VQ SI in Car Environment for Personalized Infotainment

A VQ Speaker Identification System in Car Environment for Personalized Infotainment

Javier R. Saeta  Christian Koechling ¹  Javier Hernando ²

(1) Robert Bosch, GmbH
Hildesheim, Germany

(2) TALP Research Center
Universitat Politecnica de Catalunya, Barcelona, Spain
VQ SI in Car Environment for Personalized Infotainment

Outline

• VHE Middleware project
• System design
• Speaker recognition system
  • Architecture
  • Design parameters
• Database & Scenarios
• Tests
• MPEG-7
• Conclusions & Future work
VQ SI in Car Environment for Personalized Infotainment

VHE - Middleware project

- Smart Watch
- Mobile Phone
- Chipcard
- PDA
- Fingertip
- Voicepattern
VQ SI in Car Environment for Personalized Infotainment

System design

- Recognition with a few words in a car environment
- Not a security matter, not physical access
- The user is, in principle, the car driver
- Limited number of users:
  - a single person
  - a family
  - a group of people
Speaker recognition system. Architecture:

- Feature extraction
- Pattern matching (VQ)
- Decision
- Enrollment
- Speaker models

You are...

- Algorithm for 10-20 users of a car
- Fast recognition
- Low complexity
- Text independent
Speaker recognition system. Design parameters:

- Sampling rate: 22,050 kHz, with 16 bits
- Preemphasis: $H_{\text{pre}} = 1 + 0.97z^{-1}$
- Framing every 10 ms, Hamming windowed
- 12 MFCC + Energy + $\Delta + \Delta \Delta$
- Euclidean distance
- Vector quantization, LBG algorithm and 16 codewords
VQ SI in Car Environment for Personalized Infotainment

Outline

• VHE Middleware project
• System design
• Speaker recognition system
  • Database & Scenarios
  • Tests
• MPEG-7
• Conclusions & Future work
Database

Database of 23 speakers

Training sentences

21
2

- Short sentence: 1-2 words
- Medium sentence: 6-8 words
- Long sentence: 21 words
VQ SI in Car Environment for Personalized Infotainment

Scenarios

Scenario I
Office
• SNR ~ 20 dB

Scenario II
Car
• Variable SNR
• SNR ~ 7 dB at 50 km/h
Parameters adjustment. Number of codewords.

**Scenario I**

L-M (Training with a Long sentence-Test with a Medium length sentence)

**Scenario II**

MN-MN (Training with a Medium sentence with car noise-Test with a Medium length sentence with car noise)
Parameters adjustment. Threshold value.

![Graph showing parameters adjustment and threshold value for short and medium sentences.](image-url)
VQ SI in Car Environment for Personalized Infotainment

Testing the system. Scenario I.

L-M (Training with a Long sentence-Test with a Medium length sentence)
M-L (Training with a Medium sentence-Test with a Long sentence)
M-S (Training with a Medium sentence-Test with a Short sentence)
VQ SI in Car Environment for Personalized Infotainment

Testing the system. Scenario II (1)

Some additional SNR:

- *Parking*: 13.21 dB
- *Accelerating*: 8.49 dB
- *Stopped with windows opened*: 10.94 dB
- *With indicators*: 10.29 dB
- *With windows opened at 170 km/h*: 1.27 dB
VQ SI in Car Environment for Personalized Infotainment

Testing the system. Scenario II (2)

- Car at 50 km/h
- SNR ~ 7 dB

Scenario I
L-M (Training: Long sentence- Test: Medium sentence)

Scenario I / Scenario II
L-MN (Training: Long sentence- Test: Medium sentence with car noise)

Scenario II
LN-MN (Training: Long sentence with car noise- Test: Medium sentence with car noise)

- It is better to train and recognize in the same noisy background
VQ SI in Car Environment for Personalized Infotainment

