語音事件偵測器的設計與估算
On Designing and Evaluating Speech Event Detectors

Author:Jinyu Li and Chin-Hui Lee

Professor:陳嘉平

Reporter:吳柏鋒

Outline

- Introduction
- Detector design
- Experiment

摘要

- 簡介
- 偵測器的設計
- 實驗

Introduction

- Event detection is a critical component of a recently proposed automatic speech attribute transcription (ASAT) paradigm.

- A good detector needs to effectively detect speech attributes of interest while rejecting extraneous events.

簡介

- <u>事件偵測</u>在最近提出的ASAT中是一個很重要的元件

- 一個好的偵測器必須能夠偵測出有效的語音屬性並拒絕無關的事件

Introduction

- These "<u>events</u>" are usually low level speech attributes related to information required to form higher level "<u>evidences</u>", so they can then be combined to detect phones, words and sentences.

簡介

- 這邊的"事件"通常都是指低階的語音屬性，透過相關資料可以做轉換形成高階的"證據"，並將多個證據作結合成為偵測音素、字、和句子

Introduction

- We compare <u>frame</u> and <u>segment</u> based detector by their properties in detecting manners of articulation.

- Proposed new mechanism to evaluate detector performance.

簡介

- 我們由<u>音框基礎偵測器</u>與<u>音段基礎偵測器</u>兩者在發音行為上的特性作為比較之依據
- 並提出一個新的估算偵測器效能的方法

Detector design

- **Frame and segment based detector:**

(1) frame based detector:

realized with artificial neural networks (ANNs).

(2) Segment based detector:

built by combining frame based detectors, or with segment models, such as hidden Markov models (HMMs)

偵測器設計

- **音框基礎與音段基礎偵測器:**

(1)音框基礎偵測器

實現在ANN上

(2)音段基礎偵測器

可以由結合音框基礎偵測器或者使用音段模型，例如HMM

Detector design

(1)Frame based detector

- ANN based detectors is that the output scores can simulate the *a posteriori probabilities of an attribute given the* speech signal.

- Strictly speaking, these event “detectors” can also categorize each speech frame into one of the competing attributes.

偵測器設計

(1)音框基礎偵測器

- ANN為基礎的偵測器，主要會將語音訊號的提供的屬性轉換成為一個事後機率

- 嚴格來說，這些事件偵測器可將每個音框分類到對應的屬性

Detector design

(2) Segment based detector

- We train two HMMs, one for a target event and the other for all other competing events to decode the utterance.

- The segments that are recognized with the target label mapping to the detected target events.

偵測器設計

(2)音段基礎偵測器

- 我們訓練兩個HMMS，一個是用在處理目標事件，而另外一個是用在其他的競爭事件，並利用他們來對句子做解碼
- 根據目標標籤對應的偵測目標事件來作為音段的辨識

Detector design

- **Performance measurement**
  - For the fricative event, the reference and detected strings are listed as ref.mlf and det.mlf

ref.mlf	det.mlf
#!MLF!#	#!MLF!#
"*/si1039.lab"	"*/si1039.rec"
0 8 fricative -3.1	0 13 fricative -13.1
8 13 fricative -10.0	13 21 non -12.1
13 21 non -12.1	21 24 fricative -16.0
21 30 fricative -34.2	24 30 fricative -18.2
30 41 non -19.0	30 41 non -19.0
41 45 fricative -3.0	41 45 fricative -3.0
.	.

偵測器設計

➢效能偵測

- 在摩擦音事件中， 參照與偵測字串為 ref.mlf 和det.mlf

ref.mlf	det.mlf
#!MLF!#	#!MLF!#
"*/si1039.lab"	"*/si1039.rec"
0 8 fricative -3.1	0 13 fricative -13.1
8 13 fricative -10.0	13 21 non -12.1
13 21 non -12.1	21 24 fricative -16.0
21 30 fricative -34.2	24 30 fricative -18.2
30 41 non -19.0	30 41 non -19.0
41 45 fricative -3.0	41 45 fricative -3.0
.	.

Detector design

- In HTK, the detection false alarms (FA) and hits can be obtained with its HResults tool.
- If the reference label lies between the start and end times of a recognized segment with the same label, we regard it as a hit.
- For every detected segment, we consider it as a FA only if the center of the recognized segment falls into the reference segment with a different label.

偵測器設計

- 在HTK中，我們可從HResult tool中獲得誤警報(FA)和擊中(hit)
- 如果辨識音段在開始到結束這段時間內的標籤與參照標籤相同時，我們稱為擊中(hit)
- 若每個偵測音段，其中間辨識音段與落在參照音段的標籤不同時，稱為FA

Detector design

➢ Detector optimization

- In order to enhance detector performance, we can use discriminative speech parameters and models with detailed acoustic resolution and refined content dependency.

