A Suite of Mobile Applications to Assist Speaking at Right Speed

Imran Ahmed, Meghna Pandharipande, Sunil Kumar Kopparapu

TCS Innovation Labs - Mumbai, Yantra Park, Thane (West), Maharashtra, INDIA
{ahmed.imran, meghna.pandharipande, sunilkumar.kopparapu}@tcs.com

Abstract

One of the prominent reasons for ineffective communication in call center telephone conversations is primarily due to the manner in which the voice agents speak and not necessarily due to what they speak. Speaking rate, a non-linguistic aspect of speech, is a critical factor affecting intelligibility and comprehension of speech in general and specifically in call center telephone conversations. There have been attempts to monitor speaking rate of agents in a call center setup. However, there has been a paradigm shift to the conventional call centers with companies opting for home-agents. In this model the agents, unlike in the current call center setup, can operate virtually from anywhere using their mobile phones. In this paper, we present SpeakRite - a suite of Android mobile applications that can assist home-agents. While one of the components of SpeakRite analyzes the speaking rate during an ongoing telephone conversation and provides a real time feedback to assist the speaker modify his speaking rate, there are several other components that allow the home-agent to assess their speaking rate and learn to speak right themselves.

Index Terms: speaking rate, speech rate, mobile application, home-agents

1. Introduction

In voice based call centers the effectiveness of telephone conversations between the agent and the customer is determined more by the manner in which the agent speaks and is less dependent on what the agent speaks [1]. Ineffective communication in telephone conversations is especially prominent when people from different geographies or cultures converse in a language common to them like in the case of a voice based call center, where the call center voice agent and the customer can be from two different geographies. Speaking rate, a non-linguistic feature of speech is a critical factor affecting intelligibility and comprehension of speech [2, 3, 4, 5]. It is well known that speaking rate varies across native and non-native speakers of a language [4, 5, 6]. This variation of speaking rate additionally affects the comprehension of speech in native and non-native listeners. This observation seems to suggest that it is important that speakers converse at an optimal speaking rate for an effective conversation. Good communicators continuously monitor their speaking rate. They consciously adjust their conversational pace to get the message across effectively and efficiently. However monitoring and maintaining the speaking rate at the desired levels, may be hard for an average person who is not conscious of the rate at which he is speaking or in a very practical setup is in an emotional state that does not allow him to concentrate on the speaking rate.

Automatic monitoring of speaking rate by analyzing speech in real time can help speakers speak at the right speed and make conversations effective. In [7] a server based speaking rate monitoring tool was proposed which assists call center voice agents, in real time, to maintain an optimal speaking rate. As discussed in [7] the signal processing of speech happens on a server making it real time. However, driven by a number of business and socioeconomic parameters there has been a paradigm shift in what we know as call centers today in the form of virtual call centers or home-agents [8, 9]. A home-agent, unlike a call center voice agent, works solely from home, with no office space at a company facility. Voice agents are required to make calls on their mobile phone to converse with the customers to answer their queries. While the concept of home-agents has been practiced for more than a decade, the pace of adoption has accelerated in the past few years, however this has been restricted to non-voice agents. It is predicted [10] that more than three hundred thousand home-based agents will be working in the United States by 2013. However, the most obvious challenge with home-based agents is that a supervisor cannot simply walk over to an agent to supervise [8]. The ability to act directly and spontaneously is not possible in the changed scenario of home-agent call center paradigm requiring remote monitoring and remote coaching, preferably through an automatic process.

In this paper, we present a suite of mobile phone applications which assist home-agents go about performing their tasks efficiently. The suite of applications contain a real time speaking rate monitoring tool, which can monitor the speaking rate of the mobile phone user, first presented in [11] and discussed here for the sake of completeness, during an ongoing conversation. The suite of applications contain some off-line tools which assist the home-agent to practice to speak at the right speed and a tool that can evaluate the performance of the home-agent after a conversation, thereby allowing the home-agent to identify scope for improvement.

The rest of the paper is organized as follows: In Section 2 we briefly discuss automatic speaking rate monitoring. We present the suite of applications which assist the home-agent perform efficiently in Section 3 and conclude in Section 4.

