

Univerzita Karlova

Filozofická fakulta

Ústav informačních studií a knihovnictví

Studia nových médií

Diplomová práce

Bc. Vojtěch Jouza

Umělá inteligence a kompozice vážné hudby

Artificial Intelligence and Classical Music Composition

Praha 2020

Vedoucí práce: Mgr. Josef Šlerka, Ph.D.

Poděkování

Rád bych na tomto místě poděkoval vedoucímu práce Mgr. Josefu Šlerkovi, Ph.D. za pomoc při vypracování tohoto textu, stejně jako Mgr. Dítě Malečkové, Ph.D., Mgr. Jakubovi Fialovi a Janu Pudlákovi, se kterými jsem práci konzultoval. Tato práce by nevznikla také bez podpory a trpělivosti rodiny a přátel. Speciální dík patří Karolině Vocelové.

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 19. července 2020

Vojtěch Jouza

Klíčová slova

Turingův test, umělá inteligence, skládání hudby, klasická hudba, evaluace kreativních systémů, Rozlišovací test, skladatel, kreativita

Keywords

Turing test, artificial intelligence, composing music, classical music, evaluation of creative systems, Discrimination test, composer, creativity

Abstrakt

Tato práce se zabývá umělou inteligencí skládající klasickou hudbu a způsoby hodnocení jejího výkonu posluchači. Text poskytuje první přehled provedených experimentů vycházejících z tzv. Turingova testu a na základě analýzy dochází k jejich možným vylepšením z hlediska metodologie. V závěru navrhuje alternativní test, který se odklání od filozofických implikací Turingova testu a oproti původnímu experimentu poskytuje i prostor pro hudebně-teoretickou analýzu generovaných děl.

Abstract

This thesis deals with artificial intelligence composing classical music and ways of evaluating its performance by listeners. The text provides the first overview of the conducted experiments based on the so-called Turing test and, set up on the analysis of primary sources, it suggests possible improvements in terms of methodology. In the end, we propose an alternative test, which rejects the philosophical implications of the Turing test and, in contrast to the original experiment, also provides room for a music-theoretical analysis of the generated works.

Obsah

ÚVOD	8
1. HISTORIE ALGORITMICKÉHO KOMPOZOVÁNÍ.....	10
1.1. RANÁ HISTORIE	10
1.2. ARS COMBINATORIA.....	12
1.3. MUSIKALISCHES WÜRFELSPIEL	13
1.4. ALGORITMICKÝ SKLADATEL	15
2. VYNALEZENÉ PROGRAMY A FORMÁLNÍ METODY KOMPOZICE	17
2.1. SYMBOLICKÁ UMĚLÁ INTELIGENCE	19
2.2. STROJOVÉ UČENÍ: MARKOVY ŘETĚZCE A NEURONOVÉ SÍTĚ	22
2.3. EVOLUČNÍ METODY	24
2.4. LIMITY	25
3. HUDEBNÍ TURINGŮV TEST.....	29
3.1. KREATIVITA UMĚLÉ INTELIGENCE A PROBLÉM AUTORSTVÍ	32
3.2. VÝSTUP NEBO PRODUKCE?.....	35
3.3. EVALUACE KREATIVNÍCH SYSTÉMŮ	37
4. PROVEDENÉ TESTY	43
4.1. MUSICAL OUTPUT TOY TEST	43
4.2. MUSICAL DIRECTIVE TOY TEST.....	49
4.3. ROZLIŠOVACÍ TESTY	51
4.4. POSLECHOVÉ TESTY A DALŠÍ METODY	57
4.5. KRITIKA POUŽITÝCH METOD	60
5. DISKUZE	64
5.1. NÁVRH	64
5.2. ALTERNATIVNÍ TEST	67
ZÁVĚR	70
SEZNAM POUŽITÉ LITERATURY.....	72
SEZNAM OBRÁZKŮ	92
SEZNAM ZKRATEK	93

Úvod

Výzkum umělé inteligence přinesl od 50. let minulého století pokusy programátorů, skladatelů a hudebníků automatizovat proces skládání hudby. Motivace alespoň částečně nahradit člověka může být různá. Takovéto vynalezené programy mohou umět generovat hudbu v reálném čase a improvizovat s dalšími lidskými hudebníky, sloužit jako inspirace skladatelům při jejich práci nebo působit jako autonomní skladatel, který imituje existující hudební styly a lidské umělce nebo který komponuje hudbu ve stylu vlastním. V této práci se budeme zabývat výhradně posledním typem programu, který má plnohodnotně napodobit práci skladatele klasické hudby.

V současné době dokáží vynalezené programy generovat přesvědčivou hudbu ve stylech skladatelů, jako jsou například Johann Sebastian Bach, Ludwig van Beethoven, Fryderyk Chopin nebo György Ligeti. Od začátku vzniku těchto programů si jejich autoři kladou otázku, jestli se tato hudba může rovnat hudbě lidských skladatelů a odpověď pak někteří hledají ve známém testu Alana Turinga (1950), respektive Turingově *Imitační hře*, v tomto kontextu založené na rozlišování člověka od počítače na základě poslechu jejich hudby.

Záměrem této práce je poskytnout ucelený přehled provedených testů vycházejících z konceptu Turingova testu v oblasti skládání klasické hudby, kriticky tyto testy zhodnotit a na základě jejich analýzy navrhnout možná vylepšení s ohledem na jejich metodologii. Srovnání experimentů bude nezbytně předcházet také souhrn metod pro objektivní evaluaci kreativních systémů zaměřených na hodnocení poslechem. Diplomová práce nemá za cíl provádět vlastní experiment, záměru bude dosaženo v teoretické a následné prakticko-analytické části důkladnou analýzou primárních a sekundárních zdrojů.

V první části se budeme zabývat konceptem tzv. algoritmického komponování, tedy částečně automatickým metodám umožňujícím lidem postrádajícím odborné znalosti skládat hudbu bez použití počítačů. Ve druhé kapitole popíšeme vynalezené programy, které skládají vážnou hudbu, nabídneme jejich výčet z hlediska využitých metod a pokusíme se zachytit jejich principy a omezení. Třetí

kapitola bude věnována teoretickému konceptu Turingova testu, kreativitě a metodám vnější evaluace a ve čtvrté kapitole navážeme výčtem a komparační analýzou provedených experimentů. Na základě jejich kritiky nakonec v poslední části navrhneme možný lepší způsob implementace Turingova testu.

1. Historie algoritmického komponování

Pokud mají počítačové programy umět samostatně skládat hudbu, je nezbytně nutné proces skládání i samotnou hudbu formalizovat – i přes nebezpečí, že ji kvůli tomu neobsáhneme v celé její hloubce (Loy, 1989). Tento proces skládání je u počítače možné popsat jako postup, který má konečný sled jednoznačných, proveditelných příkazů nebo pravidel, které po určitém čase vydají výsledek. Dá se tak označit jako algoritmus (Rapaport, 2020, s. 269).

Z této definice je zřejmé, že algoritmy nemusí být nutně spojovány jen s počítači a mohou je provádět i lidé. O těchto algoritmech nicméně nemusíme uvažovat jen v kontextu počítačových programů, které ji umějí různými způsoby generovat, ani z příliš obecného pohledu, označující jakýkoliv proces lidského skládání hudby jako algoritmus. Již téměř 1000 let můžeme pozorovat u skladatelů a teoretiků explicitní využívání různých postupů, které v kreativním procesu aplikují určitá pravidla na předem daný hudební materiál, v rozsahu od jednotlivých not až po celé takty a fráze (Ferreira a Machado, 2008). Takový proces pak lze označit jako tzv. *algoritmické komponování* (Nierhaus, 2009, s. 2), které vývojově přímo předchází komponování strojovému.

Tato kapitola přinese stručný historický přehled formalizujících přístupů ke skládání hudby lidskými skladateli. Jak uvidíme níže, tyto metody nefungují samy o sobě, ale je u nich zásadní zásah lidského skladatele – ať už v procesu přípravy hudebního materiálu, na který se určitá pravidla aplikují, výběru nejlepších možností, určení formálních základů nebo i následné úpravě „produkované“ hudby dle estetických či jiných stylově formálních pravidel.

1.1. Raná historie

Prvním z raných příspěvků k algoritmickému komponování je práce benediktínského mnicha Guida d'Arezza (též známého jako Guido Aretinus), významného hudebního pedagoga a teoretika, žijícího na přelomu 10. a 11. století, který je především znám pro svůj podíl na vývoji moderní notace a solmizace (Miller, 1973). Ve svém nejslavnějším díle, traktátu *Microgohus* (D'Arezzo, 1876),

se nicméně věnuje i metodě převodu jakéhokoliv textu na melodii, která byla původně určena k učení improvizace a skládání.

Metoda d'Arezzy je založená na tabulce, která přiřazuje extrahované samohlásky v jednotlivých slabikách daného textu k výškám tónů. Aby jednotlivý text neměl po převodu vždy jen jednu určenou melodii, je možné každou samohlásku spárovat se třemi tóny. Počet možností je tedy založený na délce textu, různých melodií může být konkrétně 3^n , kde se N rovná celkovému počtu slabik (Loy). Tímto výběrem možností skládání nekončí, d'Arezzo své studenty přiměl k aplikování dalších pravidel z hudební teorie na tuto „generovanou“ melodii, aby zněl výsledek muzikálně. Cílem této metody bylo spíše předat studentovi kreativitu a cit, než jej naučit formální pravidla (Miller, s. 242).

Jako další příklad algoritmického komponování můžeme považovat tzv. *izorytmické* skladby 14. století, jejichž princip, uplatňovaný především v izorytmickém motetu, popsal Philippe de Vitry ve svém opus magnum *Ars nova* (pravděpodobně z roku 1321), dalším významným představitelem tohoto stylu skládání je Guillaume de Machaut (Nierhaus). Skladatelský proces je založen na vytvoření dvou složek, respektive schémat. Těmi jsou soubor intervalů (*color*) a předpis rytmu (*talea*). Obě schémata po spojení tvoří danou skladbu, v jejím rámci se ale mohou vícekrát opakovat. Většinou mají tyto složky jinou délku, při jejich opakování tedy dochází k vzájemnému posunu a tudíž ke stále jiné melodii – do doby, než se cykly obou schémat znovu setkají (Cope).

Z tohoto pohledu se za částečně algoritmické skládání dá považovat i kánon, jehož principy se dají pozorovat ve skladbách už ve 13. století (Sadie, 1980). Kánon tvoří skladatelé tím, že se k hlavnímu hlasu se zpožděním postupně přidává více hlasů, které imitují hlavní melodickou linku. Z laického pohledu se tento přístup může zdát lehce napodobitelný, podle Loye (1989) je však pro skladatele obtížné současně dostát tomuto formalizujícímu přístupu a zároveň všem harmonickým pravidlům mezi jednotlivými hlasy.

1.2. Ars combinatoria

Od 17. století se hudební teoretici a skladatelé začali zabývat systematickým uspořádáním předem daného hudebního materiálu, který se označuje jako *Ars combinatoria* (Umění kombinatoriky). Tento způsob skládání je založen na kombinaci, variaci a permutaci jednotlivých prvků – např. not, rytmu, předem složených frází – do větších celků a dá se tedy aplikovat na harmonii, melodii, rytmus, strukturu frází či celé skladby (Ratner, 1970). Tímto způsobem je pak například možné složit z pár předem daných intervalů velký počet melodických frází.

„Všichni teoretici, kteří se zabývají permutací, od Mersenna po Galeazziho, to dělají z praktického důvodu – aby osvobodili představivost studenta. Metoda je to mechanická; hudební materiál je prostý a je ho málo; ale možností je nemyslitelné množství“ (Ratner, s. 350).

Podle Ratnera byla *Ars combinatoria* v hudební konstrukci dokonce modus operandi celého 18. století. Je tomu tak především díky stylové průzračnosti skladeb baroka a klasicismu, jejichž styl je formálně jasně definovatelný. Jako příklad mohou sloužit kodifikovaná pravidla kontrapunktu, respektive vedení hlasů ve vícehlasu, které souborně popsal Johann Joseph Fux (1725). Do této doby je Fuxovo dílo *Gradus ad Parnassum* nejvlivnějším pojednáním o striktním umění kontrapunktu, které ovlivnilo další generace hudebníků, jako byli například Johann Sebastian Bach, Wolfgang Amadeus Mozart, Joseph Haydn, Ludwig van Beethoven a Paul Hindemith (Mann, 1965).

Jedním z prvních teoretiků se začal explicitně zabývat kombinatorikou v hudbě Athanasius Kircher (1650). Ve svém díle *Musurgia Universalis* popisuje *Arca Musarithmica*, systém na automatické skládání čtyřhlasého kontrapunktu, který funguje na základě propojování vybraných kombinací předpřipravených melodických postupů a rytmů. Kircher se nicméně nedá pokládat za prvního, kdo s tímto přístupem přišel, jelikož se silně inspiroval v práci *Harmonie universelle* Marina Mersenna (1636), který podobným způsobem vypracoval tabulkový systém kombinování dvou souborů 720 hudebních frází a sady rytmů (Murata, 1999). Podobné metody navrhli například i Andrea Bontempi (1660) ve svém díle *New*

Method of Composing Four Voices, by means of which one thoroughly ignorant of the art of music can begin to compose nebo Wolfgang Printz (1696) ve *Phrynis Mytilenaeus oder der Satyrischer Componist* (Cope, 1996).

1.3. Musikalisches Würfelspiel

Prvním praktickým využitím kombinatoriky v nakládání s hudebním materiálem byla v 18. století oblíbená *Musikalisches Würfelspiel* (Hudební hra v kostky). Tato hra využívá prvku náhody ke generování hudby z předpřipravených možností, oproti předchozím metodám je tak skládání nezávislé na hodnocení a úpravě výstupu. Tímto způsobem je možné skládat menuety, valčíky, polonézy a další skladby ať už hodem kostkou, točením káčou nebo jen výběrem náhodných čísel (Hedges). Výsledky sice nejsou podle Ratnera (1970) geniálními hudebními díly, zábavu spojenou s těmito hrami označuje za až „dekadentní“, skladby jsou ale nicméně stylově platné především díky tomu, že jsou založené na kvalitních hudebních základech.

Za původní hru je považována „*Der allezeit fertige Menuetten- und Polonaisencomponist*“ (Vždy připravený skladatel menuetů a polonéz) Johanna Philippa Kirnbergera z roku 1757. Ta také silně ovlivnila své nástupce, kterých jen do roku 1812 vzniklo minimálně dalších 20 (Hedges, 1978). „*Každý, komu je známá kostka, čísla a dokáže zapsat noty, může složit tolik z výše uvedených menších skladeb, kolik touží,*“ píše v úvodu svého díla Kirnberger (1757, s. 3). Jeho hra je založena na fixní harmonické složce a melodické složce, kterou hráč generuje házením dvěma kostkami. Po jednom hodu vždy vybírá pro další takt jednu z jedenácti možných předpřipravených frází a tvoří tak postupně skladbu. Pro polonézu, která má čtrnáct taktů, tak existuje 11^{14} možných variant, pro menuet a trio dokonce 11^{32} (Ratner).

Na tuto hru přímo navázal neznámý autor v *Ludus melothedicus* (1758), který za pomoci devítistranné káči nechal podobným způsobem komponovat dvojhlasné menuety a tria o osmi taktech. Oproti Kirnbergerovi ale nekombinoval předpřipravené takty, ale jednotlivé noty (Hedges). Podobnou metodu využívá i Carl Philipp Emanuel Bach (1754 - 1778) v díle *Einfall einen*

doppelten Contrapunct in der Octave von sechs Tacten zu machen ohne die Regeln davon zu wissen, výsledkem hry je pak skladba o šesti taktech ve dvouhlasém kontrapunktu, kterou je možné hrát i vertikálně převrácenou (Hedges).

Musikalisches Würfelspiel

Table of Measure Numbers

Part One									Part Two								
	I	II	III	IV	V	VI	VII	VIII		I	II	III	IV	V	VI	VII	VIII
2	96	22	141	41	105	122	11	30	2	70	121	26	9	112	49	109	14
3	32	6	128	63	146	46	134	81	3	117	39	126	56	174	18	116	83
4	69	95	158	13	153	55	110	24	4	66	139	15	132	73	58	145	79
5	40	17	113	85	161	2	159	100	5	90	176	7	34	67	160	52	170
6	148	74	163	45	80	97	36	107	6	25	143	64	125	76	136	1	93
7	104	157	27	167	154	68	118	91	7	138	71	150	29	101	162	23	151
8	152	60	171	53	99	133	21	127	8	16	155	57	175	43	168	89	172
9	119	84	114	50	140	86	169	94	9	120	88	48	166	51	115	72	111
10	98	142	42	156	75	129	62	123	10	65	77	19	82	137	38	149	8
11	3	87	165	61	135	47	147	33	11	102	4	31	164	144	59	173	78
12	54	130	10	103	28	37	106	5	12	35	20	108	92	12	124	44	131

Table of Measures



Obrázek 1: Příklad "Mozartovy" hudební hry v kostky (Bača a Oberholtzer)

Další metodu výběru hudebního materiálu, náhodnou volbu čísla od 8 do 48, využívá Piere Hoegi ve své hře *A tabular system whereby the art of composing minuets is made so easy* (Hedges). Další slavné hry jsou pak přisuzované Wolfgangovi Amadeovi Mozartovi (*Anleitung zum Componieren von Walzern so viele man will vermittelst zweier Wurfel*) a Josephu Haydnovi (*Gioco Filarmonico, 1793*) (Ratner), jejich autorství však není potvrzeno a zůstává tak sporné (Nierhaus, Cope, Hedges).

1.4. Algoritmický skladatel

V hudbě 20. století se setkáváme s novými technikami skládání, které znamenají razantní zvýšení podílu racionality, především pak oproti subjektivistickému přístupu romantického skladatele (Leeuw, 2005). Prvky algoritmického skládání můžeme najít především v minimalismu, seriální, aleatorické a stochastické hudbě. Skladby těchto stylů se totiž vždy vyvíjejí jako proces podle předem daných pravidel skladatele.

Za počátek extrémně racionálního kompozičního myšlení se dá považovat *dodekafonie* Arnolda Schoenberga (Hons, 2010). Tato technika je založená na plném využití chromatiky a odproštění se od tonálních center předchozích staletí. Základem skladby je série dvanácti tónů (odtud *serialismus*), z nichž se standardně žádný neopakuje (Brindle, 1969). Na tuto sekvenci je pak možno aplikovat další možné postupy (jako například kvartová, kvintová obměna, račí obraty obou) a z těchto transformovaných sekvencí pak utvořit řady, které tvoří zdroj výšek tónů pro celou kompozici (Loy).

Racionalita pak „kulminuje“ zejména ve skladbách tzv. *totálních serialistů*, jako byli v 50. a 60. letech například Pierre Boulez, Karlheinz Stockhausen a další, kteří neaplikují metody serialismu jen na výšku tónů, ale i na rytmus, dynamiku, témbro apod. (Simoni, 2003). Z komponování se tak spíše stává „organizace“ všech parametrů skladby (Hons, s. 264). Georgi Ligeti (1958) například v analýze Boulezova prvního klavírního cyklu píše o procesu „zautomatizování“ strukturních vztahů. Boulez (1993), který se následně této techniky vzdal, se o serialismu vyjádřil následovně: „*Pro mě a mou generaci byl velice nápomocný, protože nám dal velmi striktní disciplínu (...) Je to vlastně úplně stejné jako psaní striktního kontrapunktu; pomáhá vám to najít řešení tam, kde si myslíte, že už žádné není.*“

Oproti do jisté míry přesně determinovaným skladbám totálního serialismu stojí aleatorická, na neurčitosti založená hudba, jejíž daná složka je záměrně přenechána náhodě – ať už při skládání či samotném provedení (Sadie, s. 341). Neurčitosti při provedení skladeb se věnovali například Karlheinz Stockhausen v díle *Klavierstück IX* (1956) a Pierre Boulez v *Klavírní sonátě č. 3* (1956 - 7). Obě skladby jsou založené na tom, že nechávají na interpretovi výběr a pořadí provedení předem

zkomponovaných částí. Nejznámějším představitelem aleatorických postupů při kompozičním procesu, ze kterého většinou vzejde již neměnná skladba, je John Cage. Ten například užil házení mincí společně s využitím Čínské knihy proměn k určení atributů skladby, jako jsou výšky a délky tónů, dynamiky a další ve svých dílech *Music of Changes* (1951) a *Music for Piano* (1952 – 6), kde tento proces obohatil ještě o zapisování not na nedokonalosti notového papíru (Cage, 1973).

Z pohledu algoritmického skládání je mimořádnou osobností hudby 20. století experimentální skladatel Iannis Xenakis, který byl nejen průkopníkem tzv. stochastické hudby (Xenakis, 1966), vycházející z aleatorické, ale který jako první začal ve skladatelském procesu využívat i počítač.

Xenakis vidí jakoukoliv skladbu jako konstrukci, která je shromážděním kauzálních vztahů a posloupností, obecnou kauzalitu tak podle něj nemůže přinést deterministický serialismus, ale teorie pravděpodobnosti aplikovaná na hudbu. Podle Xenakise (1992, s. 4) je stochastická skladba „*asymptotickou evolucí směrem ke stabilnímu stavu, určitému cíli*“, tedy procesem změny podobnému Markovovému procesu (Sadie). Xenakis (1992) ve svých dílech aplikoval různé procesy z matematiky a fyziky, například Poissonovo rozdělení v *Achorripsis* (1957), Teorii her ve skladbách *Duel* (1958 - 1959) a *Stratégie* (1962), Markovovy řetězce u *Analogique A* a *Analogique B* (1958 - 1959) či Gaussovu funkci a Brownův pohyb v orchestrální skladbě *Pithoprakta* (1955 - 1956).

V této době již začaly vznikat první programy, jejichž autoři experimentovali s automatizací procesu skládání. Svůj vlastní program fungující na počítači IBM 7090 vyvinul i Xenakis v roce 1962, s pomocí kterého složil skladby *ST/10-1,080262*, *ST/48-1,240162* a *Morsima-Amorsima (ST/4-1,030762)* (Ames, 1987). Od 50. let se začínáme setkávat nejen s vývojem programů asistujících skladatelům, jako tomu bylo v případě Xenakise, ale v historii algoritmického skládání také s prvními pokusy o úplné nahrazení osoby skladatele.

2. Vynalezené programy a formální metody kompozice

Jako první příklad kompozice hudební skladby s pomocí počítače se nejčastěji uvádí *Illiad Suite* pro smyčcové kvarteto z roku 1956 (Nierhaus, Fernandéz a Vico, 2013), k čemuž použili jeho autoři Lejaren Hiller společně s Leonardem Isaacsonem počítač Illiac z University of Illinois (Hiller a Isaacson, 1959). Skladba byla generována pomocí systému založeném na pravidlech v kombinaci s Markovovými řetězci (Fernandéz a Vico, s. 517). Hiller a Isaacson nicméně nebyli první, kdo ke kompozici použil počítač. Sami zmiňují dva experimenty, respektive programy Davida Caplina a Dietricha Prinze z roku 1955, které ale nejsou formálně zdokumentovány. Jejich nejranější program implementoval kombinaci hudebních prvků z „Mozartovy“ *Musikalisches Würfelspiel* s generátorem náhodných čísel z počítače Ferranti Mark I (Ariza, 2011). Na rozdíl od Xenakise či Caplina a Prinze však Hiller a Isaacson s počítačem nepracovali jen jako s pomůckou, ale jako plnohodnotným pokusem o simulaci práce skladatele (Loy).

