

Learning Sign Language Using Speech Recognition Technique

Basim Alhadidi ¹, Hussam Nawwaf Fakhouri ² and Rasha Awamleh ³.

Information Technology Department

^{1,3} Al Balqa' Applied University

² The university of Jordan

Jordan

Abstract

Sign language is a means of communication that uses facial expressions, hand movement, arms or body gestures to express a speaker's thought instead of sound. Learning sign language by computer saves time and effort as well as it guarantees accurate results.

This paper proposes an unsupervised implemented algorithm for learning sign language by computer system using speech recognition which attempts to replace the long courses to learn sign language. It also proposes a way for using a sign language database to search and find a sign word . The main motivation behind this research is the necessity to change all traditional learning ways and create new ones.

Keywords

Sign language , sign dictionary , speech recognition

1 Introduction

Learning new languages is difficult and it takes a long time especially if this language contains Signs. Sign language is mainly used by deaf or mute people as well as other related people such as interpreters, teachers, friends, and families of deaf people. As there are some people who suffer from such disabilities in the community, sign language may play a critical role to include them in our community.

It is even easier for those who use sign language to communicate regardless of their backgrounds than normal people from different cultures.

Sign language in this respect provides an easy access to the international deaf community. This paper aims to develop the techniques and methods of learning sign language by computer speech recognition (John Wilry&Sos, 1999, Lawrence,1993).

2. Used Materials and Method

We used Matlab software to implement the algorithm because Matlab is a high-performance language for education and research. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation and it also has toolboxes for signal processing , neural network, image processing , database ... etc

To test the software we used headset Microphone in order to convert the sound into an electrical signal. Headset Microphone allows the ambient noise to be minimized by allowing you to have the microphone at the tip of your tongue all the time , we also needed 16 bite sound card , windows environment operating .

And the method of how to apply the proposed implemented algorithm first starts by recording the electrical sound signals by the user using any recording program such as windows recording application after that adding these sounds to the sound folder and the database so that it can be read by the matlab software , then using matlab functions that is found in matlab toolbox to read the sound file after that applying the Short-Time Fourier Transform in order to extract features of the human voice that represent spectrum in order to Compute local & global distances by Dynamic time wrapping after that the next step of the algorithm will be Calculating the minimum coast in order to compare the recorder sound that is entered

by the user and the sound files in the database and if there is a match the algorithm will search for the right video file to be displayed .

3. Proposed Algorithm, Discussion

We will start discussion the algorithm by Speech Recognition flow chart shown in fig.1.

