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

László Czap[1*]

[1] Department of Automation and Infocommunication, Faculty of Mechanical Engineering and Informatics, University of Miskolc, H-3515 Miskolc, Egyetemváros, Hungary
* Corresponding author, e-mail: 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. It was developed as part of a research project that was carried out jointly by the University of Debrecen and the University of Miskolc within 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 that are already known. The idea of the Speech Assistant came 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 was to create a complex system to assist the speech production improvement of hearing impaired children by the visualization of speech sound and articulation. In addition, the system has many other features (such as prosody display, automatic assessment and knowledge-based systems implementation), which subsequently allow individual practice not only on computers but also on mobile devices. However, it is important to note that the 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. The online evaluation system developed for measuring the progress of children in speech production is also shown in this paper.

Keywords

speech of hard of hearing people, online evaluation system, reference speech database, speech production training

1 Introduction

Children who are deaf and hard of hearing need special attention in speech tuition, 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 all 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 proven to help speech recognition [2]. If one-modal detection is difficult then perception strongly relies on the other one [3]. Since speech production is based on perception an obvious assumption would therefore be that visual modality will promote the beneficial effect of acquiring the correct productions. Image information processing of hard of hearing people is smoother and even more experienced than that of people with normal hearing. The human brain integrates acoustic and visual signals in all people regardless of whether or not they have hearing impediments. The poorer the quality of the acoustic signal, the more we rely on visual modality [4].

To the best of our knowledge there has not been any other speech enhancement project that has covered such a long period and which has as large vocabulary. Similar projects only have training periods of approximately a few weeks; ten weeks [5], twenty-one weeks [6]. Our training period was two years long and involved sixty children.

Thirty of them had practised with visual support while the other thirty formed a control group who had extra lessons without the Speech Assistant. Both groups had twenty-five minute lessons twice a week. Other speech training that have been published involved vocabulary lists ranging from twenty-four to several hundred words: twenty-four items [7], one hundred words [8] one hundred and four test words [6], four hundred and eighty key words [9], two hundred and seventy short Swedish everyday sentences, one hundred and thirty-eight symmetric VCV and VCC{C}V words and forty-one asymmetric C1VC2 words [10]. Our application works with three thousand and thirty-one words, three hundred and seven oppositional word pairs and five hundred and ninety-three other items (see 4.1).

One of the main objectives of the project TÁMOP-4.2.2.C-11/1/KONV was to create a method for automatic assessment of speech samples.

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 we also had 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 The 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 an audio-visual representation of the speech process. This is done by the visual representation of the sound images of speech on the one hand and of the articulation of speech on the other all set in a training framework system. The 3D head model with its transparent face can visualize the motion of the tongue 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 for 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 Fig. 1.

The development of the Speech Assistant system [11] began in September 2013 with the participation of thirteen teachers specialized in dealing with children who are deaf and hard of hearing. Sixty 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 speak 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].

This 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 Info-communication at the University of Miskolc.

3 Development of an Online Evaluation System

In the course of learning the pronunciation of the reference words, which are presented by the server or the teacher, the pupil tries to imitate them using their current speech pattern. 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 generating feedback. The evaluation can apparently be formed by comparing it with the earlier results, since the same pronunciation can be one pupil's success as well as being another's failure. Verification of automatic

Fig. 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).

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

3.1 The Concept of the Evaluation System

With regards to the implementation of the system, a primary goal was to ensure accessibility. Thus the most obvious solution is a web-based application that provides 24-hour 'anywhere and anytime' access (see Fig. 1). The system depicted on Fig. 2 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.

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

Various features have been implemented through PHP functions. Users can be managed entirely by an administrator having supervisor rights. Supervisors can perform operations such as creating new user accounts, disabling

users, deleting user accounts, managing user rights, changing user profiles.

3.2 Structure of the Reference Database

First of all, for automatic evaluation, first of all, the speeches of hearing-impaired children were 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, and vowel-consonant formula etc.

