

Real Time Speaking Rate Monitoring System

Meghna Abhishek Pandharipande & Sunilkumar Kopparapu TCS Innovations Lab – Mumbai, Maharastra, India.

2011 IEEE International Conference on Signal Processing, Communications and Computing, Sept 14-16, 2011, Xi'an Garden Hotel, Xi'an China.

Introduction

A metric to measure the rate at which speech is spoken

Why is the Need to Measure Speaking Rate?

- Fast Speaking Rate makes conversation unintelligible
- Whereas Slow Speaking Rate makes conversation boring
- Monitoring of Speaking Rate can assist call center agents speak at the desired rate
- Several aspects of a call conversation can be analyzed based on speaking rate
 - Without actually having to convert speech to understandable text (speech to text)

What Influence Speaking Rate?

- Speaking rate definitely varies depending on the *emotional state* of the person when s/he is speaking
- Regional and cultural influences thespeaking rate

How is it Measured?

• Measured in *number of words spoken per minute* or *number of syllables spoken per*

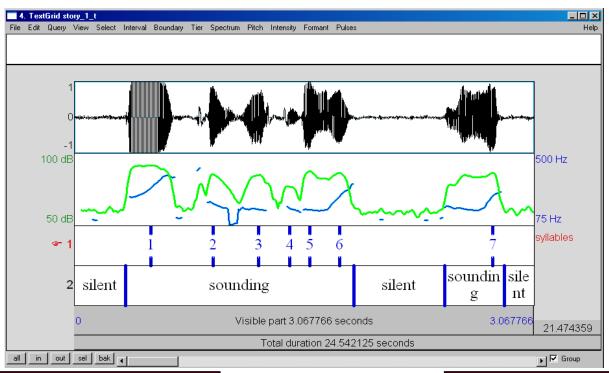
Speaking Rate

- As per studies, Average English-speaking rate
 - between 130 and 200 Words Per Minute (WPM). applies to 90% of the English-speaking population.
 - complex material, a speaking rate of 130 to 145 WPM may be required and
 - average complexity, a speaking rate between 145 and 175 is optimal
 - simple material, many listeners can accommodate over 175 WPM.
- Monitoring Speaking Rate of a call center agent in real time is essential and can improve CSI (Customer Satisfaction Index) by several notches

Syllable Nuclei Detection Algorithm - Nivja Jong

- The algorithm describes sequence of actions
 - Find syllable nuclei using intensity (dB) and voicedness.
 - Intensity is used to find peaks in the energy countour
 - vowel within a syllable (the syllable nucleus) has higher energy than surrounding sounds
 - Then use intensity contour to make sure that the intensity between the current peak and the preceding peak is sufficiently low.
 - With this procedure, we delete multiple peaks within one syllable
 - Finally, we use voicedness to exclude peaks that are unvoiced, which is required to delete surrounding voiceless consonants that have high intensity
 - sound files that are quite noisy should be filtered so that the frequency range is speech-band limited.

Detected Syllable Nuclei using Algorithm



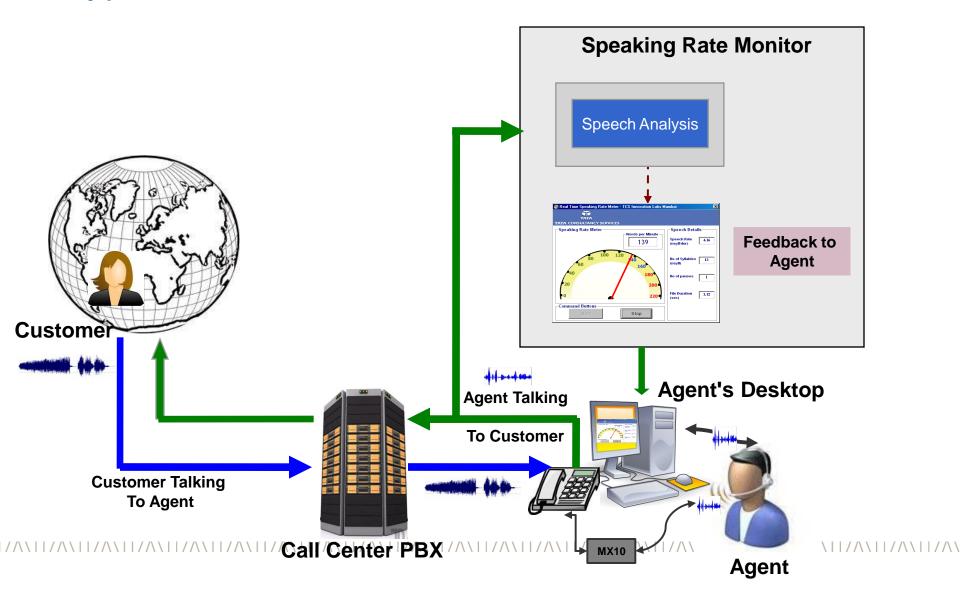
 $SR_{(wpm)} = \Psi \times SR_{(sps)} \times 60$