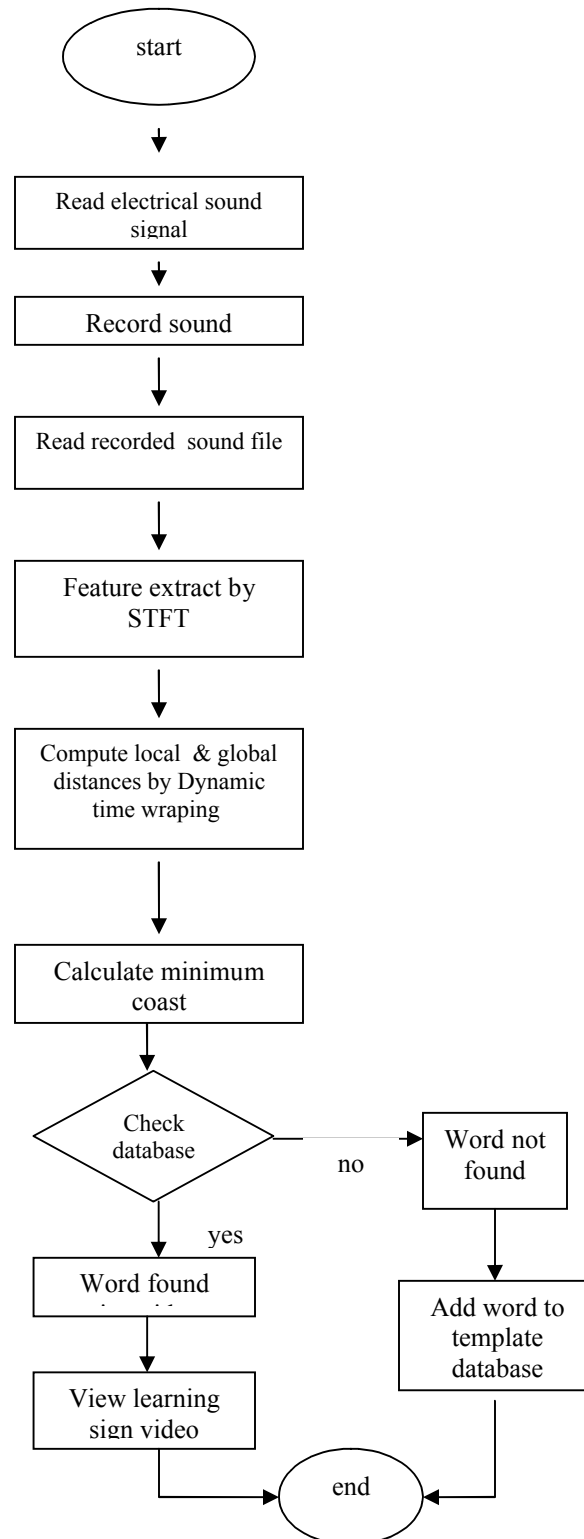


Fig.1

The first step is to read the electrical signal that is converted from the recorded sound file that was recorded from microphone using the **function** wavread from matlab toolbox as shown in equation1

```
f1=wavread('filename') ..... (1)
```

After the first step that cut the speech signal into frames with overlap and this resulted in matrix where each column is a frame of N samples from original speech signal. We applied Windowing and transformation for the sound signal into the frequency domain. This process is done using the Short-Time Fourier Transform (STFT). We used Short-Time Fourier Transform because there are Features in the human voice that represent sepctrum so for every word that is said it is devided into short-time segments (frames) that is represented in the time graph by amplitude and time of the signal as shown in fig .2. (Daniel Jurafsky, James Martin, 2000, matlab documentation), also because as the short-time Fourier transform is defined by Transforming the signal into the frequency domain or time-dependent. shown in fig. 3 , it also maps a signal into a two-dimensional function of time and frequency. After that we applied sliding window A technique that called windowing the signal in order to analyze a small section of the signal at a time. STFT represent a sort of compromise between the time and frequency-based views of a signal. It provides some information about both when and at what frequencies a signal event occurs And determine size for the time window as shown in fig . 5 (John Wilry&Sos, 1999, matlab toolbox)

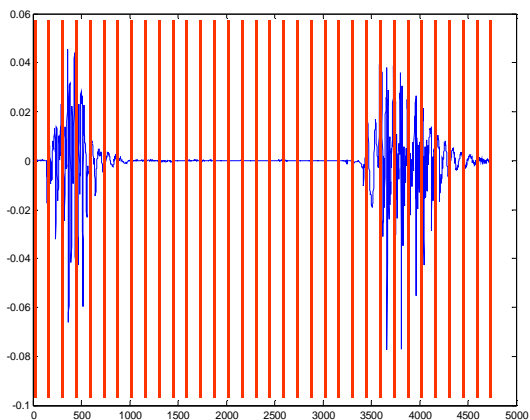


Fig.2

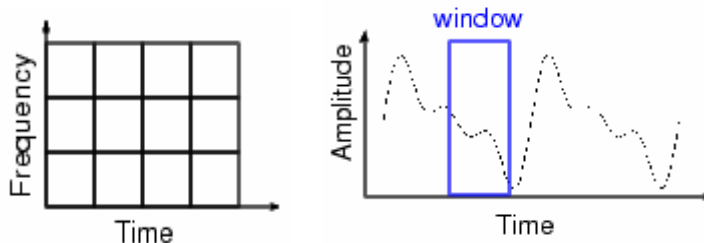


Fig.3

After that we compared different speaking speeds by applying Dynamic time warping Because it is a method that finds an optimal match between two given sequences with certain restrictions called the tested data template. The recognition process consists of matching the incoming speech with the stored template. If the sample has the lowest distance then it is the recognized word,

The next step is to study the Speech characteristics for a the speaker that is normally found in differences of energy , time, frequency and pitch. The extracted features containing characteristic information as shown in fig. 4,5. (C.H. Lee, 1996)

