ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

Audio and Image Processing Matching and a Speech Therapy Application for Hearing-Impaired People

Alperen KAÇAR[1,*], Ali GÜNEŞ[2]

P.G. Student, Department of Computer Engineering, İstanbul Aydın University, Turkey[1]

Professor, Department of Computer Engineering, İstanbul Aydin University, Turkey[2]

ABSTRACT: We are living in the era of smart systems, which can learn and make decisions on its own. It is seen that technological developments are used in many different fields in order to transform all segments of the business world and the society with practical and theoretical studies. Computers now have the ability and capacity to do more than what we tell them to do. In many areas, computer systems develop fast solutions and are extremely successful considering their levels of speed and accuracy. Smart systems that can think and make decisions like humans in the fields of education, health, defense and security aim to produce solutions for the needs of people. The purpose of this study is to support the development of speech and language acquisition of hearing-impaired individuals. In this context, it is aimed to realize a software which provides visual feedback and extracts the characteristics of the individual's audio information and visible lip movements during the speech by using audio and image processing methods. In the future, these technologies are seen as an important operational step in improving the language and speaking skills of hearing-impaired individuals.

KEYWORDS: Audio and image processing; speech recognition; computer assisted speech therapy

I. INTRODUCTION

For humans, the most important of all the sounds in the world is the speech sounds [17]. The realization of human thoughts verbally in a complex way is described as speech [8].In order to express themselves, people have the ability to produce many different speech sounds consisting of words. This ability results from complex interactions between many parts of the body such as brain, lungs, larynx, vocal cords and all the mobile articulators. Speech is a complex actual behavior with physiological and psychological aspects, requiring a general skill [8]. The characteristics of the speech sounds are determined according to the differences in their acoustic properties such as volume, power, timbre and duration [17].

Expression of the thoughts verbally in an organized and symbolic way through speech is a tool which enables the recognition of the human ideas. A major part of our body participates in this process directly or indirectly during the speech.

The speech perception process means the understanding of the hearing of the ear. Hearing plays an important role in effective communication with our environment. Human ear has a completely complex structure. The sounds that occur around us result from invisible vibrations circulating in the air. While the weakest sound detected by the human ear is 0 dB as shown in Figure 1, normal speech is approximately 60 dB. The sound perception threshold for people with hearing difficulties is around 140 dB. The difference between the loudest and weakest sounds people can hear is around 120 dB in a million width ranges. There are only about 120 high sound levels that can be detected between the lightest whisper and the loudest thunder [26].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

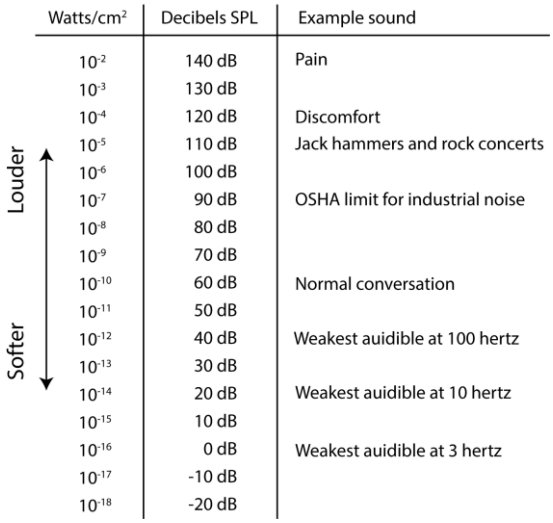

Figure 1. Sound intensity [26]

In Turkey, approximately 1000-2000 children are born with hearing impairment every year, or experience hearing disorder before starting to learn a language. Children with severe hearing impairment benefit from hearing aids at a minimum level. Children aged 18-48 months typically produce abnormal sounds. However, they have almost no vocabulary compared to their peers. Although this problem is attempted to be solved with the hearing aid, the hearing aid is insufficient for improving speech qualities of older children and adults and for more fluent interaction [5].

The number of hearing-impaired individuals is significant in the world and in Turkey. Because hearing-impaired individuals are a part of the world which hears, their training should be planned to solve their communication problems [16]. According to the researches, two out of every thousand children have congenital hearing problems or hearing problems resulting from different causes after birth [6]. In the communication system, language and speech are integrated concepts. The language existing in communication is a system consisting of basic units such as signs and words [4].

Lack of auditory feedback for hearing-impaired individuals causes speech impairment. Therefore, hearing-impaired people cannot exhibit speech development despite having the proper speech production mechanism. Children born with hearing impairment must rely on the visual cues of the phonetic features of speech to learn speech imitations and improve their speech movements. Although hearing-impaired individuals can imitate lip movements, they cannot make adequate progress in correct sound production and toning. Speech training systems try to ensure that speech is performed by visualizing the feedbacks such as basic voice frequency levels, lip and vocal tract movements. The aim of this study is to enable the hearing-impaired individuals to learn the speech processes correctly and to contribute to the development of learning and communication skills through a software that provides visual feedback as in speech training systems. The study can be described as a multi-functional assistant system between the trainer, the student and the parent. Thus, a system will be provided to the families and educational institutions for the development of hearing-impaired individuals.

