

Abstrakt

Tato práce se zabývá vícejazyčnou syntézou řeči. Porovnali jsme tři odlišné modely, které jsou založeny na Tacotronu. Tyto modely se liší především v přístupu ke sdílení informací a parametrů mezi jazyky. Dva z nich navazují na současné systémy pro vícejazyčnou konverzi textu na řeč. První využívá plně sdíleného enkodéru a doménově specifického klasifikátoru, který je modifikovaný za účelem odstranění informací, které závisí na syntetizovaném hlase, z enkodéru. Druhý model používá separátní enkodér pro každý jazyk. V této práci navrhujeme nový přístup, který kombinuje nejlepší z obou zmíněných metod. Díky technikám metaučení umožňují efektivní sdílení parametrů při zachování flexibility. Tyto tři modely porovnááme na dvou úlohách. Jedna z nich se zaměřuje na sdružené vícejazyčné učení na deseti jazycích a odhaluje možnosti porovnávaných modelů sdílet znalosti mezi jazyky. Druhá se zabývá syntézou vět, které obsahují výrazy z několika různých jazyků. Dokládáme, že náš nový přístup umožňuje efektivní sdílení informace mezi jazyky a že dle subjektivního hodnocení produkuje přirozenější řeč bez častých přerůků a chyb ve výslovnosti.