多語言聲學特徵對於在語音辨識系統上新語言之辨識

Multilingual Acoustic Features For Porting Speech Recognition Systems To New Language

Author: Sebastian Stiiuker

Professor:陳嘉平

Reporter:吳柏鋒

Outline

- Introduction
- Multilingual acoustic modeling using ML-MIX
- Articulatory features
- Experiments

摘要

- 簡介
- 使用ML-MIX作多語言聲學模型
- 發音特徵
- 實驗

Introduction

- For rapidly porting speech recognition systems to new languages, the techniques without the need for extensive linguistic or phonetic knowledge about the new language and without the need for large amounts of training materials.

- addition of articulatory features (AF), such as place and manner of articulation, can improve the performance of ASR systems in a multilingual way

簡介

- 希望能夠有技術讓語音辨識系統快速移植到新語言上，也就是無需進行擴展語言或新的語言的音素知識(phonetic knowledge)且不需要大量的訓練材料。

- 並提出加入articulatory feature (AF) 方法
  例如:place 和manner articulatory
  可以用來改善在多語言時的ASR系統辨識

Multilingual acoustic modeling using ML-MIX

- Multilingual Automatic Speech Recognition (ML-ASR) which defines multilingual recognition systems.

- The systems that are capable of simultaneously recognizing languages which have been presented during training.

使用ML-MIX作多語言聲學模型

- 在此定義多語言辨識系統為Multilingual Automatic Speech Recognition (ML-ASR)

- 該系統有能力在訓練時，同時作語言的辨識

Multilingual acoustic modeling using ML-MIX

- For finding a phoneme set common to all languages, phonemes are identified by their symbol in the International Phonetics Alphabet (IPA).

- In the technique ML-MIX, phonemes from different languages that share the same IPA symbol share one model. This model is then trained by pooling all the training data from the different languages.

使用ML-MIX作多語言聲學模型

- 從International Phonetics Alphabet (IPA)中音素的標記，找出一組所有語言通用的音素合

- 在ML-MIX技術中，使用不同語言的音素所共享相同的IPA 標記與模型(此模型主要是集合所有語言的訓練資料)

Articulatory features

- The articulatory features are integrated into the recognition process by using a flexible stream based approach to linearly.

- Additively combine the scores from the AF detectors and the emission probabilities from the phonetic HMM at the state level.