2. Automatic Speaking Rate Monitoring

In [7] a tool that computes speaking rate in real time was proposed. The tool assists the call center agent to maintain an optimal speaking rate in real time by providing a just in time feedback to the agent. They used the count of the syllables detected in speech as a measure to compute the speaking rate. The algorithm described in [12], modified to work for real time, was used to detect syllable nuclei in spoken speech. The syllables in spoken speech were detected as:

**Step 1.** Identify all the intensity peaks that are preceded and followed by an intensity dip.

**Step 2.** Of all the intensity peaks detected in Step 1, retain only those intensity peaks that are above a certain intensity threshold and mark them.
Step 3. Discard the intensity peaks, after Step 2., that are unvoiced. The voiced intensity peaks are the syllable nuclei.

Once the speaking rate is computed in terms of number of syllables per second (sps), we can compute the speaking rate in words per minute (WPM) using a multiplication factor, namely, \( S_{\text{WPM}} = \gamma \times S_{\text{sps}} \times 60 \), where \( S_{\text{WPM}} \) is the speaking rate in words per minute, \( S_{\text{sps}} \) is the number of syllables per second and \( \gamma \) is a constant that captures the average number of syllables per word; this depends on the language being spoken [13]. For English language \( \gamma = 1.5 \) as suggested in [14]. It must be noted that since the measure is based on count of syllables in speech, same technique of computing speaking rate can be applied across different languages. Thus the speaking rate measurement tool can be used across call centres serving in different languages. The same syllable detection procedure when implemented on a Android mobile phone with a 650 MHz processor imposes enormous computational load and fails to operate in real time. In order to enable real time operation on a mobile phone, the syllable detection algorithm was modified as discussed in [11].

3. SpeakRite Mobile Application Suite

SpeakRite is a suite of mobile applications useful for a voice home-agent. The tools available for the home agent are:

1. Tool to compute the speaking rate in real time even while the speaker is in a live conversation. Figure 1 shows a screen-shot of the tool monitoring the speaking rate of the user, while a call is active.

2. SpeakRite tool set can also provide a complete analysis of the speaking rate for the entire duration of a recorded call. Figure 2 shows a screen-shot of the SpeakRite showing the variations in speaking rate for the entire call duration.

3. SpeakRite allows the user to manually define a desired speaking rate as a reference. This reference can be used to give a feedback to the user in terms of how he is speaking relative to the set speaking rate. Optionally, the reference value can be remotely set by the call center or the user can record any speech and set its average speaking rate as the reference.

4. SpeakRite gives an option to modify user’s speaking rate. In this option, the user is asked to read a pre scripted text displayed on the mobile phone screen. As the user speaks the speaking rate is analyzed. If the average speaking rate of this recording is different from the pre-defined reference then the speech recording is time-scaled, using WSOLA time-scale modification technique [15], such that the resulting speech has an average speaking rate equal to the reference speaking rate. This time-scaled speech is played back to the user to give him a feedback of his speech at the desired reference speaking rate. The WSOLA time-scale modification is implemented on Android using the SoundTouch Audio Processing Library [16]. Figure 3 shows a screen-shot of the SpeakRite application depicting the user recording before and after time-scale modification.

5. SpeakRite allows the user to self train. In this mode, a moving ticker text is displayed on the user’s mobile phone screen and the user is asked to read out the text as it is scrolling. The speed of the scroll is varied to allow the user to practice speaking at different speaking rates. This facilitates providing the user an instant feedback of his current speaking rate by movement of the speedometer needle on the mobile phone screen. Figure 4 shows a screen-shot of the SpeakRite application useful for training.

6. SpeakRite application also facilitates the users to check their progress as they train to speak at a particular rate. Figure 5 shows a screen-shot of the SpeakRite application depicting the distribution of the speaking rate of the user before and after training. This enables users to monitor their own progress.
4. Conclusion

For effective communication during a telephone conversation it is important that the speakers converse at an optimal speaking rate and learn to adjust their conversational pace to make themselves intelligible and comprehensible. With the advent of voice based home-agents it is necessary to provide the home-agents with a suite of tools that will allow them the necessary support to enable them perform their task efficiently. SpeakRite is a suite of mobile phone applications on an Android platform that provides this support. The suite consists of both real time and off-line applications. The real time application allows monitoring the speaking rate of the speaker and also gives an instant feedback of the current speaking rate. This real time automatic monitoring of Speaking Rate can sensitize speakers to their current speaking rate thereby allowing them to speak at the desired speaking rate. The suite has applications that provide off-line analysis of several aspects of speech which help the home-agent to practice, train and monitor his progress in achieving to speak at the desired speaking rate.

5. References