

AN ALGORITHM OF GENERATION OF ALTERNATIVE PHONETIC TRANSCRIPTIONS FOR SPONTANEOUS RUSSIAN SPEECH RECOGNITIONS

Irina Kipyatkova

Saint-Petersburg State University of Aerospace Instrumentation,
Saint-Petersburg, Russia

Abstract

For increasing accuracy of automatic speech recognition it is necessary to create alternative transcriptions of words that allows to take into account variability of spontaneous speech. An algorithm of creation of alternative transcriptions by applying rules of reduction and assimilation in spontaneous speech is proposed in this work. Testing of the developed module was made with the text corpus of the catalogue “Yellow Pages of Saint-Petersburg”. The distributions of frequency of use of the phonetic rules in stems and endings, and by a type of formation were experimentally obtained.

I. INTRODUCTION

One of the main difficulties of automatic speech recognition is variability of word pronunciation. In spontaneous speech some sounds can be reduced up to complete disappearance or assimilated with changing their sounding depend on phonetic context. These phenomena of spontaneous speech must be taken into account for increasing of recognition accuracy.

At the stage of recognition system training it is necessary to create a vocabulary of recognition words with phonetic transcriptions [1]. Word transcriptions are created by basic transcription rules [2]. But often in spontaneous speech transcriptions of pronounced words do not coincide with canonical transcriptions, that significantly decreases word recognition rate of recognition. For improving spontaneous speech recognition in addition to basic transcriptions it is necessary to create alternative transcriptions. There are several ways of creation of alternative transcriptions by using:

1. free phoneme decoder;
2. Viterbi search with free choice between alternative transcriptions;
3. rules of reduction and assimilation.

In the first and second methods alternative transcriptions are created using statistically hand-labeled speech corpuses. Therefore large speech corpuses are required for application these methods and

incompleteness of training data can lead to creation of limited set of alternative transcriptions.

Current work presents third method and alternative transcriptions are generated from basic transcriptions using rules of within word and across word reduction and assimilation. On the base of the rules described in the work [3] the rules for taking into account reduction and assimilation were formulated for speech recognition. Using these rules the software for generation of alternative transcriptions was developed.

II. ALGORITHM OF GENERATION OF ALTERNATIVE TRANSCRIPTIONS BY REDUCTION AND ASSIMILATION RULES

On the base of the investigations results obtained in the field of experimental phonetics [4] several rules that will sufficiently accurately describe possible deviations concerned with sound assimilation and reduction within word and at conjunction of two words in phonetic transcriptions can be formulated. In this part the algorithm which combines basic transcribing rules and rules of assimilation and reduction will be described for generation of word alternative transcriptions and for extending of recognizer’s vocabulary.

Fig. 1 shows general algorithm of word transcribing by reduction and assimilation rules [5].

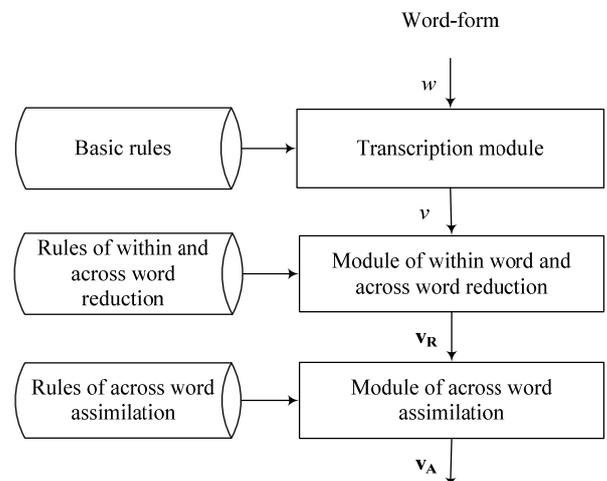


Fig. 1. Process of word transcription taking into account assimilation and reduction phenomena

The transcription module analyzes basic rules and transforms an input word \mathcal{W} in a consequence of phonemes. An obtained transcription \mathcal{V} comes into the block of within word and across word reduction where with the help of within word and across word reduction rules a set of alternative transcriptions \mathbf{V}_R of the given word is obtained. Since for one word several rules can be applied, the transcriptions for all possible variants of word pronunciations are obtained at the output of the module. Then these alternative transcriptions are transformed into alternative transcriptions of the word \mathbf{V}_A in the module of across word assimilation. The obtained set of transcriptions must theoretically contain all variants of pronunciations that can take place in spontaneous speech of native Russian speakers.