- For MCE training, the LLR is defined as:

$$LLR(O) = \log L(O|\Lambda_0) - \log L(O|\Lambda_1)$$

偵測器設計

➢ 偵測器最佳化

- 我們可以利用含有詳細聲學分析與內容相依的<u>個別的語音參數</u>和<u>模型</u>來加強偵測器效能

- 針對MCE(minimum classification error)訓練

  定義LLR為:

$$LLR(O) = \log L(O|\Lambda_0) - \log L(O|\Lambda_1)$$

Experiment

- Use the **TIMIT** database

- We design a experiment for comparing frame based ANN and segment HMM detectors for manner of articulation attributes.

實驗

- 使用TIMIT資料庫
- 我們設計一個實驗，比較以音框為基礎的ANN和音段HMM的偵測器

Experiment

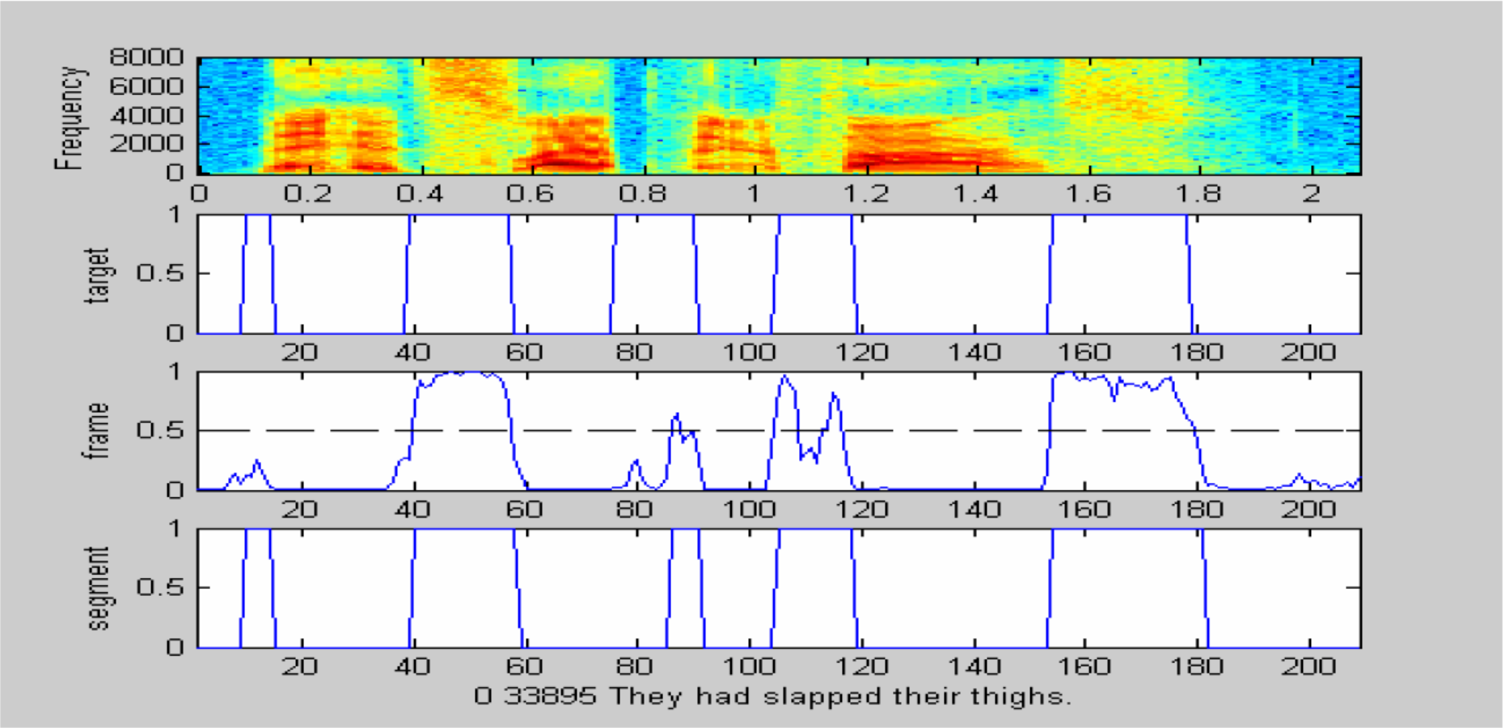

Figure 1 Detection curves of the ANN frame and HMM segment based detectors for the fricative attribute

Experiment

- **Detection with known segment boundaries**

  -sum over all frame scores of the ANN detector to smooth noisy outputs to some extent.

- **Detection with unknown segment boundaries**

  -use the two competing HMMs to decode each utterance into the target and non-target segments, and compute the FA, FR and error rates.

實驗

- **在音段界線已知情況下測量偵測器效能**
  -將ANN產生的所有音框分數作加總，來做平移到一個程度

- **在音段界線未知情況下測量偵測器效能**
  -使用兩個HMM來對句子做解碼，而分成目標音段與非目標音段，並計算FA、FR和錯誤率