# A method of reducing the search for words in speech recognition based on phonetic coding algorithm

Post graduated student: J.M. Du

Professor: V.S. Vykhovanets

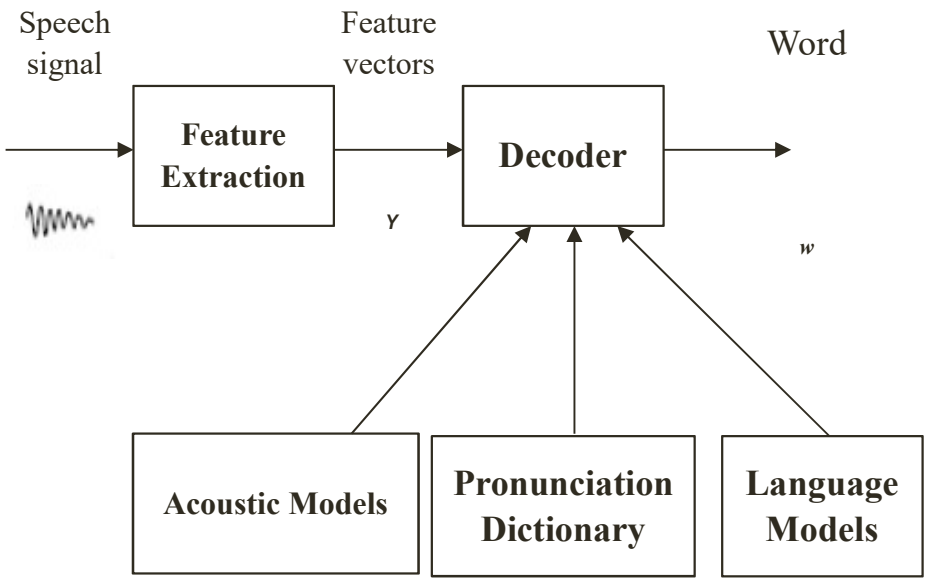
# Introduction
– Problem

- In the recognition process, the decoder will use these 3 files to compute – acoustic model**, pronunciation dictionary** and language model.
- For the growing volume of dictionary, the decoder which scans all possible sequences of words will **need a huge time.**

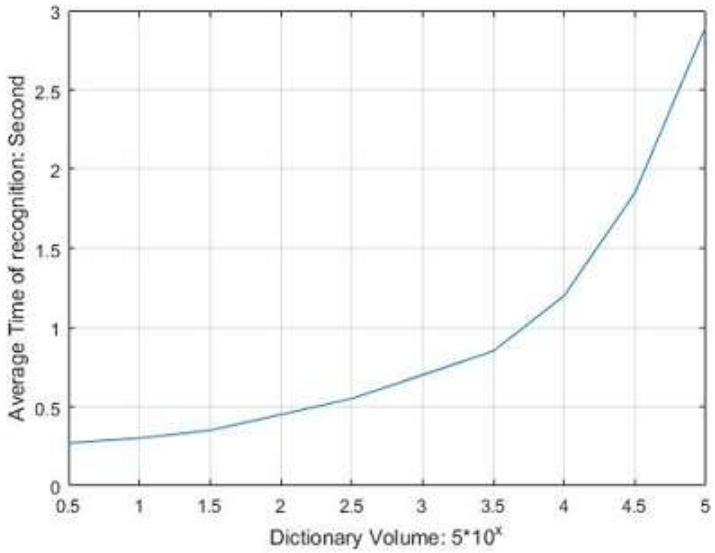
# Introduction

- This method reduces the volume of dictionaries used in the **decoding process via pre-processing**, which is based **on acoustic feature**, to release the burden of search.

