Consonant Voicing Investigation

Chris Chen

August 2019

1 Abstract

Obstruct consonant could be detected by energy difference profiles. In this project, we investigate a glorithm to detect three types of landmarks, [stop closure], [stop release], [fricative closure]. As a result, Overall detection rates for stop closure and release are 71.3% and 82.3% for obstruent landmarks in TIMIT, and fricatives yield 83.2% and 81.9% respectively.

2 Acoustic parameters

The parameters used in the proposed refinement method are specified as RMS energy, band RMS energy, spectral center of gravity, and voiced probability[7, 6, 5, 3, 4, 1, 2]. Additional parameters are applied to solve the problems in the previous landmark detection algorithm. Earlier study of landmark detection uses the peak amplitude of each sub-band as sub-band energy, because the peak amplitude for the range of the formant frequency represents the movement of formant approximately. Otherwise, the RMS energy averages the peaks and nulls in the power spectrum, thus it is not adequate to detect the abrupt change in the sub-band. However, the RMS energy efficiently monitors the changes of energy in perspective. Thus, the RMS energies are added to the proposed system to improve the performance of landmark detection. RMS energy is additionally used to detect the landmark of stop closure, and band RMS energy between 5.0 to 8.0 kHz is used to detect obstruent burst landmarks. Abrupt changes are calculated by the peaks of the extreme values from the first order difference of each RMS energies. Spectral center of gravity and voiced probability are described in detail next.

3 Experiment

Database is subdivided into development set and test set from the TIMIT corpus. The development set is composed of 100 sentences included in the sx groups of TIMIT training set. It is used to develop the landmark detection algorithm. For the performance evaluation, 840 sentences in the sx groups of TIMIT test set are selected. The numbers of tokens in the transcription of TIMIT test set are distributed as: fricative closure and release (3301), stop burst (3417), and stop closure (4068). The performance evaluation is tested with 30 ms error boundary and hand correction for the test result is not included in the evaluation. Overall, 84% of landmarks are detected within 30ms and 32% of landmarks are inserted. The insertion rate includes both the neutral and insertion in the test. The results for each landmark are shown in Fig. 4. For the result of obstruent closure and release landmark, strident fricatives such as /s/, /sh/, /z/, and /zh/ are matched to the landmarks over 92%. 98% of landmarks are detected for the affricates. For the stop consonants, voiced stops have less performance than the unvoiced stops. Unvoiced stops are matched to the landmark over 90% and landmarks are detected with around 80% for voiced stops. Especially, the bilabial voiced stops such as /b/ have lowest detection rates with 71%. To summarize result, voiced fricatives and stops show lower performance than unvoiced consonants because the parameters of voiced consonants located in the vicinity of vowel are smoothed and the result of vowels affect the decision of voiced consonants. 86% of stop closure landmarks are detected by the test. Voiced stop closure has less performance than the unvoiced stop closure. The results of stop closure are influenced by the types of stops. Bilabial stop closure has the highest performance and alveolar stop closure has the lowest performance.

References

[1] Fredericka Bell-Berti. Control of pharyngeal cavity size for english voiced and voiceless stops. *he Journal of the Acoustical Society of America*, 57(2):456–461, 1975.

[2] Court S. Crowther1 and Virginia Mann1. Native language factors affecting use of vocalic cues to final consonant voicing in english. *The Journal of the Acoustical Society of America*, 92(2):711–722, 1992.

[3] Xiang Kong, Jeung-Yoon Choi, and Stefanie Shattuck-Hufnagel. Analysis of distinctive feature matching with random error generation in a lexical access system. *The Journal of the Acoustical Society of America*, 138(3):1780–1780, 2015.

[4] Xiang Kong, Preethi Jyothi, and Mark Hasegawa-Johnson. Performance improvement of probabilistic transcriptions with language-specific constraints. *Procedia Computer Science*, 81:30–36, 2016.

[5] Christine H. Shadle. The acoustics of fricative consonants. *RLE Technical Report No. 506, MIT, Cambridge,MA*, 1985.

[6] Kenneth N. Stevens. Modelling affricate consonants. *Speech Communication*, 13(1):33–43, 1993.

[7] Kenneth N. Stevens. Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, 114(4):872–1891, 2002.