Language and Speech in Hearing-Impaired People

The language and speaking way of people help them communicate with the rest of the society. According to Chomsky, language acquisition is genetically innate in human beings [7]. Hearing loss in infancy and childhood has a significant impact on communication, education and quality of life. Language acquisition and speech training are essential for effective communication. Any damage on these factors will make it difficult or prevent people from communicating effectively. Hearing loss in children delays speech and language development; As a result, different problems can be faced in the development of social, behavioral and academic skills.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

The techniques of developing hearing-impaired people's language and speaking skills are considered as teaching first the language and then speaking in a developmental sequence [27]. Linguists emphasize that it is important to acquire the language in early childhood and that lagging behind in language development due to any reason will lead to a lot of problems at later ages [13].

II. COMPUTER AIDED SPEECH THERAPY

Problems such as incomprehensibility of the words uttered during the speech, difficult expression of the letters and incorrect production of letters are described as speech disorders. Besides, problems such as swallowing letters and syllables occur during the speech. Eliminating speech barriers in children's pronunciation of words with speech therapy methods can help improve their language development and speaking skills. Speech therapy is a training that focuses on language development and developing comprehension and expression skills.

Traditional speech therapists work to develop appropriate methods and techniques for solving a child's problems such as coordinating the mouth (articulation, fluency, volume regulation), language comprehension and expression to form words and sentences. Phonetic assistance of the hearing-impaired people is usually performed through the observation of the mouth of a teacher or orientation of the movements of the vocal organs by a trainer. To help the speech training of the hearing-impaired people, many researchers have developed various speech training aids [14][22][24].

For nearly 70 years, researchers have been unable to achieve successful results in using technology to improve the speech success [5]. The development of technology and innovations in signal processing have been the driving force for new speech training applications. With the help of computer-assisted speech therapy applications, speeches of hearing-impaired children can be improved by providing visual and auditory feedbacks. However, their usage has not become widespread, because their feedbacks are scarce and difficult to understand. The most important feature of the computer aided speech training is to make the voices of the hearing-impaired individual visible. This is why different training programs have been developed to provide the hearing-impaired individuals with visual feedbacks for the development of the speech perception [18].

III. SYSTEM ARCHITECTURE

The system is designed to visualize the speech of the hearing-impaired individuals and to allow them to see the audio information of a trainer and repeat it. The hearing-impaired person is expected to watch the pre-recorded training video for the specified words and then make a speech close to it. The system has two working threads, the audio front end and the image front end, as shown in Figure 2. The image processing section lists the lip images of the spoken word. In the audio processing section, speech recognition and comparison of the spoken word with the voice of the trainer are performed.

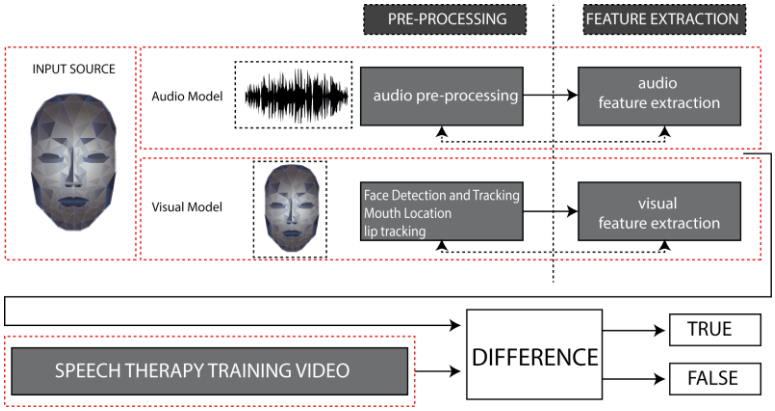

Figure 2. System architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

Image Front End

Image processing is a technique that can process any image to extract real information, that is, useful information from the image. The pre-processing and visual feature extraction on the images of the introduction video are shown in Figure 3 and explained in the subsections.

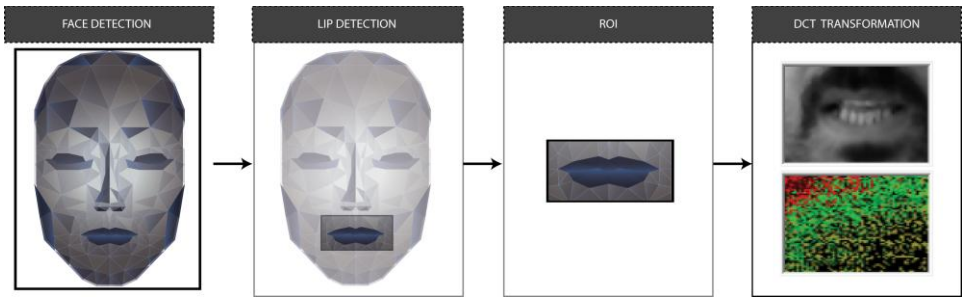

Figure 3. Visual feature extraction process

The image front end of the system is realized in two stages: pre-processing and feature extraction.

Pre-processing includes cleaning the introduction image, face location detection, mouth location detection and lip movement tracking. This stage is of great importance for the solidity of the results obtained from the image processing. A video is a sequence of images displayed fast consequently. Images obtained via the camera and the introduction image are optimized with settings such as noise cancellation, sharpness adjustment and lighting.

After image cleaning, face perception is performed to extract the lip area. Since lip images are very small parts that move continuously, face perception is firstly required to limit the area where the lips should be searched. After the face is detected, cropping is performed from the specified image for further processing.