The current database consists of exactly two thousand, three hundred and fifty-five words (some words occur multiple times, but the announcers and their intelligibility are different) which have been evaluated by thirteen professional SEN (Special Educational Needs) teachers with phonetic expertise and twenty-three unprepared students. The basis of the five-scale rating for evaluation is determined by the teachers. All of the teachers have only rated the students of other schools to avoid any bias resulting from the recognition of the speakers. Results have been recorded via the web application developed for this purpose [13].

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 Fig. 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.

Fig. 2 Concept of the online evaluation system

ratings

datafield name	datafield type
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

Fig. 3 Data table for storing evaluation

Period. Polytech. Elec. Eng. Comp. Sci., 62(4), pp. 126–133, 2018 | 129

Czap

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.

3.3 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 impaired , 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 non-hearing impaired (5): at most one or two speech sound defects may occur.

Before the evaluators submitted their scores they could listen to the speech samples repeatedly thereby eliminating 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 type of problem.

3.4 Using the Developed Evaluating System

After submitting the user name and password of the account created by the supervisor on the login interface user data is 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 then 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. The system indicates to the user how many samples he or she has already rated. Teachers in Budapest have evaluated one thousand, four hundred and forty-one speech samples, teachers in Eger have evaluated two thousand, and forty-three samples and teachers in Debrecen have evaluated one thousand, two hundred and twenty-six samples. Lay students have had to score all the samples; meaning two thousand, three hundred and fifty-five 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 'Submission of scores' after which the next audio file is loaded for rating.

We have used the online evaluation system for two purposes:

- For the automatic assessment of speech samples in development produced during the lessons a speech quality scale is needed as a reference. Subjective evaluators (specialized teachers and unprepared students) listened to two thousand, three hundred and fifty-five pre-recorded speech samples of different levels of utterance and scored them according to the five-grade scale described in Section 3.3.

- For the assessment of the speech production progress during the two-year training the evaluators listened to the recordings 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 this course of evaluation compared to the previous one. Namely that 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 defined in Section 3.3.

4.1 Introducing Speech Assistant into education

The testing of the system started in September 2013 with fourteen participating pedagogues. Four of them were given methodological and organizational tasks. In the research period each of the ten speech therapist pedagogues worked with six children; three of them were taught using traditional methods while the other three used the Speech Assistant software. The pupils selected for training are of different ages and stages of cognitive and speech development, having hearing impairments to varying extents. The sixty pupils involved were distributed among the group working with the talking head and the control group evenly, according to the extent of their hearing impairment, their capabilities and present joint handicaps.

Before using the system registration is required during which time only some basic information must be entered to get 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 Fig. 4).

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 the case of a pupil being deleted, his/her samples won't be automatically deleted from the server.

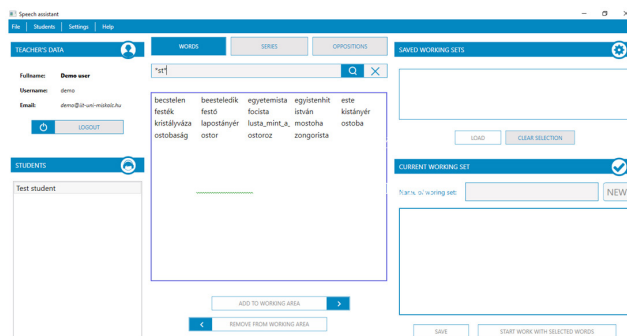The system contains a training word database of 3031 words. These were selected by the participating

Fig. 4 Word selection in the Speech Assistant system

pedagogues based on their previous experiences when 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 a 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)

By the end of the second year the set of word samples are augmented with three hundred and seven oppositional word pairs (e.g. bab-pap, bont-pont, lombos-lompos) and five hundred and ninety-three fixing word sequences (e.g. bu-be-bi-..., ub-eb-ib-..., ubu-ebe-ibi-...) [14].

4.2 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 following areas.
- Correction, fixing and automation of speech sounds.
- Long-short sounds, voiced-unvoiced sound pairs, correct pronunciation of sounds in consonant obstructions.
- Awareness of prosodic elements.
- Improvement of speech hearing, differentiating after hearing.