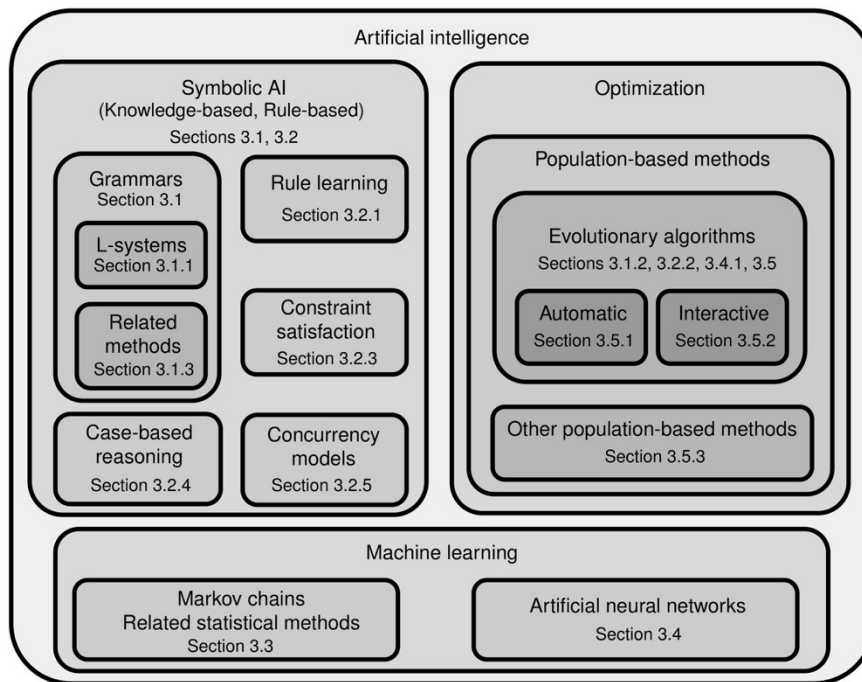


The image shows a musical score for "I. EXPERIMENT NO. I" by L.A. Hiller, Jr. and L.M. Isaacson. The score is for Violin I, Violin II, Viola, and Cello. It is marked "PRESTO" and includes dynamic markings like "p" and "f". The score is divided into two systems, with the second system labeled "(A)".

Obrázek 2: První část Illiac Suite (Povilioniene, 2017)

Od konce 50. let minulého století pak vzniká řada programů určených pro skládání hudby, které využívají celou paletu přístupů umělé inteligence – od symbolické UI, Markovových řetězců, evolučních algoritmů až po neuronové sítě. Popisu těchto technik a programů, které je využívají, respektive jejich limitům, bude věnována velká část této kapitoly, s přihlédnutím na taxonomii systematizující algoritmické skládání, kterou navrhl Nierhaus (2009) a doplnili Fernandéz a Vico (2013).

Předtím je však potřeba popsat navržené programy z hlediska jejich účelu a definovat tak ty, které jsou pro tuto práci relevantní.



Obrázek 3: Taxonomie použitých metod Fernandéze a Vica (2013)

Hlavním záměrem je zhodnocení ohraničených, plnohodnotných skladeb, ne všechna hudba, kterou počítače můžou generovat, je tedy pro tuto práci podstatná. To jest programy, které se zabývají produkcí improvizované hudby, potažmo určené k živým, interaktivním vystoupením či obecně zvukové syntéze, pro nás nejsou významné. Zároveň je potřeba vyloučit většinu generátorů melodií. Jak trefně podotýkají Papadopoulos a Wiggins (1999, s. 4): nemá v zásadě smysl hodnotit melodii jako hudební výstup, pokud k němu nemáme harmonický kontext, jelikož většinu melodií lze nakonec obhájit příslušným doprovodem. Výjimkou může být monofonní hudba, která je jasně stylově ohraničená (jako například gregoriánský chorál) nebo koncerty pro sólové nástroje. Vyprodukovat samotnou melodii je navíc ve většině programů nesrovnatelně jednodušší, než uvažovat i harmonickou složku.

Dalším důležitým atributem zvažovaných programů je míra interakce mezi člověkem-skladatelem a programem (Rankin, 2012). Do jedné skupiny patří programy, tzv. „push button“, u kterých maximálně zadáváme požadovanou délku,

tóninu a jiné základní prvky skladby. Jinak je proces plně automatizovaný a my dostáváme přímo výstup. Na druhé straně stojí programy více či méně interaktivní. Do této skupiny patří programy, které například využívají interaktivních formálních technik, jako jsou interaktivní generativní algoritmy, kde člověk neustále vybírá nejvhodnější hudební struktury a poskytuje tak programu zpětnou vazbu. Největší míru svobody pak získává člověk u programů, které jsou zamýšlené jako spotřební software. Ty slouží jako pomůcka při skládání, respektive automatizují část skladatelského procesu. Těmto programům se ale, až na výjimečné případy, věnovat nebudeme. To je také důvod, proč nebude popsána celá škála možných technik, které jsou ke kompozici hudby v oblasti umělé inteligence používány.

2.1. Symbolická umělá inteligence

Do kategorie symbolické umělé inteligence, popřípadě „Staré dobré umělé inteligence“, jak ji nazývá John Haugeland (1985, s. 112), patří především přístupy založené na generativní gramatice, expertních znalostech a pravidlech, která vycházejí z hudební teorie. Určení těchto pravidel je možné dvěma způsoby: přenesením z hudební teorie, což vyžaduje vysoké expertní znalosti, anebo odvozením z většího množství již existujících skladeb (Fernández a Vico). Přístupy spojené se symbolickou umělou inteligencí v tomto smyslu představují hudební expertní systémy.

2.1.1. Generativní gramatika

Generativní gramatika, nebo formální gramatika, je postavená na práci se syntaxí umělého jazyka, tedy jeho formální strukturou a jejími pravidly. Hudba je v tomto kontextu vnímána, stejně jako jazyk, jako na sobě do jisté míry závislé a provázané symboly. Pokud jsou tedy využita vhodná hudební syntaktická pravidla, je díky nim možné generovat nové a formálně správné „jazykové“ konstrukce, tzn. skladby (Nierhaus).

Ranným využitím tohoto přístupu byla práce Garyho M. Radera (1974), který pro generování hudby kombinoval pravidla hudební teorie a heuristické řešení problémů. Rader touto metodou použil jen ke skládání melodií, stejně jaké mnozí další autoři (viz například Cruz-Alcázar & Vidal-Ruiz, 1998; Hamanaka a kol.,

2008 a další). Komplexní skladby vytvořili například S. R. Holtzman (1981), který reprodukoval Schoenbergův přístup ve skladbě pro harfu, Mariateresa Storino a kol. (2007) v programu LEGRE skládající hudbu ve stylu Giovanniho Legrenziho nebo Donya Quick (2010 a 2014), která ve svém programu Kulitta použila schenkerovskou redukční analýzu (více viz Spurný, 2012), metodu pro určování obecné harmonické struktury, k vytváření kontrapunktické hudby z období baroka (fugy a chorály).



Obrázek 4: Příklad sonáty ve stylu Sergeje Prokofjeva, EMI (Cope, 1996)

Patrně nejznámější program na skládání hudby, který vytvořil skladatel David Cope (1996, 2000, 2005), pracuje s několika metodami, zacházející s hudbou podobným způsobem jako s jazykem. Na těchto metod základě vytvořil Cope program *Experiments in Musical Intelligence* či *EMI* (1992), který je schopný napodobovat styl různých skladatelů, jako jsou například Johann Sebastian Bach, Fryderyk Chopin nebo Béla Bartók. EMI (a pozdější verze SARA a ALICE) pracuje s množstvím skladeb daného autora nebo hudebního stylu a k jejich imitaci užívá rekombinace; nejdříve rozkládá každou jednotlivou skladbu na bázi latentní sémantické analýzy, identifikuje opakující se struktury, signatury, a poté používá tyto struktury při reprodukci. Aby byly reprodukovány ve smysluplném významu, EMI užívá *Augmented Transition Network*, konečný automat do hloubky analyzující jazyk (Woods, 1970). V současné době byla nejnovější verze programu

překřtěna na Emmy (či Emily Howell), která skládá hudbu ve vlastním stylu. Cope do programu EMI vložil totiž jen skladby vyprodukované v předchozích letech samotným programem, aby zjistil, jak bude znít její ryzí styl (Cope, 2005).

Specifický typ formální gramatiky, původně vytvořeným k umělému napodobení růstu květin, představuje tzv. Lindenmayerův systém, zkracován také jako L-system. Ten vytváří soběpodobné struktury, na základě setu pravidel opakovaně „překládá“ prvotní znak a z něj vycházející exponenciálně rostoucí rekurzivní vzorce (Nierhaus). Kromě generování melodií jej využil především Nigel Morgan (2007) ke složení plnohodnotné sedmivěté skladby Heartstone pro dechový orchestr, perkuse a klavír.

2.1.2. Znalostní systémy založené na pravidlech

Znalostní systémy založené na pravidlech (Rule-Based Systems) a omezeních vycházejí ve velké většině případů z hudební teorie a představují tak přirozený způsob, jak automatizovat kompoziční proces. I proto je u dále zmíněných příkladů časté, že se zabývají jasně definovanými styly a technikami, jako je kontrapunkt a serialismus (Fernandéz a Vico).

Příkladem je raný program Stanleyho Gilla (1963), který využívá Schoenbergova pravidla dodekafonie ke kompozici díla pro housle, violu a fagot. Tento program funguje na bázi heuristického přístupu - generuje „náhodné“ struktury, které následují co největší možné množství pravidel a poté s využitím zpětného sledování eliminuje zbývající chybně určené a následně vybírá ty nejsprávnější. Další programy využívají obecně uznávaná pravidla k harmonizování čtyřhlasu (Rothgeb, 1968; Thomas, 1985; Steels, 1986), pro komponování skladeb na základě klasického kontrapunktu (Gjerdingen, 1988; Schottstaedt, 1989), menuetů raného klasicismu (Löthe, 1999) nebo hudby ve stylu Witolda Lutosławského (Krupowicz, 2015). Do této skupiny systémů patří také ty, které se pravidla z části učí nebo je časem mění (Schwanauer, 1993; Spangler, 1999; Morales & Morales, 1995).

Jedním z přístupů, který začal postupně v 80. a 90. letech převládat, je definování algoritmického komponování jako problému formální logiky - splnitelnosti omezujících podmínek (tzv. constraint satisfaction problem – CSP), který je založen

na heuristice a kombinatorickém vyhledávání možností (Fernandéz a Vico), jehož hlavní myšlenkou je radikální omezení vyhledávacího prostoru „identifikací kombinací proměnných/hodnot, které porušují podmínky“ (Russell a Norvig, 2016, s. 202). Jedním z prvních, kdo tento přístup využil při generování striktního dvojhlasého kontrapunktu a harmonizaci byl Kemal Ebcioğlu (1980; 1988) se svým programem CHORAL, kterého následovali další autoři věnující se harmonizaci (Tsang a Aitken, 1991; Pachet a Roy, 1995; Phon-Amnuaisuk, 2002; Yilmaz a Telatar, 2010; Evans a kol., 2014 a další), Schoenbergově technice (Wiggins, 1998; Anders a Miranda, 2009) nebo skladbám ve stylu György Ligetiho (Chemillier a Truchet, 2001).

Další metodou může být také případové usuzování (case-based reasoning), které je založeno na tom, že systém nehledá nové, unikátní řešení, ale odvozuje jej z předešlých případů a jejich úspěšných řešení. Tento přístup byl použit například ke generování barokní hudby (Pereira a kol., 1997; Ribeiro a kol., 2001) a harmonizaci (Sabater a kol., 1998).

2.2. Strojové učení: Markovovy řetězce a neuronové sítě

Markovovy řetězce jsou znázorněním typu stochastického procesu s konečným stavem, při kterém pravděpodobnost určité události závisí v čisté formě jen na současném stavu (Gagniuc, 2017). V kontextu hudebních programů je za tento stav, resp. základní jednotku, považována většinou jedna nota. Častější je nicméně využívání Markovových modelů n -tého řádu, u kterých jde závislost budoucího stavu do minulosti n -stavů (např. Triviño-Rodríguez a Morales-Bueno, 2001; Collins a kol., 2016). Pravidla pro určení dalšího stavu jsou nejčastěji odvozena buď přímo z hudební teorie nebo je algoritmus vytrénován na již existující hudbě (Fernandéz a Vico).

Markovovy řetězce využil například Dan Ponsford a další (1999) při generování skladeb ve stylu saraband, dvorního tance 17. století, Tsubasa Tanaka a kol. (2010) pro imitaci kontrapunktu nebo Marcus T. Pearce a Geraint Wiggins (2007) ke generování melodie ve stylu luteránských hymnů. Jiný druh Markovových řetězců, tzv. skrytých (se skrytými stavy), využili jako první Mary Farbood a Bernd Schoner

(2001), kteří tuto techniku implementovali do programu skládající hudbu ve stylu Giovanniho Pierluigihho da Palestriny s využitím kontrapunktu 1:1. Skryté Markovovy řetězce byly použity také při harmonizaci ve stylu J. S. Bacha (Biyikoglu, 2003; Allan, 2002; Allan a Williams, 2004) a dalších skladatelů (Schulze, 2009; Yanchenko a Mukherjee, 2017).

Jednou z nejoblíbenějších metod ve sféře strojového učení jsou kromě Markovového modelu především umělé neuronové sítě, které vycházejí ze svého předobrazu v biologických procesech (Russel a Norvig, s. 727 - 728). Tvoří je série vzájemně propojených umělých neuronů, které si předávají informace a na základě přenosových funkcí je zpracovávají. Díky trénování na vytvořené hudbě se mohou učit, u skládání hudby se jedná především o učení s učitelem (supervised learning). Tyto metody jsou tak vhodné především pro imitaci a analýzu hudebních stylů (Fernandéz a Vico).

Neuronové sítě byly použity pro harmonizaci (Bellgard a Tsang, 1992), k modelování kontrapunktu (Adiloglu a Alpaslan, 2007) nebo skládání celých skladeb (Kalegasi, 2017). Hybridní přístup použil Hermann Hild a kol. (1992) ve svém programu HARMONET, kteří ke čtyřhlasé harmonizaci ve stylu chorálů J. S. Bacha použili dvě neuronové sítě v kombinaci se systémem založeným na pravidlech. Z tohoto programu také vycházeli další autoři, jako Johannes Feulner a Dominik Hörnel (1994) v programu MELONET, který generoval chorály ve stylu Johanna Pachelbela. Z HARMONETu vycházeli také Dominik Hörnel a Thomas Ragg (1996), kteří využili neuronové sítě ve spojení s evolučními algoritmy.

V současné době jsou také využívány další druhy neuronových sítí, například konvoluční (Huang a kol., 2017) a rekurentní neuronové sítě (Boulanger-Lewandowski a kol., 2012; Gaëtan Hadjeres a kol., 2017), respektive jeho typy využívající tzv. Long Short Term Memory (LSTM), díky kterému se systém učí dlouhodobé závislosti a v důsledku tak i celkovou hudební strukturu (viz Lyu a kol., 2015; Huang a Wu, 2016; Jaques, 2016; Liang, 2016; Johnson, 2017; Kotecha, 2018), nebo podobný, jednodušší model Gated recurrent units (viz Chung a kol., 2014; Staněk, 2017).

2.3. Evoluční metody

Další z metod, které se inspiřují biologickým světem, jsou tzv. evoluční algoritmy. Ty jsou založené na generaci řešení, respektive populace organismů, která se postupně proměňuje díky neustálému generování, evaluaci a variaci. V darwinistickém pojetí jde vlastně o přežití nejsilnějšího, tedy nejvhodnějšího řešení podle tzv. hodnotící funkce (fitness function), která určuje, které organismy z dané generace přežívají nebo se dál množí a které spolu se svým genomem vymírají. Existují obecně dva hlavní přístupy k těmto funkcím: buď může být automatická nebo interaktivní (Russell a Norvig, s. 127). Třetí typ funkce, využívající metod strojového učení, se zatím u generování plnohodnotných skladeb nevyskytuje, nicméně byl použit u skládání melodie (viz Bellinger, 2011).

Automatická hodnotící funkce může být vyjádřena jako rozdíl mezi ideálním a produkovaným výstupem, nebo jako pravidla a omezení z hudební teorie, respektive suma vážených (žádaných či naopak nepřipustných) atributů, jako mohou být například počet not nebo akordů v určité tónině nebo tzv. paralelní kvinty a oktávy. Právě tento přístup je u automatické hodnotící funkce využit nejčastěji, například při melodické harmonizaci (McIntyre, 1994; Horner a Ayers, 1995; Phon-Amnuaisuk a kol., 1999; Maddox a Otten, 2000; De Prisco a kol., 2010), komponování podle pravidel kontrapunktu (Polito a kol., 1997; Gwee, 2002; Garay Acevedo, 2004; Donnelly a Sheppard, 2011) anebo soudobé klasické hudby (Dahlstedt, 2007; Browne a Fox, 2009).

Interaktivní generativní algoritmus je založen na ohodnocující funkci, kterou představuje sám uživatel (nebo nezávislí posluchači). Ten hodnotí každou generovanou populaci poslechem a dle svých estetických měřítek určuje, jak bude vypadat generace následující. Různé typy posluchačů od laiků po experty využil pro generování Jeffrey Power Jacobs a James A. Reggia (2012) pro evoluci barokní hudby podle pravidel kontrapunktu, naopak program Vox Populi Artemis Moroni a kol. (2000) ovládá přes grafické rozhraní jen jeden uživatel, který přímo určuje podobu dalších generací.

Kombinace automatického a interaktivního přístupu se dá nalézt v práci skupiny Melomics z Univerzity v Malaze, kteří vyvinuli počítač Iamus, který je schopen

samostatně skládat soudobou klasickou hudbu. Iamus má v sobě jakýsi genom, který uchovává informace ohledně melodie, polyfonie a instrumentace. Genom funguje jako „semínko“, jehož mutováním se samotná skladba postupně vyvíjí. Nové skladby jsou tak generovány přímo z dříve složených kompozic. V jeho hodnotící funkci bylo postupně, poslechem profesionálních hudebníků a skladatelů, zohledněno téměř tisíc pravidel, které se mají postarat o estetickou hodnotu a formální správnost jeho skladeb (Sánchez-Quintana a kol., 2013). Jeho první skladba složená bez lidského zásahu byla *Opus one* v roce 2010, v roce 2012 pak natočil London Symphony Orchestra album jeho skladeb „*Transits – Into an Abyss*“ (Diaz-Jerez, 2011).



Obrázek 5: Příklad skladby *Colossus*, Iamus (Sánchez-Quintana a kol.)

2.4. Limity

I přesto, že se v některých případech mohou prolínat, existují dva odlišné přístupy v kontextu tvorby skladby počítačem: imitace určitého historického stylu či konkrétních metod daného skladatele, anebo tvorba „originální“ skladby, respektive novátorského skladatelského procesu – do té míry, v jaké může program tento cíl splnit. Stylová imitace je silně převládající metoda, což je zapříčiněno do jisté míry danými a kodifikovanými zvyklostmi a pravidly určitého stylu, které se dají přenést do prostředí počítačového softwaru.

Jak je vidět výše, vícehlasý kontrapunkt a harmonizace jsou příklady nejreflektovanějších uzavřených problematik. Za tím je především ten

důvod, že se jedná o přesně definovaný obor, jehož pravidla jsou silně formalizována např. již zmíněnými Johannem Josephem Fuxem nebo Heinrichem Schenkerem (Allan, 2002). Podle Nierhause (s. 263) je na druhé straně za nízkým počtem publikovaných programů směřujících k originální kompozici do jisté míry i to, že za těmito programy spíše stojí skladatelé, kteří je užívají ve vlastních dílech a buď není publikování primární součástí jejich práce, nebo své postupy nechtějí zveřejňovat.

Za největší výzvu počítačů se dá označit způsob prohledávání prostoru možných hudebních skladeb. Již u melodické hudební fráze bez rytmu v jednom dvanáctitónovém systému (oktávě) o délce deseti tónů totiž dostáváme obrovský vyhledávací prostor o velikosti 12^{10} možných řešení (Garay Acevedo, 2004), je tak potřeba tento prostor (nebo využitý algoritmus) vhodným způsobem omezit. Formálně správné struktury však musejí být také generovány takovým způsobem, aby měli hudební smysl (Steedman, 1984). I přes to, že se hudbě někdy přezdívá „univerzální jazyk“, je potřeba mít na paměti, že její obsah či význam je oproti přirozenému jazyku těžko definovatelný.

Jedním z nejpobulárnějších přístupů jsou metody symbolické umělé inteligence, které jsou sice poměrně efektivní, vyžadují nicméně časově náročnou implementaci velkého množství pravidel a omezení, respektive velkou znalostní bázi tvořenou expertem (Coats, 1988). Největší výzvou je pro metody založené na pravidlech tato omezení „*oslabit nebo modifikovat kvůli nejednoznačným a „fuzzy“ vlastnostem struktury (...) forem hudby*“ (Lewis, 1991, s. 214). Jak ale bylo znázorněno na příkladu EMI (Cope, 1996), je možné za pomoci formální gramatiky úspěšně pracovat s hudební syntaxí, především pak při imitaci stylů.

U Markovových řetězců se setkáváme hned s několika limity. U modelů nižších řádů pozorujeme velkou náhodnost generované hudby, vyšší řády zase často kopírují celé sekce originálních skladeb (Nierhaus). Podle Kaana M. Biyikoglu (2003) umí Markovovy řetězce generovat kvalitní přechody mezi akordy a dají se tak dobře využít při harmonizaci melodie.

Stejně jako u dalších metod strojového učení, neuronových sítí (ale i evolučních metod), je ale jejich omezením slabá schopnost řešit globální hierarchické vztahy (Mozer, 1994). To v zásadě znamená, že tyto metody umí dobře zachytit lokální strukturu, v kontextu více taktů už však hudba generovaná s jejich pomocí nikam nevede (Lewis). Oproti Markovovým řetězcům a znalostním systémům, jejichž výstup je do jisté míry predikovatelný, neuronové sítě umí nicméně generovat překvapivá řešení problémů (Nierhaus).

U evolučních metod je zásadní otázka hodnotící funkce. Automatická funkce je podle Nierhause (2009) velmi těžká na implementaci, až na jednoduchá pravidla hudební teorie. Oproti tomu interaktivní funkce, závislá jen na subjektivní evaluaci, je vyčerpávající a časově náročná pro samotného uživatele či posluchače. Každou generaci řešení totiž musí uživatel posoudit poslechem, což jej po několika cyklech může jednoduše přestat bavit (Ariza). Phon a Wiggins (1999) poukazují i na to, že posluchači jsou náchylní pozitivněji vnímat hudbu, kterou již několikrát slyšeli. Tento problém se dá částečně vyřešit menší populací založené na kvalitní první generaci, která není definována náhodně, ale vychází z hudebních pravidel (jako využívá např. Iamus).

