Magic numbers vs. sound reasoning

Different scientific sub-cultures such as phonetics (and psychology) vs. automatic speech processing (including Computational Paralinguistics) have evolved within different "evaluation paradigms". By that I mean simply: "which (types of) numbers tell us how good we are?". Such numbers can be a strong threshold for "statistically significant or not", or they can be conceived as weak threshold such as "low" vs. "medium" correlation, or as classification performance.

I will describe the historic and still relevant discussion about the meaning and meaningfulness of inferential statistics (statistical significance), the different semantics of all these different numbers, and possible ways out and towards some common language measure that is both meaningful and understandable. I will conclude with some pragmatic advices on how to survive, i.e., what to do and what not to do when reporting and evaluating results.

Reference: