International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 2 Issue 12, December-2013, pp: (8-11), Available online at: www.erpublications.com

Vowel recognition for hearing impaired person

Sheetal K. Bhongle, Dr. C. Y. Patil

Instrumentation & Control Engineering Department
College of Engineering, Pune, India

Abstract: The communication process of human being is deals with the facial expression & lip moments. The hearing impaired, aside from using residual listening to communicate with other people, can also use lip reading as a communication tool. As the hearing impaired learn the lip reading using a computer-assisted lip-reading system, they can freely learn lip reading without the constraints of time, place or situation a computer-assisted lip-reading system for phonetic pronunciation recognition of the correct lip-shape with an image processing method, object-oriented language and neuro-network. This system can accurately compare the lip image of Mandarin phonetic pronunciation using self-organizing map neuro-network and extension theory to help hearing impaired correct their pronunciation. In automatic recognition of Cued Speech, lip shape and gesture recognition are required. Moreover, the integration of the two modalities is of great importance. In this study, lip shape component is fused with hand component to realize Cued Speech recognition. Using concatenative feature fusion and multi-stream HMM decision fusion, vowel recognition, consonant recognition, and isolated word recognition experiments have been conducted.

Introduction

The visual information is widely used to improve speech perception or automatic speech recognition. With lip reading technique, speech can be understood by interpreting the movements of lips, face and tongue. In spoken languages, a particular facial and lip shape corresponds to a specific sound (phoneme). However, this relationship is not oneto-one and many phonemes share the same facial and lip shape (visemes). It is impossible, therefore to distinguish phonemes using visual information alone. Without knowing the semantic context, one cannot perceive the speech thoroughly even with high lip reading performances. To date, the best lip readers are far away into reaching perfection. Cued Language (Fleetwood and Metzger, 1999)) uses hand shapes placed in different positions near the face along with natural speech lip reading to enhance speech perception from visual input. This is a system where the speaker faces the perceiver and moves his hand in close relation with speech. The hand, held flat and oriented so that the back of the hand faces the perceiver, is a cue that corresponds to a unique phoneme when associated with a particular lip shape. A manual cue in this system contains two components: the hand shape and the hand position relative to the face. Hand shapes distinguish among consonant phonemes whereas hand positions distinguish among vowel phonemes. A hand shape, together with a hand position, cues a syllable. Although many sensory substitution devices have been developed as speech training aids for the hearing impaired, most have inherent problems and fall short of expectations. Both tactile and visual aids have potential benefits in speech therapy programs for the hearing impaired, but this potential is largely unrealized. The main problems associated with these devices are difficulty in interpreting the displays, lack of consistency between displayed parameters and variables required for speech production, and generally inadequate feedback to the user for speech correction.

Concerning the face-to-face communication, the channels used by the hearing-impaired people can be classified in three categories:

- The hearing-impaired that use lip-reading, as a complement to the voice. People using hearing aid (only one million people in France) or cochlear implant (550 implantations per year) are exploiting both visual and auditory channels. Even if the auditory channel is deficient, these people can perceive some auditory residual information.
- The profoundly deaf people of the oralist category, whose auditory channel is severely damaged, can have their lip-reading ability enhanced by using the Cued Speech method, which is at the heart of this project.
- The profoundly deaf people of the gestural category use the Sign Language, which is very well known, but not considered in this project.

Speech is multimodal dimensions and in the context of automatic processing. Indeed, the benefit of visual information for speech perception called “lip-reading” is widely admitted.

However, even with high lip reading performances, without knowledge about the semantic context, speech cannot be thoroughly perceived. The best lip readers scarcely reach perfection. On average, only 40 to 60% of the phonemes of a given language are recognized by lip reading, and 32% when relating to low predicted words with the best results obtained amongst deaf participants - 43.6% for the average accuracy and 17.5% for standard deviation with regards to words. The main reason for this lies in the ambiguity of the visual pattern. However, as far as the orally educated deaf people are concerned, the act of lip-reading remains the main modality of perceiving speech. This led to develop the Cued Speech system as a complement to lip information. CS is a visual communication system that makes use of hand shapes placed in different positions near the face in combination with the natural speech lip-reading to enhance speech perception from visual input. CS is largely improving speech perception for deaf people.