Evoluční metody mohou v omezené míře u přesně specifikovaných problémů přinášet dobré výsledky, které někdy dokonce předčí i studenty skladby, u imitace stylu nebo automatizace kodifikovaných skladatelských postupů ale dosahuje i jednoduchý znalostní systém kvalitnějších výsledků (Phon-Amnuaisuk, 1999). Obecně se také tyto algoritmy potýkají s problémem způsobu výběru další populace - pokud je výběr moc elitářský, program se často spokojí jen s lokálním optimem (Phon), naopak moc široký výběr vede k reprodukování nekvalitních řešení (Burton a Vladimirova, 1999).

Výše zmíněné programy se podle Pearce a kol. (2002) potýkají ve velké míře také s metodologickými problémy. Odborné články a publikace, které se jimi zabývají, podle nich selhávají v definování motivace, respektive praktických či teoretických cílů výzkumu a způsobů, jakými jich dosáhnout. Navíc ani neumí jasně vyhodnotit, jestli těchto cílů bylo dosaženo a často se tak jen uchylují k lapidárním popisům

programů, které komponují hudbu. Z nedostatečného popisu samotného procesu tak pramení nejasnost spojená s rolí a podílu autora programu na hudebním výstupu.

3. Hudební Turingův test

Hlavní koncept, který je v souvislosti s evaluací generativních programů citován a ze kterého čerpají mnohé způsoby evaluace, je tzv. *Turingův test*, vycházející z imitační hry Alana Turinga (1950). Ta vznikla jako způsob odpovědi na otázku, jestli umí stroje, respektive digitální počítače, myslet. Turingův test je založen na textové komunikaci mezi lidským tazatelem a dvěma entitami: počítačem a člověkem. Na základě rozhovoru skrz textové rozhraní má tazatel rozhodnout, který z nich je člověk a který počítač. Pokud tazatel nedokáže obě entity rozlišit tak, jako by ve stejné hře nedokázal odlišit muže a ženu, jedná se podle Turinga o úspěch počítače. Původní otázka, jestli může počítač myslet, je tak nahrazena otázkou, jestli může existovat digitální počítač, který uspěje v imitační hře.

Koncept Turingova testu se neobejde bez kritiky, která čerpá i z toho, že Turing samotný test nepopsal nijak zvlášť detailně. Neuvedl celkový počet tazatelů a entit, jestli jsou zapojení experti v jakékoliv oblasti, celkovou organizaci, délku testu a ani počet jeho opakování (Halpern, 2006). Ve svém textu jen nastínil odhad, že kolem roku 2000 nebude průměrný pozorovatel schopen po pěti minutách ve více než 70 % případů rozeznat rozdíl mezi počítačem a člověkem. Turing také vědomě abstrahuje od teorie her, tedy možných strategií počítače a člověka směřujících ke splnění testu.

Sám Turing popisuje různé námitky, které by se vůči tomuto testu daly vznést. Jednou z nejzajímavějších je argument zformulovaný Geoffreyem Jeffersonem (1949), který tvrdí, že stroj nemůže mít nikdy vědomí:

„Dokud nebude stroj schopen napsat sonet nebo složit koncert na základě svých myšlenek a emocí, nikoli náhodným výběrem symbolů, nemůžeme přistoupit na tvrzení, že stroj rovná se mozek. To znamená, že stroji nestačí sonet jen napsat, ale musí i vědět, že jej napsal. Žádný mechanismus nemůže cítit (nejen uměle dávat najevo, což je snadné) uspokojení z úspěchu nebo smutek ze spálené elektronky, nemůže být potěšen lichotkami, zarmoucen svými chybami, okouzlen opačným pohlavím, rozzloben nebo deprimován, když není schopen dosáhnout toho, co chce“ (Tvrdý, 2011, s. 35).

Turing tento argument odmítá s tím, že v důsledku reprezentuje extrémní solipsismus. Zjistit, jestli počítač vyprodukoval něco s určitým záměrem, je totiž prakticky nemožné. Člověk by se musel stát tímto strojem, aby zjistil, jestli má vědomí. U lidí je (alespoň od určitého věku) samozřejmě konvencí o ostatních členech *Homo sapiens* předpokládat, že mají vědomí. U počítačů to zřejmě neplatí, právě proto je ale imitační hra vhodná jako kritérium. Entity, které se tedy *jeví* jako inteligentní, inteligentními jsou (French, 1990).

Podle Raye Kurzweila (2002) je hlavní myšlenkou Turingova testu to, že lidský jazyk v psané podobě umí v plném rozsahu reprezentovat lidské myšlení. Je tedy vůbec hudba vhodným médiem pro tento test? Pokud by měl Turingův test přímo testovat přítomnost myšlení, pak pravděpodobně ne, ale jak Kurzweil dodává, v testu jde spíše než o vědomí o výzkum výkonu v určité oblasti lidské činnosti.

Mark Halpern (2006) oproti tomu poukazuje na to, že schopnost systému dát správné odpovědi na lidské otázky přítomnost myšlení či vědomí nedokazuje. To zakládá na známém *Argumentu čínského pokoje* Johna R. Searleho (1980), který tvrdí, že schopnost vést konverzaci neznamena v plné míře rozumět. Searle předkládá příklad člověka, který neumí čínsky, jsou mu ale dány dostatečně kvalitní instrukce, jak mechanicky zacházet s formálními znaky. Pak se tento člověk může jevit jako někdo, kdo umí čínsky, i když má jen znalosti o formě, a ne obsahu. Tento argument rozporuje Kurzweil (2005) tvrzením, že Searle uvažuje systém, jehož určitá část sice čínsky nerozumí, jako celek ale ano. Člověk v tomto systému vlastně zastupuje procesor: „*Anglicky rozumím, žádný z mých neuronů ale nikoliv*“ (Kurzweil, 2005, s. 330).

Stevan Harnad (2000) naopak uznává argument čínského pokoje jako příklad entity, která by mohla být schopná originální Turingův test úspěšně překonat, zároveň ale nemít opravdovou inteligenci. Proto Harnad pro potřeby kognitivní vědy rozšířil Turingův test, který testuje nerozlišitelnost člověka od stroje nejen na verbální, ale i na fyzické, viditelné úrovni. Na základě této teorie vypracovává pět úrovní hierarchie Turingových testů:

- 1) "*Toy Model*" (t1) - zkoumající jednotlivé komponenty mysli nebo lidské činnosti
- 2) Turingův test (T2) - standardní Turingův test v textové verzi, chápán ne jako hra nebo test, který může být vyřešen, ale spíše jako jedno z kritérií
- 3) Totální Turingův test (T3) – fyzická a interakční nerozlišitelnost, ideální úroveň Turingova testu
- 4) Interní mikrofunkční nerozlišitelnost (T4) - všechny složky lidského těla, orgány, stejně jako vnitřní procesy, jsou stejné jako u robota
- 5) Mikrofyzická nerozlišitelnost (T5) - nerozlišitelnost je na takové úrovni, že molekuly nejsou syntetické, ale fyzicky identické s lidskými

Pokud bychom tento Harnadův Totální Turingův test převedli jen na hudbu, musel by před námi takovýto android například umět zahrát Bachovy Suity pro violoncello nebo v našem případě vlastníma rukama na notový papír nebo na počítači zkomponovat určitou skladbu. Již původní Turingův test se setkával s kritikou přílišného antropocentrismu, který jen použitím lidského jazyka a potřeby po lidských zkušenostech znevýhodňuje počítač (viz French, 1990), ale například i zvířata a mimozemské entity (Tvrdý). V případě skládání hudby se tak dá tento přístup rovnou odmítnout s tím, že je extrémní a o procesu samotného skládání nám stejně moc neřekne. Ani originální *Imitační hra* neměla za cíl zkoumat schopnost stroje imitovat řeč, v hudební verzi tak není potřeba hudbu samu o sobě *provádět* a můžeme se zaměřovat jen na skládání (Burnett a kol., 2012).

Z této hierarchie pro nás však vychází jeden důležitý poznatek, že totiž hudební Turingův test můžeme zařadit spíše na úroveň t1 zkoumající jednotlivé komponenty lidské činnosti, respektive mysli. Je to především proto, že hudba jako médium Turingova testu postrádá jednoznačně dekodovatelný obsah jako je tomu u jazyka a nemůže tak reprezentovat myšlení jako celek.

Kvůli rozsahu se zde nemůžeme zabývat všemi námitkami proti Turingově testu ani jeho obhajobě. Jak tvrdí Ariza (2009, s. 52), tato filozofická debata není klíčová, pokud Turingův test vnímáme spíše jako „*symbolické měřítko určité schopnosti*“, které navíc inspirovalo další možné formy evaluace hudebních programů.

3.1. Kreativita umělé inteligence a problém autorství

Můžeme u těchto programů, respektive strojů obecně, uvažovat o kreativitě, podle všeho výsostně lidské schopnosti, stejně jako se u nich mluví o inteligenci? Na začátku této debaty nelze přejít další z předestřených námitek Alana Turinga (1950), tzv. námitku lady Lovelace, vycházející z práce Augusty Ady King, hraběnky z Lovelace (viz Lovelace, 1842). Podle ní je stroj vždy čistě deterministický, tedy musí nutně udělat vždy jen to, co jsme mu sami řekli, aby udělal. Nemůže být tedy nikdy opravdu inteligentní či kreativní, dokud neudělá něco, k čemu nebyl naprogramován (Turing, s. 454).

Turing tento argument odmítá s tím, že můžeme teoreticky vytvořit stroj, který bude myslet za sebe a může se učit, tudíž vytvořit něco, co jsme do něj nevložili. Pokud totiž uvažujeme hypotetický stroj, který je schopen vytvořit něco opravdu originálního, je i tento stroj možné emulovat univerzálním počítačem. Podle Tvrdeho (2011) tato tautologie nepříliš dobře funguje, předkládá ale jiný argument, že totiž determinismu může v tomto pohledu podléhat i - zatím v nesporně komplikovanějším smyslu - člověk: „*Stejně jako jsou počítače determinovány svým programem a vstupními informacemi, jsou i lidé determinováni genetickou výbavou a sociokulturním prostředím, ve kterém vyrůstají a žijí*“ (Tvrdý, s. 41).

V případě evaluace takového systému či jeho výstupů je vypořádání se s touto námitkou jedním z klíčových momentů. Za jakýmkoliv „kreativním“ strojem je totiž v každém případě člověk či jeho znalost (Halpern, 2006), programátor vlastně vytváří kreativní dílo, které má samo něco vytvářet, jde tak o *metakreaci* (Whitelaw, 2004). Pokud na tuto myšlenku přistoupíme, je jakákoliv kreativita stroje kreativitou stvořitele systému či počítače. Podobně se k problému vyjadřuje Dreyfus (1979), který kreativitu přisuzuje jen programátorovi, u počítače je dle něj nemožná.

Problém autorství či kreativity se nevztahuje exkluzivně jen na počítače. Minimálně do 18. století nebyla v západní kultuře kreativita chápána jako lidská vlastnost, člověk byl v křesťanské tradici jen nástrojem boha jako jediného opravdového stvořitele, respektive prázdnou nádobou, jejímž zdrojem inspirace je vyšší moc (Runco a Albert, 2010). Koncept moderního individualistického autorství se podle

Rolanda Barthes (2006) a Michela Foucaulta (1979) zrodil až s příchodem romantismu (Bennett, 2005, s. 55). Je celkem známým faktem, že do 19. století si hudební skladatelé od Händela po Beethovena od sebe často „půjčovali“ hudbu a recyklovali celé melodie a témata (Sinnreich, 2019). Také ve východních kulturách u Hinduismu, Buddhismu, Konfucianismu či Taoismu je kreativní proces chápán spíše jako objevování či napodobování, než jako „tvorba z ničeho“ (Runco a Albert).

Jakákoliv debata o kreativitě výpočetního systému se neobejde bez definice kreativity vůbec, potažmo lidské kreativity. Konsenzus však nenajdeme ani zde (Jordanous, 2012). Obecná definice výpočetní kreativity poskytnutá řídícím výborem pro výzkum výpočetní kreativity (ICCC) zní následovně: „*Výpočetní kreativita je studium a simulace (...) chování (...), které, pokud by bylo pozorováno u lidí, by bylo považováno za kreativní* (Pease a Colton, 2011). Tato definice nám toho o samotné kreativitě neřekne mnoho, ale důležitá (a ne zcela často explicitně formulovaná) podmínka, že by měla kreativita počítačů v každém případě vycházet z kreativity, jak je chápána v kontextu lidského chování, je centrální téměř v každé debatě o výpočetní kreativitě (Ritchie, 2007).

Jako jeden z prvních příspěvků k počítačové kreativitě je článek *The digital computer as a creative medium* Michaela A. Nolla (1967). Ten tvrdí, že nazvat deterministický počítač kreativním je možná až příliš silné prohlášení. Díky počítači je možné vytvořit zcela nová díla a počítač se tak může stát *kreativním médiem*, ale jen v interakci s umělcem, který stále zůstává hlavním tvořitelem.

Významnou postavou této debaty je Margaret Bodenová (2004), podle které si můžeme pokládat čtyři otázky v kontextu vztahu počítačů a kreativity:

- 1) Mohou nám počítače pomoci pochopit, jak funguje lidská kreativita?
- 2) Mohou počítače dělat věci, které se jeví kreativně?
- 3) Mohou počítače někdy rozpoznat kreativitu?
- 4) Mohou být počítače někdy opravdu kreativní? (Boden, s. 16 – 17)

Bodenovou zajímá hlavně první otázka, na kterou odpovídá jasné ano. Aby bylo možné tuto lidskou kreativitu skrz počítač pochopit, zabývá se i dalšíma dvěma, které také s určitými podmínkami zodpovídá kladně. Dohromady tvoří vědecké, zkoumatelné otázky, na rozdíl od poslední, která je jen otázkou filozofickou a není pro Bodenovou důležitá. Spokojuje se s tím, že počítače dokáží produkovat kreativní výstupy a odhalit, co to kreativita vůbec je. Podle Bodenové se dá dělit na psychologickou a historickou. První typ kreativity se vztahuje na artefakty, které jsou novátorské ve vztahu k jednotlivci, druhá se vztahuje k artefaktům lidstvu do té doby neznámým. Historická je zároveň podmnožinou psychologické (Pearce a Wiggins, 2007).

Pro kreativitu strojů je důležité především její dělení kreativity na kombinační, průzkumnou a transformační (Boden, 2004). Kombinační kreativita produkuje neznámé kombinace známých idejí, průzkumná objevuje nové ideje v rámci kulturně daného prostoru myšlení (či konceptuálního prostoru), transformační tento daný rámec transformuje novými pravidly a jako taková je nejvyšší formou kreativity. Každá z nich je teoreticky dosažitelná umělou inteligencí, v kontextu počítačových programů v současnosti vyjma transformační, někdy ale s výbornými výsledky, jak zmiňuje Bodenová například EMI Davida Copea u průzkumné kreativity. Pro Bodenovou je zároveň důležitý pojem hodnoty, daná myšlenka musí být tedy buď užitečná, podnětná či poučná. Podle Bundyho (1994) se však kreativita obecně nedá vysvětlit čistě jen v termínech výpočetních procesů počítače a tato definice Bodenové je tak neúplná, i když vědecky použitelná. Kreativita spíše odpovídá konkrétním aspektům mnoha různých druhů výpočetních procesů.

Je u strojů kreativita v pravém slova smyslu dosažitelná, pokud stroj jako takový nemá žádné vědomí, tužby, hodnoty ani preference, a nemůže tak ani sám přesvědčivě zhodnotit či ocenit to, co dělá? Wiggins (2006) k této debatě dodává, že Bodenová i Bundy spolu souhlasí, že čistý výpočet množiny artefaktů hrubou silou, ať už komplexní či ne, není kreativní. Pro takový systém je možné, aby vygeneroval artefakt, který je nepředvídatelný, je to však spojené se schopností rozpoznat hodnotný artefakt, což už je samo o sobě kreativní proces. Zároveň tvrdí, že počítače jsou schopny sami najít správný artefakt dle formální stránky díky

vlastním procesům, ale co je hodnotný artefakt, určují zatím jen lidé, z jednoho pohledu tak stroje opravdu kreativní nejsou.

Z tohoto pohledu není opravdu samostatný a kreativní žádný z uvedených programů na skládání hudby, protože posuzování jejich výstupů probíhá skrz jejich tvůrce, popřípadě formální pravidla. Dle Graemea Ritchieho (2007) je možné u těchto programů uvažovat o samohodnocení, a pokud by měl být program posuzován jako kreativní, jeho hodnocení vlastních výstupů (podle kritérií dle Ritchieho, viz níže) by mělo korelovat s hodnocením lidí. Na druhou stranu i náhodný výběr z množiny možných kombinací artefaktů má například v hudební historii své místo (viz aleatorická hudba) a Geraint Wiggins (2006) uznává, že tato simulace lidské kreativity může být dostatečná, zejména pokud se systém řídí alespoň nějakými pravidly oproti čistě náhodnému výběru. Daná množina však musí být dostatečně komplexní. V tomto případě sám artefakt, který z ní vychází, spíše pozorovatelé vyhodnotí jako kreativní. Oproti Bundymu (1994) tvrdí Wiggins, že tato podmínka je silnější, než aby byl artefakt komplexní sám o sobě.

3.2. Výstup nebo produkce?

Z otázky kreativity počítačů můžeme nepatrně poodstoupit k empirické perspektivě, která se zaměřuje spíše na to, jestli počítače mohou za určitých okolností přijít s myšlenkami či artefakty, které se alespoň *jeví* jako kreativní, respektive do jaké míry a v jakém smyslu jsou kreativní jejich výstupy (Boden, 2004). Podobně se k výstupům těchto programů staví Ritchie (2007), který tvrdí, že kreativita jednotlivce se projevuje v jeho artefaktech, především by pak měly být brány v potaz jen empiricky pozorovatelné faktory. Interní procesy programu relevantními daty nejsou a můžeme tedy zcela ignorovat proces jeho produkce. Ritchie argumentuje, že a) kreativita lidí je obvykle posuzována podle jejich artefaktů a tak se tímto zjednodušením dávají jejich artefakty a artefakty počítačů na podobnou úroveň, tak, aby se daly posuzovat, za b) skryté a základní procesy nejsou pozorovatelné a tudíž nespolehlivé a c) je zde riziko zacyklení, pokud posuzujeme zároveň artefakt i proces jeho produkce (Ritchie, 2007).

Mnoho testů, které se zaobírají programy produkující hudbu, se tak téměř výhradně zaměřuje jen na samotný výstup programu. Colton (2008) oproti tomu přichází se zaměřením na produkci a samotný systém, jelikož posouzení systému jen na základě jeho výstupů není dostatečné – samo o sobě nám neřekne nic o použitelnosti systému (Pearce, Meredith a Wiggins, 2002). Podobně i Loughran a O’Neill (2017) dodávají, že mnoho přístupů se zaměřuje jen na finální artefakty i přes to, že počítačová kreativita je definována jako chování systému jako celku. Z tohoto pohledu je relevantní se zabývat například tím, jak a za jakých podmínek daný artefakt vzniknul, roli programátora nebo tým, jaké metody se více blíží lidským artefaktům či vnitřním procesům.

Zaměřování se jen na výstupy počítače může být nicméně přístupem, který obecně počítač při poměrování skladeb počítače a člověka z debaty *a priori* nevyřazuje. Podle Simona Coltona (2008) existuje negativní sentiment vůči počítačům a vnímání jejich předpokládané kreativity. Podle Bodenové (2010) pokud lidé zjistí, že jimi obdivované umělecké dílo je generované počítačem, mění svůj názor a cítí se být ošálení. Je například znám případ kritika, který měl napsat recenzi na veřejný koncert Copeova programu EMI během Santa Cruz Baroque Festival, odmítavou kritiku napsal ale již o dva týdny dříve, než koncert vůbec proběhl (Blitstein, 2010). Zaujatost vůči hudbě generované počítači empiricky zkoumali David C. Moffat a Martin Kelly (2006), kteří zjistili, že předsudky vůči takovýmto skladbám existují ve větší míře u hudebníků než u laiků, Philippe Pasquier a kol. (2016) nicméně v replikované studii důkazy o výrazné míře předsudků nenalezli.

Úplné oddělení „*autora*“ od skladby je ale v kontextu praktického provozování hudby vždy zjednodušením, tento úzký vztah je totiž stále nenarušený – ať už v kritice nebo koncertních programech. Jak píše Barthes (2006, s. 75): „*Výklad díla je vždy hledán u toho, kdo je vytvořil (...) Tato koncepce se velmi dobře hodí kritice, která si chce dát za hlavní úkol objevit v díle Autora (či jeho hypostaze: společnost, historii, ducha, svobodu): jakmile se najde Autor, text je „vysvětlen“.*

3.3. Evaluace kreativních systémů

Hlavní otázka se nám tedy přesunula z filozofické „*Mohou být počítače opravdu kreativní?*“ k více praktickému dotazu „*Jaké metody jsou nejvhodnější k produkování kreativního chování?*“ (Ritchie, 2007) a způsobům, jak takový kreativní systém nebo jeho výstup hodnotit.

Pokud mluvíme o evaluaci, je potřeba oddělit vnitřní, systémovou a vnější, nezávislou evaluaci. Vnitřní evaluace (resp. hodnotící funkce) je v systémech inkorporována jako součást generace nových skladeb. Jak bylo zmíněno v kapitole II, tuto roli vnitřního kritika mohou zastávat předem daná pravidla hudební teorie, naučená pravidla jednou z technik strojového učení, nebo posluchači u interaktivní evaluace. Bez těchto mechanismů sebereflexe nemůže být nikdy žádný systém dostatečně autonomní, aby mohl generovat opravdu kreativní díla (Agres a kol., 2017), zabývat se budeme ale druhým typem evaluace.

Vnější evaluace jako taková naproti tomu odpovídá na otázku, jak úspěšný je daný systém při komponování hudby (Loughran a O'Neill, 2017), při použití termínu evaluace tak bude myšlen tento význam. Vnější evaluace výstupů systému může mít podobu nezávislého hodnocení, které vychází ze stejných procesů jako u hudby lidského skladatele, jako je kritika, muzikologická analýza či v širším významu ohlas publika. Evaluace, která zohledňuje umělý zdroj komponované hudby a potřebu po objektivním a empiricky měřitelném výsledku, může mít dvě podoby: a) testy pracující s posluchačem, jeho vnímáním a subjektivními pocity a b) statistické metody.

Jak již bylo zmíněno, mnoho snah o formalizované a empirické zhodnocení kreativních systémů se až na výjimky zaměřuje právě na jejich produkci, kdy jde v první řadě o poslechové testy navazující na Turingův test.

