# A System for Automatic Stutter Removal

Suraj S. Sheth, Om D. Deshmukh, Ashish Verma
IBM Research India
odeshmuk@in.ibm.com
ASRU Topic: Speech Signal Processing

In this session we would like to demonstrate a system for automatic stutter removal. Users will have an opportunity to speak stuttered or simulated-stuttered speech (e.g., 'pa.pa.pa.pathology') and hear the corresponding stutter-free speech generated by the system (e.g., 'pathology'). The system has no ASR-dependency thus making it potentially applicable to even those languages which currently lack ASR capabilities.

Currently, assistance to stuttering subjects comes largely from (a) interactive sessions between the subjects and speech and language therapists who provide guidance and feedback to improve speaking skills, (b) devices that provide visual feedback of the speech production apparatus to the subjects, and (c) electronic devices that fit around the ear and generate frequency-altered or time-delayed versions of the spoken utterances. Our effort to automatically and directly convert stuttered speech signal into its corresponding smooth version is first of its kind. The proposed system can provide the user with personalized feedback on the type and frequency of stutter he/she commits and a stutter-free rendition of his/her own speech signal.

The proposed system has three main components: (a) Speech Segmentation: The goal here is to demarcate silence regions from speech regions and to segment the speech into syllables. The proposed syllable boundary detection method incorporates a novel 'syllable-silence synchronization' step and is shown to perform superior to existing syllable segmentation methods, (b) Stutter Detection and Removal: Given the speech units, the next step is to identify stutter regions. Patterns in stuttered speech are used to identify long silences or stops closures before stop bursts (e.g., /b/, /p/) which can potentially be stutter elements. Unusually long phones (e.g., 'lllllost') indicate phone-level stuttering. Duration of spectrally-steady regions is used to identify such phone-level stuttering instances. Syllable repetition, which is also a strong indicator of stutter, is detected by acoustically comparing consecutive syllables. Both frame-level and syllable-level features are used for syllable comparison. Stutter removal techniques depend on the type of the stutter detected. For example, in the case of syllable repetition, all but one instance of the syllable is retained. (c) Speech Signal Reconstruction: The stutter-free regions of the original speech signal are combined to construct the final stutter-free speech signal. The current system uses the Pitch Synchronous OverLap and Add (PSOLA) method for speech reconstruction.