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 for 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.

1. According to the development plan the words containing the sound to be practiced are determined depending on its position within the word.
2. The selected word is read, listened, interpreted and embedded in a sentence to make sure that it is properly understood.
3. 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.
4. 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.
5. The pupil's speech is recorded.
6. 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.
7. At the end the performance of the pupil is evaluated.

4.3 Results of the two-year speech development
From the sixty pupils selected initially for the extra speech development lessons five abandoned the training, therefore only the progress of the other fifty-five pupils could be measured. Twenty-eight of them were girls and twenty-seven were boys. Twenty-eight of them used the talking head and twenty-seven of them did not (the control group) in the development lesson that lasted for half an hour twice a week.

For measuring progress speech samples of all participating children were recorded (sixty words and thirty 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 this 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 (with regards to all the pupils the average of the average scores is 1.12, while the average of the median scores is 0.86).

Fig. 5 shows the progress of each pupil for the two-year development period.

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 two years 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 of Section 3.3.

We have evaluated the speech production progress by different grades of improvements (see Table 1 and Fig. 6)
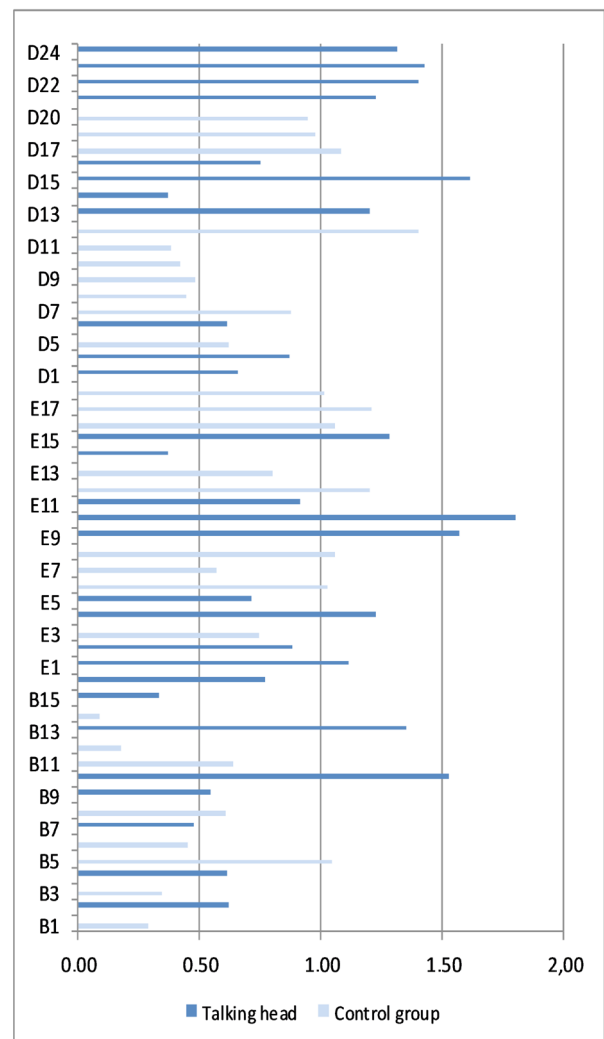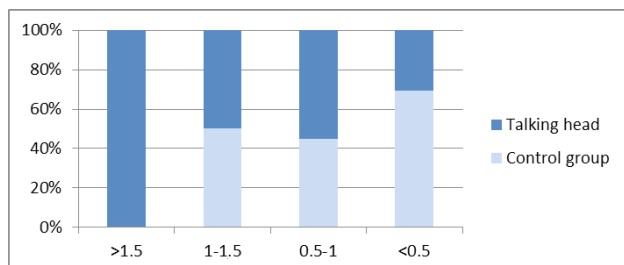

Fig. 5 Detailed results of the progress of children

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

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

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

5 Conclusion

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 people with hearing impediments. 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 attaining the least progress were taught in the control group. According to the written assessment of the teachers the motivation and attitude of the pupils was different, however the conclusion can be drawn since the groups were composed impartially.