Název a autor	Rozlišování člověka a počítače	Interakce	Výstup / Produkce	Předmět testu	Zpětná vazba	Využito při evaluaci programů
Musical Output Toy Test (Ariza, 2009)	ano	ne	výstup	vědomí, kreativita	ne	ano
Turingův test pro kreativní systémy (Boden, 2010)	ano	ne	výstup	kreativita	ne	ne
Musical Directive Toy Test (Ariza, 2009)	ano	ano	výstup	vědomí, kreativita	ne	částečně
Rozlišovací test (Pearce a Wiggins, 2001)	ano	ne	výstup i produkce	schopnost generování skladeb	ano	ano
Lovelace test (Bringsjord a kol., 2001)	ne	ne	výstup	kreativita	ne	ne
Empirická kritéria pro přisouzení kreativity (Ritchie, 2007)	ne	ne	výstup	kreativní chování	ne	ne
SPECS (Jourdanous, 2012)	ne	ne	výstup	kreativita	ano	ne
Kreativní trojnožka (Colton, 2008)	ne	ne	výstup i produkce	kreativita	ano	ne
IDEA (Colton a Pease, 2011)	ne	ne	výstup i produkce	kreativita	ano	ne

Obrázek 6: Typy evaluace kreativních systémů založené na poslechu (autor)

Christopher Ariza (2009) nabízí dva různé přístupy při užití Turingova testu v hudbě, které sám podle hierarchie Stevena Harnada dává na úroveň t1 (Toy Test). Prvním z těchto testů je *Musical Directive Toy Test*. V něm lidský agent komunikuje s lidským skladatelem a počítačem a zadává jim, jak by měla daná skladba vypadat (nálada, délka, tónina apod.). Poté dostává od skladatelů kompozice ve formě not, syntetické hudby či přímo nahrávek a má za úkol vybrat, které dílo složil počítač, respektive člověk. *Musical Output Toy Test* je naproti tomu bez interakce, tazatel dostává od skladatele a počítače již hotová umělecká díla, na jejichž podobu neměl přímý vliv. U obou testů je možné, aby tazatel obdržel více než jedno umělecké dílo od každé entity.

Bodenová (2010) uvažuje o vlastní verzi Turingova testu pro nezávislé programy produkující umělecké artefakty, který se dá považovat za *Musical Output Toy Test*. Jeho podmínkou by mělo být, že lidský agent má zhruba pět minut na posouzení

daného uměleckého díla. Tento program by testem prošel, pokud by jeho výstup byl:

- 1) nerozlišitelný od skladby lidského skladatele, nebo
- 2) měl stejnou estetickou hodnotu

Tuto druhou podmínku Bodenová upravuje tak, aby se estetická hodnota poměřovala ne s geniálními díly lidské historie, ale s průměrnými díly člověka – stejně jako v původním Turingově testu stačí, aby měl testovaný počítač literární schopnosti běžného člověka. Podle ní například EMI Davida Copea nedokázala o svých výsledcích přesvědčit hudební experty, pokud soutěžila přímo s díly imitovaných skladatelů, v porovnání s díly současných skladatelů imitujících skladby dávných mistrů ale uspěla.

Marcus Thomas Pearce a Geraint Wiggins (2001) přišli s *Rozlišovacím testem* (Discrimination Test). Ten je podobný popsanému *Musical Output Toy Test* co do míry pasivity tazatele, opouští ale koncept Turingova testu a vyhýbá se tak jeho implikacím směrem k inteligenci či vědomí daného systému. U tohoto testu dostává tazatel určitý data set, ve kterém má za úkol vybrat skladby, které jsou podle něj složeny počítačem. Pokud jsou skladby počítače označovány jako skladby lidského skladatele v četnosti vyšší, než je náhodný výběr, můžeme říci, že jsou od těch lidských nerozlišitelné.

Rozlišovací test je nicméně jen poslední částí konceptuálního rámce, skrz který chtějí Pearce a Wiggins *objektivně* hodnotit kompozice počítače. Před samotným testem je nezbytným krokem definice cílů programu, nastavení automatického „kritika“ systému, který není založen na pravidlech, ale na některé z metod strojového učení, a samotné generování hudby, která tímto hodnocením kritika projde.

Ve své pozdější práci se Pearce (2005, s. 188 - 189) detailněji zabývá rolí posluchačů-tazatelů a jejich odborností. Cituje definici kreativity Teresy Amabile (1996), podle které je produkt kreativní, jen pokud jej za kreativní považují – nezávisle na sobě - odborníci v dané oblasti. Na této definici je také postavena

empirická metodologie tzv. technika konsenzuálního posuzování (consensual assesment technique), která počítá s tím, že by tazatelé měli být zkušení odborníci v příslušné oblasti, schopni: a) na ostatních odbornících nezávislého hodnocení, b) u artefaktu posuzovat technické provedení, stejně jako estetické kvality a originalitu, c) artefakty posuzovat relativně s ohledem na ostatní prezentované v rámci experimentu a d) posuzovat artefakty v náhodném pořadí. Tato technika byla v hudební oblasti například použita při hodnocení hudby studentů skladby, vychází z nich ale také autoři při evaluaci skladeb počítače.

Na základě námítky Lady Lovelace navrhli Selmer Bringsjord a kol. (2001) tzv. *Lovelace test*. Podle nich je stroj kreativní jen a pouze v tom případě, pokud je zcela autonomní, tedy pokud jeho stvořitel / programátor neumí vysvětlit vznik artefaktu, který tento stroj vyprodukoval. Zároveň se nesmí jednat o chybu tohoto stroje a stroj tedy může podobný artefakt vytvořit znovu. Sami autoři jsou však k takové možnosti autonomie strojů silně skeptičtí, jelikož podle nich vyžaduje kreativita vnitřní pohled z první osoby, tedy vědomí.

Otázkou, jak nejlépe zhodnotit kreativní systémy obecně bez odkazování se na Turingův test, se zabývalo více autorů. Ritchie (2007) také navrhuje postup se zaměřením na výstup programu, ale v první řadě nezkoumá, jestli je počítač schopen napodobit lidského skladatele do té míry, že “zmate” člověka. Člověk hodnotí výstupy podle 18 formálních kritérií zaměřujících se na atributy, jako jsou a) kvalita, b) typičnost produkovaného artefaktu v kontextu žánru či stylu a sekundárně také c) originalita. Výsledkem pak není odpověď na otázku, jestli je daný systém kreativní, spíše jestli je počítač vůbec schopný uspět v úkolu vytvářet ten správný artefakt a do jaké míry se choval kreativně (Ritchie, 2001).

Podle Anny Jordanous (2012), která Ritchieho návrh sama implementovala při evaluaci programu na jazzovou improvizaci (Jordanous, 2010), jsou představené atributy neurčité a praktická využitelnost není příliš velká. Sama proto přišla se Standardizovaným postupem pro evaluaci kreativních systémů (Standardised Procedure for Evaluating Creative Systems), který je třístupňovým procesem vycházejícím z její definice 14 oblastí kreativity. Mezi ně patří: aktivní zapojení a vytrvalost, schopnost vyrovnat se s neurčitostí, znalost oboru, obecná inteligence,

generování (nových) výstupů, nezávislost, záměr a emocionální účast, originalita, rozvoj, komunikace a interakce, spontaneita a podvědomí, experimentování a variace, sebereflexe a generování hodnotných výstupů. Hlavními kroky tohoto procesu jsou:

- 1) Identifikace definice kreativity, které by měl daný systém vyhovovat, aby byl posouzen jako kreativní
- 2) Odvození z této definice ty oblasti kreativity, které budou testovány a jaké váhy jim budou přiděleny
- 3) Samotné testování výkonu systému v daných oblastech kreativity, ideálně jako kombinace kvalitativního a kvantitativního výzkumu

Test výkonu v kreativních oblastech provádějí lidé na výstupech daného systému či systémů, a kromě možného srovnání jejich výsledků mohou dát jednotlivým systémům i potřebnou zpětnou vazbu, což je z hlediska výše popsaných způsobů evaluace opomíjeným atributem hodnotících prostředků (Jordanous, s. 257- 258).

Jak již bylo zmíněno, Colton (2008) uhýbá z hodnocení samotného výstupu a přichází s konceptem Kreativní trojnožky (The Creative Tripod), který popisuje tři vlastnosti samotného systému, které by měl mít, aby byl posouzen jako kreativní. Jde o a) dovednost b) pochopení a c) představivost: „*bez dovednosti by nemohl nic produkovat; bez pochopení by nemohl produkovat nic hodnotného; a bez představivosti by mohl přinejlepším produkovat jen imitace děl jiných lidí*“ (Colton, s. 4). Všechny tyto vlastnosti se mohou projevit nejen u samotného počítače, ale i u posluchače a programátora, kteří se dohromady na projevu kreativity podílí. U tohoto přístupu však není například jasné, jestli má každá „noha“ stejnou váhu a pokud ano, jestli tomu tak je i v jiných oblastech lidské činnosti (Jordanous, 2012).

Simon Colton a Alison Peaseová (2011) přišli s dalším modelem evaluace výstupů kreativních systémů, tzv. IDEA modelem (Iterative Development Execution Appreciation). Ten pracuje s „ideálním“ publikem, které by mělo být schopno určit v jaké míře na ně a jejich náladu měl produkový artefakt vliv a jaké kognitivní úsilí museli na jeho hodnocení vynaložit. Z těchto hodnot pak model IDEA odvozuje ukazatele, jako jsou popularita, ošklivost, provokativnost nebo

indiference. Danému publiku by měly být poskytnuty co nejdetailnější informace o produkci dané skladby a fungování softwaru, aby mohlo program a jeho dílo co nejlépe zhodnotit.

Další možnosti evaluace jsou statistické metody, které většinou měří určitou vzdálenost (či podobnost) dané skladby s požadovanou skladbu či stylem skrz porovnání distribuce různých atributů, jako jsou výšky tónů, rytmická struktura, celkový počet disonancí atd. Na otázky ohledně estetické kvality výstupu dle lidských měřítek nicméně odpovídat neumí.

Úkrok od zaměřování se na lidský poslech hudebních artefaktů činí i Róisín Loughran a Michael O'Neill (2016, 2017), kteří nabízí další cestu k více objektivní evaluaci. Tvrdí, že autonomní, statistická evaluace by byla nadřazená tradičním testům vycházejících z Turingova testu. Tato evaluace by se především neměla řídit lidskými standardy, jelikož dostatečně nerozumíme ani kreativitu lidské. Lidské hodnocení je náchylné k předpojatosti, subjektivnímu hodnocení a zaujatosti, navíc skrze něj omezujeme potenciál autonomních systémů: „*Pokud odmítneme artefakty hned, co se nám znelíbí, nebo je neshledáme příjemnými, jak máme čekat, že nás kreativní systém někdy opravdu překvapí?*“ (2016, s. 13).

Přístup Loughran a O'Neilla nicméně stále zůstává v teoretické rovině, které zatím nenabízí oproti výše zmíněným přístupům praktické využití. Způsoby hodnocení kreativních systémů pocházející z prací autorů Bringsjorda a kol., Margaret Bodenové, Simona Coltona a Alison Peaseové nebo Anny Jordanous sice poskytují rámec pro empirické zkoumání, ale, pokud je nám známo, žádný z autorů programů skládajících hudbu jej nevyužil a v následující kapitole tak budou popsány buď jen provedené testy vycházející z Turingova testu a Rozlišovacího testu anebo další poslechové testy neodkazující na žádnou z uvedených metod.

4. Provedené testy

O napodobení Turingova testu v hudbě se za posledních 30 let pokusila řada autorů, kteří se neomezili jen na práci skladatele jako takového, ale i na interaktivní komponování v reálném čase, tedy improvizaci (viz Aucouturier a Pachet, 2003), aranžování originálních skladeb pro určitý nástroj (viz Huang a kol., 2012) nebo na samotnou hudební interpretaci, např. při hraní na klavír (viz Schubert a kol., 2017; Hiraga a kol., 2004). Tato kapitola bude nicméně věnována jen programům, které hudbu skládají podobně jako lidští skladatelé v užším slova smyslu. V první části této kapitoly budou popsány testy přímo či nepřímo navazující na práci Alana Turinga, další část bude věnována dalším provedeným poslechovým testům a alternativním metodám.

4.1. Musical Output Toy Test

Řadě experimentů zkoumajících rozeznatelnost lidské a počítačové hudby byl podroben program *Experiments in Musical Intelligence* Davida Copea. Tyto testy patří podle Arizy (2009) k *Musical Output Toy Test*, jelikož se přímo odvolávají na Turingův test a zároveň je hodnocen jen výstup daného programu bez jakékoli interakce mezi ním, potažmo lidským skladatelem a posluchačem / posluchači. Jedny z prvních provedl přímo Cope (1996), i když, jak ostatně sám přiznává, neměly tyto experimenty nijak přísné vědecké parametry a způsob jejich zachycení je spíše anekdotický.

Cope (2001) testy označuje termínem *Hra* (The Game), inspirující se pravděpodobně v Turingově *Imitační hře*. *Hra* má několik pravidel: prvním je obecně výběr hudby lidských skladatelů, která je méně známá. To má dva důvody: taková hudba většinou není tím nejlepším, co daný skladatel složil a nedává mu to tak ve *Hře* velkou výhodu. Zároveň tyto skladby většinou na poslech lidé nepoznají a nemusejí být z tohoto testu ve velké míře diskvalifikováni. Lidské skladby nicméně musí mít alespoň průměrnou kvalitu, jelikož „*smíšení slabých skladeb člověka a skvělých skladeb počítače by posluchače jen oklamalo*“ (Cope, 2001, s. 13).

První z testů proběhl v roce 1991, kdy dal Cope za úkol osmnácti studentům, aktivním hudebníkům, aby rozpoznali v souboru 25 hudebních nahrávek části skladeb Wolfganga Amadea Mozarta od těch, které v jeho stylu zkomponoval program EMI. Cope vybral části méně známých Mozartových sonát, skladby EMI ve stylu Mozarta se signaturami, tedy reprezentativní výstup Copeova programu a také skladby EMI bez signatur. Studenti předem neznali poměr skladeb člověka / počítače. Mozartovy skladby rozeznalo správně 67 % studentů, zároveň jako hudbu Mozarta označilo 60 % skladby EMI se signaturami a 38 % skladby EMI bez signatur.

Další experiment provedl Cope v roce 1992 v rámci konference Americké asociace umělé inteligence (American Association for Artificial Intelligence - AAAI) v San Jose. U tohoto testu využil jen skladby EMI se signaturami a vybral ze svého programu ty nejlepší. Zde označovalo publikum skladby počítače jako skladby lidského skladatele v četnosti mezi 40 až 60 %, což podle Copea dokazuje, že minimálně pro laika je nemožné rozeznat umělého Mozarta od pravého. Cope (1996, s. 81 – 82) mluví o vysokém počtu účastníků tohoto experimentu, který byl „téměř“ dva tisíce, později (Cope, 2001, s. 13) píše až o pěti tisících. Podobný experiment s díly Johanna Sebastiana Bacha proběhl i na konferenci AAAI v roce 1998 v Madisonu (Mostow a Rich, 1998).

Cope dále píše o experimentu Douglase Hofstadtera s experty na Eastman School of Music na Univerzitě v Rochesteru (New York), kde před publikem, ve kterém byla také řada skladatelů a hudebních teoretiků, hrál Hofstadter na klavír živě skladby lidské a EMI. V experimentu Bachovy hudby EMI nepochodila, ale u „Chopinových“ mazurek bylo správných odpovědí zhruba 50 %. Tento výsledek ve *Hře* je obecně pro Copea přijatelný a ukazuje, že posluchači nedokáží poznat opravdový rozdíl. Hofstadter tento experiment zopakoval i na Univerzitě v Massachussets, kde pro publikum hudebníků, skladatelů a vědců z oblasti IT ve „většině“ vybíralo skladby EMI (Silva, 2003, s. 66).

V roce 2000 zorganizovala BBC1 experiment *Live Lab* v rámci programu *Tomorrow's World*, při kterém se několik tisíc účastníků zapojilo do tří verzí Turingova testu umístěného na webových stránkách. Jedním z nich byl klasický

Turingův test s chatboty (ALICE, MABEL), další byl výtvarný (AARON) a poslední byl hudební test (EMI). Hudebního experimentu se účastnilo 1653 diváků, z nichž 42 % označilo skladbu EMI jako hudbu člověka, a 31 % účastníků, respektive 27 % hlasovalo pro zbylé dvě kompozice nespécifikovaných skladatelů (Bloomfield a Vurdubakis, 2003, s. 32).

Prvním z akademicky pojatých a také zachycených testů EMI (a programu ILLIAC) je již zmíněná práce Davida Moffata a Martina Kellyho (2006). Ti zkoumali primárně předsudky vůči počítačové hudbě, jednou z jejich hypotéz však bylo i to, že lidé jsou schopni rozeznat jazzovou a klasickou počítačovou hudbu od lidské a zároveň že v této disciplíně budou lepší experti spíše než laici. Autoři pro test vybrali 6 nahrávek skladeb o délce jedné minuty, z čehož polovina byly skladby lidských autorů (Dmitrij Šostakovič: Symfonie č. 3 - Allegro non troppo, Johann Sebastian Bach: Koncert č. 5 v F dur - Largo a Art Ensemble Chicago: Full Force) a druhá polovina skladby počítačových programů (ILLIAC: Smyčcový kvartet č. 4, EMI: Koncert: II – ve stylu J. S. Bacha a VOYAGER: Duo 3).

Pro experiment bylo vybráno 10 hudebníků, respektive lidí s formálním hudebním vzděláním a 10 laiků ve stejné věkové skupině. Ve skupině lidí s hudebním vzděláním data ukazují, že přesně v 50 % případů nebyli schopni skladby počítače rozpoznat. U laiků je tento poměr překvapivě v neprospěch počítačových programů, 63,3 % odpovědí odhalilo počítač. Bohužel Moffat a Kelly kombinují dohromady data tří různých programů, nedají se z nich tedy vyčíst jednotlivé výsledky, zároveň jsou ve stejné tabulce uvedeny sečtené výsledky pro jazzovou a klasickou hudbu dohromady. Jakékoliv závěry je tedy těžké odvozovat. Test navíc neproběhnul přímo dle pravidel původního Turingova testu – účastníkům nebylo explicitně zadáno, ať rozeznají lidské skladby od skladeb počítače, a navíc neznali jejich poměr v souboru všech skladeb.

Pro evaluaci svého modelu Multiattribute Prediction Suffix Graph (MPSG), založeným na skládání čtyřhlasých chorálů za pomoci modelu podobnému Markovovým řetězcům, využili José Luis Triviño-Rodríguez a Rafael Morales-Bueno (2001) Kolmogorovův–Smirnovův test. Skrz něj zjistili, že MSPG dokáže generovat skladby se stejným rozdělením pravděpodobnosti jako mají skladby pro

sbor J. S. Bacha, nicméně si chtěli statistickou podobnost potvrdit i poslechovým testem. V tomto testu figurovalo 52 posluchačů, kteří měli za úkol správně přiřadit dva desetitaktové fragmenty ze skladeb MPSG a Bacha k jejich autorům. To se jim povedlo v 55 % případů. Triviño a Morales zmiňují 50 % hranici, která je pro počítač potřebná k úspěšnému splnění Turingova testu, při spočítání statistické významnosti rozdílu ($p < 0,01$) tak podle nich program prošel (Triviño a Morales, s. 73). Detailní data o výsledcích tohoto testu nicméně neposkytují, zároveň není jasné, jaká byla odbornost daných posluchačů.

Již zmíněný program BachBot (Liang, 2016) byl podroben webovému testu nazvanému *The BachBot Challenge*¹, kdy měli účastníci za úkol rozeznat Bachovy čtyřhlasé chorály od skladeb, které složil celé BachBot nebo jen určitý hlas (soprán, alt, tenor, bas) či jejich kombinaci. Experimentu se zúčastnilo 721 lidí z několika kontinentů, které Liang shromáždil především přes sociální sítě. Každý účastník měl na začátku testu zadat, jaká je jeho odbornost v hudbě, respektive jaká je jeho úroveň, co se týče formálního vzdělání:

- 1) začátečník, který nemá formální vzdělání, ale poslouchá hudbu
- 2) středně pokročilý, který hrál či hraje na hudební nástroj, nestudoval skladbu
- 3) pokročilý, který studoval hudební teorii či skladbu
- 4) expert, který hudbu vyučuje nebo je hudební teoretik

Tyto skupiny nejsou nijak zvlášť přesně rozděleny a často se mohou překrývat, popřípadě se dá rozporovat spolehlivost anonymních účastníků, nicméně se, až na menší odchylky, nakonec míra odbornosti odráží na výsledcích testu (oproti experimentu Moffata a Kellyho). Celkově účastníci rozpoznali počítač od J. S. Bacha v 59 % případů, což Liang interpretuje jen jako o 9 % lepší výsledek, než je náhodný výběr a test tak „*potvrzuje úspěch našeho projektu*“ (Liang, s. 58).

Z akademických pokusů o replikaci Turingova testu se v tomto případě podařilo do experimentu zapojit největší počet lidí, bohužel ale Liang v konečných výsledcích neoddděluje začátečníky od posluchačů s větší expertízou. Pokud celkové výsledky

¹ Dostupný zde: bachbot.com

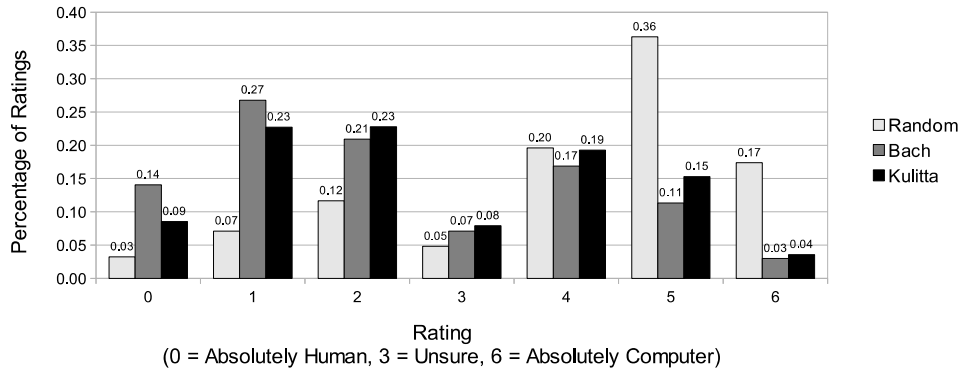
nebereme jen jako důkaz toho, že průměrný účastník počítač nerozpozná (začátečníci mají skóre jen 51 %, tedy náhodný výběr), tak při oddělení začátečníků nám vyjde číslo úspěšnosti 61 % a při započítání jen studentů skladby či opravdových expertů je celkový výsledek dokonce 68 %, což se jako úspěch programu jistě brát nedá a určitě ne jako důkaz vyřešení otázky harmonizace chorálů ve stylu J. S. Bacha, jak tvrdí Liang.

Gregory Cowart a kol. (2019) replikovali experiment s BachBotem k vlastnímu Turingovu testu, jehož se zúčastnilo 10 studentů a vyučujících z Virginia State University s různým hudebním a technologickým zaměřením, z nichž se 40 % označilo jako začátečníci a dalších 40 % jako středně pokročilí podle rozdělení Lianga (2016). Cowartův test měl v jeho kvantitativní části kombinovat výsledky níže popsaného testu s umělou inteligencí Rajona (2018) a testu s BachBotem a srovnat je s tím, jak účastníci zařadí pět skladbiček projektu JukeDeck, který se zaměřuje na popovou hudbu. Bohužel se, zcela nepochopitelně, Cowart omezuje stran Turingova testu s BachBotem jen k prohlášení, že tvůrci na svých stránkách neposkytují výsledky k jednotlivým skladbám, nýbrž jen k celkovým a že hodnota mediánu správně zodpovězených otázek je nízká. Dál se tedy tímto testem nezabývá a pracuje s výsledky dalších dvou testů.