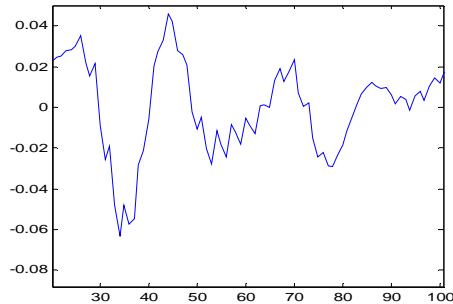


Fig.4

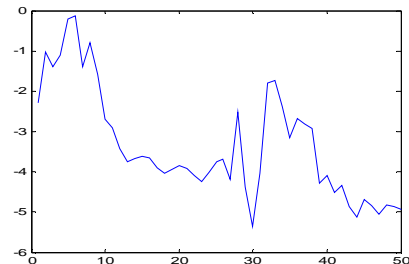


Fig. 5

The next step of the algorithm is to apply the Dynamic Time Warping (DTW), because it has Two basic Feature extraction in which the information in each time dependant signal represented in stft manner. And Distance calculation in which some form of metric has to be used in order to obtain a match path. There are two types: A- Local distance: a computational difference between a feature of one signal and another means that calculating the local distance is required. The distance measure between two feature vectors is calculated using peak distance metric. Therefore, the local distance between feature vector A of signal 1 and vector B of signal 2, (Daniel Jurafsky, 2000) B- Global distance: the overall computational difference between an entire signal and another of possibly different length (C.H. Lee, F.K. Soong and K.K. Paliwal (Eds.), Kluwer, Boston, 1996).

To obtain a global distance, time alignment must be done by the fuction shown in equation 2

$$D(I,j)=\min[D(I-1,j-1),D(I-1,j),D(I,j-1)] +d(I,j).....\text{equ2}$$

The next step is to fine the best matching sample and it is the one, which has the lowest distance path aligning the input pattern to the template. Although utterances of the same word will have different durations, will differ in the middle, due to different parts of the words spoken at different rates and also Speech is a time-dependent process so the best match will depend on the rate the user says the word . so we applied , the simple global distance score for a path to calculate the sum of local distances that go to make up the path for the sound signal. (Daniel Jurafsky, James Martin, 2000 , matlab documentation).

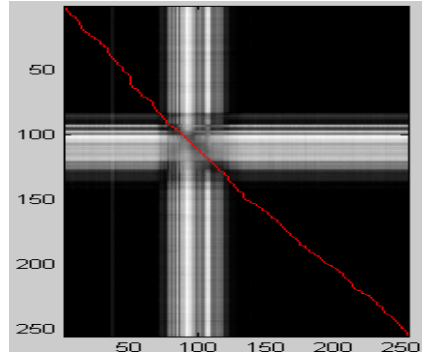


Fig. 6

The last step is to search the database for the words if the minimum path for the same word there will be a match between them shows in fig. 6 and the matched video will be displayed for the user but if there is no match as shown in fig 7 for another word recorded when compared with the word we are comparing with that is differ the signal not in a straight line this mean there a difference in utterances from the previous one or it is silent word as shown in fig 8

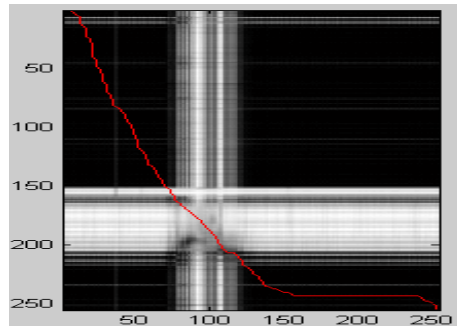


Fig.7

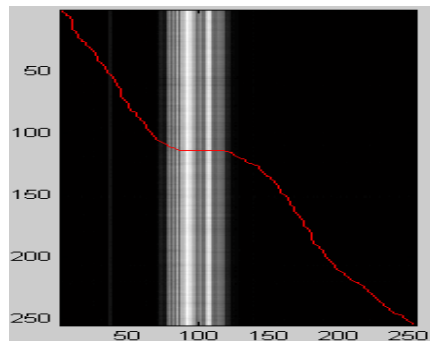


Fig.8

In order to see the results of the implemented algorithm and its performance we tested the algorithm by chosen a random sample of 29 words with x person voice (family group, day's group, numbers group) and save different pronunciations per word. We used these words as templates to compare with the words we will record. The implemented algorithm will cut the voice into segments and compare every segment with the template until the minimum cost is achieve if there is match between the word and the template then video will be displayed to show visual sign of the word that exist. Viewing the video depends on the

utterances of the word that will be compared with the template. if the minimum cost path is less than the number that is chosen according to the testd environment then there will be a match between them and the video will view the sign if the result is more than that number that were chosen then there is no match and no video will be displayed .We noticed the following constraint while testing our algorithm that the Matching paths cannot go backwards in time. Every frame in the input must be used in a matching path(Conrad F. Sabourin, 1994).