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 the highest 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.

References

[1] Katsaggelos, A. K., Bahaadini, S., Molina, R. "Audiovisual Fusion: Challenges and New Approaches", Proceedings of the IEEE, 103(9), pp. 1635–1653, 2015.
https://doi.org/10.1109/JPROC.2015.2459017

[2] Massaro, D. W., Cohen, M. M. "Evaluation and integration of visual and auditory information in speech perception", Journal of Experimental Psychology: Human Perception & Performance, 9(5), pp. 753–771, 1983.
https://doi.org/10.1037/0096-1523.9.5.753

[3] Massaro, D. W., Cohen, M. M. "Speech perception in perceivers with hearing loss: Synergy of multiple modalities", Journal of Speech, Language, and Hearing Research, 42(1), pp. 21–41, 1999.
https://doi.org/10.1044/jslhr.4201.21

[4] Hunyadi, L., Szekrényes, I., Czap, L., Sziklai, I. "Seeing the sounds?", Argumentum, 10, pp. 335–338, 2014. [online] Available at: http://argumentum.unideb.hu/2014-anyagok/angol_kotet/hunyadil.pdf [Accessed: 14 September 2015]

[5] Massaro, D. W., Light, J. "Improving the Vocabulary of Children with Hearing Loss", Volta Review, 104(3), pp. 141–174, 2004.

[6] Massaro, D. W., Light, J. "Using Visible Speech to Train Perception and Production of Speech for Individuals With Hearing Loss", Journal of Speech, Language, and Hearing Research, 47, pp. 304–320, 2004.
https://doi.org/10.1044/1092-4388(2004/025)

[7] Al Moubayed, S., Beskow, J., Öster, A-M., Salvi, G., Granström, B., van Son, N., Ormel, E., Herzke, T. "Studies on Using the SynFace Talking Head for the Hearing Impaired", In: Proceedings of Fonetik'09: The XXIIth Swedish Phonetics Conference, Stockholm, Sweden, 2009, pp. 140–143.

[8] Siciliano, C., Williams, G., Beskow, J., Faulkner, A. "Evaluation of a Multilingual Synthetic Talking Face as a Communication Aid for the Hearing Impaired", In: Proceedings of the 15th International Congress of Phonetic Sciences: 15th ICPhS, Barcelona, Spain, 2003, pp. 131–134.

[9] Beskow, J., Engwall, O., Granström, B., Nordqvist, P., Wik, P. "Visualization of speech and audio for hearing impaired persons", Technology and Disability, 20(2), pp. 97–107, 2008.

[10] Czap, L., Varga, A. K., Illés, B. "Concept of a Speech Assistant System", In: 2013 Fourth World Congress on Software Engineering, Hong Kong, China, 2013, pp. 207–211.
https://doi.org/10.1109/WCSE.2013.37

[11] Vicsi, K., Szaszák, Gy. "Automatic Segmentation of Continuous Speech on Word Level Based on Supra-segmental Features", International Journal of Speech Technology, 8(4), pp. 363–370, 2005.
https://doi.org/10.1007/s10772-006-8534-z

[12] Pintér, J. M. "A beszédminőség automatikus értékelése" (Automatic Assessment of Speech Quality), PhD Thesis, University of Miskolc, 2015. (in Hungarian)

[13] Czap, L., Pintér, J. M. "Intensity feature for speech stress detection",
In: Proceedings of the 2015 16th International Carpathian Control
Conference (ICCC), Szilvasvarad, Hungary, 2015, pp. 91–94.
https://doi.org/10.1109/CarpathianCC.2015.7145052

[14] Kovács, Sz., Tóth, Á., Czap, L. "Fuzzy model based user adaptive
framework for consonant articulation and pronunciation therapy
in Hungarian hearing-impaired education", In: 2014 5th IEEE
Conference on Cognitive Infocommunications (CogInfoCom),
Vietri sul Mare, Italy, 2014, pp. 361–366.
https://doi.org/10.1109/CogInfoCom.2014.7020479