Podobně jako Liang i Rajon (2018) využil pro sběr odpovědí online prostředí. Jeho Turingova testu s 5 skladbami umělé inteligence a 5 od lidského skladatele se zúčastnilo 85 účastníků, u kterých se však nedozvídáme jejich vztah k hudbě a Rajon jej pravděpodobně ani nezjišťoval. Z těchto účastníků rozeznalo správně skladby umělé inteligence od lidského skladatele 46 %, číslo je tak blízké náhodnému výběru. Tyto výsledky ale po poslechu použité hudby překvapením nejsou. Rajon píše, že lidská hudba je vyextrahovaná ze skladeb J. S. Bacha, bohužel při poslechu není jasné, v jakém stylu „lidská“ hudba (a v důsledku také ta od umělé inteligence) má být. Použitá je jen melodie, která je zrychlená, k ní jsou přidány laicky složené akordy na každou dobu, a navíc jsou zde použity bicí. Nepoužitelnost těchto výsledků dokazuje také to, že Rajon vyextrahoval části poměrně známých skladeb z Bachova *Dobře temperovaného klavíru I.*

Poměrně ojedinělý přístup k Turingovu testu zvolila Donna Quick (2014) při evaluaci svého programu Kulitta. Ve svém testu využila Quick kromě svého programu a chorálů J. S. Bacha i stochastický generátor hudebních frází, zjednodušené verze Kulitty neschopné tvorby hierarchických struktur. U těchto programů předpokládala, že by klasickým Turingovým testem neprošly, proto účastníci neměli za úkol rozlišit skladby počítače od člověka, ale na Likertově škále zadat stupeň přesvědčení o jejich původu. Quick zmiňuje, že Likertovu škálu použila kvůli tomu, že „soupeřem“ Kulitty není jen obyčejný člověk, jako by tomu bylo u klasického Turingova testu (nebo při využití žijícího autora), ale „expert“ - J. S. Bach. *„Předpokládejme, že oba [programy] by v Turingově testu neuspěly a byly by jako stroje označeny ve 100 % případech. To ale neznamená, že si oba algoritmy vedly při testu stejně“* (Quick, s. 156). Klasický binární test by tak v detailu neukázal, jak na tom Kulitta doopravdy je.

Quick uskutečnila dva testy na skupinách o 121, respektive 116 účastnících, které autorka oslovila na webu Amazon Mechanical Turk. Mezi účastníky byli jak hudební experti, tak laici. U první skupiny využila Quick Kulittu s předem vloženými pravidly postupu akordů, ve druhé skupině byla Kulitta, která se učila přímo na chorálech J. S. Bacha. Samotné testování proběhlo v online prostředí. Účastníkům bylo po seznámení s hudbou lidských (J. S. Bach, A. Schoenberg) a počítačových skladatelů (neidentifikovaný L-systém a EMI) přehráno na 40 hudebních frází (bez možnosti opakování), které měli za úkol hodnotit na Likertově škále od 0 – Určitě člověk až po 6 – Určitě počítač. Hudební fráze byly převedeny do MIDI, ve čtyřech hlasech, ve 4/4 taktu bez dynamiky a změn tempa, každá fráze měla délku 10 vteřin.



Obrázek 7: Výsledky testu s Kulittou (Quick, 2014)

Výsledky ukazují, že rozdíly mezi jednotlivými zdroji hudby jsou u každé skupiny statisticky velmi významné ($p < 0,01$). U Bachových chorálů a „náhodného“ generátoru dopadly u obou skupin dle očekávání: Bachovy skladby hodnotili účastníci nejvíce jako lidské, generátor jako počítačové. Kulitta se umístila někde uprostřed, podle Quick se ale překvapivě umístila na lidské části stupnice, s průměrem 2,67 pro první skupinu a 2,8 pro druhou (viz Obr. 7), což dokazuje, že by klasickým Turingovým testem mohla projít. Pro tento test se v žádné skupině nepotvrdila korelace mezi hudebním vzděláním a správnými odpověďmi, data k tomu ale Quick neuvádí.

4.2. Musical Directive Toy Test

Donna Quick poznamenává, že především experti mohou být při testu ovlivněni tím, že dané skladby J. S. Bacha předem znají. Proto by bylo nejlepší mít nové (či nově objevené) skladby od samotného Bacha. Zde se však již pomalu dostáváme k druhému typu Turingova testu, *Musical Directive Toy Test* (Ariza, 2009), kdy tazatel či posluchač dává dvěma entitám zadání k nové hudbě, kterou poté porovnává. O tomto typu testu v ryzí podobě nejsou, co se týče hudební kompozice, žádné záznamy. Existují pokusy s využitím nově složené hudby od lidských skladatelů s jistým zadáním, jehož stejné podmínky se dávají jak skladateli, tak počítači, v následujících případech jde však o situace, kdy zadavatelem není hodnotitel, nýbrž samotný programátor a hodnotitelé jsou tak znovu ve stejné pasivní roli jako u MOTT.

Steve Larson na Univerzitě v Oregonu provedl hudební test s EMI, kde pro své publikum připravil hudební Turingův test. Publiku dal na výběr ze tří skladeb: skladby EMI, Bachovy a své vlastní. Jako počítačovou označilo publikum jeho vlastní skladbu a hudbu EMI jako Bachovu (Kurzweil, 1999, s. 116). Více se však o testu nedozvídáme.

Inspiraci Triviňem a Moralesem zmiňuje Walter Schulze (2009), který při vzniku svého programu SuperWillow využil podobné techniky. Zároveň k jeho evaluaci provedl online experiment, pro který nechal složit nové skladby ve třech stylech od studentů s různým hudebním vzděláním a zkušenostmi. Pro každý styl nechal generovat hudbu přímo z nových kompozic. Testu se zúčastnilo 440 posluchačů, studentů a vyučujících ze Stellenbosch University (JAR), kteří celkově označili 36 % skladeb SuperWillow jako hudbu člověka. Tento „částečný“ Turingův test zkombinoval Schulze s průzkumem posluchačských preferencí, a i přes neúspěch svého programu v testu píše o tom, že 68 % posluchačů buď nedokázalo rozeznat počítač od člověka nebo více preferovalo hudbu počítače.

Bohužel program SuperWillow je spíše prototypem, proto měli oslovení skladatelé za úkol složit hudbu úměrnou k jeho omezeným možnostem. Museli například složit jen skladbu o cca 8 taktech v C dur, aby byla celou dobu ve 4/4 taktu a navíc bez pomlky. Zároveň není jasné, k jakým hudebním stylům by měla uvedená hudba patřit, jelikož Schulze označuje jednotlivé styly jen jmény studentů podílejících se na experimentu. Pokud by měla být tato hudba označena jako klasická, tak jen jako velmi primitivní².

Přístup využívající nové složené hudby skladatelů přímo pro Turingův test použili také Tim Murray Browne a Charles Fox (2009), kteří experimentovali s evolučními algoritmy. Svému programu dali jako zadání čtyři motivy z Preludia a Fugy č. 4 v c moll, BWV 849 J. S. Bacha s tím, že výsledná skladba musí být pro klavír se třemi hlasy ve 4/4 tempu v délce 32 taktů. Skladatel Zac Gvirtzman dostal identické zadání, zároveň měl na kompozici dva dny a předem neznal výstup počítače. Výsledná hudba, v obou případech připomínající značně atonální fugu³, byla

² K poslechu zde: <http://superwillow.sourceforge.net>

³ K poslechu zde: <http://c4dm.eecs.qmul.ac.uk/papers/2009/MurrayBrowne09evo-data>

nahrána profesionálním klavíristou a nahrávky byly představeny 22 náhodně vybraných studentům hudby z Univerzity v Oxfordu s tím, aby rozeznali skladbu počítače. Z tohoto počtu zodpovědělo 9 studentů správně, 12 nesprávně a 1 odmítnul odpovědět, z čehož Browne a Fox odvozují, že jsou skladby navzájem nerozeznatelné.

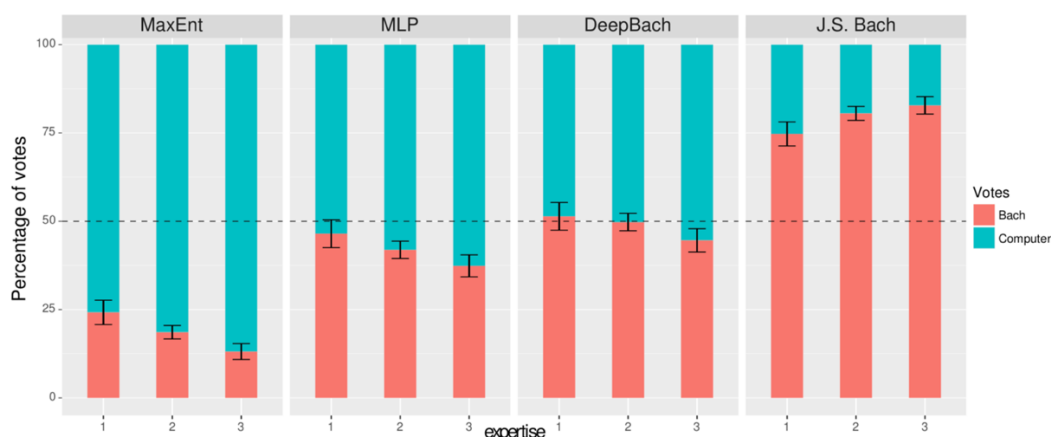
U výše popsaných testů je v naprosté většině autory hlášen úspěch nebo alespoň výsledek blízky úspěchu svého programu v překonání hudebního Turingova testu. Všichni tito autoři zmiňují ve větší či menší míře práci Alana Turinga, co ale mají ve své práci společné, je to, že výsledek svého programu v testu komunikují ne jako důkaz přítomnosti vědomí (alespoň nikdy ne explicitně), nýbrž jako ukazatel kvality složené hudby a programu jako takového.

4.3. Rozlišovací testy

Další skupinou experimentů jsou poslechové testy s rozlišováním hudby počítače a člověka, jejichž autoři nezmiňují práci Turinga a z jeho myšlenek nevycházejí ani do té míry, že by „úspěšnému“ programu přisuzovali jakékoliv vědomí. Forma Turingova testu je pro ně jen východiskem k empirickému zkoumání jejich programu, podobně jako u výše zmíněných autorů. Identicky zachází s těmito poslechovými testy autoři, kteří také nepíšou o inspiraci v Turingovi, ale namísto toho staví na konceptu *Rozlišovacího testu* Pearce a Wigginsa (2001), který aplikují na design svých experimentů.

Gaëtan Hadjeres a kol. (2017) tímto rozlišovacím testem nazvaným „Bach nebo počítač“, který proběhl online s účastí 1272 lidí, hodnotil výsledky programu DeepBach v porovnání s jednoduchým modelem vícevrstvé dopředné neuronové sítě a méně komplexním programem MaxEnt. Všem programům byl zadán sopránový hlas z Bachových chorálů, které měly programy znovu harmonizovat podle sebe. Účastníci byli rozděleni podle úrovně expertízy na tři skupiny: a) občasný posluchač vážné hudby (261), b) pravidelný posluchač či hudebník (646) a c) student skladby nebo profesionální hudebník (365).

V testu pak měli účastníci za úkol rozlišit, jestli přehraná syntetizovaná hudba pochází od J. S. Bacha či od počítače. Výsledky (viz Obr. 8) v první řadě ukazují, že čím vyšší je komplexita programu, tím spíše jej účastníci označili jako lidského skladatele, u pravého Bacha je pak tento poměr nejvyšší. Zároveň se potvrdilo, že přesnost odpovědí ve všech případech stoupá s úrovní odbornosti. DeepBach si tedy vedl nejlépe, jeho výsledek zhruba 50 % u všech skupin Hadjeres a kol. (2017) hodnotí bez dalších komentářů jako úspěch oproti jednodušším modelům.



Obrázek 8: Výsledky testu „Bach nebo počítač“ (Hadjeres a kol.)

Poslechové testy, ve kterých měli posluchači rozeznat skladby programu LEGRE od původních árií Giovanniho Legrenziho a dalších barokních skladatelů, provedli Mariateresa Storino a kol. (2007) v několika různých verzích. Cílem bylo zjistit, jestli LEGRE dokáže zachytit styl Legrenziho, proto nebylo účastníkům v žádné fázi explicitně sděleno, ať rozeznají počítač, ani jim nebylo řečeno, že nějaké skladby počítač vůbec složil. Účastníci kromě binárních odpovědí měli také určit na 7 bodové škále míru jistoty s danou odpovědí. Všechna hudba jim byla prezentována skrz nahrávky zpěvačky a jejího doprovodu, respektive skrz notový zápis u třetího experimentu. Nahrávky trvaly od 30 sekund do jedné minuty.

První tři experimenty, které se zaměřovaly na odlišení programu od skladeb Legrenziho, obsahovaly před samotným rozlišováním i fázi učení, ve které se účastníci seznámili s jeho původní hudbou poslechem a ve třetím experimentu analýzou tištěných not. Prvního experimentu se zúčastnilo 28 hudebníků ve věku od 30 to 45 let s vysokou odborností; buď se jednalo o profesionální hudebníky se zaměřením na Barokní hudbu nebo pokročilé studenty cembala nebo bassa

continua. Hudebníci byli rozděleni na dvě skupiny, první skupině bylo přehráno 10 skladeb od LEGRE a 10 skladeb od Legrenziho v párech, druhé skupině 20 skladeb v náhodném pořadí. Celkové výsledky prvního experimentu ukazují, že posluchači dokázali rozeznat skladby LEGRE od Legrenziho významně nad úroveň náhodného výběru ($p < 0,001$) s poznatkem, že první skupina, které byla prezentována skladba LEGRE a ihned poté skladba Legrenziho (a naopak) si vedla lépe (76,25 % správných odpovědí) oproti skupině druhé (60,69 % správných odpovědí).

Druhého experimentu se zúčastnilo 15 nehudebníků a 15 hudebníků, kteří neměli žádnou nebo malou zkušenost s barokní hudbou. Podmínky tohoto experimentu jinak byly stejné, jako u první skupiny předešlého experimentu. Zde dokázali hudebníci rozlišit obě entity s výsledkem 60,58 %, nehudebníci vybírali téměř náhodně s výsledkem 54,66 %. Storino a kol. u obou experimentů zmiňují, že žádná či slabá korelace mezi odpověďmi a hodnocením jistoty naznačuje, že nikdo z účastníků nevěděl, co přímo ovlivnilo jeho rozhodnutí a že tak bylo spíše intuitivní. V experimentu číslo tři využili autoři 9 účastníků s vysokou mírou odbornosti, kteří hráli minimálně 10 let na nějaký nástroj a zároveň studovali alespoň 8 let skladbu. Měli za úkol od sebe rozlišit LEGRE a Legrenziho stejně jako v případě druhého experimentu, nicméně skrz analýzu zapsaných not, což se jim celkově podařilo v 62,2 % případů, tedy s jen o něco lepším výsledkem, než uměli hudebníci v experimentu č. 2.

Poslední experiment porovnával hudbu LEGRE s Legrenzim a dalšími barokními skladateli, Luigim Rossim a Domenicem Gabriellim. Účastníky bylo v tomto případě 27 hudebníků, kteří měli 10leté formální hudební vzdělání, nicméně žádnou významnou odbornost v barokní hudbě. Těmto hudebníkům byly nejdříve v první fázi puštěny skladby od LEGRE, poté v testovací fázi jen 18 árií lidských skladatelů. Zde měli účastníci za úkol hodnotit, jestli daná árie pochází od skladatele z první fáze. Zde autorům vyšlo, že skladby Legrenziho účastníci hodnotili jako od programu LEGRE v 49,3 % případech a skladby Gabrielliho v 16 %, což by byl pro program sám o sobě dobrý výsledek, nicméně Rossiho skladby označili jako skladby od LEGRE účastníci v 51,2 % případů. Celkové výsledky hodnotí Storino a kol. tak, že jedině nehudebníci u druhého experimentu nebyli

schopni přesvědčivě rozeznat rozdíl mezi počítačem a lidským hudebníkem a nepovedlo se tak zachytit všechny stylové charakteristiky daného skladatele.

Pro evaluaci dvou modelů pro komponování klasické hudby v určitém stylu využili Collins a kol. (2016) *Rozlišovací test* Pearce a Wigginse (2001) v kombinaci s technikou konsenzuálního posuzování (Amabile, 1996). Dvě skupiny, které byly tvořeny 16 náhodnými posluchači klasické hudby a 16 experty mající formální hudební vzdělání se zaměřením na hudbu 19. století, měly za úkol rozlišit mazurky F. Chopina od hudby počítačových programů Racchman-Oct2010, jeho další verze Racchmaninof-Oct2010 vylepšené o využívání replikace hudebních struktur a EMI, zařazený jako měřítko úspěchu.

V první fázi byly účastníkům představeny dva úryvky Chopinových mazurek, u kterých měli verbálně popsat charakteristiky, které byly vlastní oběma skladbám. Po této části přišla hlavní fáze, ve které měli účastníci rozlišovat mezi originálními Chopinovými mazurkami, jinou hudbou lidských skladatelů (mazurky od Clary Schumann nebo Davida Kinga, ale i Chopinovy skladby v jiných stylech) a hudbou generovanou zmíněnými počítačovými programy. Posluchačům bylo představeno 32 MIDI úryvků o zhruba 15 taktech, u kterých kromě rozlišení jejich původu měli hodnotit do jaké míry jsou stylově blízké originálním mazurkám na škále od 1 do 7 a zároveň subjektivně určit jejich estetické kvality.

Kromě těchto poskytnutých informací účastníci zaškrtovali, jestli již danou skladbu slyšeli, každý úryvek pak mohli ohodnotit i slovně. Před touto částí měli posluchači možnost vyzkoušet si experiment na dvou úryvcích, z nichž jeden pocházel ze Chopinovy mazurky. Tuto mazurku ze zkušební fáze, společně s dvěma úryvky z první, učící fáze, si měli možnost posluchači přehrát kdykoliv během celého testu. Výsledky samotného rozlišovacího testu podle Collinse a kol. ukazují, že 8 z 16 posluchačů klasické hudby volilo lépe, než by byl náhodný výběr, stejně tak 15 z 16 expertů. Pokud se na výsledky podíváme detailněji (viz Obr. 9), tak vidíme, že je Collins neudává jako průměr správných odpovědí za jednu skupinu, ale vždy jednotlivě podle skladeb. Z těchto dat vidíme, že obě skupiny účastníků správně rozlišovalo jak mazurky Chopina, tak další skladby lidských autorů a programů Racchman-Oct2010 a Racchmaninof-Oct2010.

Jediným překvapením je tak EMI, jehož skladby byly klasifikovány správně jen v 14,6 % případů pro obě skupiny, většinou byly označovány jako originální Chopinovy mazurky. V průměru se také potvrdila odbornost expertů, kteří až na výjimky vybírali lépe než posluchači. Celkově se podle Collinse a kol. nepotvrdil rozdíl mezi jeho dvěma programy, jelikož si vedly podobně jak u odlišovacího testu, tak v testu jejich stylistického úspěchu a nedokázaly být úspěšné jako EMI.

Stimulus	Mean Style Success		Distinguish Correct (%)		Classed Ch. Mazurka (%)	
	C'Goers	Experts	C'Goers	Experts	C'Goers	Experts
Chopin Mazurkas						
1	4.56	5.38	31.3	68.8	31.3	68.8
2	5.63	5.56	68.8	81.3	68.8	81.3
3	4.13	4.38	37.5	43.8	37.5	43.8
4	5.75	5.75	81.3	81.3	81.3	81.3
5	5.13	5.82	56.3	81.3	56.3	81.3
6	4.19	4.88	43.8	62.5	43.8	62.5
Human Other (7 Clara, 8 Ch. Prelude, 9 Brahms, 10 Coupn, 11 Schnbg, 12 King)						
7	5.63	6.13	0.0	0.0	81.3	93.8
8	3.94	3.25	62.5	68.8	18.8	25.0
9	3.06	2.00	81.3	87.5	12.5	6.3
10	2.56	1.56	81.3	81.3	6.3	6.3
11	1.19	1.38	68.8	81.3	0.0	0.0
12	3.06	2.69	31.3	68.8	12.5	0.0
Computer Based: EMI						
13	4.75	5.88	25.0	6.3	43.8	87.5
14	5.38	5.13	12.5	25.0	62.5	56.3
15	5.19	4.88	18.8	18.8	50.0	50.0
16	5.25	5.50	12.5	25.0	56.3	75.0
17	5.75	6.00	6.3	6.3	87.5	87.5
18	5.25	5.63	12.5	6.3	75.0	68.8
Computer Based: System A						
19	3.25	3.31	62.5	56.3	18.0	6.3
20	4.75	4.38	25.0	62.5	56.3	25.0
21	2.81	2.69	75.0	81.3	6.3	0.0
22	2.75	2.38	50.0	68.8	18.8	0.0
23	2.75	2.63	62.5	87.5	6.3	0.0
24	3.13	3.19	68.8	93.8	18.8	0.0
Computer Based: System B						
25	2.00	1.81	75.0	81.3	0.0	0.0
26	2.94	2.69	75.0	68.8	0.0	0.0
27	2.25	2.75	68.8	50.0	12.5	12.5
Computer Based: System B*						
28	3.25	2.88	43.8	81.3	25.0	0.0
29	2.94	3.06	87.5	75.0	6.3	6.3
30	2.69	2.63	56.3	68.8	12.5	6.3
31	2.75	2.89	50.0	87.5	0.0	0.0
32	2.50	2.75	81.3	93.8	12.5	0.0

Obrázek 9: Výsledky testu Collinse a kol. (2016)

Rozlišovací test jako východisko pro empirickou evaluaci programu Eigenfeldta a Pasquieria (2010) založeném na Markovových řetězcích zmiňují Adam Burnett a

kol. (2012), kteří rozdělili posluchače na dvě skupiny: formálně vzdělané hudebníky nebo skladatele, kteří měli zkušenost s analýzou klasické hudby a laiky s minimální zkušeností s hudbou. Tito účastníci neměli zároveň na výběr mezi dvěma možnostmi, ale přiřazovali skladby počítači nebo skladateli na čtyřstupňové škále, díky které autoři měřili i míru jistoty s danou odpovědí. Touto stupnicí chtěli autoři zároveň vyřadit neutrální odpověď, aby účastníci volili svůj nejlepší odhad.

Celkově se experimentu zúčastnilo 87 lidí, kterým bylo prezentováno 10 skladeb počítače a 10 lidských skladatelů. Podobně jako u původního *Rozlišovacího testu* Pearce a Wiggins (2001)⁴ prezentuje Burnett a kol. výsledky (viz Obr. 10) nejen jako průměr správných odpovědí za jednotlivé skupiny, ale také odděleně u lidských a počítačem generovaných skladeb. Z těchto výsledků vyplývá, že účastníci byli schopni poznat lidskou hudbu významně lépe než při náhodném výběru, naopak u skladeb počítače se jejich výběr podobal spíše náhodnému. To potvrzuje i větší míra jistoty u skladeb člověka než u hudby počítače. Rozdíly mezi skupinami hudebníků a laiků se celkově neukázaly jako významné.