Face Location Detection

Face detection is an important pre-processing step for many facial applications (face recognition, lip reading, age, gender and race recognition, emotion prediction). The accuracy of these applications depends on the reliability of the face detection processing step. The purpose of face detection is to determine whether there is a face in any image and to return the position and size information of the face, if there is any. In general, the leading factors that influence the reliability in face perception and determination are the parametric differences which come together with posing, orientation and lighting conditions. Different algorithms and methods are used to solve this. Some of the most successful methods are to use a 20 × 20 (or more) pixel observation window on the image for all possible positions, scales, and orientations. Some of these methods are support vector machines and artificial neural networks. Some researchers use skin color to detect the face in color images.

For the detection of the mouth area, it is necessary to detect the face area in each frame from the video. The approach proposed by Viola & Jones in their article “Fast Object Detection is an effective method of object perception (Viola & Jones, 2001). This method was used for face detection and the system was trained with many positive and negative images. Accurate identification of the face area facilitates the process of identifying the mouth area.

There are 4 basic stages in [28] algorithm as shown in Figure 4.

1. Integral image.
2. Haar feature extraction.
3. AdaBoost training.
4. Cascade classifier.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

This method allows an image to be analyzed without having to examine each pixel.

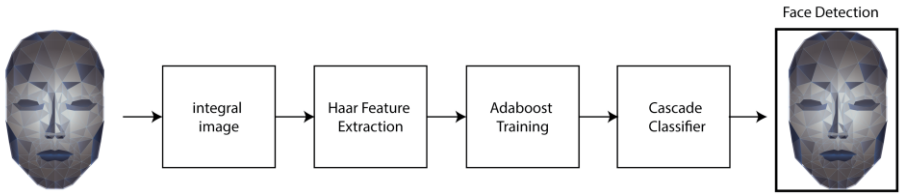

Figure 4. Basic principle of Viola and Jones's algorithm

Mouth Location Detection

To detect the mouth area, first the face area should be determined. After the face area is detected with the Viola-Jones algorithm, the mouth area is extracted from the ratio information of the face area. Although the Viola-Jones algorithm yields successful results in the detection of the face area, as shown in Figure 5, it cannot give good results in the detection of the mouth area.

Figure 5. After the mouth detection with the Viola-Jones algorithm

Since the mouth area is the region where speech information is visually extracted, it is a more important process step. The extraction of the correct visual information depends on the detection of the correct area. As shown in Figure 6, the face ratio information provides mathematical and logical information for lip extraction.

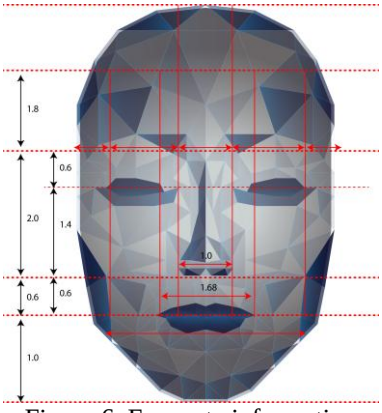

Figure 6. Face rate information

The extraction of the lip area is based on this logical information.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

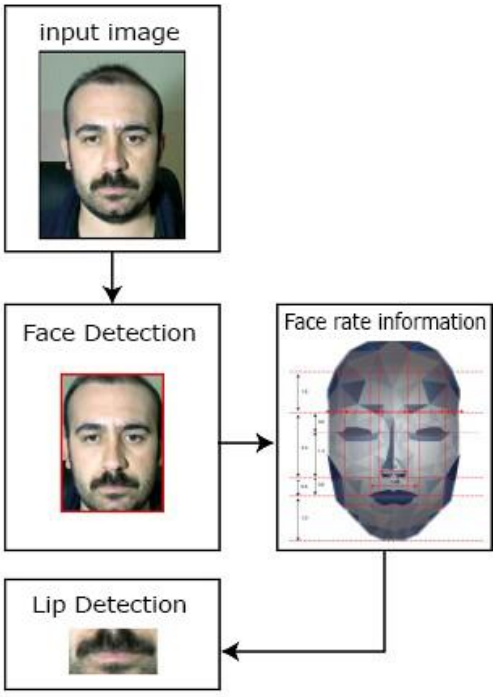

Figure 7. Lip detection

In our study, the steps taken for lip area detection are as follows:

1. Face detection is performed with the Viola-Jones algorithm.
2. After the face detection, the specified area is surrounded by a square.
3. The coordinate information of the area in the square is obtained.
4. According to the coordinate information, mouth area extraction is performed by making use of the face ratio information.

As shown in Figure 8, images obtained after mouth detection become ready for the feature extraction.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