發聲特徵

- 主要使用以stream為基礎，將articulatory feature整合到辨識過程中

- 將AF偵測器產生的分數結合起來與音素HMM在state level產生的機率相加起來

Articulatory features

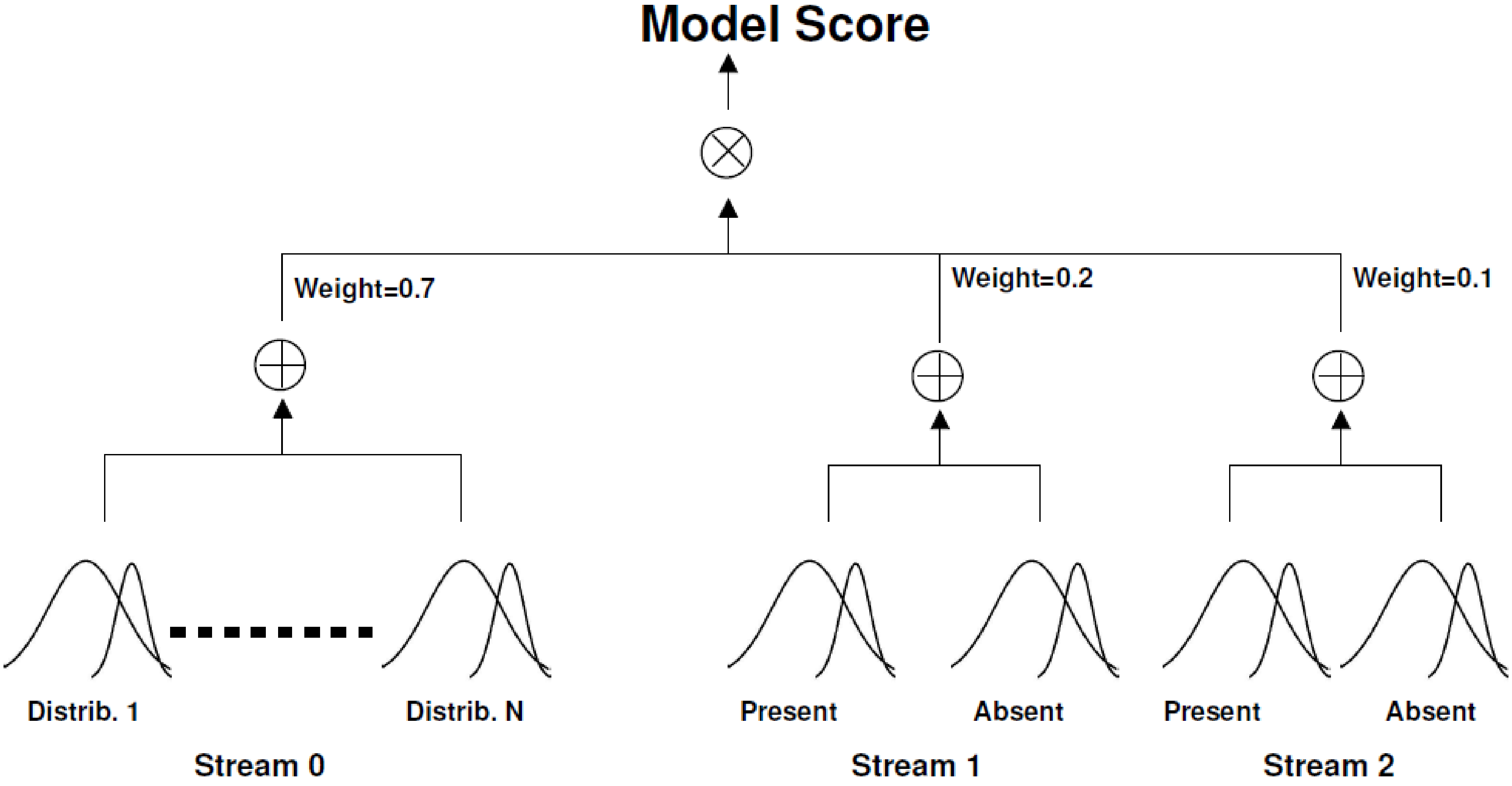


Figure 1 - Stream based architecture for integrating the articulatory feature models

Articulatory features

- The AF can be modeled in a multilingual way.

- The share factor, that measures the overlap between different languages, was also shown to be larger for AF than for phonemes.

- This indicates that AF might be very suitable for multilingual modeling and porting ASR systems to new languages.

發聲特徵

- AF可以被定義在多語言模型中。
- 我們測量出所有不同語言中，以AF的overlap為共享因子的範圍會比音素更大
- 所以我們可以看出AF較適合用在多語言模型與ASR系統對於新語言之辨識上

Experiment

- Use the **GlobalPhone** corpus

- GlobalPhone is an ongoing data collection effort that now provides transcribed speech data that was collected in an uniform way in 18 languages.

- The work presented the four language English (EN), German (GE), Russian (RU), and Spanish (SP) were used.

