

Generování hudby je jedna z nejtežších úloh pro umělou inteligenci. Přístupy založené na strojovém učení již dokázaly přesvědčivě napodobovat skladatelskou práci na kratších hudebních úsecích, nicméně většina současných přístupů nedokáže vytvořit delší koherentní skladby. Tato práce chce ukázat, jak můžeme ovládat modely hlubokého učení z vnějšku, abychom docílili dlouhodobé koherence vygenerovaných skladeb s více hlasy. Pracujeme s klasickou hudbou s instrumentací od klavíru, přes smyčcový kvartet až po symfonické skladby. Ovladatelnost generovacího procesu spočívá ve využívání hudební formy, která je v abstraktním slova smyslu obecným popisem pro podobnosti a nepodobnosti hudebních pasáží a jejich rozložení ve skladbě. Pro dosažení této ovladatelnosti, (1) navrhujeme pseudometriku pro porovnávání hudební podobnosti s pomocí již existujících metod, (2) extrahujeme hudební formu z trénovacích dat pomocí shlukovacího algoritmu nad podobnostními hodnotami, (3) trénujeme tři modely, které generují podobné, nepodobné a lokálně koherentní hudební fragmenty, a (4) navrhujeme způsob, jak využívat extrahované hudební formy společně se třemi modely pro generování delších koherentních skladeb z hudebních fragmentů. Ukazujeme, jak si v generování hudebních fragmentů vedou modely "transformer", když jim předkládáme data s různorodými možnostmi výstupů (různé varianty pro jeden vstup). Opakovaný běh modelu pro generování variant se stejným vstupem může být využit jako inspirační balíček pro hudební skladatele. Abychom vyhodnotili úspěšnost generovacího systému, experimentovali jsme s adaptací "Fréchet inception distance" pro hudbu, kde se ukázalo, že v určitém smyslu jsou vygenerované skladby podobné těm reálným. Nicméně, nejrelevantnější ohodnocení skladeb záleží převážně na tom, zda-li je systém schopný generovat skladby, které opravdu zaujmou posluchače (nebo hráče na hudební nástroj). Na základě vygenerovaných ukázek věříme, že systém je úspěšný.