

Speech Recognition: What's Left?

Michael Picheny¹

Abstract: Recent speech recognition advances on the SWITCHBOARD corpus suggest that because of recent advances in Deep Learning, we now achieve Word Error Rates comparable to human listeners. Does this mean the speech recognition problem is solved and the community can move on to a different set of problems? In this talk, we examine speech recognition issues that still plague the community and compare and contrast them to what is known about human perception. We specifically highlight issues in accented speech, noisy/reverberant speech, speaking style, rapid adaptation to new domains, and multilingual speech recognition. We try to demonstrate that compared to human perception, there is still much room for improvement, so significant work in speech recognition research is still required from the community.

¹ IBM T. J. Watson Research Center