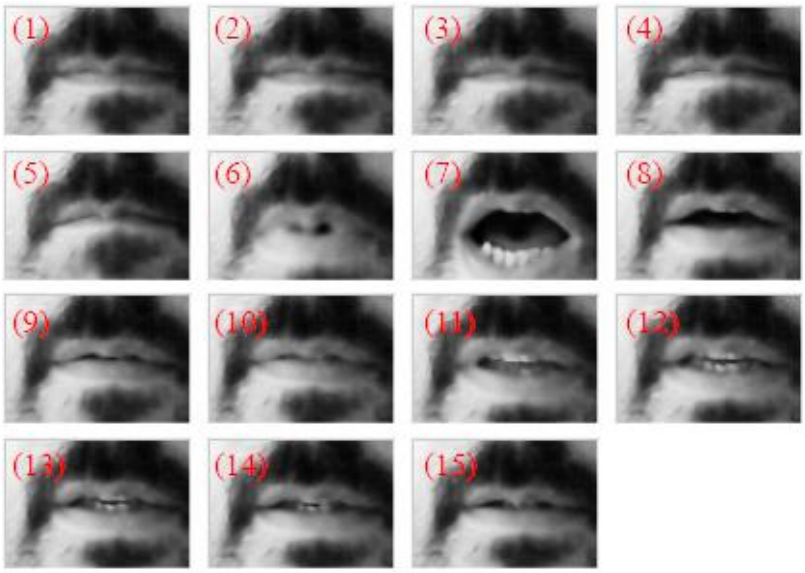

Figure 8. Lip images captured during the speech

Audio Front End

Sounds are signals that change over time in the real world. Their meaning is all related to this time variability. Therefore, to facilitate understanding, comparison, modification and re-synthesis processes, audio analysis is performed to take advantage of the distinctive features of time-varying sounds [23]. The audio processing process to benefit from the related features consists of two stages: pre-processing and feature extraction.

Pre-processing is the first part of the speech sensor and can be interpreted as the level of preparation before the feature extraction. In order to achieve optimal speech recognition, useful speech information should be appropriately extracted and represented in the pre-processing section. Pre-configuration and processing of the raw sound is an important processing step. Typically, for a clean audio file, various undesirable parts such as coughs, hiccups and disturbing peaks are removed to normalize the sound.

Feature Vector Extraction

Generally, short-term analysis methods are used in the analysis of audio signals, because most of the audio signals are stable for short periods of time. For example, acoustic characteristics are captured in the audio signal framed at 20 ms. intervals.

In general, the most distinctive acoustic characteristics in each frame of audio signals are [11];

- Audio: this feature indicates the height of the audio signal associated with the amplitude of the signals. It is also commonly named as energy and intensity in the audio signal.
- Pitch: this feature represents the vibration velocity of the audio signals to be represented by the basic frequency or simultaneously corresponds to the basic period of the audio signals.
- Timbre: this represents the meaningful content of the audio signals characterized by a waveform within a period of basic audio signals like a vowel in Turkish.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

The exemplified audio information is divided into frames and a more efficient operation is performed. Attributes are obtained with the extraction of the acoustic characteristics in the audio signal. MFCC is one of the most commonly used attribute extraction methods in speech recognition.

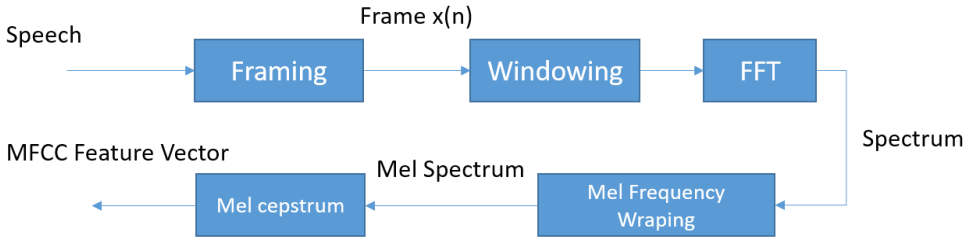

Figure 9. Feature vector extraction

Mel-Frequency Cepstral Coefficients

The first step in any speech recognition system is feature extraction. The Mel-frequency cepstral coefficients (MFCC) perform the conversion of the sound in the time domain to the frequency domain, enabling access to information contained in the speech signal. MFCC takes into account human perception sensitivity according to frequencies and is therefore one of the best techniques for speech recognition.

MFCC is often used for feature extraction in both speech and speaker recognition systems. As shown in Figure 10, a machine learning approach is applied on the basis of the features acquired from the recorded speakers for speaker recognition.

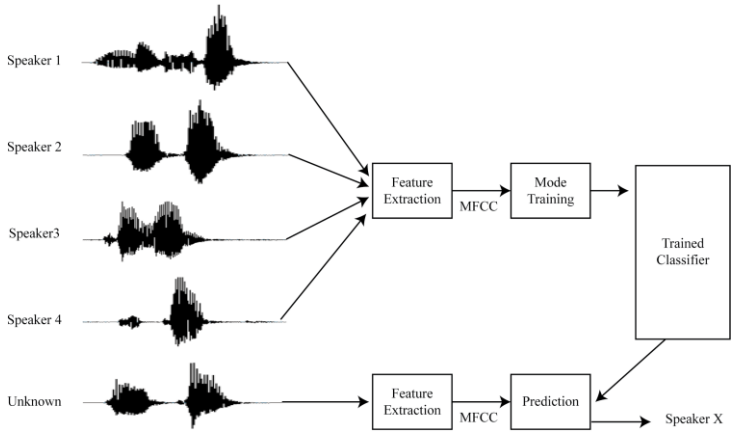

Figure 10. The Approach scheme used for speaker recognition.

