Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Czech Republic

ZÁPADOČESKÁ UNIVERZITA V PLZNI

P. Císař, J. Trojanová, M. Železný

MLMI 07

Visual Hybrid Parameterization

INTRODUCTION

Visual part of the audio-visual speech system improves speech recognition when the acoustic part of it is damaged e.g. by noise.
The most common method for visual parameterization is image-based and shape-based parameterization.
Aim of our work is to experiment with both parameterization and to choose a suitable combination of both approaches with using information provided by human lip-reading experts.
The parameterization should include description of the lip shape and objects inside a mouth (teeth and a tongue).

ABSTRACT

We present design of the visual speech parameterization for audio-visual speech recognition.
Iinformation from human lip-reading expert was used for development of new parameterization.
Experiments were performed on database UWB-05-HSCAVC.
The hybrid parameterization was compared with DCT parameterization.
The results with our parameterization improve audio-visual speech recognition by 4%.

CONCLUSION

We introduced new hybrid parameterization. This parameterization is based on description of the lip shape and objects inside the mouth.

The result obtained with our parameterization shows that information from inner objects of the mouth improve audio-visual speech recognition.

We found parameterization which improve audio-visual speech recognition by 4% in contrast to DCT parameterization which is mostly used for visual part of speech.

PRE-PROCESSING

Database UWB-05-HSCAVC
-contains 100 people
-each saying 200 sentences

Type of data
- static data - each frame
- dynamic data - sequence of frames

Region of interest (ROI)
- convert to Cr/Cb color space
- treshold image using information obtained from background color
- find eyes by template matching
- distance of eyes used to find ROI

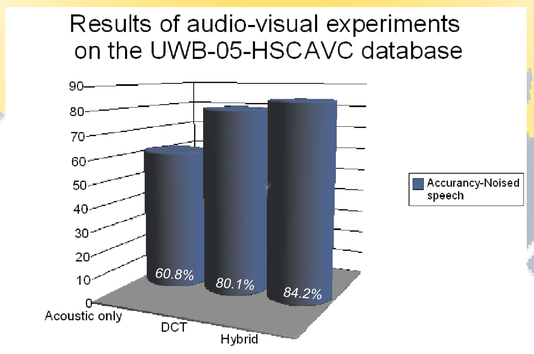
Finding region of interest (ROI):
a) background sample,
b) tresholded image
c) find eyes
d) from distance of eyes => ROI

EXPERIMENTS

- made on database UWB-05-HSCAVC
- audion signal damaged by noise
- training - 150 diferent sentences
- testing - 50 same sentences
- HMM multimodal speech recognition system with zero-gram language model

- base line visual DCT parameterization

Results are compared
- acoustic only 60.8%
- acoustic + DCT 80.1%
- acoustis + Hybrid 84.2%

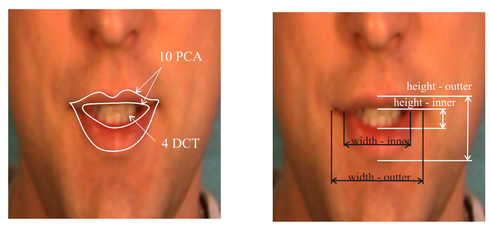


PARAMETERIZATION

Basic parameters of the lips
- height of the lips inner and outer
- width of the lips inner and outer

Shape-based paramatererization
- active shape model
- 32 points controled by principal component analysis (PCA)
- lips shape model is made of 10 PCA

Inner lip parameterization
- relation between teeth and tongue is very important for lip-reading
- discrete cosine transformation (DCT) coefficients used for inner objects



a) point model connected with splines
b) basic parameters of the lips

HYBRID PARAMETERIZATION

Final hybrid parameterization contains 18 visual parametrs:

4 basic lip properties (height, width of lips)

10 PCA parameters (shape of the lips)

4 DCT coefficients (mouth inner objects)