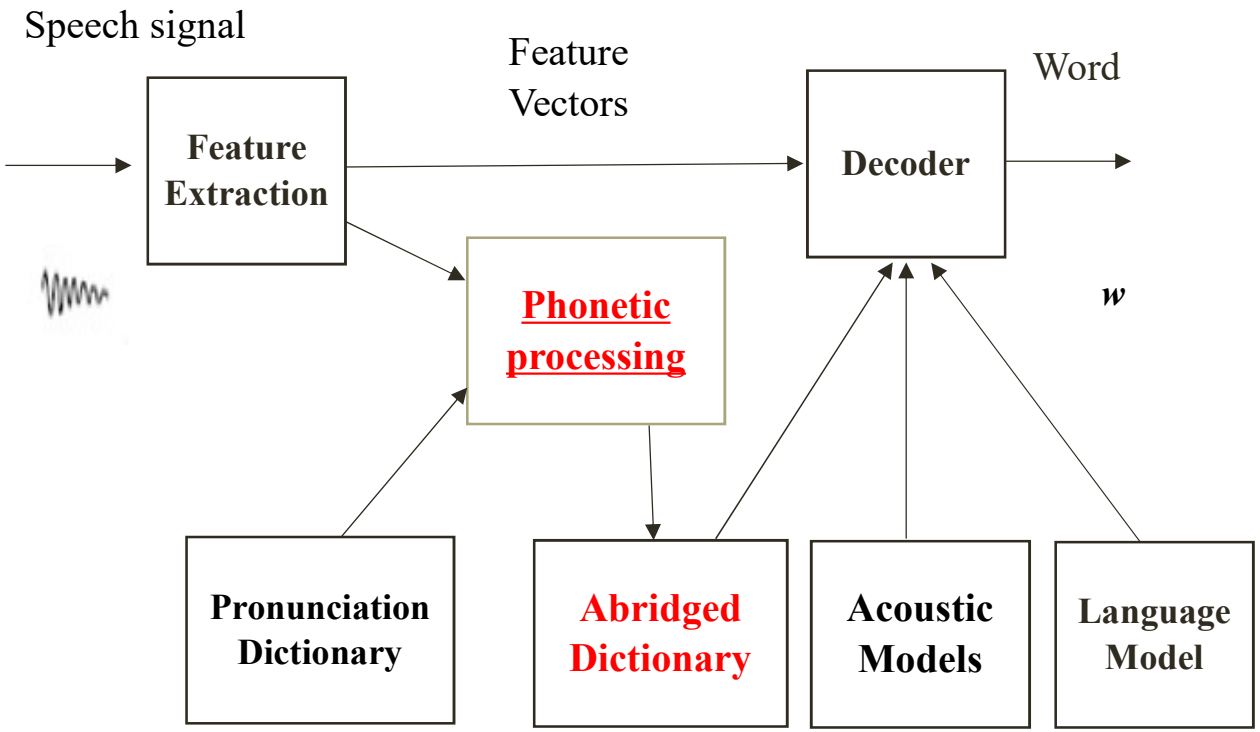
# Theory




**The architecture of recognition system**



# Theory


Speech signal
Feature Extraction
Feature Vectors
Decoder
Word
w
Phonetic processing
Pronunciation Dictionary
Abridged Dictionary
Acoustic Models
Language Model

# Theory
## – Phonetic coding algorithm

- The most famous phonetic coding algorithm is **SoundEx**.

| Num | Letter |
|---|---|
| 1 | B, P, F, V |
| 2 | C, S, K, G, J, Q, X, Z |
| 3 | D, T |
| 4 | L |
| 5 | M, N |
| 6 | R |

| Num | Letter |
|---|---|
| 1 | Б, П |
| 2 | Ф, В |
| 3 | Ж, З, С, Х |
| 4 | К, Г, |
| 5 | Ц, Ч, Ш, Щ |
| 6 | Д, Т |
| 7 | Л, Й |
| 8 | М, Н |
| 9 | Р |

- **«Lee» -> «L000»**
- **«Gauss» -> «G200»**
- **«Ashcraft » -> «A226»**

# Theory
– Rules

- **Phonetic rules for allophones**
- 1) The **length** of the sequence of allophones and the length of the transcription of the words

| Word | Transcription | Phoneme |
|---|---|---|
| **дедушка** | dj e1 d u0 sh k a0 **-7** | bj dj e0 d u0 sh kj a0 **-8** |
| **бабушка** | b a1 b u0 sh k a0 **-7** | b a1 mj rj sh k a1 mj **-8** |
| **дед** | dj e1 d **-3** | dj a0 t **-3** |
| **ресница** | rj e0 s nj i1 c a0 **-7** | e0 z n1dj z a1 m **-8** |
| **поплакать** | p o0 p l a1 k a0 tj **-8** | k a0 p l a1 g j a0 **-8** |