The Mel-frequency correlates the perceived frequency of a pure tone to the actual frequency measured. People are better at recognizing small changes in low frequencies than they are at high frequencies. The Mel-frequency scale ensures that the sound characteristics are closely matched to what people hear. The MFCC compresses the information about the vocal tract to a small number of coefficients on the basis of the cochlea understanding. The MFCC calculation steps are shown in Figure 11.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

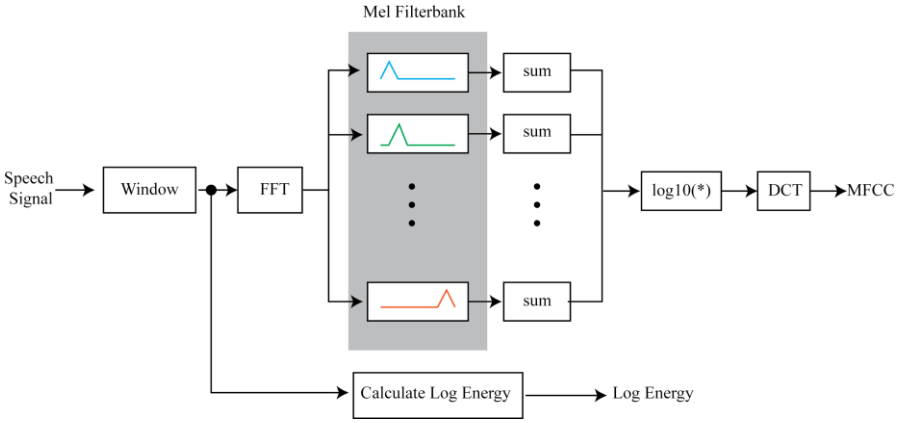

Figure 11. MFCC calculation steps.

As stated in Figure 12, the Mel filter places the first 10 triangular filters in a linear way and the rest logarithmically.

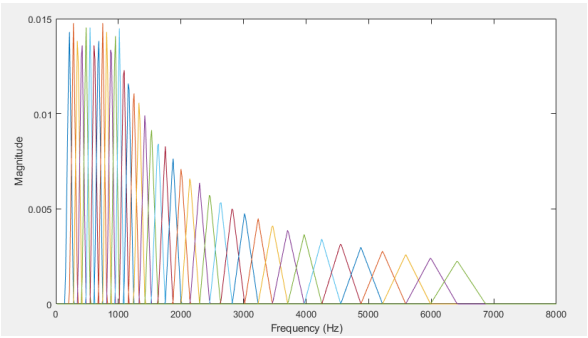

Figure 12. Mel filterbank image.

Speech signal is dynamic in nature and changes over time. However, the extraction of the sound characteristics in speech and speaker recognition systems are performed by assuming that the sound is constant in 20-40 ms. windows.

IV. COMPARISON OF SPEECH SOUNDS

HMM Classifier

Hidden Markov Model (HMM) is a popular statistical method used for modeling a wide variety of time-series data. Natural language processing (NLP) methods give successful results in tagging the speech and phrasal discrimination [3]. HMMs is commonly used in signal processing, especially in speech recognition systems.

Speech recognition systems are examples for interdependent applications. For example, letters following each other in a word have a dependent structure; while it is a high possibility that "a" comes after "b" in Turkish, it is less possible for "f" to come after it. In speech recognition, it is assumed that sound is comprised of basic allophones and has a certain order. It can be stated that words can be put in a certain order depending on the syntactic and semantic rules of that language and these orders correspond to sentences [2].

HMM is a finite machine changing its status in a course of time. In HMM-based speech recognition systems, it is assumed that the vector series corresponding to a word is produced by a Markov chain. In Figure 13, the training diagram of the HMM acoustic model is presented.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

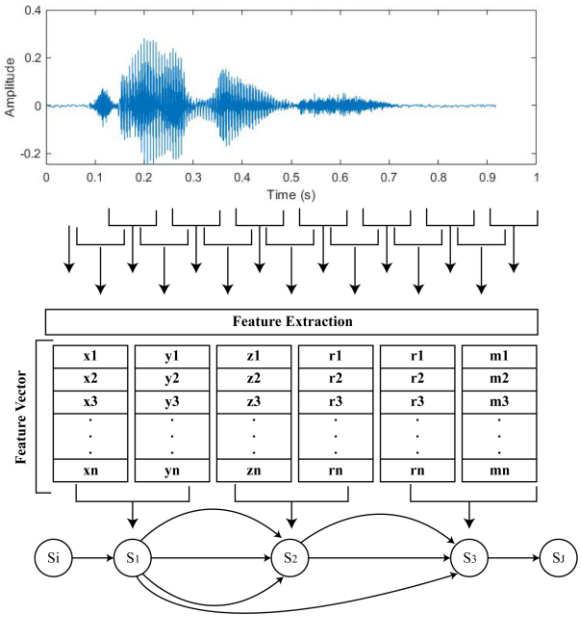

Figure 13. The training diagram of the HMM acoustic mode.

Dynamic Time Warp (DTW)

A time series is a group of observations performed consecutively in time. DTW is an algorithm used for measuring the similarity between two time series which could change over time. The values of a time series do not have any meaning on their own, but their graphic demonstration can give us information for different application areas. Real-time heartbeat values, speech patterns, DNA structures, handwriting and even shapes can be converted to a variety of time series.

The speech process is realized with obtaining the desired information from a speech signal. Speech recognition systems based on matching the acoustic patterns is based on the principle of comparing the measurements which change in time. The difference observed in speech signals which are created when the same word is uttered by different people can be seen in Figure 14. Moreover, differences occur in speech signals again even if the same person utters the same word in same durations. These differences can be observed in images in Figure 15.

