI: 10.1109/CogInfoCom.2014.7020479