In the domain of Automatic Speech Recognition (ASR), Inverse Text Normalization (ITN) is applied after the speech recognition step to transform recognized verbalized text into written form. This process includes converting verbalized numbers into digits, formatting dates and monetary amounts, and applying correct capitalization and inserting punctuation marks. As ITN systems serve as post-processing modules for ASR outputs, integrating the original audio input as an additional signal into the ITN system is also possible. In this thesis, we explore the impact of the speech signal on the performance of ITN neural models and create a dataset for training and evaluating speech-informed ITN models. Our best model demonstrates a significant improvement in the precision and recall of inserting periods, commas, and question marks, as well as in adding letter casing, when compared to the text-only baseline. Improvements are also observed in less frequent punctuation symbols, though they are not statistically significant.