Where, N = number of syllables spoken per unit time

Where conversion factor, $\Psi = 1.5$

Tsilent = duration for which no speech was detected

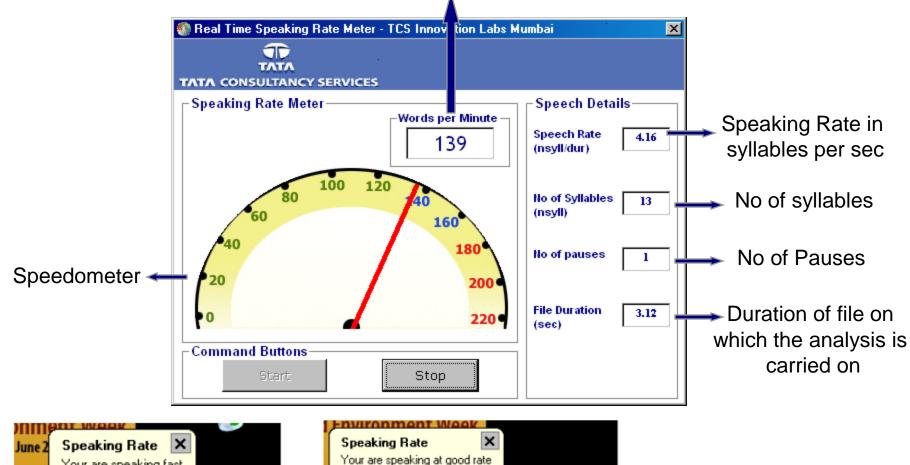
A Typical Call Center Senario



Speaking Rate Monitoring Real Time – Agent

Interface

Speaking Rate as words spoken per minute





//\\||





Experimentation and Results

- Level 1: Performance Accuracy was tested on a text "The first one believed in faith, he thought"
 - Method 1: Using the Algorithm
 - No of syllables = 8
 - Method 2: *Manually* calculating the number of syllables in the text
 - Th/e/f/i/rst /one/ b/e/l/ie/v/e/d in f/ai/th, he th/ou/ght
 - No of syllables = 8
- Level 2: Performance Accuracy: A short paragraph in English was asked to be spoken by 10 persons with 3 different Speaking Rate
 - The number of syllables identified using algorithm for these 30 spoken speech was within 10% of the actual number of syllables, present in the text

Challenges

- For Enabling Real Time Speaking Rate Monitoring System we had the challenge of computing the speaking rate not only accurately but also quickly.
 - Experimented with several duration of speech signal and converged to the fact that analyzing a 5 sec speech sample every 1 second gave the desired results of accuracy
- Note: The longer the duration of the speech analyzed the better is the accuracy of detection of syllables and hence computation of speaking rate.
- Further the delay of 5 sec was found to not cause any feel of a delay by the agent.

Conclusions

- The speaking rate monitoring system was used on the agents desktop in the floor of a call center and gave a real time feedback to the agent.
- The feedback allowed the agent to change the speaking rate on the fly
- Speaking Rate icon on taskbar/popup messages has a significant effect on CSI in an Agent-Customer interaction
- Use Cases
 - Analysis based on Speaking Rate can *identify calls with low CSI* (agent speaking fast) and also where the *customer is unhappy*.

Thank You.

meghna.panddharipande@tcs.com

Tata Consultancy Services- India.