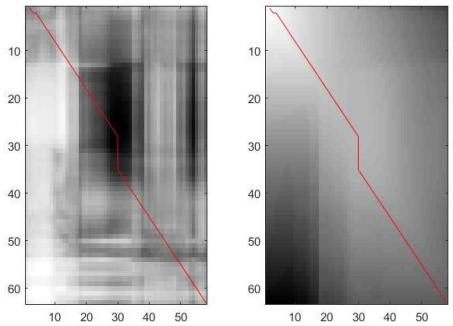

Figure 14. Utterance of the same word by different people with DTW

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

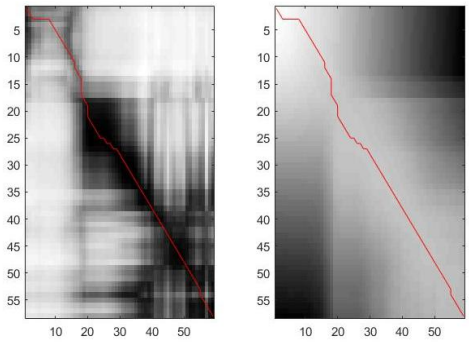

Figure 15. Demonstration of the utterance of the same word by the same person with DTW

A time alignment should be performed in order to obtain a global distance between the two speech patterns. Features that correspond to two speech patterns can be displayed in the same place on a common time-scale with the dynamic time warp method. Thus, the similarities between the signals are highlighted. Dynamic time warp (DTW) can be applied to solve problems related to speech spectral sequence comparison. As shown in Figure 16, the Dynamic time warp method makes reference comparisons of sound patterns by compressing or expanding vocal expressions.

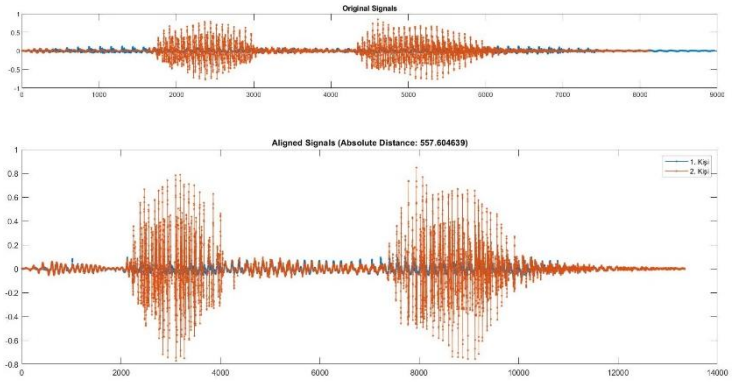

Figure 16. Expanding DTW on a common time scale

In isolated word recognition systems, the acoustic pattern or template of each word in phrases is stored as a time feature. Recognition is the process of comparing the acoustic patterns of the word to be recognized and then selecting the word that best matches the recognized word. This means a comparison of the time compliances of two different patterns.
In this study, pattern matching process was tried to be carried out by comparing the information existing in the memory with the coming speech information.

V. DISCUSSION, CONCLUSION AND RECOMMENDATIONS

Turkey has been experiencing serious difficulties and deficiencies in the training of the current hearing-impaired individuals. In classic speech therapy trainings, it is quite difficult especially for the trainer to constantly control the lip movements and voice output of the student continuously and this affects the quality of the training. Immediate and meaningful visual feedback should be provided to the child in order to teach and explain whether the word spoken during the speech is correct or incorrect. Considering these deficiencies in this field, a Turkish computer-aided speech therapy software was achieved in this study for speech therapy trainings of the hearing-impaired individuals. The system was designed to visualize the speech signals of hearing-impaired individuals and to allow the hearing-impaired person to see and repeat the audio information. Hearing-impaired individuals will first watch a sample training video

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

recorded for designated words and then give a speech close to it. Our software consists of two system sections: audio processing and image processing. Each part includes two steps: pre-processing and feature extraction. In the image processing section, visual features of the face location detection, mouth location detection and lip movements are demonstrated. In the audio processing section, audio signals are cleaned and the audio information features are obtained. Speech recognition Hidden Markov Models (HMM) are often used, but dynamic programming techniques such as the Dynamic Time Warp (DTW) model are more successful as the sample words contain timing differences between templates. Therefore, HMM and DTW techniques were used together in this study. HMM and DTW speech distance results are shown in the appendix.

It is thought that this software will contribute to the speech therapy process of many hearing-impaired individuals, because the software provides visual feedback to both the speech therapist and the hearing-impaired person by visualizing the data obtained from images and audio information. Giving feedbacks with graphs can motivate the hearing-impaired individual for a long-term and effective training during the process of learning speaking. It can also increase the efficiency of the education process by assisting the special education specialist. In addition, systems developed and to be developed for individuals in need of special education may solve the problem of lack of trained manpower in this field. The fact that these systems are consistent, portable and cost-efficient can be counted among their advantages. Based on these justifications, it can be put forward that our study can fill this gap

Table 1. Results of the word "abi (brother)" obtained by DTW for trainers

	E1	E2	E3	E4	E5
E1	0	53,77	40,26	11,95	21,83
E2	53,77	0	28,36	47,71	49
E3	40,26	28,36	0	35,69	31,80
E4	11,95	47,71	35,69	0	24,40
E5	21,83	49	31,80	24,40	0