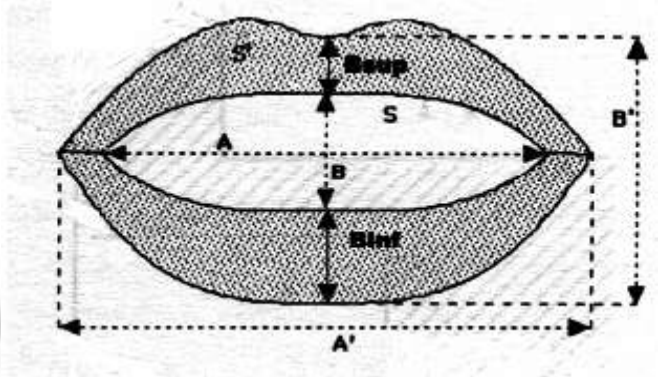
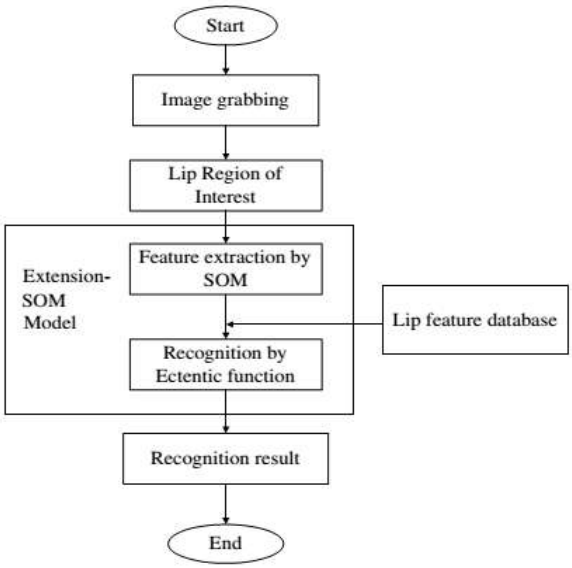


Figure: Showing the measurement of lip movement

To evaluate the contribution of both the upper and lower lip pinching to recognize CV syllables in a HMM recognition test and more precisely to better modelize the C4 consonant group. So, the HMM recognition test is based this time on eight lip parameters. In addition to the six parameters used in the previous experiment, the pinching of upper and lower lips (respectively Bsup and Binf) is measured at one point (more precisely in the mid-lips).



MATLAB is a high-performance language for technical computing especially those with matrix and vector formulations, in a fraction of the time it would take to write a program in a scalar non interactive language such as C++ or so on. It integrates computation, visualization, and Programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.

- Step1. Read image from input device (Camera)
- Step2. Resize all the images to fit 150x140 pixels (optimal size).
- Step3. Find the edges (boundaries).For this two filters were used. For the x direction x = [0 -1 1]. For the y direction y = [0 1 -1] shows two images of the result with the x-filter and y-filter" .

- Step 4. Dividing two resulting matrices (images) dx and dy element by element and then taking the atan (tan−1) to get gradient orientation.
- Step 5. Feature extraction from image .
- Step 6. Classification using neural network.
- Step 7. Vowel recognition.

