Vowel recognition of English and German language using Facial movement (SEMG) for Speech control based HCI

Sridhar P Arjunan, Dinesh K Kumar, Wai C Yau  
School of Electrical and Computer Engineering, RMIT University  
GPO Box 2476V, Melbourne, VIC 3001, Australia

Hans Weghorn  
Information technology, BA-University of Cooperative Education  
Rotebhlplatz 41, 70178 Stuttgart, Germany

s3099587@student.rmit.edu.au  
dinesh@rmit.edu.au
Introduction

- Research on alternative methods of computer control has focused on three types of body functions: speech, electrical brain activity and eye movements.

- The aim is to convert and analyse the activities of muscle movements during speech in order that it can be recognised by machine to control.

- One fundamental challenge facing the development of speech recognition technology is to make the systems robust in noisy acoustic environments.
Facial movement identifiable by muscle

- The face can communicate varied personal information including subjective emotion, communitive intent, and cognitive appraisal.
- The facial musculature is a three dimensional assembly of small, independently controlled muscular slips performing a variety of complex and important facial functions such as speech, mastication, swallowing and mediation of motion.

Facial muscles relates to speech

- Speech is one of the important functions of the facial musculature.
- A preliminary study was performed to ascertain the presence of speech information within the muscle activity signals from facial muscles (Chan et al).

Objective

- To develop a technique for a reliable speech control based HCI by classification of features of facial electromyography without any audio signals.

Issues on this objective:
- The difference in patterns for different individuals
- The variations in style and speed of speech of individuals (Inter subject variations)
- The variations in the muscle activity during speaking different languages (other than the native language)
- Determining method to extract features of facial EMG during speech for classification

A preliminary work has been done on this objective to recognise speech using facial muscle activity which makes this speech recognition ‘Voiceless’
What is Electromyography?? EMG!!!!

- Electromyography (EMG) is the study of muscle function through the inquiry of electrical signals the muscles emanate.
- EMG represents the level of muscle activity so that it can be used to recognize human movement.

(Source: AMLAB Manual)
**Methodology – Experiments**

- A four channel, portable, continuous recording MEGAWIN equipment (MEGA Electronics, Finland) was used for this purpose. Raw signal sampled at 2000 samples/second was recorded.
- The muscles selected were the right side *Zygomaticus Major, Masseter & Mentalis* and left side *Depressor anguli oris*.
- The participants were native German speakers with English as their second language.
- During the recordings, the participants spoke the German vowels and English vowels.
- The experiment was repeated for ten times. A suitable resting time was given between each experiment.
- The participants were asked to vary their speaking speed and style to get a wide based training set.
Methodology - Block Diagram

EMG Recording Equipment → Determine RMS of the EMG signal → Calculating Integral of the RMS values → Normalisation of the Integral RMS values

- Megawin, Finland

Overall recognition accuracy is 80%

Classifying the data using trained Neural Network

Training the normalised values using Neural Network Architecture

Where 's' is the window length and $f(s)$ is data within the window.

$$\text{RMS} = \sqrt{\frac{1}{s} \sum_{s} f^2(s)}$$ (1)
Raw EMG signal and its RMS

![EMG Signal and RMS](image)
Integral RMS and 3 D plot of Normalised IRMS

Integral RMS plot and 3-dimensional plot of the normalized values of different vowels a) of English Language b) German Language

s3099587@student.rmit.edu.au
dinesh@rmit.edu.au
Cluster Analysis

Silhouette plot showing the silhouette values of the clusters

English vowels

German vowels
Architecture of Neural Network

- The designed ANN used two hidden layers of Feed forward Network
- The number of neurons in the hidden layer was iteratively chosen according to the convergence of the training.
- The error rate of 0.01 was kept for the training.
Architecture of Neural Network

Vowel /a/

<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>A3</th>
</tr>
</thead>
</table>

Target \{0 0 1\}

Vowel /i/

<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>A3</th>
</tr>
</thead>
</table>

Target \{0 1 0\}

Vowel /u/

<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>A3</th>
</tr>
</thead>
</table>

Target \{1 0 0\}
## Classification Results

### Table 1. Classification results for different participants uttering English Vowels

<table>
<thead>
<tr>
<th>Vowels</th>
<th>No of Utterances used for testing</th>
<th>Correctly Classified Vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Participant 1</td>
</tr>
<tr>
<td>/a/</td>
<td>5</td>
<td>3(60%)</td>
</tr>
<tr>
<td>/e/</td>
<td>5</td>
<td>4(80%)</td>
</tr>
<tr>
<td>/u/</td>
<td>5</td>
<td>5(100%)</td>
</tr>
</tbody>
</table>

### Table 2. Classification results for different participants uttering German Vowels

<table>
<thead>
<tr>
<th>Vowels</th>
<th>No of Utterances for testing</th>
<th>Correctly Classified Vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Participant 1</td>
</tr>
<tr>
<td>/a/</td>
<td>5</td>
<td>4(80%)</td>
</tr>
<tr>
<td>/i/</td>
<td>5</td>
<td>5(100%)</td>
</tr>
<tr>
<td>/u/</td>
<td>5</td>
<td>5(100%)</td>
</tr>
</tbody>
</table>
Observations

- The results indicate that this technique can be used for the classification of vowels for the native and foreign language, in this case, German and English.
- The investigation reveals the suitability of the system for English and German, and this suggests that the system is feasible when used for people speaking their own native language as well as a foreign language.
- The promising results obtained in the experiment indicate that this approach based on the facial muscles movement represents a suitable, reliable method for classifying vowels of single user without regard to the speaking speed and style in different times for different languages.
- It should be pointed that this method at this stage is not being designed to provide the flexibility of regular conversation language, but for a limited dictionary only, which is appropriate for simple voice control systems.
Future work

- Large set of phoneme with many subjects
- Multilingual speech recognition e.g. English, German
- Classification technique using Fractal and Wavelet Networks to extract more features of EMG signal with respect to subtle human actions.
THANK YOU VERY MUCH.......