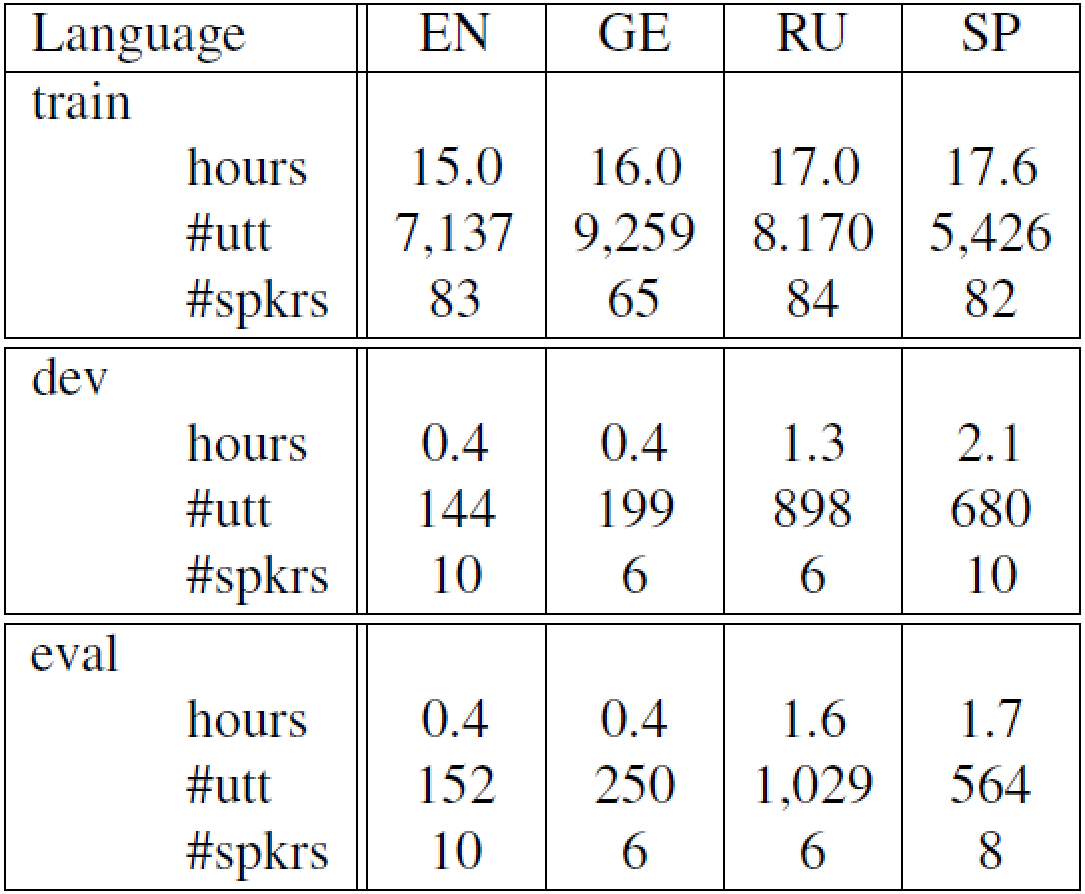
實驗

- 使用 GlobalPhone 語料庫

- GlobalPhone仍是在持續的作收集資料，目前以統一的方式對18種語言所收集的資料作語音的描述

- 此處實驗主要使用到四種語言:
  English (EN), German (GE), Russian (RU), and Spanish (SP)

Experiment



Language		EN	GE	RU	SP
train					
	hours	15.0	16.0	17.0	17.6
	#utt	7,137	9,259	8.170	5,426
	#spkrs	83	65	84	82
dev					
	hours	0.4	0.4	1.3	2.1
	#utt	144	199	898	680
	#spkrs	10	6	6	10
eval					
	hours	0.4	0.4	1.6	1.7
	#utt	152	250	1,029	564
	#spkrs	10	6	6	8

Experiment

- In order to test whether AF models can help when porting ASR systems to new languages, we examined three different scenarios.

(1) German to English

(2) English to German

(3) ML_MIX to German

實驗

- 為了測試是否AF能夠在多語言ASR系統中，對新語言之辨識有更佳的改善，這裡實做了三個實驗

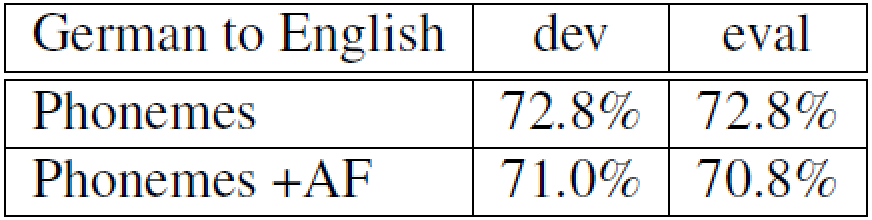
  (1)德語到英語

  (2)英語到德語

  (3)ML-MIX到德語

Experiment

(1) applied the German phoneme based models to English

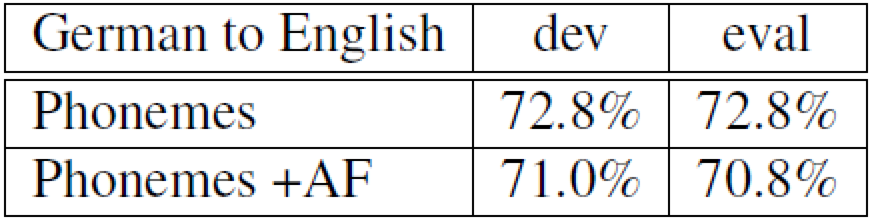


German to English	dev	eval
Phonemes	72.8%	72.8%
Phonemes +AF	71.0%	70.8%

Table 4 - WER when applying the German recognizer to the English test data, without and with Articulatory Features models

實驗

(1) 應用以德語音素為基礎模型對英語作辨識

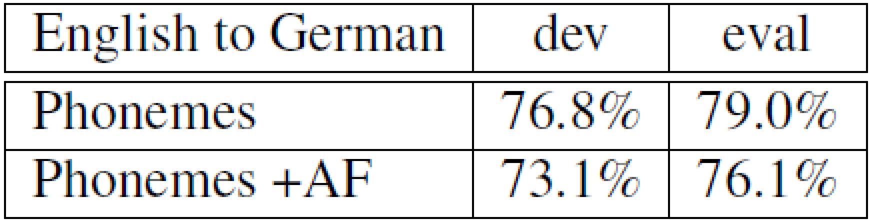


German to English	dev	eval
Phonemes	72.8%	72.8%
Phonemes +AF	71.0%	70.8%

Table 4 - WER when applying the German recognizer to the English test data, without and with Articulatory Features models

Experiment

(2) Applied the English phoneme based models to German.

