Automatic Annotation for Korean Speech Corpus Analysis

Jiwon Yun (Cornell University), Hyun Kyung Hwang (International Christian University), Seongyeon Ko (Queens College, CUNY)

Introduction

We conducted automatic annotation for a speech corpus in Korean and evaluated its performance in comparison with manual annotation.

Automatic Annotation

30 hours of read speech (24,300 sentences) of a single speaker from a speech corpus (ETRI 2006) was annotated in the automated procedure.

Step 1: Labeling

Unlike languages like English whose text-to-speech principles are not highly regular, Korean pronunciation is mostly predictable from the orthography by a relatively small number of rules. We first Romanized the transcripts of the corpus, then used a finite state machine (Yun 2005) to convert them into phonetic symbols by applying pronunciation rules of Korean.

![Labeling Diagram]

Step 2: Alignment

To align the phonetic symbols to the sound files, we used Prosodylab-Aligner (Gorman et al. 2011). This tool is applicable to any language in the world since its mechanism is independent of language-specific features.

![Alignment Diagram]

Comparison with Manual Annotation

Manual annotation was carried out by three human transcribers for part of the corpus (25 sentences) to evaluate the performance of the automatic system.

Labeling

61 tokens of mismatch out of 1306 in labeling were found, yielding 95% of agreement between manual and automatic transcriptions. Among those tokens, 23 were unanimously labeled by all the human transcribers. This suggests that only 2% of the labels can be certainly regarded as errors by the automatic system. Such high accuracy in labeling can be attributed to the fact that the speech data were quite carefully read. Most discrepancies were due to optional phonological processes or lexical processes that are not predicted by a purely rule-based system.

Alignment

The time differences of boundaries were measured for identically labeled segments. The average deviation for the automatic procedure was 16 ms, whereas the manual procedure was 6 ms. This is comparable to results from previous studies such as Wesenick and Kipp (1996) for read German speech (automatic: 18 ms, manual: 10 ms) and Pitt et al. (2005) for spontaneous American English speech (manual only: 16 ms).

Conclusion

The automatic annotation system for Korean used in this study is fairly reliable, compared to the results of human transcribers or other automatic systems in previous studies. Thus we expect that this automatic system will be easily applicable to research in phonetics after relatively small amount of hand-correcting, which will significantly reduce time and efforts in annotation task.

References


Acknowledgement

We would like to thank Hyosung Hwang and Ho-young Lee for their help in obtaining the corpus. We are also grateful to Kyle Gorman for help in the use of the Prosodylab-Aligner and useful comments on our work.

correspondence: jy249@cornell.edu