Table 2. Results of the word "bak (look)" obtained by DTW for trainers

	E1	E2	E3	E4	E5
E1	0	30,20	118,59	59,78	56,74
E2	30,20	0	30,20	54,748	43,01
E3	118,59	95,80	0	86,521	60,32
E4	59,78	54,748	86,521	0	55,635
E5	56,74	43,01	60,32	55,635	0

Table 3. Results of the word "beyaz (white)" obtained by DTW for trainers

	E1	E2	E3	E4	E5
E1	0	24,74	26,77	40,53	31,37
E2	24,74	0	14,23	36,97	29,74
E3	26,77	14,23	0	37,02	23 ,74
E4	40,53	36,97	37,02	0	44,85
E5	31,37	29,74	23,74	44,85	0

Table 4. Student and trainer speech distance values of the words included in the software

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

WORDS	Trainer 1	Trainer 2	Trainer 3	Trainer 4	Trainer 5
	Student	Student	Student	Student	Student+
One	34,97	39,16	57,38	42,47	35,57
Two	72,43	49,83	59,28	65,85	60,19
Three	123,90	84,76	83,09	131,21	76,36
Four	110,45	68,77	75,00	74,03	89,10
Five	64,61	81,34	77,36	103,73	39,14
Six	38,03	25,40	51,86	35,81	18,66
Seven	49,86	40,35	30,17	80,58	41,55
Eight	89,04	45,04	60,34	73,93	26,34
Nine	61,92	20,08	55,74	45,73	53,98
Ten	30,98	38,04	36,39	83,97	46,75
White	25,25	47,01	32,25	33,97	30,62
Black	45,62	26,21	72,52	40,13	31,04
blue	24,12	35,49	44,31	22,18	22,91
Purple	75,21	34,77	92,76	37,90	54,91
Red	83,91	71,70	52,40	60,44	29,61
Pink	60,71	49,85	43,52	67,17	58,44
Grey	55,52	47,66	29,48	18,77	20,18
Orange	69,77	46,83	75,73	34,45	49,34
Mother	76,15	57,24	82,58	60,39	46,13
Father	27,15	19,28	32,46	67,63	54,47
Grandfather	13,89	53,93	26,20	54,94	27,35
Brother	42,53	52,22	45,43	40,96	53,62
Uncle	29,62	27,98	14,62	44,90	55,90
Aunt	70,78	51,50	66,32	71,60	39,67
Aunt	91,06	42,20	55,89	87,51	91,75
Grandmother	71,15	46,12	65,93	41,87	15,75
Grandmother	37,95	39,19	46,85	44,90	45,41
Hello	29,02	18,26	61,20	14,67	35,01
Come	43,05	30,63	22,67	57,09	50,86
Go	48,99	43,04	42,16	37,85	32,24
Remain	91,24	45,81	76,88	23,10	57,14
return	45,13	26,02	71,06	15,31	27,69
Look	40,94	69,77	38,58	74,31	88,16
Plug	42,13	77,12	21,56	83,64	47,73
Car	16,83	28,93	40,83	41,49	33,62

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

Train	39,86	29,32	44,48	46,09	39,52
Truck	79,15	70,68	83,78	64,82	38,05
Bus	88,96	16,35	29,58	51,03	47,13
Airplane	142,14	102,65	90,67	60,64	95,24

Table 5. Student and trainer speech distances of the words with realized training procedure via HMM

	Brother	Look	White
One	-24733.177	-18225.2312	-17195.3418
Two	-20806.5403	-13803.702	-12875.5111
Three	-22489.3205	-14558.0416	-12721.7663
Four	-21237.2322	-11424.4435	-12661.8109
Five	-22727.8317	-14505.2292	-13668.7915
Six	-22176.7696	-9614.722	-11889.2332
Seven	-22564.9758	-15648.2083	-14271.8969
Eight	-20534.4728	-13896.9983	-13934.913
Nine	-28084.4104	-10865.0184	-15960.2735
Ten	-27065.9424	-11977.205	-15841.5259
White	-22304.8377	-11129.2684	-11281.0969
Black	-19613.3852	-8942.4953	-12632.2699
blue	-15405.4994	-10129.8667	-11503.0207
Purple	-16320.5036	-14271.8969	-21891.8691
Red	-20190.7974	-11538.9703	-11786.0773
Pink	-21217.0709	-10320.8107	-11942.1191
Grey	-18377.6444	-12637.2963	-13371.708
Orange	-17580.2401	-9807.2508	-10514.9832
Mother	-21088.2739	-10615.355	-13133.9053
Father	-31066.6009	-10962.1437	-16320.5036
Grandfather	-21479.8385	-13435.7713	-12689.7563
Brother	-19829.966	-10750.2228	-16879.8424
Uncle	-17101.8713	-9449.4167	-12080.4356
Aunt	-21891.8691	-12765.8692	-12657.0537
Aunt	-28687.8451	-11488.5566	-17877.5624
Grandmother	-21163.9315	-10985.0956	-12351.15
Grandmother	-21437.3281	-10289.0122	-13615.3899
Hello	-20630.8123	-10212.6016	-11471.3165
Come	-23291.1687	-13030.6117	-12821.3888
Go	-23208.2054	-17691.7511	-17657.9288
Remain	-27167.4986	-10382.4301	-15205.3389