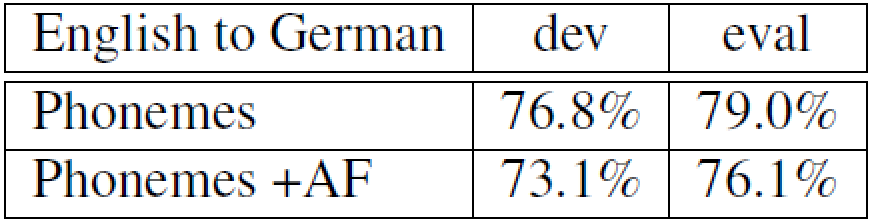


English to German	dev	eval
Phonemes	76.8%	79.0%
Phonemes +AF	73.1%	76.1%

Table 5 - WER when applying the English recognizer to the German test data, without and with Articulatory Features models

實驗

(2) 應用以英語音素為基礎模型對德語作辨識



English to German	dev	eval
Phonemes	76.8%	79.0%
Phonemes +AF	73.1%	76.1%

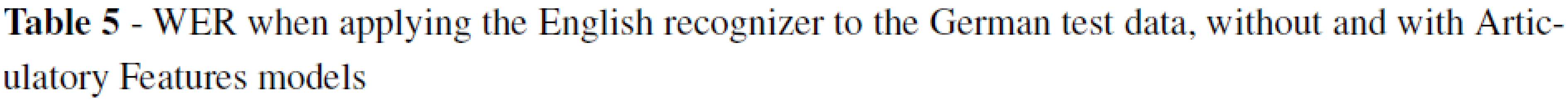
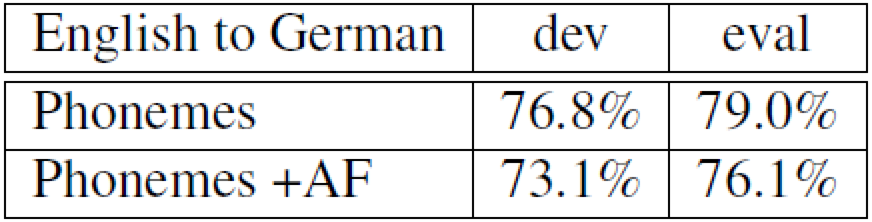


Table 5 - WER when applying the English recognizer to the German test data, without and with Articulatory Features models

Experiment

(3) Applied the ML-MIX phoneme based models to German.



English to German	dev	eval
Phonemes	76.8%	79.0%
Phonemes +AF	73.1%	76.1%

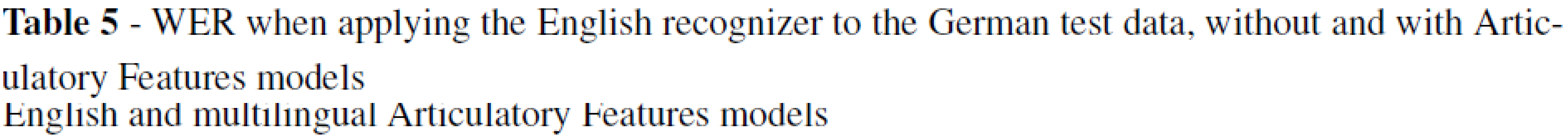
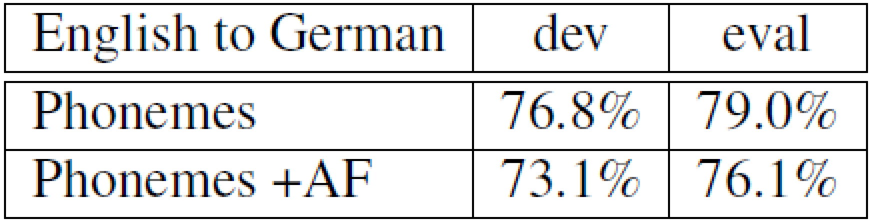


Table 5 - WER when applying the English recognizer to the German test data, without and with Articulatory Features models
English and multilingual Articulatory Features models

實驗

(3) 應用以ML-MIX音素為基礎模型對德語作辨識



English to German	dev	eval
Phonemes	76.8%	79.0%
Phonemes +AF	73.1%	76.1%

Table 5 - WER when applying the English recognizer to the German test data, without and with Articulatory Features models