Outline

• VHE Middleware project
• System design
• Speaker recognition system
• Database & Scenarios
• Tests
• MPEG-7
  • Introduction
  • Profile definition
  • Example
• Conclusions & Future work
VQ SI in Car Environment for Personalized Infotainment

System design

- Recognition with a few words in a car environment
- Not a security matter, not physical access
- The user is, in principle, the car driver
- Limited number of users:
  - a single person
  - a family
  - a group of people
MPEG-7: Introduction

- Standard for the Multimedia Content Description Interface
- Description of audiovisual information: pictures, video, speech, audio, graphics, 3D models and synthetic audio
- Descriptors, Description Schemes, Coding Scheme and DDL
- Digital libraries, multimedia directory services, broadcast media selection...
VQ SI in Car Environment for Personalized Infotainment

Definition of a user profile using MPEG-7

• Through MPEG-7, it is possible to establish user preferences
• MPEG-7 uses XML language
• MPEG-7 lets connection between a user and an Internet music database
• You will be able to receive at your own car information related to music: credits, lyrics, videos, next tours, awards, merchandising...
• It is expected the system could be capable in a future of ‘learning‘ the user preferences
Definition of a user profile using MPEG-7. Example:

```xml
<UsagePreferences allowAutomaticUpdate="false">
  <FilteringAndSearchPreferences>
    <ClassificationPreferences preferenceValue="100">
      <Language>en</Language>
      <Genre csName="Escort2_4:Content" csLocation="http://www..." csTermId="1.4">
        <Label xml:lang="en">Jazz</Label>
      </Genre>
    </ClassificationPreferences>
    <ClassificationPreferences preferenceValue="120">
      <Genre csName="Escort2_4:Content" csLocation="http://www..." csTermId="1.2">
        <Label xml:lang="en">Classical</Label>
      </Genre>
    </ClassificationPreferences>
  </FilteringAndSearchPreferences>
  <PreferenceCondition>
    <Place>
      <PlaceName xml:lang="en">Tokyo</PlaceName>
      <Country>JP</Country>
    </Place>
  </PreferenceCondition>
</UsagePreferences>
```
Definition of a user profile using MPEG-7. Example:

```xml
<CreationPreferences>
  <Creator preferenceValue="180">
    <Role><Label>singer</Label></Role>
    <Individual>
      <Name>
        <GivenName>Michael</GivenName>
        <FamilyName>Jackson</FamilyName>
      </Name>
    </Individual>
  </Creator>
  <Creator preferenceValue="200">
    <Role><Label>singer</Label></Role>
    <Individual>
      <Name>
        <GivenName>U2</GivenName>
      </Name>
    </Individual>
  </Creator>
  <DatePeriod>
    <TimePoint>1998-01-01</TimePoint>
    <Duration>P1035D</Duration>
  </DatePeriod>
</CreationPreferences>
```
VQ SI in Car Environment for Personalized Infotainment

Outline

• VHE Middleware project
• System design
• Speaker recognition system
• Database & Scenarios
• Tests
• MPEG-7
  • Introduction
  • Profile definition
  • Example
• Conclusions & Future work
Conclusions

• The system works reasonably well in a noisy environment
• It is hard to obtain high recognition rates in mismatched training and test conditions
• The length of the training sentences affects the robustness whereas the length of the testing sentences affects also the computational cost
• It is possible to define a user profile using MPEG-7 in order to obtain information from an Internet music database
Future work

• Emotional and physical state (Lombard effect, driving effect...), families and change of voice over time
• Possible use of other recognition techniques instead of vector quantization
• Use of an array of microphones inside the car
• Create a mechanism to join the speaker recognition with the loading of the personal profile
• The system can also be enlarged for regular passengers
A VQ Speaker Identification System in Car Environment for Personalized Infotainment

Javier R. Saeta    Christian Koechling ¹    Javier Hernando ²

(1) Robert Bosch, GmbH
Hildesheim, Germany

(2) TALP Research Center
Universitat Politecnica de Catalunya, Barcelona, Spain