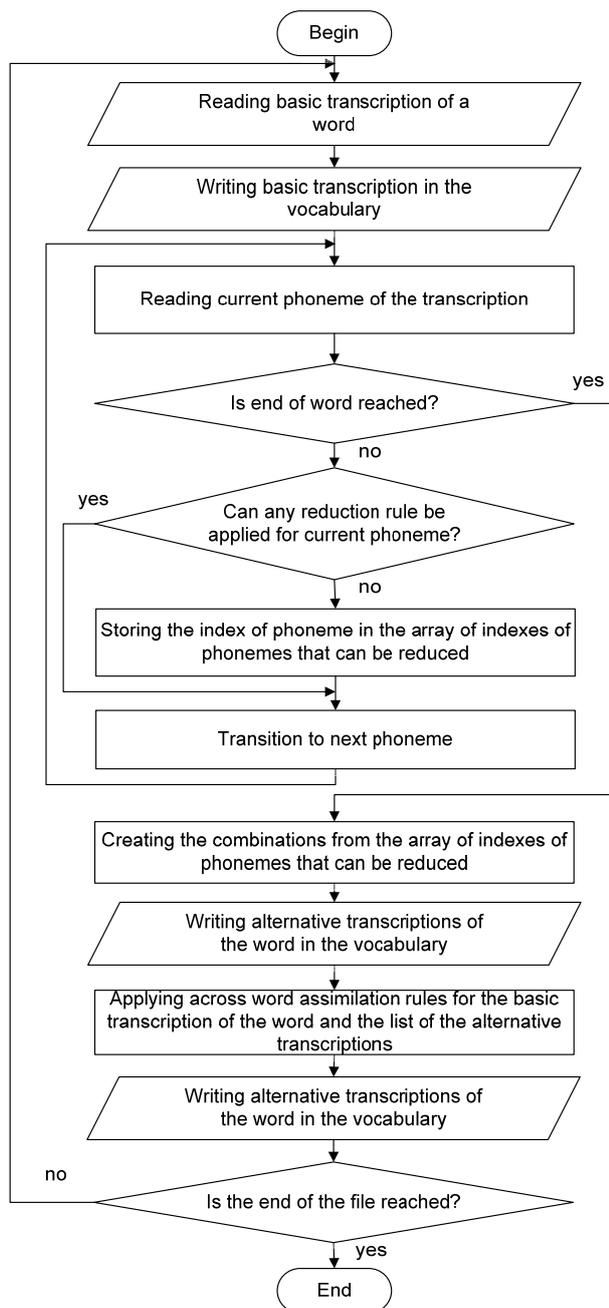


Fig. 2. The algorithm of creation of alternative phonetic transcriptions

Since this work is directed on creation of alternative transcriptions let's consider two last modules in detail. The list of basic word transcriptions is given to the input of the algorithm (Fig. 2). In the block "reading word transcriptions" each basic word transcription is loaded from the input file. The basic transcription is written in the vocabulary.

The block "checking reduction rules" verifies, whether is it possible to apply some of the reduction rules [3] to the current phoneme. If yes, then an index of current phoneme is written in the array of indexes of phonemes that can be reduced and transition to the next phoneme is made. If no reduction rule can be applied, then transition to the next phoneme is made. When end of the word is achieved all acceptable combinations from the array are created. Then alternative transcriptions are created by removing from the basic transcription those phonemes, indexes of which exist in the given combination, and alternative transcriptions are stored in the vocabulary. Then possibility of applying the rules of across word assimilation for basic transcription and alternative transcriptions is checked. Obtained alternative transcriptions are stored in the vocabulary of alternative transcriptions. Then transition to next word is made. The listed above operations are carried out until the end of the original file. A file, which contains basic and alternative word transcriptions, is created as a result of program work.