<i>Experience</i>	<i>mean</i>	<i>t-score</i>	<i>p</i>	<i>df</i>
musicians (H)	6.550 (1.82)	3.808	0.0012	19
musicians (C)	5.300 (1.92)	0.698	0.4936	19
nonmusicians (H)	6.477 (1.51)	8.003	< .0001	66
nonmusicians (C)	5.089 (1.99)	0.369	0.7136	66

Obrázek 10: Výsledky testu Burnetta a kol. (2012)

Skladbám generovaným systémem v tomto experimentu nicméně chybí několik důležitých charakteristik, které tvoří hudbu jako takovou. Mezi tyto charakteristiky patří především melodické postupy a komplexnější rytmus, program by se tak dal nejlépe popsat jako generátor harmonických postupů, který tvoří štaktové skladbičky. Aby byl porovnatelný s lidskou hudbu, byly části skladeb lidských autorů z období Romantismu (Dvořák, Chopin, Brahms ad.) vypreparovány, aby

⁴ V této práci není prezentován detailněji, jelikož se zabývá zkoumáním programu, který generuje hudbu ve stylu *Drum & Bass*.

z nich vymizela melodie a zbyla jen basová linka a příslušný akord. Mezi další limity této studie, který nicméně autoři přiznávají, patří také odlišný počet hudebníků (19) a laiků (66), což může vést k nepřesným výsledkům při jejich porovnávání.

4.4. Poslechové testy a další metody

Zmíněné přístupy vycházející především z Turingova testu nebo Rozlišovacího testu nejsou jediným způsobem, jak autoři přistupují k evaluaci svých programů. Jednou z metod je již popsané statistické porovnávání výstupů se vzorovými skladbami, které využili například Triviño a Morales (2001) nebo Alex Chilvers a Menno van Zaanen (2008), kteří použili metodu křížové validace k otestování výstupů svého programu na vzorku 230 Bachových chorálů.

Tuto metodu k evaluaci výsledků svého programu využili i Richmond a Rahal (2018), kteří k ověření úspěchu provedli i test s pěti experty, kteří měli za úkol generované skladby rozřadit podle hudebního stylu (klasicismus, baroko ad.) a na stupnici od 1 do 5 určit jistotu se svou odpovědí. Podobně jako další autoři, i Richmond a Rahal posluchačům umožnili komentovat hodnocení slovně. Stejně i Cheng-Zhi Anna Huang a kol. (2017) zkombinoval statistickou a lidskou evaluaci, v obou případech založené na hodnocení blízkosti skladeb dvou modelů a J. S. Bacha.

Anna K. Yanchenko a Sayan Mukherjee (2017) vyhodnocovali skladby svých modelů založených na skrytých Markovových řetězcích skládajících klavírní skladby Romantismu třema číselně vyjádřenými ukazateli, kterými jsou: a) originalita, respektive hudební předvídatelnost b) muzikalita, založená na podobném „objemu“ disonance mezi novým dílem a původními skladbami, stejně jako podobné distribuci notových výšek c) hudební struktura, tedy melodie a harmonie. Skladby, které si v těchto třech oblastech vedly nejlépe, byly následně evaluovány slovně a na škále od 1 do 3 podle oblíbenosti dvě skupinami, hudebníky a nehudebníky. Podle autorů z tohoto hodnocení vychází, že použité modely neumí zajistit dostatečný melodický vývoj ani celkovou strukturu, což je evidentní jak z posluchačské, tak číselné evaluace.

Pearce a Wiggins (2007) ve své pozdější práci nevyužili pro evaluaci *Rozlišovacího testu*, ale metodu vycházející z techniky konsenzuálního posuzování (viz Pearce, 2005) k hodnocení stylistického úspěchu tří modelů založených na Markovových řetězcích 16 experty, vyučujícími a studenty z londýnských City University, Goldsmiths, University of London a Royal College of Music. Tito experti hodnotili na škále od 1 do 7 generované a původní Bachovy melodie vycházející z luteránských hymnů v kontextu shody s důležitými stylistickými rysy, intervalové struktury, melodické formy a dalších charakteristik. Z výsledků (viz Obr. 11) vychází, že generované melodie byly v průměru hodnoceny hůře, než originální Bachovy, navíc ani jeden z modelů není schopen konzistentního generování, systém B se zvláště špatnými výsledky.

Base	System A	System B	System C	Original	Mean
249	2.56	2.44	5.00	6.44	4.11
238	3.31	2.94	3.19	5.31	3.69
365	2.69	1.69	2.50	6.25	3.28
264	1.75	2.00	2.38	6.00	3.03
44	4.25	4.38	4.00	6.12	4.69
141	3.38	2.12	3.19	5.50	3.55
147	2.38	1.88	1.94	6.50	3.17
Mean	2.90	2.49	3.17	6.02	3.65

Obrázek 11: Výsledky testu Pearce a Wiggins (2007)

Tímto však test pro Pearce a Wiggins neskončil, experti kromě číselného hodnocení melodií využívali i slovní komentáře, které autoři studie použili k specifikování pěti hlavních oblastí, ve kterých se generované melodie odlišují od Bachových: rozsah, melodická a tonální struktura, struktura frází a rytmus. Tyto oblasti byly poté u výstupů jednotlivých modelů kvantitativně analyzovány a poté z nich byly vyvozeny kritéria, na jejichž základě byl v původním testu nejúspěšnější model C vylepšen na model D, který by si podle autorů vedl v novém experimentu až srovnatelně s Bachovými skladbami.

Existují také jen čistě poslechové testy, u kterých posluchači hodnotí estetickou kvalitu programu, jako například v práci Allena Huanga a Raymonda Wu (2016), kde 26 dobrovolníků s nespécifikovanou expertízou hodnotilo výstupy tří programů na stupnici od 1 do 10 (1 představovalo náhodné dílo, 5 hudebně věrohodné a 10

dílo skladatele). Identickou metodu zvolil i Jan Staněk (2017) k evaluaci pěti skladeb své neuronové sítě na 38 lidech, u kterých se také nezabýval jejich odborností.

Kromě pokusů o kvantitativní, měřitelné hodnocení programů můžeme u dalších autorů vyzorovat spíše tendenci ke kvalitativnímu hodnocení, ke kterému patří expertní evaluace a analýzy skladeb. Tento způsob nicméně není nikdy centrem zveřejněných prací, spíše je mu věnováno několik vět v závěrech a limitech programů a jen výjimečně je takovéto analýze věnováno více prostoru.

Například Hild a kol. (1992) píše jen o tom, že jejich HARMONET posoudili hudební profesionálové jako „improvizujícího varhanistu“ (Hild a kol, s. 272), stejně jako Hörnel a Ragg (1996), kteří se zmiňují o „vysoké kvalitě“. Podobně jen pár poznámek od expertů zveřejnili Kamil Adiloglu a Ferda N. Alpaslan (2007), Palle Dahlstedt (2007) nebo Francisco Pereira a kol. (1997), o něco delší analýzu například Dorien Herremans a Elaine Chew (2010), Nikhil Kotecha (2018) nebo Wiggins (1998), který využil vyučujícího z hudební fakulty Univerzity v Edinburghu, aby známkoval program podle stejných měřítek, jako studenta skladby prvního ročníku.

Není nám známá žádná hlubší muzikologická analýza hudebních děl počítače, kterou by inicioval a zveřejnil autor programu. Steven Jan (2018) nicméně ve vlastní práci, kde zkoumal hudební analýzu strojem generované hudby a rozdílné přístupy vůči lidským skladbám, detailně zpracoval klavírní dílo *Colossus*, které v roce 2010 složil Iamus.

K tomuto expertnímu hodnocení se dá řadit i odborná recenze, jako způsob na autorovi programu nezávislé a jím nepublikované evaluace skladeb, které již vstoupily do provozovací praxe – ať už při živém koncertu nebo na nahrávce. Takto byly předmětem recenzí skladby většinou pokročilejších programů, jejichž díla se dostala do koncertního provozu, jako jsou EMI (Silva, 2003; Boden, 2010) nebo Iamus (Ball, 2012), v českém prostředí například AIVA (Žáček, 2019).

4.5. Kritika použitých metod

Jak již bylo řečeno, ani původní *Imitační hra* není Turingem nijak detailně popsána, proto není divu, že Turingův test autoři jeho hudební verze nereplikují doslovně a slouží jim spíše jako inspirace. S tím, že médiem Turingova testu je hudba, také vyvstává několik nových metodologických otázek, které je třeba vyřešit. Je to například téma hudebních ukázek, které jsou v poslechových testech prezentovány. Pokud je nám známo, tak jen David Cope využil živé provedení skladeb profesionálním klavíristou, a jen zřídka autoři nechali hudbu pro potřeby testu nahrát (viz Browne a Fox, 2009 a Storino a kol., 2007). Nejčastěji využívají autoři skladby ve formátu MIDI, což se dá interpretovat jako potřeba po co největší uniformitě prezentované hudby (Collins a kol., 2016), nicméně je to záměr diskutabilní, jelikož touto redukcí hudba ztrácí svůj obsah. Sama totiž není tvořena jen frekvencí ale také dynamikou, barvou a samotnou interpretací umělce.

Problematické jsou ale především tendence ke zjednodušování hudby, jako je například omezování hudebních ukázek na 4/4 tempo (viz Quick, 2014; Burnett a kol., 2012) a jednu tóninu (viz Schulze, 2009), případně jejich laické úpravy (Rajon, 2018). Časté je také jejich zkracování, kdy se naprostá většina hudebních ukázek vejde do škály od 10 do 60 vteřin. Tato omezování by byla v klasickém Turingově testu těžko představitelná, například zákazy používání souvětí nebo komunikace omezená jen na několik desítek vteřin by v testu v akademické literatuře pravděpodobně neprošly. Je to bezesporu prostředkem, jak vůbec nějaký test s nedokonalými programy provést, nicméně mluvit poté o úspěchu programu či dokonce o zvládnutí Turingova testu, není ničím jiným než zkreslováním skutečnosti.

Hudba jako médium Turingova testu je příčinou další významné odchylky od původního testu. V klasickém testu je třeba, aby použitému jazyku zapojené osoby rozuměly a uměly číst, nicméně není pro ně třeba žádného jiného formálního vzdělání. Protože ale hudba není nositelem žádného objektivního významu, je nutné, aby v testu figurovaly osoby, které pro ni mají minimálně přirozený či vytrénovaný cit a dokáží sluchem dostatečně formálně analyzovat její strukturu. Laici jsou sice schopni intuitivního posouzení kreativních artefaktů, nicméně

experti jsou vždy v lepší pozici k vysvětlení svých postojů, jejich soud tak má větší váhu a je více informativní (Boden, 1994).

Autor testu	Program	Lidský skladatel	Test	Zdroj hudby	Délka jednotlivých skladeb	Hodnocení experty	Hlášen úspěch programu
Cope (1996, 2001); Mostow a Rich (1998)	EMI	Mozart, Cope, Bach	MOtT	neuveдено	neuveдено	ne	ano
Douglas Hofstadter (Silva, 2003)	EMI	Bach, Chopin	MOtT	živé provedení	neuveдено	ano	jen u skladeb Chopina
BBC (Bloomfield a Vurdubakis, 2003)	EMI	neuveдено	MOtT	neuveдено	neuveдено	ne	ano
Moffat a Kelly (2006)	EMI, Illiac	Bach, Šostakovič	MOtT	nahrávka	60 sekund	ano	jen u skupiny expertů
Triviño-Rodríguez a Morales-Bueno (2001)	MPSG	Bach	MOtT	MIDI	10 taktů	neuveдено	ano
Liang (2016)	BachBot	Bach	MOtT	MIDI	15 - 35 sekund	ano	ano
Cowart a kol. (2019)	BachBot	Bach	MOtT	MIDI	15 - 35 sekund	ano	neuveдено
Rajon (2018)	LearnerDeep	Bach	MOtT	MIDI	11 - 15 sekund	neuveдено	částečně ano
Quick (2014)	Kulitta, L-systém	Bach	MOtT s Likertovou škálou	MIDI	10 vteřin	ano	částečně ano
Steve Larson (Kurzweil, 1999)	EMI	Bach, Larson	neúplný MDtT	neuveдено	neuveдено	neuveдено	ano
Schulze (2009)	SuperWillow	Schulze, Burger, Dalton	neúplný MDtT	MIDI	max. 8 taktů	ne	ne
Browne a Fox (2009)	Evoluční algoritmus	Gvirtzman	neúplný MDtT	nahrávka	32 taktů	ano	ano
Hadjeres a kol. (2017)	DeepBach, MLP, MaxEnt	Bach	rozlišovací test	MIDI	12 taktů	ano	jen u programu DeepBach
Storino a kol. (2007)	LEGRE	Legrenzi, Rossi, Gabrielli	rozlišovací test	živé provedení, noty	30 - 60 sekund	ano	ne, částečně jen u laiků
Collins (2016)	Racchman-Oct2010, Racchmanino f-Oct2010, EMI	Chopin, C. Schumann, King a další	rozlišovací test s Likertovou škálou	MIDI	15 taktů	ano	jen u EMI
Eigenfeldt a Pasquiera (Burnett a kol., 2012)	Markovovy řetězce	Dvořák, Chopin, Brahms a další	rozlišovací test se škálou jistoty	MIDI	8 taktů	ano	částečně ano

Obrázek 12: Srovnání provedených testů vycházejících z Imitační hry (autor)

V hudebních verzích Turingova testu jsou v naprosté většině přítomny skupiny expertů a někdy i kontrolní skupina laiků, problém je však často s nedostatečnou definicí toho, kdo je přesně expert (jestli se jedná „jen“ o hudebníka nebo o hudebního teoretika, který se zabývá analýzou skladeb) a ještě častěji i se

zdůvodněním výběru té které expertní skupiny. Například student jazzové hudby, ať již skladatel nebo hudebník, nemusí o klasické hudbě vědět skoro nic, a stejně ho považují Moffat a Kelly (2006) za experta. Výjimkou jsou Storino a kol. (2007) a Collins a kol. (2016), kteří dokonce rozlišují i mezi různými disciplínami a využívají experty zaměřujících se konkrétně na hudbu baroka, respektive 19. století.

U expertů se nicméně může v rámci testu stát, že poznají přehrávanou skladbu. Nedají se (a neměly by se) tedy tímto způsobem nikdy zařazovat geniální skladby známých skladatelů, ale spíše ty méně známé či nové skladby skladatelů současných. To nicméně autoři reflektují a zveřejňují i názvy použitých skladeb, což se nicméně nedá říct o zveřejňování skladeb svých programů.

Některé testy se metodologicky vzdalují od Turingova testu například tím, že autoři využívají více entit než jen dvou, nebo že posluchači nemají za úkol rozlišovat, která z entit je člověkem a jaká počítačem, ale spíše hodnotí stylové podobnosti skladeb a poté je zařazují. Binaritu Turingova testu vědomě narušuje také Donna Quick (2014), která dává posluchačům možnost na Likertově škále hodnotit i míru jejich jistoty. Další početná skupina testů je ke své škodě naprosto nedostatečně popsána, kdy autoři poskytují o výsledcích testu neúplná data, nedefinují svá metodologická východiska a postupy. V horším případě nejsou vědecky pojaté a jejich popis je jen anekdotický.

Zásadní otázkou při evaluaci programů skrz Turingovy testy a testy z něj vycházející, je míra autonomie těchto programů. Nemáme na mysli filozofickou debatu nastíněnou v minulé kapitole, tedy jestli autor programu je i autorem hudby, kterou tento program vyprodukuje, ale míru zapojení programátora do samotného procesu generování hudby. U těchto programů se jen zřídka dozvídáme, jak přesně byly skladby složeny, tyto procesy jsou obvykle nedostatečně popsány nebo úplně skryté. Například podle Pacheta (2006) není dostatečně objasněna interakce programu EMI a Davida Copea. Sám Cope (2001) nicméně přiznává, že do děl EMI někdy zasahovat musí a například i album *Bach by Design* z roku 1994 s počítačem generovanými skladbami je vydáváno pod jménem jeho i EMI. I to považuje Ariza (2009) jako důkaz toho, že výše zmíněné testy EMI jsou spíše testy EMI a zároveň Davida Copea. Nedostatečná autonomie programů znamená, že současné testy

nejdou nikdy “soubojem” počítače a člověka, ale lidí, kteří jen užívají různé nástroje (Ariza, 2009).

Celkově je externí, objektivní evaluace něčím, co autoři programů provádějí jen ve zlomku případů. V drtivé většině se autoři nezabývají evaluací vůbec nebo jí věnují jen několik vět, případně nezmiňují objektivní kritéria pro úspěch programu (Pearce a Wiggins, 2001). Dalším problémem je, že hudební artefakty neposuzují opravdoví experti, ale jen autoři programů subjektivním hodnocením (Papadopoulos a Wiggins, 1999). Neschopnost dostatečné evaluace vychází dle Pearce, Meredith a Wiggins (2002) v první řadě z absence definování cílů a motivace výzkumu. Také Jourdanous (2011) dochází k tomu, že celková evaluace kreativních systémů obecně není reflektována systematicky.

5. Diskuze

K hodnocení hudebních skladeb a programů, které je generují, se od začátku jejich vzniku využívaly obdoby Turingova testu, ve kterých není médiem testu jako primární komunikační nástroj text, nýbrž hudba. Tato změna přináší několik problémů. Hudba jako taková kromě své struktury nenesení oproti textu žádný objektivní význam. V původním Turingově testu člověk rozhoduje, jestli je text a konverzace smysluplná a jestli jej tedy mohl napsat jiný člověk, lidé mluvící určitým jazykem se na smysluplnosti nebo náhodnosti daného textu mohou relativně dobře shodnout. Naopak v hudebním testu je podle Arizy (2009) největší rozdíl v tom, že různí posluchači, ať už poučení nebo laici, poskytují velmi rozlišné hudební soudy, a z člověka, který má určovat, co je dílo počítače a co člověka, se stává primárně kritik uměleckého díla. Test, který původně zkoumá přítomnost vědomí, se tímto transformuje na jakousi „*posluchačskou anketu*“ (Ariza, s. 49).

Do testování se dostávají všelijaké předsudky, například o tom, jak asi zní hudba generovaná počítačem nebo hudba lidská. Posluchač ale nemá při hudebním Turingově testu nárok na tázání doplňujících otázek, opravdová nehudební interakce mezi jednotlivými entitami a posluchači neexistuje. Pokud je použita hudba v testu jediné měřítko pro úspěch, pak je tento experiment více než původní test náchylný k tomu, aby byl testovaný člověk oklamán. Tím jsou hudební obdoby Turingova testu neúměrně jednodušší. Hudební Turingův test také podporuje jen čistou imitaci stylu (Colton & Pease, 2011) tím, že odměňuje programy, které spíše, než by vytvářely nové artefakty v rámci transformační kreativity, napodobují lidské chování (Ritchie, 2007), což je sice v zásadě nekreativní, ale v souladu s původními východisky *Imitační hry*.

I přes všechny tyto výhrady ale mají podobné testy, pokud jsou dostatečně metodologicky pojaté, „*výhodu přinášet empirické, kvantitativní výsledky, které mohou být hodnoceny intersubjektivně*“ (Pearce a Wiggins, 2007, s. 2).

5.1. Návrh

Pokud bychom pro objektivní evaluaci chtěli přímo následovat koncept Turingova testu a aplikovat ho na hudbu, je třeba se s těmito námitkami vyrovnat. Hudební

Turingův test, v jehož rámci chceme odpovídat na otázku, jestli je daný stroj schopný myslet, musí alespoň částečně testovat jeho autonomii. Klasická hudební verze testu testuje prakticky jen výstup v podobě hudby, tímto způsobem je tak možné testovat i systémy úplně závislé na programátorovi / skladateli a člověk pak může být oklamán i programem velmi primitivním. Testem tak může projít i třeba hudba ze hry *Musikalisches Würfelspiel*.

Podstatu Turingova testu oproti *Musical Output toy Test*, u kterého jde jen o čistě poslechový test, lépe vystihuje *Musical Directive toy Test*, při kterém dochází k interakci mezi počítačem a člověkem v reálném čase. U tohoto testu je zároveň hudba zadávána žijícímu lidskému skladateli, čímž se dá vypořádat s genialitou mimořádných skladatelů minulosti a výběrem hudbu pro test, kterou mohou posluchači znát již předem. Zatím provedené testy v oblasti skládání klasické hudby se však opravdovému MDtT nepřibližují, proto je třeba, aby tato interakce a zadávání samotného úkolu probíhali opravdu mezi posluchačem a počítačem, respektive člověkem, a ne aby hudbu zadával počítači i člověku sám autor daného programu.

Samotného testu by se měli na straně hodnotitelů účastnit opravdoví experti, tedy zkušení hudební teoretici a skladatelé, jejichž odbornost musí být dostatečně popsána. Zároveň by autoři programů měli zdůvodnit výběr konkrétních skupin a v ideálním případě, pokud program generuje skladby v určitém stylu, přihlédnout ve výběru expertů i k jejich specializaci. V případě, že je použit panel těchto opravdových odborníků, není nutně třeba velikého počtu osob, nicméně pokud je test prováděn webovou formou, musí být odbornost účastníků, potažmo jejich identita, alespoň ověřitelná.

Autor programu by měl definovat omezení pro to, jaké skladby mohou a budou v testu použity. To jest, jestli bude zkoumán určitý styl nebo forma (symfonie, smyčcový kvartet apod.), každopádně by tato omezení neměla vést k redukování hudby a zjednodušování samotného testu tak, aby musely být výsledné skladby jen v tónině C dur nebo 4/4 taktu. Tato omezení nicméně mohou být navržena hodnotiteli, kteří mohou zadávat daným entitám úkoly, jako je právě tónina, celková délka, forma, kontext ad.

Důležitou podmínkou reflektovanou již autorem daného programu by měla být dostatečně dlouhá délka úryvku, která by měla být významně delší, než je jen 10 až 60 sekund, aby hodnotitel mohl posoudit její celkovou formu ve větším kontextu a to, jestli prezentovaná hudba vůbec někam směřuje. Při následné prezentaci hudby hodnotitelům by podmínka nezjednodušování měla být dodržena také. Hodnotitelům by měla být hudba prezentovaná v plnohodnotné formě buď jako živé provedení s umělci, popřípadě jako záznam hudebního výkonu. V ideálním případě by měli hodnotitelé k dispozici také kompletní notový materiál.

Samozřejmostí musí být dostatečné popsání cílů testu, motivace autora pro vývoj samotného programu, určení kritérií pro úspěch a role autora při samotném generování hudby. Zároveň by měla být v akademické práci prezentována kompletní data z průběhu testu, stejně jako i zkomponovaná hudba obou entit.

I přes tyto podmínky je zde nebezpečí, že bychom se stále nedostatečně nevyořadali s problémem autonomie a podle vzoru Turingova testu neodpovídali na jeho prvotní otázku, tedy jestli stroj myslí, ale jen testovali určitý výsek lidské činnosti. Proto by tento test byl stále označen jen jako derivát Turingova testu, dle Stevana Harnada a Christophera Arizy „*toy model*“, respektive „*toy test*“. Jak už bylo řečeno, původní test naznačuje vztah mezi schopností vést smysluplnou konverzaci a myšlením, z tohoto důvodu by bylo nutné do hudebního Turingova testu implementovat i původní textový Turingův test.

