Abstract: Vowel-like region based acoustic-phonetic analysis for phone recognition

Biswajit Dev Sarma

In landmark-based speech recognition approach, acoustic-phonetic information is extracted from the regions around certain landmarks. In a similar direction, a vowel-like region (VLR) detection based approach is proposed in this thesis for phone recognition, where VLRs are considered as landmarks. Vowels, semivowels and diphthongs are treated as VLRs and the rest category of sounds as non-vowel-like regions (non-VLRs). Basic signal characteristics of VLR and non-VLR are different. VLRs are used as carriers with one or more non-VLRs supporting around them. The VLRs are produced by exciting the vocal tract that has a wide open configuration or moderate constriction. On the other hand, in the production of non-VLRs, the vocal tract has a narrow constriction or a complete closure configuration. Hence, it is not proper to treat VLRs and non-VLRs in a similar fashion by computing same set of features out of them. In the first step, VLRs are detected by using excitation source and vocal tract system information. Next, vowel-like and non-vowel-like sounds are recognized separately by exploring different acoustic-phonetic features suitable for them. Analysis of vocal tract constriction (VTC) is made for different sound units and an evidence is extracted. The VTC evidence indirectly contains information about different cues in the non-VLRs such as frication, burst, voice bar, etc. and therefore is useful for recognition of non-VLRs. Similar acoustic-phonetic analysis is done for VLRs. Acoustic-phonetic features related to vowel height, roundedness and frontness are analyzed and used for vowel recognition. Although VLRs and non-VLRs have different characteristic features, some portion of the VLRs may be useful for the non-VLRs. For example, in stop consonant-vowel unit recognition, the transition region has information about the stop consonant unit. Hence, apart from VLRs and non-VLRs, the transition regions are also analyzed in this thesis. Important acoustic cues such as, frication noise, and transient burst contain dominant aperiodic components. Source and vocal tract information are explored for the detection of such dominant aperiodic component regions (DARs). Detected DARs and predicted transition regions are used for the selection of non-VLR features around the VLR onsets and offsets.