Question

ASR performance degrades pretty uglily in noisy or mismatched environments. But human does it quite well in a seemingly effortless manner. So the combined process of speech production/recognition must be robust to noise. Specifically, here we can see two levels of noise robustness: speech is distinguishable from noise and speech units are mutually distinguishable in the presence of noise. The second one is very difficult for ASR.

From the knowledge of phonetics, is there any evidence that the phoneme set is intrinsically discriminative and robust to noise? If not, at what level does the discriminativity or robustness become evident?

Question

The gap between the performance of speech recognition by human and by machine underlines that fact that there is still much to be learned from human study. There are arguments that attribute such superiority to linguistic knowledge. However, even in non-sense phone sequence recognition, human leads by an order of magnitude in error rate. It appears that the current acoustic processing, mainly on the spectral domain, needs to be reinvented.

From the perspective of linguistics or phonetics, what are the most important things for the perception of phones or syllables?