

From Lord Rayleigh to Shannon: How do we understand speech?

Jont B. Allen
AT&T Labs - Research
Florham Park NJ

Oct. 19, 2001

Abstract

In 1908 Lord Rayleigh reported on his speech perception studies using the “acousticon” (a commercial sound system produced in 1905), demonstrating that he was well aware of the importance of the bandwidth in speech perception. It was the development of the telephone that both allowed and pushed mathematicians and physicists to develop the science of speech perception. Critical to this development was probability theory. One of their main tools was the confusion matrix which estimates the probability $P(H|S)$ of hearing phoneme h_i when speaking phoneme s_j , ($h_i \in H$ and $s_j \in S$ represent members from sets H and S respectively).

From 1910 to 1950 speech perception was extensively studied by telephone research departments throughout the world. However it was the work of Harvey Fletcher in 1921 that made the first major breakthrough. By 1930 millions of dollars were being spent on speech perception research at the newly created Bell Labs. The key was his quantification of the transmission of information, as characterized by the error patterns. The full and final theory was not published until 1950, following Fletcher’s retirement.

During WWII the Harvard Acoustics Lab took on this problem. The next breakthroughs were provided by George Miller and his colleagues. Miller used concepts from information theory developed at Bell Labs by Claude Shannon to quantify speech entropy. While these studies provide key insight into speech perception, they do not take the final elusive step that would allow us to build robust automatic speech recognition (ASR) machines.

Regardless of what you read in the popular press, ASR is still an unsolved problem. I will attempt to pass along some wisdom I have learned over the years on what we now know about human speech recognition (HSR). It is hoped that by learning more about HSR we might make ASR more robust to noise and filtering.

My talk will be in three parts. In part one I will describe the 30 years of work by Fletcher and his colleagues which resulted in the “articulation index,” a widely recognized method for characterizing the information bearing frequency regions of speech.

In part two I shall describe the work of George Miller. Miller studied the importance of varying the source entropy (randomness) in speech perception. He did this by controlling for both the cardinality (size of the test corpus) and the signal to noise ratio of the speech samples.

In part three I shall describe my recent experimental work in build more robust speech recognition systems. One goal is to make a system that works as well as human listeners in decoding degraded (filtered and noisy) nonsense speech sounds.