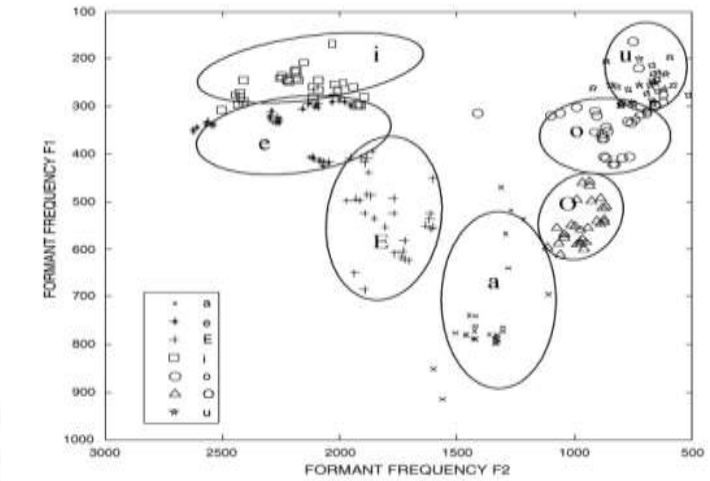


Figure:Classification of vowels using neural network for low tone

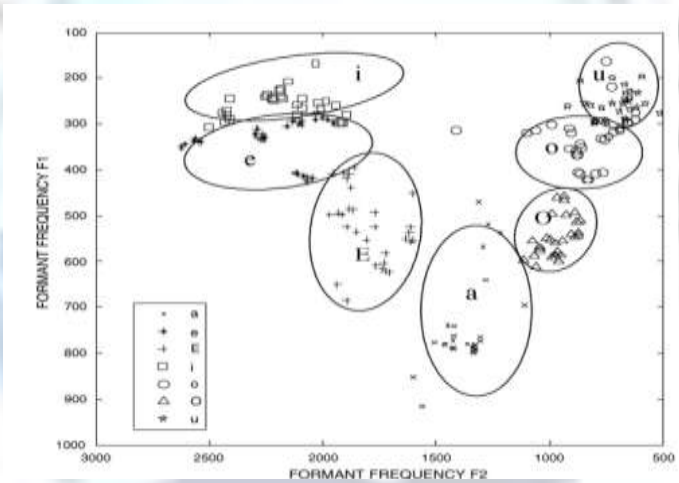


Figure: Classification of vowel using neural network for mid tone

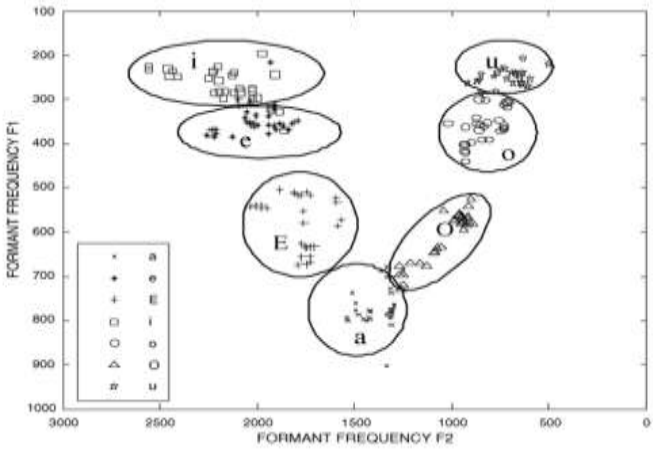


Figure: Classification of vowel by using neural network for lower tone

In this way we can classify the vowel based on lip movement. The speech features used for the development of the ASR systems are the fundamental frequency (F0), and the first two formant frequencies (F1 and F2). The two systems were developed using FL and ANN MatLab toolboxes. The performance of the two systems has been evaluated with the use of percentage accuracy and confusion matrixes. The ANN system appears to have higher accuracy than the FL system on train data set, however, the FL system performed better on the test data.

Conclusion

The development of a practical ASR system for SY language based on ANN approach will require larger amount of language resources and computational time. Such language resources, e.g. speech database and annotated speech files, are not yet widely available. Also, the interpretation of ANN model presents some challenges. The FL based model, however, facilitates the extraction of pertinent rules that could be of benefit in further scientific study of the language. To this end, it seems that a FL based approach has an edge over the ANN approach for the development of practical ASR system for the sign language. Recognition of syllables, words and continuous speech are the area of further research work, in which the principle of this work could be extended. Our aim is to carry out experiment using a neuro-fuzzy model with the aim to integrate the benefits of the two modelling approaches for hearing impaired persons.

References

[1]. S.K. Pal, D.D. Majmder, “Fuzzy sets and decision making approaches in vowel and speaker recognition”, IEEE Transactions on Systems, Man and Cybernetics(1977) 625–629.
[2]. R. De Mori, R. Gubrynowicz, P. Laface, “Inference of a knowledge source for the recognition of nasals in continuous speech”, IEEE Transactions on Acousitc Speech and Signal Processing 270 (5) (1979) 538–549.
[3]. Zimmer A., Dai, B., and Zahorian, S, (1998) “Personal Computer Software Vowel Training Aid for lhe Hearing Impaired”, lntemational Conference on Acoustics, Speech and Signal Processing, Vol6, pp. 3625-3628 .
[4]. Zahorian S., and Jagharghi, A., (1993) “Spectral-shape features versus formants as acoustic correlates for vowels”, J. Acoust. Soc. her. Vo1.94, No. 4, pp. 1966- 1982.
[5]. SI Evangelos et al., (1991) “Fast endpoint Detection Algorithm for Isolated word recognition in office environment”, lntemational Conference on Acoustics. Speech, and Signal Processing, pp. 733-736.