# Theory
– Rules

- **Phonetic rules for allophones**
- 2) The **number of vowels** in the sequence of allophones and the relevant number in the sequence of transcriptions

| Word | Transcription | Phoneme |
|---|---|---|
| **дедушка** | dj ***e1*** d ***u0*** sh k ***a0*** **- 3** | bj dj ***e0*** d ***u0*** sh kj ***a0*** **- 3** |
| **бабушка** | b ***a1*** b ***u0*** sh k ***a0*** **-3** | b ***a1*** mj ***rj*** sh k **a1** mj **-3** |
| **дед** | dj ***e1*** d **-1** | dj **a0** t **-1** |
| **ресница** | rj ***e0*** s nj ***i1*** c ***a0*** **-3** | ***e0*** z n1dj z ***a1*** m **-2** |
| **поплакать** | p ***o0*** p l ***a1*** k ***a0*** tj **-3** | k ***a0*** p l ***a1*** g j ***a0*** **-3** |

# Theory
– Rules

- **Phonetic rules for allophones**
- 3) **The first consonant** is one of the most important features of a word.
  - a) Superfluous first consonant.
  - b) Absence of the first consonant
  - c) Other first few consonants

| Word | Transcription | Phoneme |
|---|---|---|
| дедушка | **dj** e1 d u0 sh k a0 | **bj dj** e0 d u0 sh kj a0 |
| бабушка | **b** a1 b u0 sh k a0 | **b** a1 mj rj sh k a1 mj |
| дед | **dj** e1 d | **dj** a0 t |
| ресница | **rj** e0 s nj i1 c a0 | e0 z n1dj z a1 m |
| поплакать | **p** o0 p l a1 k a0 tj | **k** a0 p l a1 g j a0 |

# Theory
– Rules

- According to artificial statistics, the following ranges of errors are shown in table.

| Length of phonemes | Range of transcription |
|---|---|
| 1 | [1, 4] |
| 2 | [1, 4] |
| 3 | [1, 5] |
| 4 | [2, 7] |
| 5 | [2, 8] |
| 6 | [3, 9] |
| 7 | [3, 10] |
| 8 | [4, 11] |
| 9 | [5, 13] |
| 10 | [6, 15] |
| 11 | [8, 17] |
| 12 | [9, 18] |
| 12+ | [9, 18+] |

| Number of vowels | Range of transcription |
|---|---|
| 1 | [0, 2] |
| 2 | [1, 3] |
| 3 | [2, 4] |
| 4 | [2, 5] |
| 5 | [3, 6] |
| 6 | [4, 8] |
| 7 | [5, 9] |
| 8 | [6, 9] |
| 9 | [7, 10] |
| 9+ | [8, 10+] |