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 7, July 2019

return	-21008.6113	-11091.7082	-12780.0796
Look	-21626.8137	-9388.0675	-13290.1098
Plug	-23799.3772	-9024.3653	-13390.303
Car	-23330.2139	-9355.7674	-13567.6145
Train	-19585.0025	-13592.1841	-12622.4957
Truck	-19580.6428	-9997.3055	-14521.2724
Bus	-17697.5532	-12029.405	-12822.2859
Airplane	-23125.2829	-9307.6774	-12275.0518

REFERENCES

1. Akçamete, G. (1993). İşitme Engellilerde Dil ve Konuşma. *Özel Eğitim Dergisi, 1*(3), 2-9.
2. Alpaydın, E. (2018). *Yapay Öğrenme* (4. b.). İstanbul, Sarıyer: Boğaziçi Üniversitesi Yayınevi.
3. Arfeen, Z., & Aggarwal, J. (2016). Speech Recognition based on Hidden Markov Model Toolkit (HTK) with BODO Language. *International Journal of Advanced Research inComputer Science and Software Engineering*, *6*(9), 383-388.
4. Baykoç, D. N. (1987). 12-30 aylık Türk çocuklarında dilin kazanılması. *Çocuk Gelişimi ve Eğitimi Dergisi*(2), 36-38.
5. Bernstein, L. E., Goldstein, M. H., & Mahshie, J. J.(1988). Speech training aids for hearing-impaired individuals: 1. Overview and aims. *Journal of Rehabilitation Research and Development, 25*(4), 53-62.
6. Cengiz, D. U., Kolcu, D., & Ercan, M. K. (2016). İşitme Engelli Çocuklarda Dil Kazanımı ve Konuşma Eğitimi. *International Congress on Woman and Child Health and Training.* Kocaeli.
7. Chomsky, N.(2001). *Dil ve Zihin.* (A. Kocaman, Çev.) Ankara: Ayraç Yayınevi.
8. Gerçeker, M., Yorulmaz, İ., & Ural, A. (2000). Ses ve Konuşma. *K.B.B. ve Baş Boyun Cerrahisi Dergisi*, 71-78.
9. Hudgins, C. V., & Numbers, F. C.(1942). An investigation of the intelligibility of the speech of the deaf. *Genetic Psychology Monographs*(25), 289-392.
10. Ibrahim, Y. A., Odiketa, J. C., & Ibiyemi, T. S. (2017). Preprocessing Technique in Automatic Speech Recognition for Human Computer Interaction : An Overview. *Annals. Computer Science Series, 15*(1), 186-191.
11. Jang, R.(2019, Ocak 20). *Audio Signal Processing and Recognition.* http://mirlab.org: http://mirlab.org/jang/books/audioSignalProcessing/ adresinden alındı
12. Kaleka, J. S. (2010). solated Word Recognition using Dynamic Time Warping. *Proceedings of the International Conference on Circuits, Systems, Signals*, (s. 293-295).
13. Kol, S. (2011). Erken Çocuklukta Bilişsel Gelişim ve Dil Gelişimi. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, 1-21.
14. Mahdi, A. E. (2008). Visualisation of the Vocal-Tract Shape for a Computer-Based Speech Training System for the Hearing-Impaired. *The Open Electrical & Electronic Engineering Journal, 2*(1), 27-32.
15. McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*(264), 746-748.
16. Neumeyer, L., Franco, H., Digalakis, V., & Weintraub, M. (2000). Automatic scoring of pronunciation quality. *Speech Communication, 30*(2-3), 83-93.
17. Neyman, L., & Bogomilsky, M. R. (2001). *Anatomi, işitme ve konuşma organlarının fizyolojisi ve patolojisi.*
18. Nickerson, R., & Stevens, K. N. (1973). Teaching Speech to the Deaf: Can A Computer Help? *IEEE Transactions on Audio and Electroacoustics, 21*(5), 445-455.
19. Nilsson, N. J. (2018). *Yapay Zeka Geçmişi ve Geleceği.* (M. Doğan, Çev.) İstanbul: Boğaziçi Üniversitesi Yayınevi.
20. Petajan, E. D. (1984). Automatic lipreading to enhance speech recognition. *in Proc, Global Telecomm*, (s. 265-272). Atlanta.
21. Rao, P. (2007). Audio Signal Processing. B. Prasad, & S. R. Mahadeva içinde, *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks.* Springer.
22. Resmi, K., Sardana, H., Kumar, S., & Chhabra, R. (2011). Graphical Speech Training system for hearing impaired. *2011 International Conference on Image Information Processing.*
23. Rocchesso, D. (2003). Introduction to Sound Processing. California: Creative Commons.
24. Sang, H. P., Dong, J. K., Jae, H. L., & Tae, S. Y. (1994). Integrated Speech Training System for Hearing Impaired. *IEEE Transactions on Rehabilitation Engineering, 2*(4), 189-196.
25. Schafer, R. W., & Rabiner, L. R. (1975). Digital representations of speech signals. *Proceedings of the IEEE, 63*(4), 662-677.
26. Smith, S. W. (1999). *Digital Signal Processing.* San Diego, California: California Technical.
27. Türk, O., & Arslan, L. M. (2004). Konuşma Terapisine Yönelik Konuşma Tanıma Yöntemleri. *Signal Processing and Communications Applications Conference*
28. Viola, P., & Jones, M.(2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 511-518.
29. Wankhede, S. N. (2014). Designing visual Speech Training Aids for Hearing Impaired Children. *International Journal of Advanced Research in Electrical,Electronics and Instrumentation Engineering, 3*(4), 8726-8733.