A Real-Time Implementation of a Speaker Verification System

Author – Paul Faggion
Supervisor – Dr. E. Ambikairajah

Speaker Verification

The technique of speaker verification involves the authentication of a speaker from his/her voice print. A speaker verification system must decide if a speaker is the person who makes the claim.

1. Feature Extraction

Features of a voice signal characterise the speaker. They are significant parameters that differentiate one speaker from another. Speaker verification systems try to extract significant features, like pitch period, so a speaker characterisation can be made.

Pitch contours are a good indication if the speaker is genuine. Below is a standard pitch contour for a male speaker:

![Pitch Contour for the phrase “We-were-away-a-year-ago”](image)

2. Warping

The templates and the newly acquired input speech have to be the same length for comparison. Two warping techniques are employed for maximum results. The first is a linear interpolation which resamples the input parameters to be the same length as the template. The other is Dynamic Time Warping (DTW) which warps the input parameter contour by correlating phonetic content between the input and the template.

Below is a linearly warped input pitch contour which is then subtracted from the reference contour. If the difference is close to zero then the

![Results](image)

System Specifications

The system, implemented on MATLAB 5.3, was designed to work for male speakers who have a pitch period between 60 and 250Hz. One pitch contour and six Cepstral Coefficients are extracted from the speech and compared against pre stored templates. This yielded seven parameters for comparison.

Three text dependent systems were implemented in this thesis where two operate in real time and one in non real time. The real time systems verify from the pitch contour, whereas the non real time system uses pitch comparisons as well as six cepstral coefficients.

Four speakers participated in testing. The verification time was 15 seconds for the real time and 60 seconds for the non real time system.

<table>
<thead>
<tr>
<th></th>
<th>Real Time System</th>
<th>Non Real Time System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genuine Verification</td>
<td>80%</td>
<td>65.7%</td>
</tr>
<tr>
<td>Impostor Verification</td>
<td>60%</td>
<td>27.5%</td>
</tr>
</tbody>
</table>

Conclusion

The more features that can be extracted from the speech sample, the tighter the system is in accepting the user to be the claimant. However the trade off is lower self-verification rates and longer times to verify.