

V oblasti automatického rozpoznávání řeči (ASR) se po rozpoznání řeči používá inverzní normalizace textu (ITN), která transformuje rozpoznáný verbalizovaný text do písemné podoby. Tento proces může zahrnovat převod verbalizovaných čísel na číslice, formátování dat, peněžních částek a aplikaci správné kapitalizace a vkládání interpunkce. Vzhledem k tomu, že ITN systémy se používají pro zpracování výstupů ASR, je také možné do ITN systému integrovat původní zvukový vstup jako další signál. V této diplomové práci zkoumáme vliv zvukového signálu na výkonnost neuronových ITN modelů a vytváříme datovou sadu pro trénování a vyhodnocování ITN modelů používajících řeč. Náš nejlepší model vykazuje výrazné zlepšení přesnosti a úplnosti při vkládání teček, čárek a otazníků, stejně jako při přidávání velkých písmen ve srovnání s výchozím modelem, který pracuje pouze s textem. Pozorujeme také zlepšení u méně častých interpunkčních znamének, i když toto zlepšení není statisticky významné.