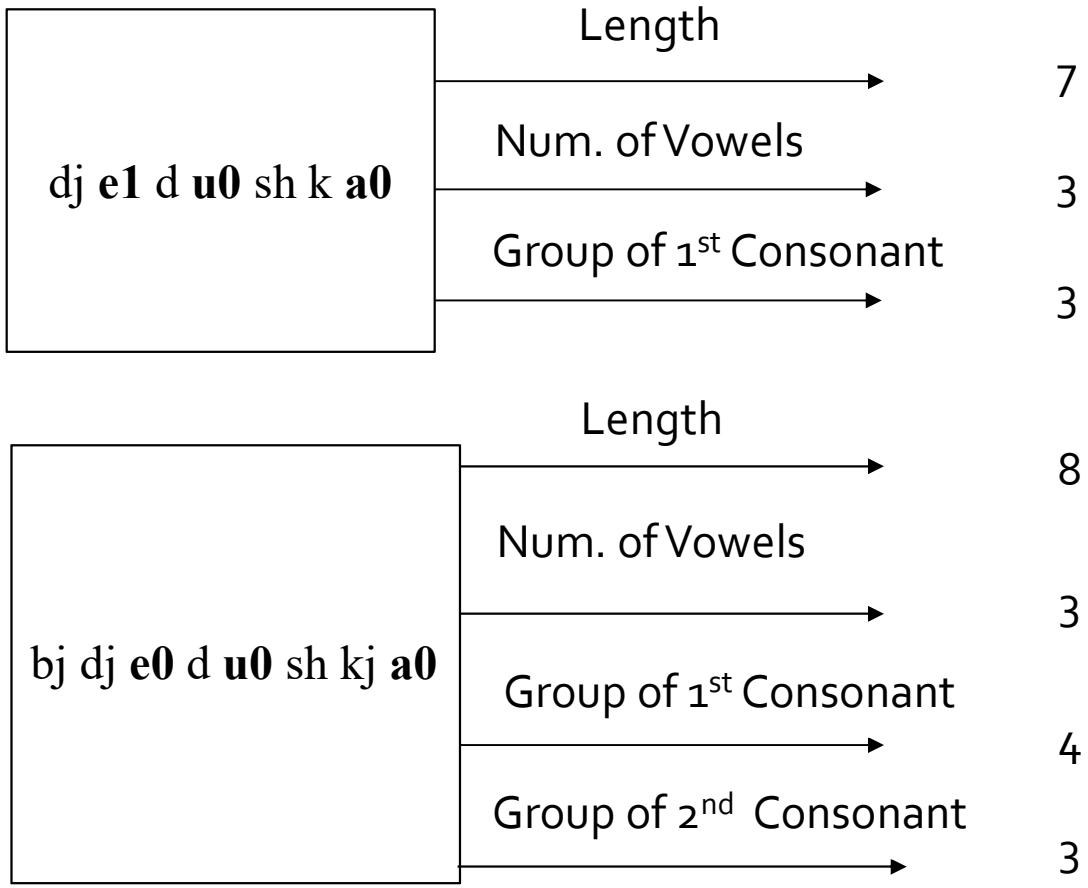
# Theory
– Rules

- According to cases a), b), and c), the consonants is regrouped

| Group | First consonant of phonemes | Possible first consonant of transcription |
|---|---|---|
| 1 | z, zj, s, sj, c, ch, sch, zh, sh | t, tj, z, zj, s, sj, c, ch, sch, zh, sh |
| 2 | p, pj | b, bj, p, pj, k, kj, t, tj |
| 3 | d, dj | d, dj, k, kj, z, zj, zh |
| 4 | b, bj | d, dj, b, bj, p, pj |
| 5 | k, kj | p, pj, b, bj, k, kj, g, gj |
| 6 | g, gj | k, kj, g, gj |
| 7 | m, mj, n, nj | m, mj, n, nj |
| 8 | t, tj | z, zj, zh, d, dj, k, kj, t, tj |
| 9 | vowels | z, zj, zh,p, pj, b, bj,p, pj, k, kj, g, gj, m, mj, n, nj, t, tj |