Table 1: **Testing and Validation**

	Test file	Minimum cost value
1	family	17.3383
2	father	21.4689
3	mother	24.3662
4	dad	14.0565
5	mam	20.3989
6	sister	21.2367
7	brother	14.7058
8	boy	54.8171
9	girl	53.0881
10	day	51.8623
11	saturday	24.8351
12	sunday	22.0040
13	moday	22.1573
14	tuesday	16.6635
15	wednesday	16.3265
16	thursday	40.4180
17	friday	17.8724
18	number	54.2557
19	zero	21.2147
20	one	13.5579
21	two	9.5038
22	three	9.8857
23	four	17.5501
24	five	22.1559
25	six	14.6983
26	seven	22.0773
27	eight	15.9612
28	nine	21.0276
29	ten	30.2510

4. Results

We observed that using different random words that is really different in the vowel sound get's the a good results using the algorithm and it also gave a good match , also choosing words that have the best cost 'minimum number' from the testing tables to use it in the software implementation will give a clear identification of the words to find the matched word when applying the program. We noticed that a headset microphone reduce's the noise around the speaker also it allow the speaker to have the microphone at the tip of the tongue all the time at same distance throw speacking and this is important to the accuracy of the results.

The best matching sample in our algorithm will be the one for which there is the lowest distance path aligning the input pattern to the template. Moreover, the simple global distance score for a path will be simply the sum of local distances that go to make up the path. fig 6 shows an example of DTW grid (Daniel Jurafsky, James Martin, 2000 , matlab documentation).

We tried to develop a method using speech recognition . We calculated the minimum distance between the entry word and the template that stored, by using the Dynamic time warping. We had made a brief

statements about our assumptions: The system is a speaker dependent ,The speaker must speak close to the microphone we test the words by headset microphone and table's microphone the distance between this last microphone's head and the mouth of who speak is 2 inches or three fingers .

The speaker should only speak the words that is stored as templates because it is recognised to the system , the environment of our testing is in silent with minimum noise , In spectrogram which indicates how the pitch changes, is visible during the vowels, but not during the consonants .(Lawrence Rabiner ,1993)

We notice the following problem in our test As shown in Fig.9 for example the word Saturday vowel is cut in the spectrogram also the word "aturday" as shown in Fig.10 .that the implemented algorithm will accept it because similar to the word Saturday with only simple difference in the spectrogram analysis.

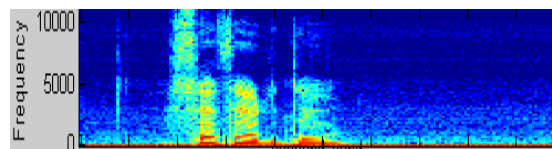


Fig.9

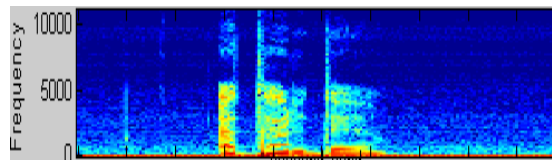


Fig.10

When we test words that have the same start or the same utterance the implemented algorithm didn't give good result there was conflicts (i.e. Saturday, Sunday and Monday or with six and seven)(C.H. Lee, F.K. Soong and K.K. Paliwal (Eds.), Kluwer, Boston, 1996 , matlab documentation).

5. Conclusion

This paper proposed an unsupervised implemented algorithm for learning sign language using speech recognition ,using this algorithm as shown in the tested results will save a time and effort in learning sign languages.

6. Acknowledgment

Authors would like to thank the presidency of both Al-Balqa' Applied University and the university of Jordan for encouragement of research and researchers at the both universities, also the authors would like to thank to the deanship of The Prince Abdullah Bin Gazi Faculty of Science and Information Technology.

7. References

- Daniel Jurafsky, James Martin, 2000 Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition.
- Lawrence Rabiner , Biing-Hwang Juang ,1993: Fundamentals of Speech Recognition: 1/e for (Prentice Hall Signal Processing Series).
- John Wilry&Sos, 1999 Rulph chasseur ,digital signal processing .
- John Wiley&Sons LtD,Baffin LaneChichester ,1997.Udo Zolzer, Digital Audio Signal Processing .

C.H. Lee, F.K. Soong and K.K. Paliwal (Eds.), Kluwer, Boston, 1996 Automatic Speech and Speaker Recognition: Advanced Topics.

X.D. Huang, Y. Ariki, M.A. Jack. Edinburgh c1990, Hidden Markov models for speech recognition, Edinburgh University Press.

Matlab 7 Image Processing Toolbox and documentation , Signal Processing Toolbox.

Conrad F. Sabourin, 1994, Computational Linguistics in Information Science.