ISSN NO: 2770-0003
Date of Publication: 24-02-2022

Linguistic Features of Speech in Russian and Uzbek

Usmanova Dilfuza Tavfiqovna
Teacher of Chirchik higher tank command engineering institution, Tashkent
head teacher of the department of languages
+998973452869
Dilfuza1969t@mail.ru

Abstract: The conversion of text to speech is seen as an analysis of the input text to obtain a common underlying linguistic description, followed by a synthesis of the output speech waveform from this fundamental specification. Hence, the comprehensive linguistic structure serving as the substrate for an utterance must be discovered by analysis from the text. The pronunciation of individual words in unrestricted text is determined by morphological analysis or letter-to-sound conversion, followed by specification of the word-level stress contour

Key words: Strings, pronunciations, phrase-level parsing, pronominal reference

Introduction

There is an increasing need for computers to adapt to human users. This means of interaction should be pleasant, easy to learn, and reliable. Since some computer users cannot type or read, the fact that speech is universal in all cultures and is the common basis for linguistic expression means that it is especially well suited as the fabric for communication between humans and computer-based applications. Moreover, in an increasingly computerized society, speech provides a welcome humanizing influence. Dialogues between humans and computers require both the ability to recognize and understand utterances and the means to generate synthetic speech that is intelligible and natural to human listeners. In this paper the process of converting text to speech is considered as the means for converting text-based messages in computer-readable form to synthetic speech. Both text and speech are physically observable surface realizations of language, and many attempts have been made to perform text-to-speech conversion by simply recognizing letter strings that could then be mapped onto intervals of speech. Unfortunately, due to the distributed way in which linguistic information is encoded in speech, it has not been possible to establish a comprehensive system utilizing these correspondences. Instead, it has been necessary to first analyze the text into an underlying abstract linguistic structure that is common to both text and speech surface realizations.

Research methodology

Once this structure is obtained, it can be used to drive the speech synthesis process in order to produce the desired output acoustic signal. Thus, text-to-speech conversion is an *analysis-synthesis system.* The analysis phase must detect and describe language patterns that are implicit in the input text and that are built from a set of abstract linguistic objects and a relational system among them. It is inherently linguistic in nature and provides the abstract basis for computing a speech waveform consistent with the constraints of the human vocal apparatus. The nature of this linguistic processing is the focus of this paper, together with its interface to the signal processing composition process that produces the desired speech waveform. For any text-to-speech system, the process by which the speech signal is generated is constrained by several factors. The *task* in which the system is used will constrain the number and kind of speech voices required (e.g., male, female, or child voices), the size and nature of the vocabulary and syntax to be used, and the message length needed. Thus, for restricted systems such as those that provide announcements of arrivals and departures at a railroad station, the messages are very short and require only limited vocabulary, syntax, and range of speaking style, so a relatively simple utterance composition system will suffice. In this paper, however, it is assumed that the vocabulary, syntax, and utterance length are *unrestricted* and that the system must strive to imitate a native speaker of the language reading aloud. For the *language* being used, the linguistic structure provides many constraining relationships on the speech signal.

ISSN NO: 2770-0003
Date of Publication: 24-02-2022

Discussion

Much of the linguistic analysis used by text-to-speech systems is done at the word level, as discussed above. But there are many important phonological processes that span multiple word phrases and sentences and even paragraph level or discourse domains. The simplest of these constraints is due to syntactic part of speech. Many words vary with their functioning part of speech, such as "wind, read, use, invalid, and survey." Thus, among these, "use" can be a noun or verb and changes its pronunciation accordingly, and "invalid" can be either a noun or an adjective, where the location of main stress indicates the part of speech. At the single-word level, suffixes have considerable constraining power to predict part of speech, so that "dom" produces nouns, as in "kingdom," and "ness" produces nouns, as in "kindness." But in English, a final "s," functioning as an affix, can form a plural noun or a third-person present-tense singular verb, and every common noun can be used as a verb. To disambiguate these situations and reliably compute the functioning part of speech, a dynamic programming algorithm has been devised (Church, 1988; DeRose, 1988; Jelinek, 1990; Kupiec, 1992) that assigns parts of speech with very high accuracy. Once again, this algorithm relies on a statistical study of a tagged (marked for part-of-speech) corpus and demonstrates the remarkable capabilities of modern statistical techniques.

Results

Contemporary text-to-speech systems are available commercially and are certainly acceptable in many applications. There is, however, both much room for improvement and the need for enhancements to increase the intelligibility, naturalness, and ease of listening for the resultant synthetic speech. In recent years much progress has followed from the massive analysis of data from large corpora. Modern classification and decision tree techniques (Brieman et al., 1984) have produced remarkable results where no linguistic theory was available as a basis for rules.

Conclusion

In general, the use of standard algorithmic procedures, together with statistical parameter fitting, has been very successful. To further this process, large tagged databases are needed, using standard techniques that can be employed by many diverse investigators. Such databases are just beginning to be developed for prosodic phenomena (Silverman et al., 1992), but they can also be extremely useful for enhancing naturalness at the segmental level. While these statistical techniques can often extract a great deal of useful information from both texts and tagged phonetic transcriptions, the quest for appropriate linguistic models must be aggressively extended at all levels of representation. Where good models are available, such as for morphemic structure and lexical stress, the results are exceedingly robust. Linguistic descriptions of discourse are much needed, and a more detailed and principled prosodic theory that could guide both analysis and synthesis algorithms would be exceedingly useful. Of course, for some tasks, such as the conversion of abbreviations and standard symbols, there is relatively little linguistic content, and statistical techniques will have to bear the brunt of the task.

References

1. Allen, J. (1992) "Overview of Text-to-Speech Systems," in S. Furui, and M. Sondhi, eds., Advances in Speech Signal Processing, Marcel Dekker, New York. pp. 741790.
2. Allen, J., M. S. Hunnicutt, and D. H. Klatt (1987), From Text to Speech: The MITalk System, Cambridge University Press, London.
3. Bachenko, J., and E. Fitzpatrick (1990), "A Computational Grammar of Discourse-Neutral Prosodic Phrasing in English," Comput. Linguist., 16:155-170.
4. Brieman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984), Classification and Regression Trees, Wadsworth and Brooks, Monterey, Calif.
5. Browman, C. P., and L. Goldstein (1989), "Articulatory Gestures as Phonological Units," Phonology, 6(2):201.
6. Campbell, W. N. (1992), "Syllable-Based Segmental Duration," in Talking Machines: Theories, Models, and Designs, G. Bailly, C. Benoit, and T. R. Sawallis, eds, Elsevier, New York, pp. 211-224.

ISSN NO: 2770-0003
Date of Publication: 24-02-2022

7. Carlson, R., and B. Granstrom (1986), "Linguistic Processing in the KTH Multilingual text-to-speech system" in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pp. 2403-2406.