V tomto testu by dvě entity, počítač a člověk, komunikovaly s hodnotitelem pomocí textového rozhraní. Hodnotitel by s oběma vedl plnohodnotnou konverzaci, která by směřovala k zadání skladby. Poté, co by jim byl poskytnut dostatečný čas ke kompozici, poskytly by hudbu v daném formátu hodnotiteli, který by skladby analyzoval. Zároveň by ale mohl pokládat jednotlivým skladatelům doplňující otázky, popřípadě by po nich chtěl jejich interpretaci hudebního díla nebo anotaci, která se standardně používá například do programů při koncertních provedeních. Na základě estetických měřítek a schopnosti skladatelské entity svoji hudbu obhájit, stejně jako vést konverzaci, by hodnotitel rozhodl, která z nich je počítačem a která člověkem.

Oproti klasickému Turingově testu by tento kombinovaný test musel trvat delší dobu díky tomu, že čas na složení skladby musí být dán oběma entitám v lidském měřítku. Samotný poslech a analýza skladby by také zabrali minimálně několik desítek minut. V ideálním případě, kdy každý hodnotitel zadává právě jednu skladbu, se časová obtížnost zvyšuje s každým hodnotitelem. Těžší jsou i celkové podmínky pro úspěch počítače, který nejen že musí být schopný generovat poměrně kvalitní skladby, ale musí být schopný projít i klasickým Turingovým testem. Dle Harnadovy škály se tak dá tento test zařadit někde mezi Turingův test a Totální Turingův test.

Tento hudební Turingův test se původnímu testu přibližuje nejvíce, nicméně odpovídá na otázku, jestli je stroj schopen myšlení, znovu jen skrz klasický Turingův test (Ariza, 2009). Pokud by takový program neuměl komponovat kvalitní klasickou hudbu, ale komunikoval na úrovni člověka, stejně by testem mohl projít. Skladatelské schopnosti tak v tomto testu nejsou důležité, pokud však nezměníme naši počáteční otázku na vhodnější „*Mohou stroje skládat hudbu?*“. Poté je možné změnit médium Imitační hry z textu čistě jen na hudbu, Turingův test však ztrácí svoji esenci. I sám Alan Turing (1950, s. 434) píše o tom, že hodnotitel „*nesmí požadovat praktickou demonstraci.*“ Tento test nicméně v čisté podobě postrádá schopnost informovat o objektivních i subjektivních vlastnostech generovaných skladeb, jejichž analýza může být důležitá jak pro autory programů, tak pro akademickou veřejnost. Proto je vhodnější hudební test dále transformovat a původní Turingův test brát jen jako nepřímou inspiraci.

5.2. Alternativní test

Alternativou může být test, který po vzoru *Rozlišovacího testu* Pearce a Wiggins (2001) opouští implikace Turingova testu ohledně myšlení strojů i jeho omezující design a který nás do jisté míry také zbavuje problémů způsobených tím, že jen zaměníme médium testu z textu na hudbu. Protože odpovídáme zpočátku na jinou otázku, tedy jestli mohou stroje skládat hudbu, není nutná interakce mezi počítačem a člověkem. I proto se dá do testu zahrnout i předešlá tvorba počítače a člověka, nejen tvorba nová či nově vzniklá na objednávku jak hodnotitele, nebo programátora. Pokud uvažujeme takový test, který po počítači nežádá důkaz o jeho

myšlení, je nicméně stále potřeba reflektovat otázku jeho autonomie. Jeho autor by měl detailně popsat, jak přesně program skládá hudbu, jaká je jeho motivace pro vytvoření takového programu a pokud opravdu má cíl vytvořit nezávislý program, tak jaká je role autora na výsledném artefaktu, tzn. jak probíhá zadání, výběr skladeb a zásahy do kompozičního procesu v jakékoliv fázi.

Obecně se testu může zúčastnit více entit (počítačů, skladatelů) s více skladbami. Skladby v náhodném pořadí pak od sebe odlišují odborníci v dané oblasti, kteří se nezávisle na sobě musejí shodnout na tom, jestli je daný výstup kreativní a jeho autor je tak počítač, nebo člověk. V ideálním případě by experti byli seznámeni s tím, co jsou schopny v současné době počítačové programy generovat za hudbu a s jakými problémy se potýkají. Stejně jako v případě upraveného hudebního Turingova testu by použitá hudba měla být co nejméně zjednodušována, expertům by tedy měla být prezentována ve formě nahrávek nebo živých vystoupeních a notového materiálu. Skladby by měly být také co nejméně zkracovány, aby se měli hodnotitelé možnost seznámit s jejich globální strukturou.

Experti v dané oblasti hudbu dále zkoumají a oproti Turingovu testu poskytují i kvalitativní analýzu, obecně zohledňující kvalitu formální struktury a estetickou hodnotu. Tato analýza může být také formálně rozdělena do více sekcí, jako můžou být např. melodická, harmonická a rytmická struktura, instrumentace apod., které experti nejen slovně popisují, ale i známkují. Tato část testu může sloužit jako zdůvodnění jejich rozhodnutí o původu hudby, ukazatel výkonu mezi jednotlivými skladbami a entitami, nebo i jako zdroj odborných stanovisek k dalšímu vylepšování programu.

Tato kvalitativní část testu má oproti jakémukoliv experimentu vycházejícímu striktně z Turingova testu tu výhodu, že není omezena jen na binární přijetí či odmítnutí dané skladby, které navíc není zdůvodněné a my se tak nedovídáme, na jakém základě daný expert rozhodl tak, jak rozhodl. Rigorózní hudební analýza umí poskytnout informace nejen o estetické hodnotě obecně přijímané v odborném diskurzu, ale především o organizaci hudebního materiálu, tedy popsat konzistenci a kontinuitu uvnitř skladby i mezi skladbami z jednoho zdroje. Jak si všímá Jan (2018), odborná diskuze se spíše zaměřuje na generování a estetickou evaluaci

počítačové hudby než na hudební analýzu těchto skladeb a její specifika oproti lidským skladbám. Tato analýza založená na hudební teorii může být cesta, kterou se může evaluace programů více přiblížit hudební praxi, kdy je program skrz tento pohled viděn spíše jako student skladby, který se zlepšuje díky zpětné vazbě od svého učitele.

Závěr

Tato práce se zabývala umělou inteligencí, určenou k nezávislému skládání klasické hudby a možnostem její vnější, objektivní evaluace. Popsali jsme možné metody hodnocení kreativních systémů, které se objevují v odborné literatuře. Nejvíce nás zajímaly ty, které se dočkaly praktického využití, tedy těch, které testují především výstup daného programu v podobě hudby a zároveň vycházejí z konceptu Turingova testu, který proti sobě v pomyslném souboji staví lidskou a umělou entitu. Mezi tyto testy patří především Musical Output Toy Test, jako test bez interakce mezi posluchačem a danými entitami, Musical Directive Toy Test, u kterého je hudba zadávána jako úkol a Rozlišovací test, který se vyhýbá filozofickým implikacím původního Turingova testu a jde u něj především o odlišení zdroje hudby. Naším záměrem bylo v primárních zdrojích nejdříve najít a posléze poskytnout přehled a výsledky provedených experimentů, které na tyto poslechové testy přímo nebo nepřímo odkazují. Zároveň jsme si také vytyčili za cíl testy srovnat, odhalit jejich nedostatky a navrhnout možné změny.

Přehled vynalezených programů v této oblasti se neobešel bez konstatování, že velká většina autorů těchto programů nedostatečně definuje cíle vlastního výzkumu, a tudíž nepopisují ani způsoby, jak jich dosáhnout. Obecně se autoři téměř neuchylují k vnějšímu, nezávislému hodnocení svých programů a buď se ve svých textech otázkám evaluace nevěnují vůbec, nebo jen v několika málo větách popíší úspěch svého programu sami.

V rámci této práce jsme poskytli výčet prací těch autorů, kteří se pokusili o objektivní evaluaci programů skládajících hudbu skrz koncepty založené na *Imitační hře*. Zde jsme zjistili, že velká část autorů u těchto experimentů nedostatečně popisuje metodologii, východiska, ani samotné výsledky testů. Co mají společného je to, že v drtivé většině hlásí úspěch či částečný úspěch, nicméně hodnotu těchto výsledků devaluje fakt, že se často nedovídáme nic o roli autora programu na finálním artefaktu, a tak není zcela jasné, do jaké míry je testován program či jeho autor.

Oproti reálně provedeným testům jsme navrhli, aby byla hudba prezentovaná posluchačům co nejbližší její přirozené podobě, tedy jako živé provedení či jako nahrávka, ne jako syntetizovaná hudba ve formátu MIDI. Zároveň by měly být testované úseky několikanásobně delší oproti těm u provedených testů, kde je nejdelším úsekem pro skladbu 60 sekund. Díky tomu mohou posluchači posoudit i globální hierarchické vztahy, jelikož u generované hudby je nebezpečí, že sice dokáže kvalitně napodobit lokální strukturu, ale již ne delší hudební kontext. Dalším diskutovaným tématem je role expertů, kteří programy hodnotí. Musí být dostatečně popsána jejich odbornost, ale také zdůvodněn výběr dané skupiny v rámci rozmanitých hudebních profesí a formálních vzdělání.

Hlavní výtkou směrem ke konceptu hudebního Turingova testu je argument Christophera Arizy (2009), který tvrdí, že změnou média z textu na hudbu ztrácí test výpovědní hodnotu ohledně schopnosti myšlení – pokud ji tedy vůbec někdy měl. Test, který by sestával jak z komponování hudby, tak textové debaty mezi hodnotitelem a entitami, by byl nicméně pro současné programy nejen neúměrně obtížnější, ale zároveň by jím mohl projít i program, který hudbu skládat vůbec neumí. I přesto, že se ukázalo, že Turingův test je pro autory programů jen inspirací a nikdo z těch, kteří hlásili úspěch, explicitně nemluví o tom, že by měl jejich program schopnost myšlení, je jeho koncept z filozofického pohledu problematický.

Z tohoto důvodu jsme došli k alternativnímu testu, který spíše nabízí odpověď na otázku „Umí počítač skládat hudbu nerozlišitelnou od lidské?“ a následně hledá východisko z binární odpovědi skrz kvalitativní část, ve které je hudba odborně analyzována. Rozsah této práce nedovolil se tímto tématem zabývat hlouběji, nicméně zaměření se na tuto hudebně-teoretickou analýzu a její specifika při rozboru skladeb generovaných počítačem by mohla být středem zájmu další návazné práce.

Seznam použité literatury

ADILOGLU, Kamil a ALPASLAN, Ferda N., 2007. *A machine learning approach to two-voice counterpoint composition*. Knowledge-Based Systems, **20**(3), 300-309.

AGRES, Kat, FORTH, Jamie a WIGGINS, Geraint, 2015. *Evaluation of musical metacreation and musical creativity*. ACM Comput. in Ent. **100**(10). DOI: <http://dx.doi.org/10.1145/0000000.0000000>.

ALLAN, Moray, 2002. *Harmonising chorales in the style of Johann Sebastian Bach*. Diplomová práce, University of Edinburgh.

ALLAN, Moray a WILLIAMS, Christopher, 2004. *Harmonising Chorales by Probabilistic Inference*. In: Proceedings of the 17th International Conference on Neural Information Processing Systems, December 2004, 25-32.

AMABILE, Teresa, 1996. *Creativity in Context*. Boulder, Colorado: Westview Press.

ANDERS, Torsten a MIRANDA, Eduardo R., 2009. *A computational model that generalises Schoenberg's guidelines for favourable chord progressions*. In: Proceedings of the Sound and Music Computing Conference 2009.

AMES, Charles, 1987. *Automated Composition in Retrospect: 1956-1986*. Leonardo, **20**(2), Special Issue: Visual Art, Sound, Music and Technology (1987), 169-185.

ARIZA, Christopher, 2009. *The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems*. Computer Music Journal. (33). Massachusetts Technology. 48-70.

ARIZA, Christopher, 2011. *Two pioneering projects from the early history of computer-aided algorithmic composition*. Computer Music Journal, **35**(3), 40-56.

AUCOUTURIER, Jean-Julien a PACHET, Francois, 2003. *Representing Musical Genre: A State of the Art*. Journal of New Music Research, **32**(1), 83-93.

BAČA, Trevor a OBERHOLTZER, Josiah Wolf. *Abjad. Mozart: Musikalisches Würfelspiel* [online]. [cit 2020-07-01]. Dostupné z: http://abjad.mbrsi.org/literature_examples/mozart.html.

BALL, Philip. 2012. *Algorithmic rapture*. Nature. Macmillan Publishers Limited. č. 488.

BARTHES, Roland, 2006. *Smrt autora*. Aluze: časopis pro literaturu, filosofii a jiné. Olomouc: Univerzita Palackého, **10**(3).

BELLEGARD, Matthew I. a TSANG, Chi Ping, 1992. *Harmonizing music using a network of Boltzmann machines*. In: Proceedings of the Annual Conference of Artificial Neural Networks and their Applications, France, 321-332.

BELLINGER, Earl, 2011. *Little Ludwig, an evolutionary learning machine for musical composition*. J. ACM 1, 1, Article 1 (May 2011).

BENNETT, Andrew, 2005. *The author*. London: Routledge. ISBN 0-415-28163-6.

BIYIKOGLU, Kaan M., 2003. *A Markov model for chorale harmonization*. In: Proceedings of the Triennial ESCOM Conference.

BLITSTEIN, Ryan, 2010. *Triumph of the cyborg composer*. Miller-McCune Magazine.

BLOOMFIELD, Brian P. a VURDUBAKIS, Theo, 2003. *Imitation Games: Turing, Menard, Van Meegeren*. Ethics and Information Technology, **5**(1), 27-38.

BODEN, Margaret A., 1994. *What is creativity?* In: BODEN, Margaret A. (ed.). *Dimensions of creativity*. Cambridge, MA: MIT Press, 75-117.

BODEN, Margaret A., 2004. *The creative mind: myths and mechanisms*. 2nd ed. New York: Routledge, 2004. ISBN 0-415-31452-6.

BODEN, Margaret A., 2009. *Computer Models of Creativity*. *AI Magazine*. 30. 23-34. 10.1609/aimag.v30i3.2254.

BODEN, Margaret A., 2010. *The Turing test and artistic creativity*. *Kybernetes*, 39(3), 409-413.

BOULANGER-LEWANDOWSKI, Nicolas, BENGIO, Yoshua a VINCENT, Pascal, 2012. *Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription*. In: Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012.

BRINDLE, Reginald Smith, 1966. *Serial composition*. London: Oxford University Press. ISBN: 978-0193119062.

BRINGSFJORD, Selmer, BELLO, Paul a FERRUCCI, David, 2001. *Creativity, the Turing Test, and the (Better) Lovelace Test*. *Minds and Machines*, 11, 3-27

BROWNE, Tim Murray a FOX, Charles, 2009. *Global Expectation-Violation as Fitness Function in Evolutionary Composition*. In: Giacobini M. et al. (eds) *Applications of Evolutionary Computing. EvoWorkshops 2009. Lecture Notes in Computer Science*, vol 5484. Springer, Berlin, Heidelberg.

BUNDY, Alan, 1994. What is the Difference between Real Creativity and Mere Novelty? *Behavioral and Brain Sciences*, 17(3), 533-534. doi:10.1017/S0140525X0003572X.

BURNETT, Adam, KHOR, Evon, PASQUIER, Philippe a EIGENFELDT, Arne, 2012. *Validation of Harmonic Progression Generator Using Classical Music*. In: Third International Conference on Computational Creativity (ICCC 2012). 126-133.

BURTON, Anthony R. a VLADIMIROVA, Tanya, 1999. *Generation of musical sequences with genetic techniques*. Computer Music Journal, **23**(4), 59-73.

CAGE, John, 1973. *Silence: Lectures and Writings by John Cage*. Wesleyan: University press. ISBN 0-8195-6028-6.

COATS, Pamela K., 1988. *Why expert systems fail*. Financial Management, **17**(3), 77-86.

COLLINS, Tom, Robin LANEY, Alistair WILLIS a Paul H. GARTHWAITE, 2016. *Developing and evaluating computational models of musical style*. Artificial Intelligence for Engineering Design, Analysis and Manufacturing. **30**(1), 16-43. DOI: 10.1017/S0890060414000687. ISSN 0890-0604.

COLTON, Simon, 2008. *Creativity versus the perception of creativity in computational systems*. Proceedings of AAAI symposium on creative systems.

COPE, David, 1996. *Experiments in Musical Intelligence*. Madison, Wisconsin: A-R Editions. ISBN 0-89579-314-8.

COPE, David, 2000. *The Algorithmic Composer*. Madison, Wisconsin: A-R Editions. ISBN 0-89579-454-3.

COPE, David, 2001. *Virtual Music: Computer Synthesis of Musical Style*. London: The MIT Press. ISBN 0-262-53261-1.

COPE, David, 2005. *Computer Models of Musical Creativity*. Cambridge, MA: MIT Press. ISBN-10: 026253410X.

COWART, Gregory, WILLIAMSON, Dane, FARHAT, Naha a LEE, Joon-Suk, 2019. *Do Humans STILL Have a Monopoly on Creativity or Is Creativity Overrated?*. In: KUROSU, Masaaki (ed.). *Human-Computer Interaction: Perspectives on Design*. 21st HCI International Conference, HCII 2019 Orlando, FL, USA, July 26–31, 2019, Proceedings, Part I. Springer. ISBN 978-3-030-22645-9.

CRUZ-ALCÁRAZ, Pedro P. a VIDAL-RUIZ, Enrique, 1998. *Learning regular grammars to model musical style: Comparing different coding schemes*. Proceedings of the 4th International Colloquium on Grammatical Inference 1998, 211-222.

DAHLSTEDT, Palle, 2007. *Autonomous evolution of complete piano pieces and performances*. In: Proceedings of the European Conference on Artificial Life.

DE PRISCO, Roberto, ZACCAGNINO, Gianluca a ZACCAGNINO, Rocco, 2010. *EvoBassComposer: a multi-objective genetic algorithm for 4-voice compositions*. In: Proceedings of the Genetic and Evolutionary Computation Conference.

DIAZ-JEREZ, Gustavo, 2011. *Composing with Melomics: Delving into the Computational World for Musical Inspiration*. Leonardo Music Journal. 21. 13-14. DOI: 10.1162/LMJ_a_00053.

DONNELLY, Patrick a John SHEPPARD, 2011. *Evolving Four-Part Harmony Using Genetic Algorithms*. In: DI CHIO, Cecilia, Anthony BRABAZON, Gianni A. DI CARO, et al., ed. Applications of Evolutionary Computation. Berlin, Heidelberg: Springer Berlin Heidelberg, 273-282. Lecture Notes in Computer Science. DOI: 10.1007/978-3-642-20520-0_28. ISBN 978-3-642-20519-4.

DREYFUS, Hubert L., 1979. *What computers can't do*. New York: Harper Row, revised edition. ISBN: 9780262041348.

EBCIOĞLU, Kemal, 1980. *Computer counterpoint*. In: Proceedings of the International Computer Music Conference 1980.

EBCIOĞLU, Kemal, 1988. *An expert system for harmonizing four-part chorales*. Computer Music Journal, **12**(3), 43-51.

EIGENFELDT, Arne a PASQUIER, Philippe, 2010. *Realtime Generation of Harmonic Progressions Using Controlled Markov Selection*. In: Proceedings of the First International Conference on Computational Creativity (ICCCX), ACM Press, Lisbon, Portugal, 16-25.

EVANS, Benjamin, FUKAYAMA, Satoru, GOTO, Masataka, MUNETAKA, Nagisa a ONO, Tetsuo, 2014. *AutoChorusCreator: Four-Part Chorus Generator with Musical Feature Control, Using Search Spaces Constructed from Rules of Music Theory*. In: Proceedings of the 40th International Computer Music Conference and 11th Sound&Music Computing conference (JointICMC|SMC|2014Conference), 1016-1023.

FARBOOD, Mary a SCHONER, Bernd, 2001. *Analysis and synthesis of Palestrina-style counterpoint using Markov chains*. In: Proceedings of the International Computer Music Conference.

FERNANDÉZ, Jose David a VICO, Francisco, 2013. *AI Methods in Algorithmic Composition: A Comprehensive Survey*. Journal of Artificial Intelligence Research **48** (2013), 513-582.

FERREIRA, Isabel M. a MACHADO, J. Tenreiro, 2008. *Algorithmic Music Composition: A Survey*. Proceedings of the 2nd Conference on Nonlinear Science and Complexity. Porto, Portugal.

FEULNER, Johannes a HÖRNEL, Dominik, 1994. *MELONET: neural networks that learn harmony-based melodic variations*. In: Proceedings of the International Computer Music Conference, San Francisco, 121-124.

FOUCAULT, Michel, 1979. *What is an Author?.* In: HARARI, Josué V. (ed.). *Textual Strategies: Perspectives in Post-Structuralist Criticism*. London: Methuen, 141-160.

FRENCH, Robert M., 1990. *Subcognition and the Limits of the Turing Test*. *Mind*, **99**(393), 53-65.

GAGNIUC, Paul, 2017. *Markov chains: from theory to implementation and experimentation*. Hoboken, NJ: John Wiley. ISBN 9781119387572.

GARAY ACEVEDO, Andres, 2004. *Fugue composition with counterpoint melody generation using genetic algorithms*. In: Proceedings of the International Conference on Computer Music Modeling and Retrieval, 96-106.

GILL, Stanley, 1963. *A technique for the composition of music in a computer*. *The Computer Journal*, **6**(2), 129-133.

GJERDINGEN, Robert, 1988. *Explorations in Music, the Arts, and Ideas: Essays in Honor of Leonard B. Meyer*, Concrete musical knowledge and a computer program for species counterpoint, Pendragon Press, 199-228.

GWEE, Nigel, 2002. *Complexity and Heuristics in Rule-Based Algorithmic Music Composition*. Dizertační práce. Louisiana State University, The Department of Computer Science.

HADJERES, Gaëtan, PACHET, Francois a NIELSEN, Frank, 2017. *DeepBach: a Steerable Model for Bach Chorales Generation*. In: Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, PMLR 70.

HALPERN, Mark, 2006. *The Trouble with the Turing Test*. *The New Atlantis* 11, 42-63. Dostupné z <https://www.thenewatlantis.com/publications/the-trouble-with-the-turing-test>.

HAMANAKA, Masatoshi, HIRATA, Keiji a TOJO, Satoshi, 2008. *Melody morphing method based on GTTM*. In: Proceedings of the International Computer Music Conference, 155-158.

HARNAD, Stevan, 2000. *Minds, Machines and Turing*. *Journal of Logic, Language and Information*, **9**(4), 425-445.

HAUGELAND, John, 1985. *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press, c1985. ISBN 0-262-08153-9.

HEDGES, Stephen A., 1978. *Dice Music in the Eighteenth Century*. *Music & Letters*, **59**(2), 180-187.

