

Summary 3:

The Power of Human and Software Teamwork: A Captioning Service That Combines The Best of Both Worlds

Sam Diamantstein 101060342

COMP 3301, Sept 27, 2017

Scribe is a service that provides content captioning for deaf and hard of hearing(DHH) individuals and uniquely combines the talents of human and computers together to convert speech to text. Scribe's value is derived by providing cost-effective captioning services relative to state of the art solutions that are more expensive for comparably accurate captions.

Scribe is able to provide this cheap yet accurate service through its approach of pooling the talents of non-expert captionists that are each provided with a segment of a large piece of content, like a lecture. Through the collaboration of the multiple individuals, ideally around ten people, their individual portions are recombined by Scribes algorithms to create a completely captioned piece of content that achieves 74 percent coverage. The result is that non-experts can, through the teamwork and guidance of Scribes software, collectively create captions at a similar or better than quality to that of a professional. Behind Scribes effective blending of man and software is an interface that connects the two that relies on the following set of technologies.

In order to pool the non-expert captionists, Scribe uses Amazons Mechanical Turk which aggregates people who are willing and able to do tasks for a monetary incentive from a remote location. When an individual has been onboarded to Scribe and has learned how to use the software, they are collectively guided by the Scribes interface to turn the speech they hear into text. To achieve an accurate conversion, Scribe created TimeWarp, a tool that employs a Waveform Similarity Based Overlap and Add algorithm (WSOLA). In doing so, Scribe increases playback speeds during times where the captionists are not expected to perform speech to text conversion and slows it down when they are. TimeWarp cycles the two speeds to increase the word per minute typed and reduce latency by extension. Combining Timewarp with the staggering of captions presented, reflects the natural flow of writing. Additionally, Scribe uses visual effect of locking in words and color them green for the captionist, creating a highly tailored and effective relationship to produce high-quality captions.

The technology that underlies Scribe is impressive as it provides a service

offered at an unmatched price point and convenience. Scribe can generally be compared to two competitors, expert human captionists and automatic speech recognition software (ASR) like Dragon Speak. When compared by price, Scribe costs 8 to 12 dollars per person per hour whereas an expert would cost 120 to 220 dollars per hour. Scribe offers additional value over the expert as it is flexible, not requiring advanced notice and is able to be used on demand through the aggregation of remote Mechanical Turk employees. While ASR technologies are less expensive than their expert counterparts, it too is inflexible, requiring voice training that if withheld drops its accuracy below 50 percent. Lastly, Scribe creates text that is more accessible than its counterpart sign language, which many DHH people do not know.

In effect, Scribe has been able to blend the abilities offered by software and non-expert captionists to produce a service that for its price, is unmatched in the quality it demonstrates. The result is that a great deal of DHH individuals can now benefit from being able to more flexibly and cheaply understand and engage in the world of speech based communication.