Speech Recognition

語音辨識

2020.1.16

The Origin Of Speech Recognition

自動語音辨認技術（ASR，Automatic Speech Recognition）是一種經過讓機器經過辨認和了解的過程，把人類的語音信號轉變為相應文本的技術。

其實早在計算機創造之前，有關ASR技術的理念就產生了，而早期的聲碼器就能夠被視為是語音辨認及合成的雛形。而1920年代消費的”Radio Rex”玩具狗，可能是最早的語音辨認器，當這隻狗的名字被召喚時，它可以從底座上彈出來。

但毫無疑問60年代計算機的開展推進了語音辨認技術，這其中最重要的一個里程碑就是動態規劃技術（DP）和線性預測剖析技術（LP），後者又開展出了更成熟的動態時間歸正技術（DTW），包括矢量量化（VQ）和隱馬爾可夫模型（HMM）理論。但這些還都只是單調又晦澀的算法，換句話說，工程師看到這些玩意也一頭霧水，基本沒方法快速應用到理論裡。

所以在80年代時，著名的AT&T Bell實驗室經過努力，把本來深奧無比的HMM純數學模型工程化，為應用開發打下了重要的基石。

Radio Rex Toy

Nowadays

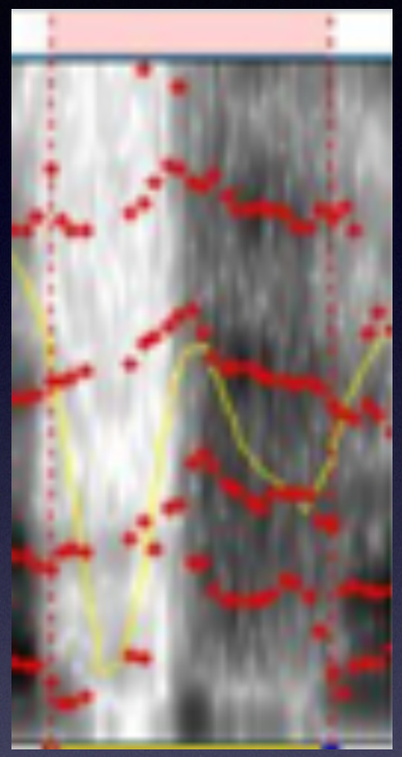
Speech recognition, which is a computer-compared acoustic feature. The best content of long speakers is converted into text. technology. Until 2012, scientists using deep neural network (DNN) speech recognition as early as the 1980s initiated research by the Massachusetts Institute of Technology's laboratory, due to the low recognition rate, and has been unable to apply it to commercial applications. Calculation method, instead of traditional Gaussian distribution calculation, the recognition rate of speech recognition has increased significantly to a level that can be commercialized, and has received the attention and attention of large international companies in recent years.

Nowadays

- 2015/5/28 Google Now 辨識錯誤率在兩年內從23%降到8% (平均每年ERR = 41%)
- 2015/6/8 Apple Siri 辨識錯誤率降到5% (相較前一年ERR = 40%)

