Channel Error Protection and OOV Detection for Distributed Speech Recognition

- FACE Workshop

Zheng-Hua Tan
Center for PersonKommunikation (CPK), Aalborg University
zt@cpk.auc.dk

Outline

- FACE project
- ETSI-DSR standard - Aurora
- DSR: Frame-based channel error protection
- DSR: FER-based out-of-vocabulary detection
- Discussion
FACE - Future Adaptive Communication Environment

FACE project

- Devices
- Networking - Adaptive and Scalable

Advanced search?
Hand-writing instead of talking?

FACE project - cont.

- Speech interaction - User-friendly

Network-based spoken language processing
Objective: Providing users with adaptive and flexible access to a given service via a variety of networks with user-friendly interfaces.

FACE

Short-range networks - WING
Wide-range cellular - CSys

UWB - A&P

Speech interaction - SMC

- WP3: Re-configurable spoken language system
  - DSR
  - SLP in networks

ETSİ-DSR standard - Aurora

Distributed Speech Recognition

- DSR client-server architecture
  - A typical DSR system consists of two functional parts: a Front-End (FE) and a Back-End (BE)

- Since IP and mobile channel are error-prone, transmission error protection is essential

The Standard
- published by ETSİ in February 2000
- aimed at dealing with the degradations of speech recognition over mobile channels
The Aurora document shows, however, for a 4dB C/I channel, the recognition rates relatively degrade from 10.0% to 16.2% for different tasks in comparison to the case of transmission without errors.

So, a more robust scheme against transmission errors is needed.

Block diagram of the front-end algorithm in Aurora

---

Channel Error Protection

- **Frame-pair architecture**: two frames are grouped together and protected by a 4-bit CRC

- increasing FER

- **Frame-based architecture**: a 4-bit CRC is calculated to protect each frame independently

- maintaining FER

- at the cost of only a marginal increase in bit-rate, from 4,800 bits/s to 5,000 bits/s
**DSR: Frame-based channel error protection - cont.**

- Frame Error Rate

The distribution of packet loss by length

- Recognition tasks - Danish digits and city names

- Four error distributions:
  - randomly distributed bit-errors

The WER for the Danish digits task decreases from 8.5% to 7.1% - an improvement of 17%

The WER for the city names task decreases from 34.2% to 30.8% - an improvement of 10%
DSR: Frame-based channel error protection - cont.

- Four error distributions – cont.
  - GSM error patterns

- UMITS statistics
  - The WER of Danish digits task decreases from 3.3% to 2.5% - an improvement of 25%
  - The WER of the city names task decreases from 26.6% to 23.9% - an improvement of 10%

DSR: FER-based OOV detection

- OOV - Out-Of-Vocabulary
  - Vocabulary size is limited for a speech recognition application

- OOV Detection
  - Reject the H0 hypothesis if
    \[ LR(O) = \frac{p(O \mid H_0)}{p(O \mid H_1)} < T \]

  - H0: represents one of the IV words
  - H1: represents OOV words modeled by one filler model
  - O is a speech signal observation
  - T is the threshold of the test

  Two kinds of errors
  - False rejection
  - False acceptance

09/12/2002
DSR: FER-based OOV detection - cont.

- Transmission errors adversely affect the distribution of the likelihood of both the IV models and the filler model in two ways:
  - increasing the standard deviation
  - shifting the mean

- A FER-dependent threshold for OOV detection
  - the threshold is modelled as a fourth-order polynomial function of the FER

- Experiments
  - IV: Danish digits; OOV: city names
  - randomly distributed bit-errors
  - Aimed at a constant FR

- Results

False rejection rate vs. AWGN channel BER values

Recognition rate vs. AWGN channel BER values

False acceptance rate vs. AWGN channel BER values
Discussion

• The work that has been done so far
  - a frame-based error protection scheme
    • is able to better maintain the recognition rates, compared to the ETSI-DSR frame-pair scheme
    • at the cost of only a marginal increase in bit-rate, from 4,800 bits/s to 5,000 bits/s for which there is plenty of bandwidth available in GSM and higher bandwidth wireless channels
  - a FER-based OOV detection
    • has been proved successful in maintaining a constant false rejection rate across a range of transmission error rates

• Further research
  - The principle of exploiting information about the channel can be used to adapt dialogue, grammars and vocabularies - re-configurable spoken language system.

Thanks!