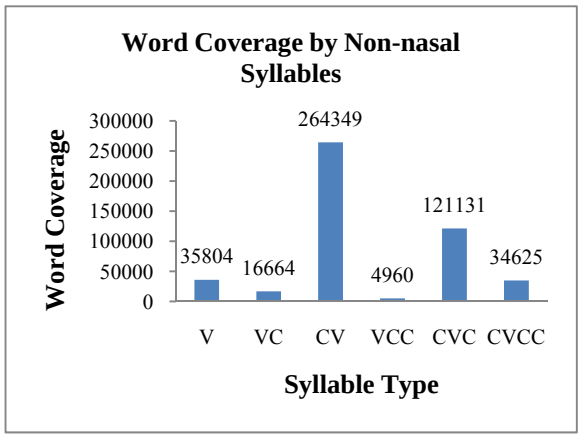

Figure 6. Word coverage by non-nasal syllables

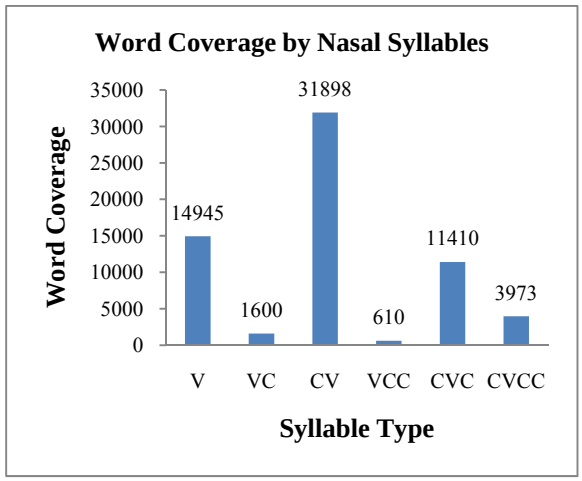

Figure 7. Word coverage by nasal syllables

4. Conclusions

It has been observed from the above results that the statistical analysis of the Punjabi syllables over the Punjabi corpus plays a vital role in selection of syllables for the speech database. Results show that large number of syllables are not occurring even once in the corpus of about 104 million words and are declared as invalid syllables. Also, other large number of syllables are having comparatively very little frequency of occurrence and are ignored for the final selection. Also for improving the quality of speech database the syllables have been analyzed statistically for the three (staring, middle and end) positions in the words of the corpus. This statistical analysis helped to select a relatively small syllable set (of about first ten thousand syllables, that are about 0.86% of total syllables) of most frequently occurring syllables having cumulative frequency of occurrence less than 99.81%, out of 1156740 available syllables. The results of this statistical analysis will also be very much helpful for the implementation of the other syllable based NLP systems.

5. References

[1] N. Kalyani, K.V.N. Sunitha, "Syllable analysis to build a dictation system in Telugu language", *International Journal of Computer Science and Information Security*, Vol. 6, No. 3 (2009), pp.171-176.

[2] P. Singh, *Sidhantik Bhasha Vigeyan*, 4th Edition, Madan Publications, Patiala India (2002), pp. 371-372.

[3] E.V. Raghavendra, S. Desai, B. Yegnanarayana, A.W. Black, K. Prahallad, "*Global syllable set for building speech synthesis in Indian languages*", in proc. of IEEE Workshop on Spoken Language Technologies (2008), Goa, India.

[4] Narayana, M.L., Ramakrishnan, A.G., "*Defining syllables and their stress in Tamil TTS corpus*", in proc. of Workshop in Image and Signal Processing (WISP-2007), IIT Guwahati, India (2007), pp. 92-95.

[5] R.A. Krakow, "*Physiological organization of syllables: a review"*, Journal of Phonetics, vol. 27 (1999), pp. 23-54.

[6] R.K. Joshi, K. Shoff, S.P. Mudur, "*A phonemic code based scheme for effective processing of Indian Languages*", in proc. of 23rd Internationalization and Unicode Conference (2003), Prague.

[7] H. Lee, C. Seong, "*Experimental phonetic study of the syllable duration of Korean with respect to the positional effect*", in proc. of 4th International Conference on Spoken Language Processing, Philadelphia, PA, USA, (1996), pp. 1193-1196.