系統構成

- 聲學特徵：聲學特徵的提取與選擇是語音辨識的一個重要環節。聲學特徵的提取既是一個資訊大振幅壓縮的過程，也是一個訊號解卷過程，目的是使圖形劃分器能更好地劃分。

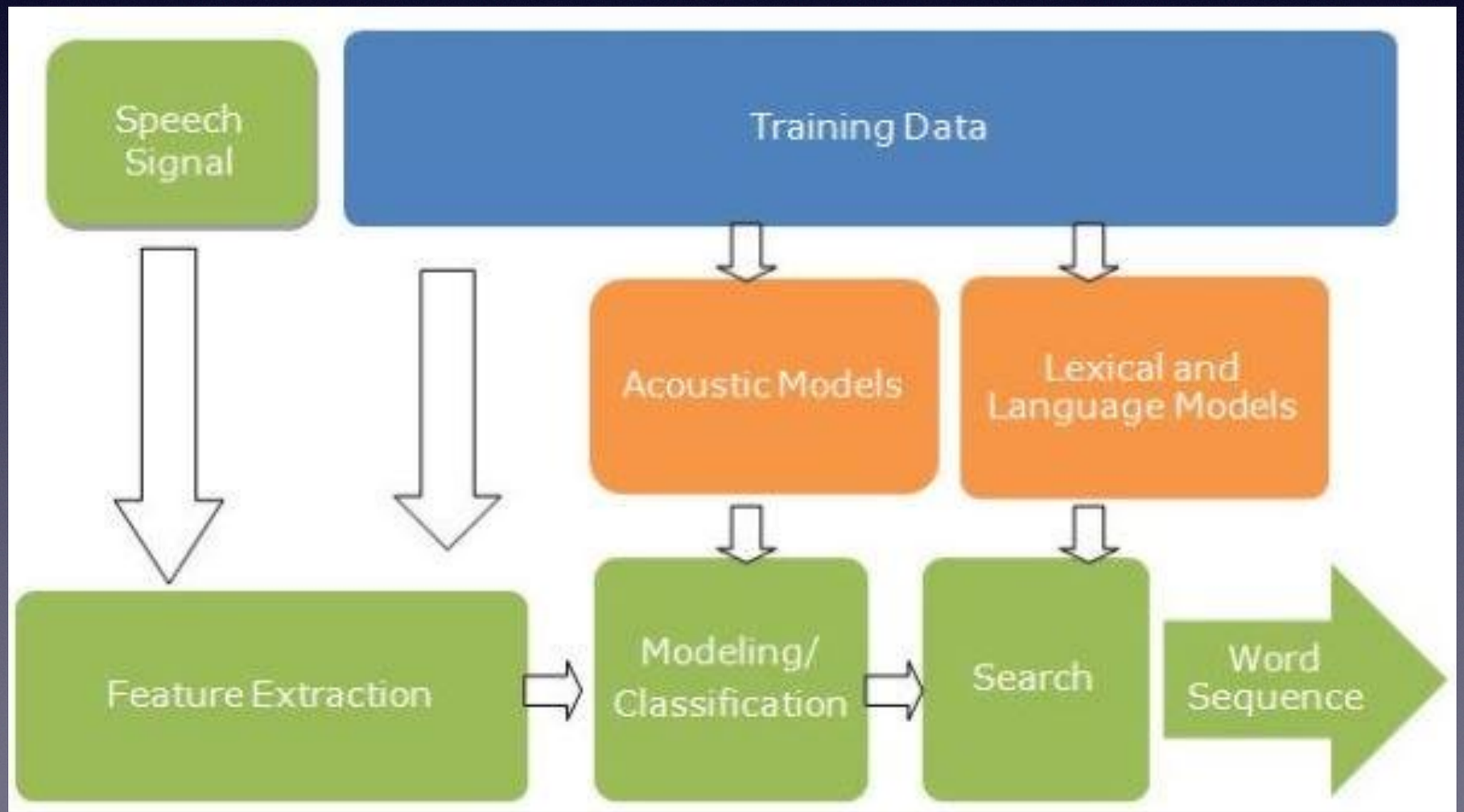


- 語言模型：語言模型是用概率統計的方法來揭示語言單位內在的統計規律。

- 系統實現：語音辨識系統選擇辨識基元的要求是，有準確的定義，能得到足夠資料進行訓練，具有一般性。英語通常採用上下文相關的音素建模，漢語的協同發音不如英語嚴重，可以採用音節建模。系統所需的訓練資料大小與模型複雜度有關。模型設計得過於複雜以至於超出了所提供的訓練資料的能力，會使得效能急劇下降。

原理

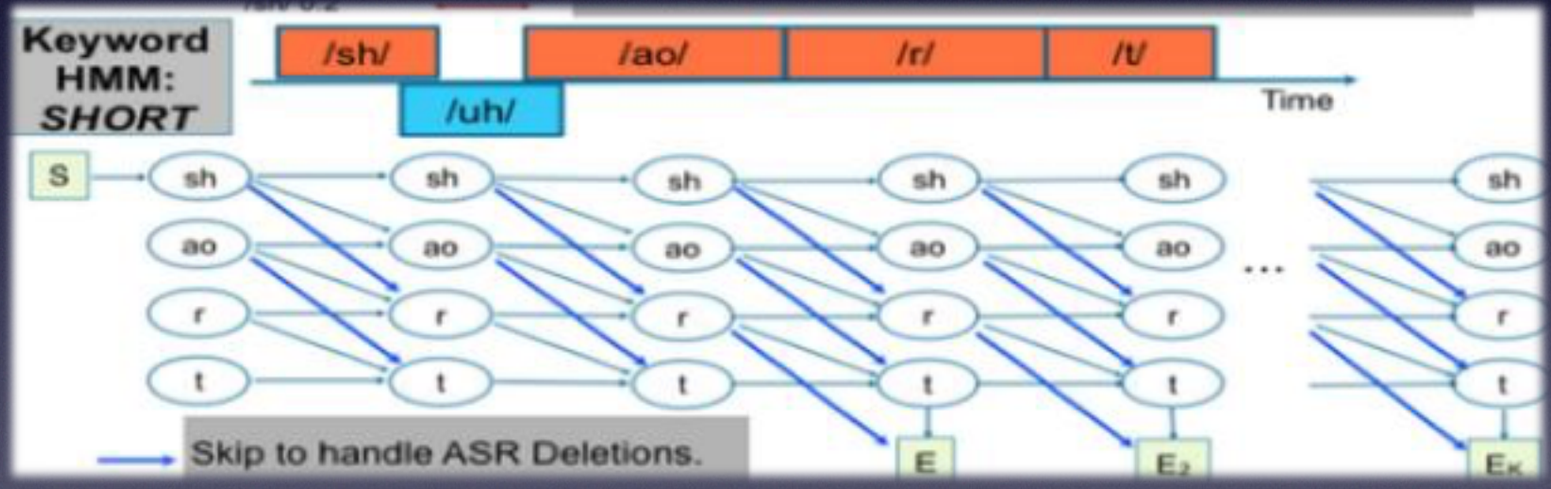
不管是三星的**S-voice**、**Goole**的**Now**、蘋果的**Siri**，在原理在實質上沒有幾差別，就是語音輸入後，進行特徵提取，將提取的特徵值放進模型庫裡，再不停地訓練和匹配，最終解碼得到結果。



原理

所謂模型鍛鍊就是指依照一定的原則，從大量已知語音形式中獲取一個最具特徵的模型參數。而形式匹配則相反，是依據一定原則，將未知語音形式與模型庫中的某一個模型取得最佳匹配。

最後的解碼過程又能夠分動態解碼網絡和靜態解碼網絡兩種：動態網絡會編譯一個狀態網絡並構成搜索空間，把單詞轉換成一個個的音素後將其依照語序拆分狀態序列，再依據上下文分歧性準繩將狀態序列停止銜接。



而靜態網絡普通是針對一些特殊詞（孤立詞）的辨認網絡，它的構造就簡單多了：先將每條特殊詞擴展成HMM狀態序列，然後再計算得分，選擇得分最大的作為辨認輸出結果。由於靜態網絡是依據聲學機率計算權重，不需求查詢言語模型機率，因而解碼速度很快。

這樣的一個流程大致上就是語音辨認技術的主要原理。

、

- 醫療領域
- 智慧車載
- 智慧家居
- 教育領域