

## Abstrakt

Modely pro práci se sekvenčními daty založené na neuronových sítích v posledních letech výrazně přispěly ke zlepšení kvality systémů pro syntézu řeči. Zatím však nebyl navržen systém, který by se dal rychle a efektivně natrénovat, byl schopen rychlé syntézy a zároveň generoval hlas ve vysoké kvalitě. V této diplomové práci představujeme systém pro syntézu řeči založený na neuronových sítích, který je schopen generovat spektrogramy v reálném čase, má nízké nároky na výpočetní kapacity a je rychle trénovatelný. Náš systém sestává ze dvou neuronových sítí – ze studenta a z učitele. Učitel je využit k aproximaci zarovnání mezi textem, z něhož má být generována mluvená řeč, a mezi korespondujícím spektrogramem. Student využívá získaná zarovnání pro efektivní syntézu spektrogramů z fonetické reprezentace vstupního textu. Obě sítě využívají jednoduché konvoluční vrstvy. Náš systém trénujeme na anglické datové sadě LJSpeech. Systém je schopen generovat řeč v reálném čase na běžném procesoru a je možné jej rychle natrénovat na jedné graické kartě. Kvalita řeči generované z našeho systému byla ohodnocena jako signifikantně vyšší v porovnání s konkurenčními systémy.