Let's consider in detail work of the program by the example of a rule of reduction. According to this rule unaccented vowels are reduced up to entire disappearance if they take place after a soft consonant and before any consonant (Fig. 3).

Checking is carried out whether does current phoneme belong to vowels without an accent. For that this phoneme is compared with all phonemes from the array of unaccented vowels. If current phoneme is not a vowel without an accent, then transition to the next rule is carried out. If current phoneme belongs to unaccented vowels, then checking, whether a preceding phoneme is a soft consonant, is carried out. For that it is compared with phonemes from the array of soft consonants. If a preceding phoneme is not a soft consonant, then transition to the next rule is made, else checking, whether the next phoneme is consonant, is carried out. If the next phoneme is not a consonant, then transition to the next rule is made, else storing the index of the current phoneme in the array of indexes of phonemes, that can be reduced, is performed.

Altogether 14 reduction rules and 4 assimilation rules, that were formulated by phoneticians on results of spoken Russian language recordings, successfully applied for speech synthesis purpose [3] and now are adapted by author to create extended vocabulary and to increase accuracy of automatic recognition of spontaneous Russian speech, were used in this module.

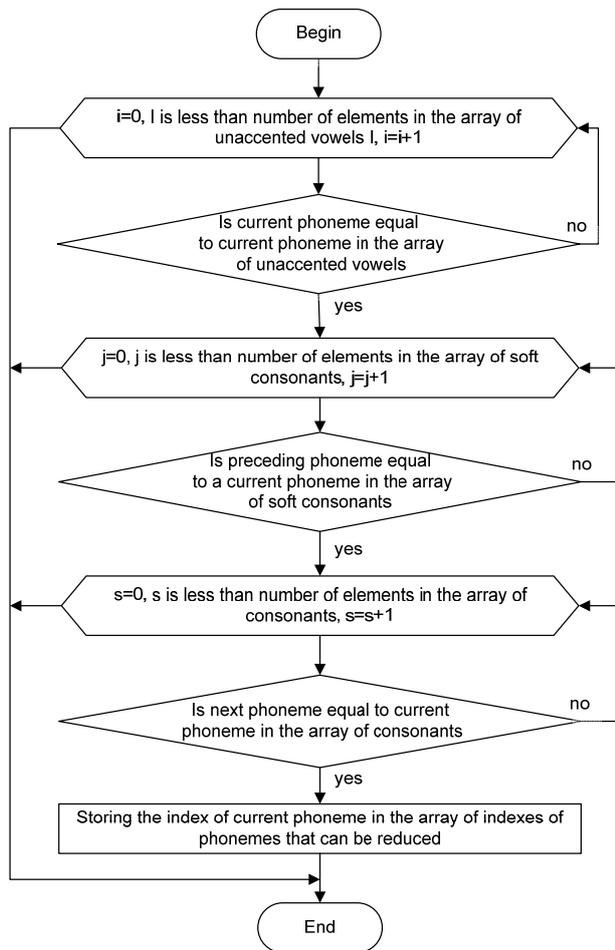


Fig. 3. The algorithm of checking of possibility of application one of the rules to current phoneme in a word

III. EXPERIMENTAL RESULTS

Testing of the developed program was carried out during creation of an extended vocabulary for “Yellow Pages of Saint-Petersburg” catalogue. A personal computer with a processor Pentium IV with clock speed of 2 GHz and RAM of 1 GB was used. Total time of work of the program was 89 s. Mean time of one word processing is 0,005 s.

At the beginning the vocabulary contained 17662 word-forms, each of them was described by one basic transcription. After application of the rules of within word and across word reduction quantity of word-forms transcriptions has increased on 180195 transcriptions and after application of the rules of across word assimilation the vocabulary size has increased on 12108 transcriptions. Distribution of basic transcriptions and transcriptions, which were obtained by the rules of within word and across word reduction and after application of the rules of across word assimilation, is shown in Fig. 4.