HERREMANS, Dorien a CHEW, Elaine, 2019. *Morpheus: Generating Structured Music with Constrained Patterns and Tension*. *IEEE Transactions on Affective Computing* 10.4 (2019), 510-523.

HILD, Hermann, FEULNER, Johannes, a MENZEL, Wolfram, 1992. *HARMONET: a neural net for harmonising chorales in the style of J.S. Bach*. In: *Proceedings of the Conference on Neural Information Processing Systems*.

HILLER, Lejaren a ISAACSON, Leonard, 1959. *Experimental Music: Composition with an electronic computer*. McGraw-Hill, Westport, Connecticut.

HIRAGA, Rumi, BRESIN, Roberto, HIRATA, Keiji a KATAYOSE, Haruhiro, 2004. *Rencon 2004: Turing Test for Musical Expression*. *Proceedings of the 2004 Conference on New Interface for Musical Expression*. New York: Association for Computing Machinery, 120–123.

HOLTZMAN, Steven R., 1981. *Using generative grammars for music composition*. *Computer Music Journal*, **5**(1), 51-64.

HORNER, Andrew a AYERS, Lydia, 1995. *Harmonization of musical progressions with genetic algorithms*. In *Proceedings of the International Computer Music Conference 1995*.

HÖRNEL, Dominik a RAGG, Thomas, 1996. *A connectionist model for the evolution of styles of harmonization*. In: *Proceedings of the International Conference on Music Perception and Cognition*.

HUANG, Jiun-Long, CHIU, Shih-Chuan a SHAN, Man-Kwan, 2012. *Towards an automatic music arrangement framework using score reduction*. ACM Trans. Multimedia Comput. Commun. Appl. 8, 1, Article 8 (January 2012).

HUANG, Cheng-Zhi Anna, COOIJMANS, Tim, ROBERTS, Adam, COURVILLE, Aaron a ECK, Douglas, 2017. *Counterpoint by Convolution*. In: 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

HUANG, Allen a WU, Raymond, 2016. *Deep Learning for Music*. arXiv preprint arXiv:1606.04930, 2016.

CHEMILLIER, Marc a TRUCHET, Charlotte, 2001. *Two musical CSPs*. In: Proceedings of the International Conference on Principles and Practice of Constraint Programming.

CHILVERS, Alex a ZAAANEN, Menno, 2008. *Chorale Harmonization in the Style of J.S. Bach, A Machine Learning Approach*. In: Proceedings of the MML 2008 International Workshop on Machine Learning and Music held in conjunction with ICML / COLT / UAI 2008. Helsinki, Finland, 27-28.

CHUNG, Junyoung, GULCEHRE, Caglar, CHO, KyungHyun a BENGIO, Yoshua, 2014. *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. In: NIPS 2014 Deep Learning and Representation Learning Workshop.

JACOBS, Jeffrey Power a REGGIA, James, 2012. *Evolving musical counterpoint: The chronopoint musical evolution system*. arXiv preprint arXiv:1207.5560.

JAN, Steven, 2018. *The Theory and Analysis of Computer-Generated Music: A Case-Study of Colossus*. In: *Proceedings of the Third Conference on Computer Simulation of Musical Creativity, CSMC2018*. Dublin, Eire, 3rd Computer Simulation of Musical Creativity Conference, Dublin, Ireland.

JAQUES, Natasha, GU, Shixiang, TURNER, Richard E. a ECK, Douglas, 2016. *Generating music by fine-tuning recurrent neural networks with reinforcement learning*. Deep Reinforcement Learning Workshop, NIPS.

JEFFERSON, Geoffrey, 1949. *The Mind Of Mechanical Man*. The British Medical Journal 1, 4616, 1105-1110.

JOHNSON, Daniel D., 2017. *Generating Polyphonic Music Using Tied Parallel Networks*. In: CORREIA, João, Vic CIESIELSKI a Antonios LIAPIS, ed. Computational Intelligence in Music, Sound, Art and Design [online]. Cham: Springer International Publishing, 2017, 2017-03-22, 128-143 [cit. 2020-04-17]. Lecture Notes in Computer Science. DOI: 10.1007/978-3-319-55750-2_9. ISBN 978-3-319-55749-6.

JORDANOUS, Anna, 2010. *A Fitness Function for Creativity in Jazz Improvisation and Beyond*. International Conference on Computational Creativity.

JORDANOUS, Anna. 2012. *A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative*. Cognitive Computation, 4(3), 246-279.

KALEAGASI, Bartu, 2017. *A New AI Can Write Music as Well as a Human Composer: The future of art hangs in the balance*. Futurism [online]. [2020-03-13]. Dostupné z: <https://futurism.com/a-new-ai-can-write-music-as-well-as-a-human-composer>.

KIRNBERGER, Johann Philipp, 1757. *Der allezeit fertige Menuetten- und Polonoisencomponist*. Berlin.

KOTECHA, Nikhil, 2018. *Bach2Bach: Generating Music Using A Deep Reinforcement Learning Approach*.

KRUPOWICZ, Stanislaw, 2015. *Witold Lutoslawski – an Algorithmic Music Composer?*. Proceedings of the 41st International Computer Music Conference, ICMC 2015, Denton, TX, USA.

KURZWEIL, Ray. 1999. *The Age of Spiritual Machines*. New York: Penguin Books.

KURZWEIL, Ray, 2002. *A Wager on the Turing Test: Why I Think I Will Win*. KurzweilAI [online]. [cit 2020-03-01]. Dostupné z: <https://kurzweilai-brain.gothdyke.com/articles/art0374.html>.

KURZWEIL, Ray. 2005. *The Singularity is Near*. New York: Penguin Books.

LEEUEW, Ton de, 2005. *Music of the twentieth century: a study of its elements and structure*. Amsterdam: Amsterdam University Press, c2005. ISBN 90-5356-765-8.

LEWIS, Jon P., 1991. *Creation by Refinement and the Problem of Algorithmic Music Composition*. In: TODD, Peter M. a D. Gareth LOY. *Music and connectionism*. Cambridge, Mass.: MIT Press, c1991. ISBN 0-262-20081-3.

LIGETI, György, 1958. *Pierre Boulez, Entscheidung und Automatik in der Structure Ia*. In Die Reihe, IV/ 1958. Wien: Universal Edition A. G.

LIANG, Feynman, 2016. *BachBot: Automatic composition in the style of Bach chorales –Developing, analyzing, and evaluating a deep LSTM model for musical style*. Diplomová práce, University of Cambridge, Cambridge, U.K., August 2016. M.Phil in Machine Learning, Speech, and Language Technology.

LOUGHRAN, Róisín a O'NEILL, Michael, 2016. *Generative Music Evaluation: Why do We Limit to 'Human' ?*. In: Computer Simulation of Musical Creativity (CSMS, Huddersfield, UK (2016).

LOUGHRAN, Róisín a O'NEILL, Michael, 2017. *Limitations from Assumptions in Generative Music Evaluation*. Journal of Creative Music Systems. 2.

LOVELACE, Ada. 1842. *Translator's Notes to an Article on Babbage's Analytical Engine*. In: TAYLOR, R, ed. *Scientific Memoirs: Selected from the Transactions of Foreign Academies of Science and Learned Societies, and from Foreign Journals*. London: printed by Richard and John E. Taylor, 3, 691-731.

LOY, Gareth, 1989. *Composing with Computers - a Survey of Some Compositional Formalisms and Music Programming Languages*. In: MATHEWS, Max a PIERCE, John, eds. *Current Directions in Computer Music Research*. Cambridge: MIT Press.

LÖTHER, Mathis, 1999. *Knowledge based automatic composition and variation of melodies for minuets in early classical style*. In: *Proceedings of the Annual German Conference on Artificial Intelligence*, 159-170.

LYU, Qi, WU, Zhiyong, ZHU, Jun a MENG, Helen, 2015. *Modelling High-Dimensional Sequences with LSTM-RTRBM: Application to Polyphonic Music Generation*. In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*.

MADDOX, Tamara a OTTEN, John, 2000. *Using an evolutionary algorithm to generate Four-Part 18th century harmony*. In: *Proceedings of the WSEAS International Conference on Mathematics and Computers in Business and Economics*.

MANN, Alfred (ed.), 1965. *Steps to Parnassus. The Study of Counterpoint*. New York: W. W. Norton & Company. ISBN 0-393-00277-2.

McINTYRE, Ryan A., 1994. *Bach in a box: the evolution of four part baroque harmony using the genetic algorithm*. In: *Proceedings of the IEEE Conference on Evolutionary Computation*, 852-857.

MILLER, Samuel D., 1973. *Guido d'Arezzo: Medieval Musician and Educator*. *Journal of Research in Music Education*, **21**(3), 239-245.

MIRANDA, Eduardo Reck, 2001. *Composing music with computers*. Boston: Focal Press, 2001. ISBN 0240515676.

MORALES, Eduardo a MORALES, Roberto, 1995. *Learning musical rules*. In: Proceedings of the International Joint Conference on Artificial Intelligence 1995.

MORGAN, Nigel, 2007. *Transformation and mapping of L-Systems data in the composition of a large-scale instrumental work*. In: Proceedings of the European Conference on Artificial Life.

MORONI, Artemis, MANZOLLI, Jonatas, ZUBEN, Fernando Von a GUDWIN, Ricardo, 2000. *Vox Populi: An interactive evolutionary system for algorithmic music composition*. Leonardo Music Journal, 10, 49-54.

MOSTOW, Jack a RICH, Chuck, 1998. *The Fifteenth National Conference on Artificial Intelligence*. Dostupné z: www.aaai.org/Conferences/AAAI/aaai98.php.

MOZER, Michael, 1994. *Neural Network Music Composition by Prediction: Exploring the Benefits of Psychoacoustic Constraints and Multi-scale Processing*. In: Connection Science - CONNECTION. 6. 247-280. DOI: 10.1080/09540099408915726.

MURATA, Margaret, 1999. *Music History in the Musurgia universalis of Athanasius Kircher*. In: O'MALLEY, John W, ed. *The Jesuits: cultures, sciences, and the arts, 1540-1773*. Toronto: University of Toronto Press, c1999. ISBN 0-8020-4287-2.

NIERHAUS, Gerhard, 2009. *Algorithmic Composition: Paradigms of Automated Music Generation*. Springer Wien New York. ISBN 978-3-211-77539-6.

NOLL, Michael A., 1967. *The digital computer as a creative medium*. IEEE Spectrum, 4(10), 89-95.

PACHET, Francois a ROY, Pierre, 1995. *Mixing constraints and objects: a case study in automatic harmonization*. In: Proceedings of the Conference on Technology of Object-Oriented Languages and Systems.

PASQUIER, Philippe, EIGENFELDT, Arne, BOWN, Oliver a DUBNOV, Shlomo, 2016. An introduction to musical metacreation. *Comput. Entertain.* **14**(2), DOI: <http://dx.doi.org/10.1145/2930672>.

PAPADOPOULOS, George a WIGGINS, Geraint, 1999. *AI Methods for Algorithmic Composition: A Survey, a Critical View and Future Prospects*. In Proceedings of the Symposium on Musical Creativity, 110-117.

PEARCE, Marcus, 2005. *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. Dizertační práce, Department of Computing, City Univesity London.

PEARCE, Marcus a WIGGINS, Geraint A., 2001. *Towards A Framework for the Evaluation of Machine Compositions*. In: Proceedings of Symposium on Artificial Intelligence and Creativity in the Arts and Sciences.

PEARCE, Marcus T., MEREDITH, David a WIGGINS, Geraint A., 2002. *Motivations and methodologies for automation of the compositional process*. *Musicae Scientiae*, **6**(2), 119-147.

PEARCE, Marcus a WIGGINS, Geraint, 2007. *Evaluating Cognitive Models of Musical Composition*. In: Proceedings of the 4th International Joint Workshop on Computational Creativity. 73-80.

PEASE, Alison a COLTON, Simon, 2011. *On impact and evaluation in computational creativity: a discussion of the turing test and an alternative proposal*. In: Proceedings of the AISB'11 convention. York, UK: AISB.

PEREIRA, Francisco, GRILO, Carlos, MACEDO, Luís a CARDOSO, Amílcar, 1997. *Composing music with case-based reasoning*. In: Proceedings of the Conference on Computational Models of Creative Cognition.

PHON-AMNUAISUK, Somnuk, TUSON, Andrew a WIGGINS, Geraint, 1999. *Evolving musical harmonisation*. In *Reproduction*, Springer Verlag Wien, 1-9.

PHON-AMNUAISUK, Somnuk, 2002. *Control language for harmonisation process*. In: Proceedings of the International Conference on Music and Artificial Intelligence, 155-167.

POLITO, John, DAIDA, Jason M., & BERSANO BEGEY, Tommaso F., 1997. *Musica ex machina: Composing 16th-Century counterpoint with genetic programming and symbiosis*. In: Proceedings of the International Conference on Evolutionary Programming, 113-124.

PONSFORD, Dan, WIGGINS, Geraint a MELLISH, Chris, 1999. *Statistical learning of harmonic movement*. Journal of New Music Research, **28**(2), 150-177.

POVILIONIENE, Rima, 2017. *Definition Problem of Algorithmic Music Composition. Re-evaluation of the Concepts and Technological Approach*. GESJ: Musicology and Cultural Science 2017, **2**(16), ISSN 1512-2018.

QUICK, Donya, 2010. *Generating music using concepts from Schenkerian analysis and chord spaces*. Tech. rep., Yale University.

QUICK, Donya, 2014. *Kulitta: a Framework for Automated Music Composition*. Dizertace, Yale University.

RADER, Gary M., 1974. *A method for composing simple traditional music by computer*. Communications of the ACM, **17**(11), 631-638.

RAJON, 2018. *AI music composition passed Turing test*. Taiwan AILabs [online]. [cit 2020-05-02]. Dostupné z: <https://ailabs.tw/human-interaction/ai-music-composition>.

RANKIN, Matthew, 2012. *A Computer Model for the Schillinger System of Musical Composition*. The Department of Computer Science Australian National University.

RAPAPORT, William J., 2020. *Philosophy of Computer Science*. New York: University at Buffalo.

RATNER, Leonard G, 1970. *Ars combinatoria: Chance and Choice in Eighteenth-century Music*. In: LANDON, H. C. Robbins a CHAMPAN, Roger, eds. *Studies in Eighteenth-century Music: A Tribute to Karl Geiringer on his Seventieth Birthday*. New York: Oxford University Press.

RIBEIRO, Paulo, PEREIRA, Francisco, FERRAND, Miguel a CARDOSO, Amílcar, 2001. *Case-based melody generation with MuzaCazUza*. In: Proceedings of the Symposium on Artificial Intelligence and Creativity in Arts and Science, 67-74.

RICHMOND, Tom Donald a RAHAL, Imad, 2018. *The Algorithmic Composition of Classical Music through Data Mining*. All College Thesis Program. 52.

RITCHIE, Graeme D., 2001. *Assessing Creativity*. In: Proceedings of the AISB'01 Symposium on AI and Creativity in Arts and Science (Wiggins, G.A., ed.) AISB.

RITCHIE, Graeme D., 2007. *Some empirical criteria for attributing creativity to a computer program*. *Minds & Machines*, 17, 67-99. <https://doi.org/10.1007/s11023-007-9066-2>.

ROSEN, Kenneth H., 2011. *Elementary number theory and its applications*. 6th ed. Boston: Addison-Wesley. ISBN 0-321-50031-8.

ROTHGEB, John, 1968. *Harmonizing the unfigured bass: A computational Study*. Dizertační práce, Yale University.

RUNCO, Mark A. a ALBERT, Robert S., 2010. *Creativity research: A historical view*. In: KAUFMAN, J. C. a STERNBERG, R. J. (ed.). *The Cambridge handbook of creativity*. Cambridge University Press, 3-19. DOI: <https://doi.org/10.1017/CBO9780511763205.003>.

RUSSELL, Stuart J. a Peter NORVIG, 2016. *Artificial intelligence: a modern approach*. Third edition. Boston: Pearson. Prentice Hall series in artificial intelligence. ISBN 1292153962.

SABATER, Jordi, ARCOS, Josep a LÓPEZ DE MÁNTARAS, Ramón, 1998. *Using rules to support case-based reasoning for harmonizing melodies*. In: Proceedings of the AAAI Spring Symposium on Multimodal Reasoning, 147-151.

SADIE, Stanley, ed., 1980. *The New Grove Dictionary of Music and Musicians*.

SÁNCHEZ-QUINTANA, Carlos, MORENO-ARCAS, Francisco, ALBARRACÍN-MOLINA, David, FERNÁNDEZ, José David a VICO, Francisco, 2013. *Melomics: A case-study of AI in Spain*. AI Magazine, **34**(3), 99-103.

SEARLE, John R., 1980. *Minds, Brains, and Programs*. Behavioral and Brain Sciences, **3**(3), 417-457.

SCHOTTSTAEDT, William, 1989. *Current directions in computer music research: Automatic Counterpoint*, Cambridge: The MIT Press, 199-214.

SCHUBERT, Emery, CANAZZA, Sergio, POLI, Giovanni a RODA, Antonio, 2017. *Algorithms can Mimic Human Piano Performance: The Deep Blues of Music*. Journal of New Music Research, DOI: 10.1080/09298215.2016.1264976.

SCHULZE, Walter, 2009. *A Formal Language Theory Approach To Music Generation*. Dizertační práce, Stellenbosch University.

SCHWANAUER, Stephan, 1993. *Machine models of music: A learning machine for tonal composition*, Cambridge: The MIT Press, 511-532.

SILVA, Patrício, 2003. *David Cope and Experiments in Musical Intelligence*. Los Angeles, California: Spectrum Press.

SIMONI, Mary, 2003. *Algorithmic Composition: A Gentle Introduction to Music Composition Using Common LISP and Common Music*. Ann Arbor, Michigan: Michigan Publishing, University of Michigan Library. Dostupné z: <https://quod.lib.umich.edu/s/spobooks/bbv9810.0001.001>.

SINNREICH, Aram, 2019. Music, Copyright, and Technology: A Dialectic in Five Moments. In: *International Journal of Communication* 13(2019), 422–439.

SPANGLER, Randall Richard, 1999. *Rule-Based Analysis and Generation of Music*. Dizertační práce, California Institute of Technology.

SPURNÝ, Lubomír, 2012. *Heinrich Schenker (1868–1935): Kapitoly z hudební teorie a analýzy*. Praha: Koniasch Latin Press. ISBN 978-80-86791-12-8.

STANĚK, Jan, 2017. *Generování hudby pomocí neuronových sítí*. Praha. Diplomová práce. České vysoké učení technické v Praze, Fakulta informačních technologií, Katedra teoretické informatiky.

STEEDMAN, Mark, 1984. *A generative grammar for jazz chord sequences*. *Music Perception: An Interdisciplinary Journal*, 2(1), 52-77.

STEELS, Luc, 1986. *Learning the craft of musical composition*. In: *Proceedings of the International Computer Music Conference 1986*.

STORINO, Mariateresa, DALMONTE, Rossana a BARONI, Mario, 2007. *An Investigation on the Perception of Musical Style*. *Music Perception*, 24. 417-432. 10.1525/mp.2007.24.5.417.

TANAKA, Tsubasa, NISHIMOTO, Takuy, ONO, Nobutaka a SAGAYAMA, Shigeki, 2010. *Automatic Music Composition Based on Counterpoint and Imitation Using Stochastic Models*. In: 7th Sound and Music Computing Conference (SMC2010), Barcelona, Spain, 21-24 July.

THOMAS, Marilyn Taft, 1985. *Vivace: A rule based AI system for composition*. In: Proceedings of the International Computer Music Conference, 267-274.

TSANG, Chi Ping a AITKEN, Melanie, 1991. *Harmonizing music as a discipline of constraint logic programming*. In: Proceedings of the International Computer Music Conference 1991.

TRIVIÑO-RODRÍGUEZ, José Luis a MORALES-BUENO, Rafael, 2001. *Using multiattribute prediction suffix graphs to predict and generate music*. Computer Music Journal, **25**(3), 62-79.

TURING, Alan M., 1950. *Computing Machinery and Intelligence*. Mind (49). 433-460. Dostupné z: <https://www.csee.umbc.edu/courses/471/pa/pers/turing.pdf>.

TVRDÝ, Filip, 2011. *Turingův test: Filosofické aspekty umělé inteligence*. Dizertační práce. Filosofická fakulta, Univerzita Palackého v Olomouci, Katedra filosofie.

WHITELAW, Mitchell, 2004. *Metacreation: Art and Artificial Life*. Cambridge, MA: MIT Press.

WIGGINS, Geraint A., 1998. *The use of constraint systems for musical composition*. In: Proceedings of the Workshop on Constraints for Artistic Applications.

WIGGINS, Geraint A., 2006. *Searching for computational creativity*. New Generation Computing.

WOODS, William A., 1970. *Transition Network Grammars for Natural Language Analysis*. Communications of the ACM. **13**(10), 591-606.

XENAKIS, Iannis, 1992. *Formalized music: thought and mathematics in composition*. Rev. ed. Stuyvesant, NY: Pendragon Press, c1992. ISBN 978-0-945193-24-1.

YANCHENKO, Anna K. a MUKHERJEE, Sayan, 2017. *Classical music composition using state space models*. arXiv preprint arXiv:1708.03822, 2017.

YILMAZ, Asim Egemen a TELATAR, Ziya, 2010. *Note-against-note two-voice counterpoint by means of fuzzy logic*. Knowledge-Based Systems, **23**(3), 256-266.

ŽÁČEK, Ivan, 2019. *PKF s Villaumem si zakoketovala s virtuální realitou*. Harmonie [online]. [cit. 2020-04-26]. Dostupné z: <https://www.casopisharmonie.cz/kritiky/pkf-s-villaumem-si-zakoketovala-s-virtualni-realitou.html>.

Seznam obrázků

Obrázek 1: Příklad “Mozartovy” hudební hry v kostky (Bača a Oberholtzer).....	14
Obrázek 2: První část Illiac Suite (Povilioniene, 2017).....	17
Obrázek 3: Taxonomie použitých metod Fernandéze a Vica (2013)	18
Obrázek 4: Příklad sonáty ve stylu Sergeje Prokofjeva, EMI (Cope, 1996)	20
Obrázek 5: Příklad skladby Colossus, Iamus (Sánchez-Quintana a kol.).....	25
Obrázek 6: Typy evaluace kreativních systémů založené na poslechu (autor)	38
Obrázek 7: Výsledky testu s Kulittou (Quick, 2014).....	49
Obrázek 8: Výsledky testu „Bach nebo počítač“ (Hadjeres a kol.)	52
Obrázek 9: Výsledky testu Collinse a kol. (2016)	55
Obrázek 10: Výsledky testu Burnetta a kol. (2012).....	56
Obrázek 11: Výsledky testu Pearce a Wiginse (2007)	58
Obrázek 12: Srovnání provedených testů vycházejících z Imitační hry (autor)....	61

Seznam zkratek

UI	Umělá inteligence
EMI	Experiments in Musical Intelligence
MOTT	Musical Output Toy Test
MDTT	Musical Directive Toy Test
SPECS	Standardised Procedure for Evaluating Creative Systems
IDEA	Iterative Development Execution Appreciation
MPSG	Multiattribute Prediction Suffix Graph
MIDI	Musical Instruments Digital Interface