Премия СПбГУ за научные труды в номинации «за фундаментальные достижения в науке»

Skrelin P. Doctor of Philology, professor, head of department and professor of the Department of Phonetics and Methods for Teaching Foreign Languages (St. Petersburg State University)
Kocharov D. Candidate of Philology, associate professor of the Department of Phonetics and Methods for Teaching Foreign Languages (St. Petersburg State University)
Kachkovskaia T. Candidate of Philology, associate professor of the Department of Phonetics and Methods for Teaching Foreign Languages (St. Petersburg State University)

Цикл работ на тему «Автоматическое членение русской речи на супрасегментные единицы»

Выходные данные работ, выдвигаемых на конкурс.

1. Kocharov D., Kachkovskaia T., Skrelin P. (2017) Eliciting Meaningful Units from Speech // Proc. Interspeech 2017, pp. 2128-2132, DOI: 10.21437 / Interspeech 2017-855; (Scopus)
2. Kocharov D., Kachkovskaia T., Mirzagitova, A., Skrelin P. (2016) Combining Syntactic and Acoustic Features for Prosodic Boundary Detection in Russian // Statistical Language and Speech Processing 2016, pp. 68-79, LNAI, Springer International Publishing, DOI: 10.1007/978-3-319-45925-7_6; (WoS, Scopus)
3. Kachkovskaia T., Kocharov D., Skrelin P., Volskaya N. (2016) CoRuSS – a new prosodically annotated corpus of Russian spontaneous speech // Proceedings of the 10th International Conference on Language Resources and Evaluation, pp. 1949-1954; (Scopus)
4. Kocharov D., Kachkovskaia T., Skrelin P. (2016) Phonetic evidence for clitic-host relations within the prepositional group in Russian // Speech Prosody 2016, pp. 198-202; DOI: 10.21437/SpeechProsody.2016-41 (Scopus)
5. Kachkovskaia T., The Influence of Boundary Depth on Phrase-Final Lengthening in Russian // Statistical Language and Speech Processing 2015, pp. 1-8, LNAI, Springer International Publishing, DOI: 10.1007/978-3-319-25789-1 13; (Scopus)

Abstract

Prosody plays a major role in the organization of speech. Prosodic phrasing helps the speaker divide information into “meaningful” units, connect these units with each other to form larger units, show semantic relations between smaller units within larger ones. It may also serve to express additional connotations. The task of automatic prosodic boundary detection plays a significant role in various aspects of speech processing, such as text-to-speech synthesis, natural language understanding and translation. Speech recordings segmented into utterances and intonational phrases can be further analyzed in terms of realizations of melodic patterns, semantic relations between adjacent phrases within the utterance and utterances within the text. As discussed by Ladd, IP in its traditional sense has the following main properties: “(i) they are the largest phonological chunk into which utterances are divided, extending from one phonetically definable boundary to the next; (ii) they are a specifiable intonational structure, including—in most versions of the theory—a single most prominent point

(primary stress, tonic, nucleus); (iii) the are phonological units which are nevertheless assumed, ideally, to match up in poorly understood way with elements of syntactic or discourse-level structure." From this it follows that automatic prosodic boundary detection should be based on both syntactic and acoustic data. Syntactic data are derived from syntactic parsing. Such analysis shows whether two adjacent words are connected with each other syntactically. Acoustic data are based on acoustic features used as prosodic boundary markers. This analysis provides information on whether a particular word is realized as phrase-final or not. The present procedure combines these two sources in one system capable of predicting boundaries of intonational phrases (IPs) in speech. The paper includes descriptions of syntactic and acoustic components separately and in combination. In practice, even in read speech, syntactic and prosodic boundaries do not always coincide. A group of closely connected words can be split further into two or more parts—due to pragmatic reasons, or when the whole phrase is too long. However, in our analysis we assume that there are such word junctures where an IP boundary is highly improbable—e.g., between a preposition and its dependent noun. This is in accordance with the principles of prosodic hierarchy, where an intonational phrase (IP) is made up of phonological phrases, and an IP boundary cannot lie inside a phonological phrase. Based on this assumption, the syntactic component is designed to predict all potential IP boundaries (with a recall close to 100 %). As a result, the text is split into short phrases—mostly 1 or 2 words long. At the next stage these syntactic boundaries are used as input to the acoustic component: it chooses among only those word junctures where an IP boundary is possible. Working with texts, we can only speak of predicting those junctures where boundaries *may* occur. In terms of boundary placement, the same utterance may be realized by different speakers in different ways with no significant change in meaning. This may be illustrated by an example from our corpus, where the phrase "along a blind long stone fence" produced by eight speakers was never pronounced as one IP: five speakers split it after the first adjective ([along a blind][long stone fence]), and three speakers—after the second ([along a blind long] [stone fence]). These realizations do not differ functionally or semantically, and listeners perceive both as neutral. This is why we propose a two-stage procedure of combining syntax and acoustics. The first stage relies on syntactic data and consists in predicting all potential prosodic boundaries based on text. In other words, this step eliminates those junctures where a boundary is virtually impossible. Now, only the potential boundaries are passed on to the next stage. At the second stage acoustics come into play: using a statistical classifier, we perform automatic classification of potential boundaries predicted at the first stage based on our set of acoustic features. The presented two-stage procedure for automatic prosodic boundary detection has shown high efficiency. Without syntactic data, acoustics alone provide the efficiency of 0.86. Syntactic pre-processing enables to eliminate from further analysis a substantial part of word junctures where a boundary is extremely unlikely. This led to an efficiency (in terms of F_1 measure) more than 0.912, precision over 0.93, and recall 0.90. This is the highest reported result for Russian and among the highest for other languages. When evaluated on the BURNC Corpus this approach yields $F_1 = 0.76$, which is comparable with the top systems designed for English. The work «Phonetic evidence for clitic-host relations within the prepositional group in Russian» presents the relation between clitics and content words within prosodic words in Russian speech expressed by vowel-reduction patterns. The paper «The Influence of Boundary Depth on Phrase-Final Lengthening in Russian» presents the influence of boundaries between various prosodic units on vowel and consonant lengthening, which is one of the most prominent boundaries cues in Russian/