Fig. 5 shows a distribution of number of transcriptions in the extended vocabulary of the text corpus “Yellow Pages of Saint-Petersburg”. Most of word-forms have 4 transcriptions (for example “авиационный” (“aviation”) → [ав'иационный], [аиационный], [ав'иационы], [аиационы]). At the same

time there are word-forms with a very large number of alternative transcriptions. For example 11 phonemes in the word-form “водонагревательного” (“waterheater”) can be reduced ([ваднагр'ивАт'ел'нава] → [анагр'Ат'н]). Thus 2047 alternative transcriptions are created for this word-form. 1057 words have not alternative transcriptions, for example, “абжур” (“lamp-shad”), “службою” (“service”), “станции” (“stations”).

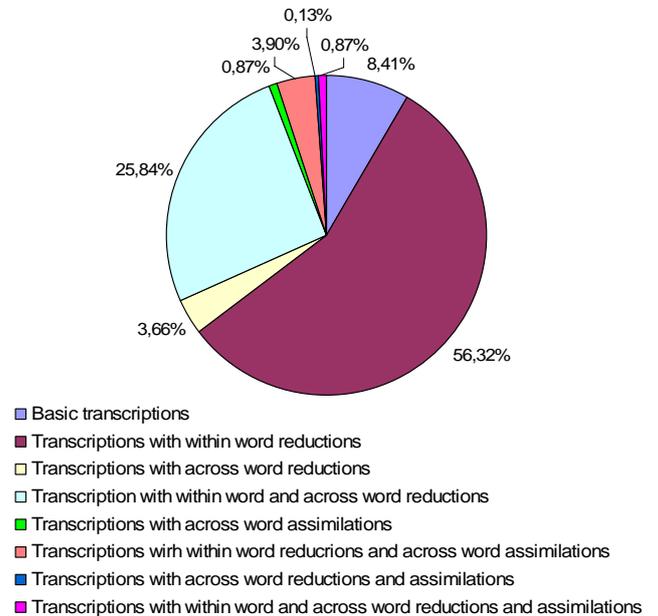


Fig. 4. Distribution of alternative transcriptions by principle of formation

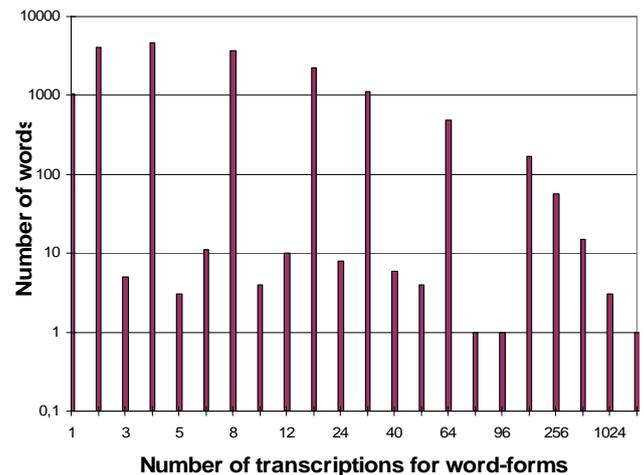


Fig. 5. Distribution of the number of transcriptions in the extended vocabulary of “Yellow Pages of Saint-Petersburg” catalogue

The number of alternative transcriptions after application of the rules of reduction corresponds to total number of possible combinations of the indexes of phonemes that can be reduced. The number of combinations C is computed by the following formula:

$$C = \frac{n!}{k!(n-k)!} \quad (1)$$

where n is the number of elements in combinations; k is the number of elements in a sample.

Thus the number of transcriptions of any word-form in the extended vocabulary is divisible by total number of combinations plus 1 (1 corresponds to basic transcription). But there are some word-forms for which both the rules of reduction and assimilation can be applied for last phoneme. In the case of phoneme reduction the rules of assimilation should not be applied. But there are a few of such word-forms, that is shown by short bars in Figure 5.

