Accessing *Style* of Spoken Speech

Sunil Kopparapu,* Akhilesh Srivastava, P.V.S. Rao

Tata Consultancy Services Limited,
Plot No 14, Sector 24, Vashi-Turbe,
Navi Mumbai 400705, India.

Keywords: Speech assessment, speech clustering, assessment metric

The ability to assess spoken speech based on the style of speaking is an important problem. In this paper, we propose a method, which robustly and accurately assess spoken speech based on the style of the speaker and classifies them into one of several predefined set of styles. The system uses a carefully selected list of 20 words and 10 sentences to assess the speaking style of a person.

The predefined style set is automatically constructed from a speech corpus, consisting of different speaking styles. We use a novel clustering technique to identify clusters of speech having similar styles. The clustering technique uses acoustic and stylistic parameters extracted from speech signal and uses the relative distances between speech files to form clusters. The centroids of the cluster is identified as that speech file that is most equidistant from all the speech files in that cluster. The cluster is assigned labels (good, average, bad) by an accent expert.

The speech of a person for speaking style is assessed by first *comparing* all the speech samples (of all the words and sentences) spoken by the person with the cluster centroids of the corresponding word or sentences. The distance of the speech file from the cluster centroids is calculated for all the spoken words and sentences using a variant of dynamic time warping algorithm. The final assessment of the speaking style of the speaker is based on a weighted sum of distances for each spoken word and sentence. Experimental results, performed on 30 speakers with varying speaking styles show that the system is able to match the judgment of the human accent expert in more than 90% of the cases.

In the full paper, we will detail the system architecture and dwell on the selection of predefined set of words and sentences; the MFCC and LPC based acoustic and stylistic parameters extracted from the speech sample to compare speech samples. We also present a metric to combine the distances of each of the spoken word (sentences) from the cluster centroids to produce a overall assessment of the speaker. We will also describe the clustering method used to find the centroids of the style clusters and present the experimental results.

---

*Contact surface: Cognitive Systems Research Laboratory, Tata Consultancy Services Limited, Plot No 14, Sector 24, Vashi-Turbe, Navi Mumbai 400705, India. phone: +91-22-56163251, fax: +91-22-27839926, e-mail: SunilKumar.Kopparapu@TCS.Com