СТЭЛ КС
Системный ИТ-Интегратор • Поставщик отраслевых и бизнес решений с 1991 года
stel@stel.ru
Россия
г. Москва
ул. Большая Почтовая, д.55/59
8 (495) 77-55-123

Multi-Pronunciation Lexicon for Russian Automatic Speech Recognition (Pilot Study)

Anna Shirokova, Boris Telesnin, Valeria Rogozhina
Stel CS, MSLU
Moscow, Russian Federation
2016

Pronunciation variation

Variations in word pronunciation have multiple sources:

- Homographs
- Orthoepic ambiguities
- Individual manner or regional accent of a speaker
- Rapid fluent (especially informal) spontaneous speech

Starting points

- Pronunciation disambiguation is crucial for STT, ASR, NLP (Schultz, Kirchhoff 2006).
- Most state-of-the-art ASR systems use phone-based representations for acoustic modeling.
- explicitly specified pronunciations allow spoken language to be modeled more accurately.
- A pronunciation-based approach includes the potential for reducing the ambiguity of a given language writing system.
- If different acoustic realizations of a word are unlikely to be covered properly by the acoustic models, a given lexical entry may be assigned multiple pronunciations to represent these significant differences.
- When adding variants, one has to consider the types of speech that will be processed in order to add pronunciation variants relevant for the actual genre and style.

Aims to achieve, questions to answer

Our work is aimed at constructing a lexicon of effective pronunciation variants on the basis of the canonical pronunciations and implementing it into the ASR system for Russian (Zulkarneev &al 2013).

We take preliminary ASR and KWS experiments:

- to roughly assess a potential profit of the explicit adding of phonetic variants for reduced tokens;
- to assess the very appropriateness of taking into account multiple pronunciations in our ASR projects;
- to analyze whether there exist trends towards ASR performance gain achieved by using such an enhanced lexicon.

Related work

Our work has been inspired by a series of researches that deals with pronunciation variation phenomena and its influence on automatic speech recognition (Adda-Decker, Lamel 1998), (Adda-Decker&al 1999).

Formal phonetic rules for Austrian German conversational speech (Schuppler&al 2014).

For the Russian language the issue of pronunciation variety has been studied in theoretical and applied aspects:

- (Bondarko&al 1988) describes the phonetic system of spontaneous speech. It includes a phonetic lexicon of 80 Russian high frequency words, which gives a number of different phonetic representations for each word.
- In (Lobanov, Tsirul'nik 2007) systematic phonetic changes in word pronunciation are generalized as formal phonetic rules.
- An algorithm for automatic generation of pronunciation variants for Russian based on the results of (Lobanov, Tsirul'nik 2007) is proposed in (Kipyatkova, Karpov 2009) and is reported to be implemented into Russian ASR system (Kipyatkova&al 2013).
- The pronunciation variety and peculiarities of reduced word forms in the ORD speech corpus of Russian everyday communication are analysed in (Bogdanova, Palshina 2010).

Building Pronunciation Variants

- Building Canonical Pronunciations (Krivnova&al 2001)
- when creating orthographic transcripts of the speech data expert phoneticians were asked to mark words with gravely reduced pronunciations (contracted forms, phone deletion) and incorrect or non-standard pronunciations (stress position, etc);
- marked words were ranked according to their frequency of occurrence in the dataset, the list of 2000 most frequent words is taken into account;
- up to four most common variants of actual pronunciations were added to the lexicon (after expert auditive analysis of speech fragments referring to the marked words).

General variation types (Bondarko&al 1988)

Those referred to consonants include:

- deletion of /j/ in word initial, word final, intervocal positions and in V/j/C contexts;
- deletion of /v, v^j, b^j, d^j/ in intervocal position;
- deletion of one of double consonants;
- deletion of word-final plosives;
- consonant cluster reduction including phonetic changes across word boundaries (strong assimilation or total deletion of phones and phone sequences);

For vowels the following observations are made:

- stronger duration reduction even in stressed syllables;
- quality reduction in unlike position to the stressed syllable;
- delabialization of /u, o/ in weak positions;
- quality reduction of /u, y/ in weak positions;
- centralization in weak position;
- vowel deletion in unstressed syllables

ASR & KWS Experiments

tests	CORR%	WER%	FA	FR%
Baseline	64.01	40.44	2.005	26.98
Test_result_2tr	64.16	40.31	2.33	25.57
Test_result_vartr	64.31	40.62	3.52	22.67

It is supposed that strongly reduced variants tend to appear when an acoustic observation is unlikely to be recognized adequately.

To reduce the number of insertions and WER the shortest transcriptions should be eliminated from the pronunciation lexicon and another set of experiments needs to be taken.

For such purpose a special technique should be applied, which would enable to track the actual system choice of a pronunciation variant in the recognition process.

Challenges

Manual processing explains the lack of statistical data and estimations for the phone deletion and other segmental changes.

Phoneset resolution: experts are limited by the phoneset of the speech recognition system so they have to approximate their actual observation and refer it to some phonetic unit in the set, therefore missing some slight phonetic differences.

Different words are likely to share the same pronunciation variants, when applying such multi-lexicon to the ASR system. It inevitably leads to a higher rate of word confusion. (Adda-Decker, Lamel 1998)

Future work

Phonetics:

- to learn potential contexts of phonetic changes localization to be able to predict their occurrence;
- to make thorough statistical assessment for the types of phonetic changes regarding contexts of appearance.

In the further research the following strategies can be implemented to effectively introduce multi-pronunciation lexicon to ASR:

- to train acoustic models within added pronunciation variants to provide model compatibility;
- to take into account less variants and to select the most effective ones;
- to assign weight to variants.

What is the optimal strategy of lexicon building?

- general phonetic modification rules applied to total lexicon;
- manual adding required pronunciations for selected words .

Thank you

СТЭЛ КС

and

Welcome to our poster!

{anna_a, telesnin_ba}@stel.ru,
mind_your_own_business@rambler.ru