Relative frequencies of appearance of reductions and assimilations in stems and endings of word-forms were calculated. For this purpose the following formulas were used:

$$R_{stem} = \sum_{i=1}^L \frac{S_{stem_i}}{P_{stem_i}} \quad (2)$$

$$R_{end} = \sum_{i=1}^K \frac{S_{end_i}}{P_{end_i}}, \quad (3)$$

where R_{stem} is relative frequency of reductions and assimilations in stems of word-forms, R_{end} is relative frequency of reductions and assimilations in endings of word-forms, S_{stem_i} is number of reductions and assimilations in a stem of i -th word-form, P_{stem_i} is number of phonemes in a stem of i -th word-form, S_{end_i} is number of reductions and assimilations in an ending of i -th word-form, P_{end_i} is number of phonemes in an ending of i -th word-form, L is number of unique stems (4790 for the given vocabulary), K is number of unique endings (225 for the given vocabulary).

After program computations it was determined that $R_{stem} = 0,213$ and $R_{end} = 0,296$. Hence reductions and assimilations occur in endings more frequently than in stems.

Also with the help of the developed program an analysis of frequency of rules applied was made. It was discovered that the following rules were the most frequently used:

1. Unaccented vowels can be reduced up to complete disappearance if they are located after a soft consonant and before any consonant (“телефон” (telephone) → [т'л'фон]).

2. Unaccented vowels can be reduced up to entire disappearance if they are located after a resonant consonant and before any consonant (“обработка” (processing) → [абрботка]).

3. Unaccented vowels at ends of words can be reduced up to entire disappearance (“камни драгоценные” (precious stones) → [камн'драгацэныйе]).

Additionally to generation of alternative transcriptions of words the developed program allows to: (1) analyze distribution of number of basic and alternative transcriptions based on transcription rules; (2) get distribution of obtained number of transcriptions according to the number of words in the given extended vocabulary; (3) calculate relative frequency of appearance of reductions and assimilations in stems and endings of words; (4) determine frequency of application of each transcription rule.

IV. CONCLUSION

Generation of alternative transcriptions when creating recognizer's vocabulary allows to take into account variability of the pronunciation of words and to avoid errors when interacting with a dialogue system or voice controlled technical objects.

The developed program for generation of recognizer's extended vocabulary aims first of all to analysis of features of spontaneous Russian speech. This program can be successfully applied in state-of-the-art telecommunication applications, in different automatized information systems and in mobile devices where speech is the most effective way of information input.

The developed program was used for creation of vocabularies for real information systems. Statement of implementation was received from St.Petersburg Institute for Informatics and Automation of RAS about application the developed program in Speech Informatics Group when executing research works of program of DITCS RAS. The developed program was registered in the Branch Foundation of Algorithms and Software (the number of the certificate is № 9525).

REFERENCES

- [1] А. Л. Ронжин, А. А. Карпов, И. В. Ли. Речевой и многомодальный интерфейсы. - М.: Наука, 2006 - (Информатика: неограниченные возможности и возможные ограничения), 173 с.
- [2] Академия наук СССР Институт русского языка «Русская грамматика». М.: Наука, 1980.
- [3] Б. М. Лобанов, Л. И. Цирульник. Моделирование внутрисловных и межсловных фонетико-акустических явлений полного и разговорного стилей в системе синтеза речи по тексту. Труды первого междисциплинарного семинара «Анализ разговорной русской речи» (AP3 - 2007).—СПб.: ГУАП, 2007. С. 57-71.
- [4] Русская разговорная речь. Под редакцией Е. А. Земской. М.: Наука, 1973. 485 с.
- [5] Ал. Б. Леонтьева, И. С. Кипяткова. Моделирование нефонемных речевых элементов и создание альтернативных транскрипций для распознавания спонтанной речи. Труды первого междисциплинарного семинара «Анализ разговорной русской речи» (AP3 - 2007). – СПб.: ГУАП, 2007, С. 77-85.