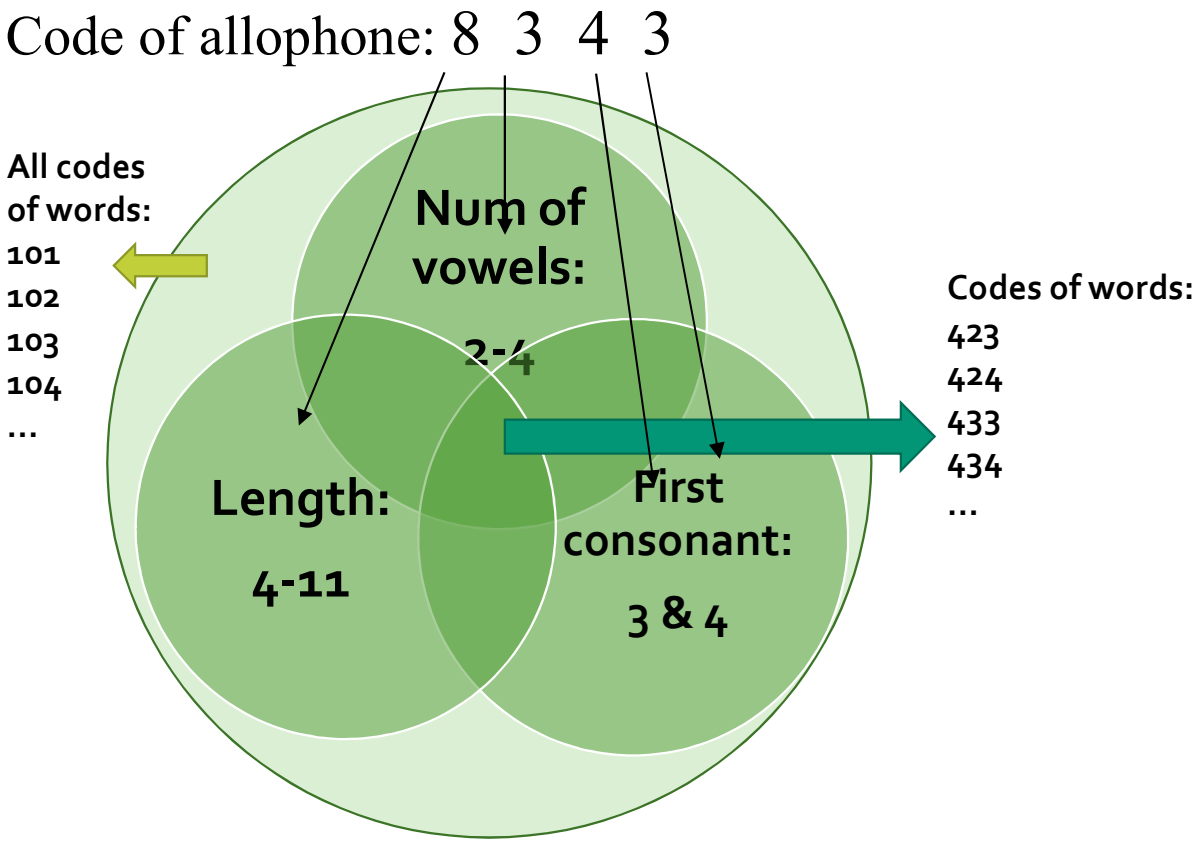
# Theory
– Coding

| Word | Transcription | Phoneme |
|---|---|---|
| дедушка | dj e1 d u0 sh k a0 | bj dj e0 d u0 sh kj a0 |



# Theory
– Coding

| Word | Transcription | Phoneme |
|---|---|---|
| дедушка | dj e1 d u0 sh k a0 | bj dj e0 d u0 sh kj a0 |
| Код | 7 3 3 | 8 3 4 3 |



# Experiment

- Table shows the translation result of **code '8-3-4-3'**. From the table, it can be found that the **code '7-3-3'** is included in the results.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 424 | 524 | 624 | 724 | 824 | 924 | 1024 | 1124 |
| 434 | 534 | 634 | 734 | 834 | 934 | 1034 | 1134 |
| 444 | 544 | 644 | 744 | 844 | 944 | 1044 | 1144 |
| 423 | 523 | 623 | 723 | 823 | 923 | 1023 | 1123 |
| 433 | 533 | 633 | 733 | 833 | 933 | 1033 | 1133 |
| 443 | 543 | 643 | 743 | 843 | 943 | 1043 | 1143 |

# Experiment

## CMU Sphinx

- In experiments, 8-hour audio material consisting of 7000 Russian words spoken by 4 announcers (2 men and 2 women) was applied.
- 5000 words are allocated for the finding of rules for constructing the method.
- 2000 word are used to analyze its effectiveness.
- **CMUSphinx** is selected as an experimental tool

--- codes, structures are mature and open.

--- ready-made acoustic model

--- Russian language model with a large dictionary (540 thousand word).

# Experiment

- In the first experiment of recognizing 2000 words by using a **classic recognition system** and the **modified system,** the relevant result of the average time of recognizing single words are shown in Table.

| Process | Classic system (second) | | Modified system (second) | |
|---|---|---|---|---|
| | Confidence interval at 0.95 | Average time | Confidence interval at 0.95 | Average time |
| Function 'Allophone' | | | [0.039,0.049] | 0.0442 |
| Phonetic processing | | | [0.34,0.51], | 0.42 |
| Decode | [2.34,3.46] | 2.90 | [0.73,1.09], | 0.91 |
| Total time | [2.34,3.46] | 2.90 | [1.21,1.55], | 1.38 |

- The modified system is noticeably faster than the classic one.

# Experiment

- In "accuracy experiments", the relevant results of the accuracy of the 2 systems are shown in Table.

| | Classic system | Modified system |
|---|---|---|
| Accuracy of recognition of single words | 56.1% | 61.7% |

- The accuracy of the modified system is higher than that of the classic one.
- Many words in the **classic system** are recognized as **word combinations (phrases)** with short words. These phrases have **more probability** in language model.
- The modified system uses the rules to determine the **length of a word in a certain range.**
- In its dictionary, there is no too-short words due to the improvement of accuracy.

# Conclusion

- One method based on the idea of the **phonetic coding algorithm** and the rules of allophones is described.
- The method shortens the search of words from the dictionary in the recognition process of **a single word in a large dictionary**.
- Experimental results show that this **method increases the average speed** of the decoder process by about 52%
- Its **accuracy is higher** than that of the classic system
- Represents **one possible direction** for speeding up the recognition process

Thank you!