From Lord Rayleigh to Shannon: How do humans decode speech?

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

Date: Mar. 18, 2003 (Tuesday)
Time: 6:15 p.m. (refreshments start at 6:00 p.m.)
Place: Room 202, ECE Building, NJIT, Newark, NJ

About the Talk

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 and blind speech testing 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 of hearing phoneme Pi when speaking phoneme Pj.

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 breakthroughs. By 1930 millions of dollars were being spent every year on speech perception research at the newly created Bell Labs. The key was his quantification of the transmission of information, as characterized by phone error patterns. Fletcher's full and final theory was not published until 1950, following his AT&T retirement.

The next breakthroughs were provided by George Miller and his colleagues at the Harvard Acoustics Lab during and following WWII. 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 robust to noise and filtering. Today ASR is based on language models which have not, and can not, give ASR the basic robustness to noise and filtering found in HSR.

I will summarize important results from the 30 years of work by Fletcher and his colleagues, which resulted in the ``articulation index.'' We now know that the AI is functionally the same as Shannon's channel capacity.

Next I summarize the speech work of George Miller. Miller showed the importance of 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. I show show how to use the singular value decomposition to convert the sound confusion matrix into a metric of distances, to define perceptual sound groupings.

Finally I show how language modeling cannot solve the speech recognition problem. The key to robustly decoding phones resides in extracting more fundamental events from the speech stream, which we call `events.' These events are transient, and can last for only 10 ms. They seem to hold the key to consonant-vowel discrimination.

About the Speaker

Jont Allen received a BS in EE from the University of Illinois in 1966, and PhD from the University of Pennsylvania in 1970. He then joined Bell Laboratories in 1970, where he was in the Acoustics Research Department as a Distinguished member of Technical Staff. From 1996-2002 he worked at AT&T Labs as a Technology Leader.In 2003 he join the ECE faculty, University of IL, UIUC.

Dr. Allen is a Fellow of the Acoustical Society of America (ASA), an IEEE Fellow, a past member of the Executive Council of the ASA, the Adm. Committee (ADCOM) of the IEEEE ASSP Society, an Editor of the ASSP transactions, the Chairman of the Publication Board of the ASSP Society, the General Chairman of ICASSP-1988. In 1986 he was awarded the IEEE ASSP 1986 Meritorious Service Award. In 1986-88 he participated in the development of the AT&T multi-band compression hearing aid, that is now sold under the Resound name. In 1990 he was an Osher Fellow at the Exploratorium museum in San Francisco. In 1991-92 he served as an international Distinguished Lecturer for the Signal Processing Society. In 2000 He received a IEEE 3d Millennium Metal for 'Outstanding achievements and contributions.'

Dr. Allen has worked for 30 years on human perception and is an expert on cochlear function, the noninvasive diagnosis of cochlear disfunction, and speech perception.

Sponsors: **IEEE Communications Society and Signal Processing Society North Jersey Chapters**
NJIT Electrical and Computer Engineering Department

For more information contact Nirwan Ansari (973) 596-3670, or Yun Q. Shi (973)-596-3501**, or check http://web.njit.edu/~ieeenj/comm.html for latest updates. Directions to NJIT can be found at: http://www.njit.edu/University/Directions.html.**