5.5 Example Approaches for Text/Speech Information Retrieval with Speech Queries

Example Difficulties in Voice Retrieval of Chinese Text Information

- **Flexible Wording Structure in Chinese Makes It Difficult to Search by Comparing the Character Strings Alone**
  - name/title

    李登輝→李總統登輝，李主席登輝
  - arbitrary abbreviation

    北二高→北部第二高速公路
  - similar phrases

    飛航班次→長榮航班
  - translated terms

    巴塞隆那→巴瑟隆納(Barcelona)
- **Word Segmentation Ambiguity**
  - 腦科→電腦科學
  - 土地公→土地公有政策

Syllable Level Feature Vectors for Text Documents to be Retrieved

- **Syllables** $V=(C(s_1),\ldots,C(s_i),\ldots,C(s_{416}))$
  - C(.) : frequency counts in the text documents
  - each syllable represents characters with meaning , somehow handling the open vocabulary problem
  - each syllable shared by many characters thus causing ambiguity
- **Syllable Pairs**

  $V=(C(s_1, s_1),\ldots,C(s_i, s_j),\ldots,C(s_{416}, s_{416}))$
  - majority of Chinese words are bi-syllabic
  - not too many bi-syllabic words share the same pronunciation
  - carrying plenty of information for retrieval
- **Inverse Document Frequency**

  $idf(P)=\log(N/N_p)$
  - N : total number of documents in the collection
  - N_p : total number of documents having the parameter P
  - the significance or importance of the parameter P for retrieval
- **Combined Features**

  $V=(\ldots,w \bullet C(s_i) \bullet idf(s_i),\ldots,(1-w) \bullet C(s_i,s_j) \bullet idf(s_i,s_j),\ldots)$

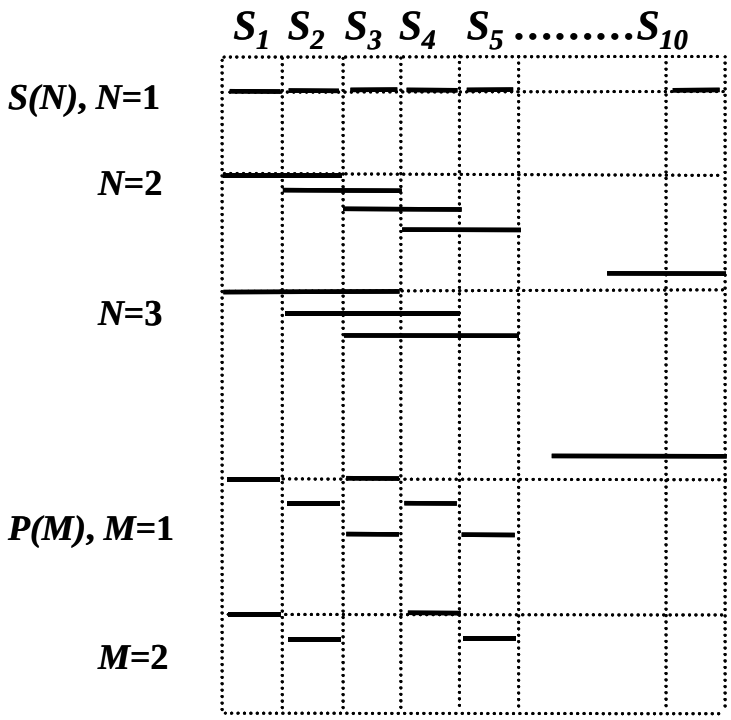
Syllable Level Feature Vectors for Text Documents to be Retrieved

- **Overlapped syllable segments with length N**

Syllable Segments	*Examples*
S(N), N=1	$(s_1)\ (s_2)\ \ldots(s_{10})$
S(N), N=2	$(s_1\ s_2)\ (s_2\ s_3)\ldots(s_9\ s_{10})$
S(N), N=3	$(s_1\ s_2\ s_3)\ (s_2\ s_3\ s_4)\ldots(s_8\ s_9\ s_{10})$
S(N), N=4	$(s_1\ s_2\ s_3\ s_4)\ (s_2\ s_3\ s_4\ s_5)\ldots(s_7\ s_8\ s_9\ s_{10})$
S(N), N=5	$(s_1\ s_2\ s_3\ s_4\ s_5)\ (s_2\ s_3\ s_4\ s_5\ s_6)\ldots(s_6\ s_7\ s_8\ s_9\ s_{10})$

- **Syllable pairs separated by M syllables**

Syllable Pair Separated by M Syllables	*Examples*
P(M), M=1	$(s_1\ s_3)\ (s_2\ s_4)\ \ldots(s_8\ s_{10})$
P(M), M=2	$(s_1\ s_4)\ (s_2\ s_5)\ \ldots(s_7\ s_{10})$
P(M), M=3	$(s_1\ s_5)\ (s_2\ s_6)\ \ldots(s_6\ s_{10})$
P(M), M=4	$(s_1\ s_6)\ (s_2\ s_7)\ \ldots(s_5\ s_{10})$



Text/Speech Information Retrieval with Speech Queries

- **For the Speech Query q**
  - acoustic recognition gives a syllable lattice
  - constructing a similar feature vector $V_{\mathbf{q}}$ for q, $C(s_{\mathbf{i}})$ including acoustic recognition scores, etc.
- **Relevance Measure**
  - $R(d,q) = cos(v_{\mathbf{d}}, v_{\mathbf{q}}) = \frac{v_{\mathbf{d}} \bullet v_{\mathbf{q}}}{|v_{\mathbf{d}}|\,|v_{\mathbf{q}}|}$
- **Automatic Keyword Extraction for the Text Documents Helpful**
  - keyword spotting from the speech queries
- **Speech Information Retrieval**
  - speech information recognized as syllable lattices, character/word sequences or lattices
  - similar feature vectors constructed
- **Character/Word Level Information Similarly Used and Integrated**
- **Language Understanding Helpful**