

**Akademie múzických umění v Praze
FILMOVÁ A TELEVIZNÍ FAKULTA**

Filmové, televizní a fotografické umění a nová média
Obor zvuková tvorba

DIPLOMOVÁ PRÁCE

AI & Mastering

BcA. Jakub Trš

Vedoucí práce: MgA. Petr Neubauer
Přidělovaný akademický titul: MgA.

Praha, 5/2023

**The Academy of Performing Arts in Prague
Film and TV School**

Film, Television and Photographic Arts and New Media
Department of Sound

MASTER'S THESIS

AI & Mastering

BcA. Jakub Trš

Thesis / Dissertation supervisor: MgA. Petr Neubauer
Academic title: MgA.

Prague, 5/2023

P r o h l á š e n í

Prohlašuji, že jsem magisterskou práci s názvem

AI & Mastering

vypracoval(a) samostatně pod odborným vedením vedoucího práce a s použitím pouze uvedené literatury a pramenů a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu. Souhlasím s tím, aby práce byla zveřejněna v souladu se zákonem a vnitřními předpisy AMU.

Praha, dne

.....

Jakub Trš

Poděkování

MgA. Petr Neubauer – za vysoce trpělivé vedení, navýsost odborné připomínky a hluboce užitečné podněty;

rodiče – za trpělivost a laskavou podporu;

Thomas Vingtrinier, Martin Ledvina, Michal Pekárek, Ondřej Ježek, Vojtěch Zavadil – za jejich čas věnovaný diskusím k tématu;

MgA. Adéla Matochová – za obětavou korekturu.

Abstrakt

V uplynulé dekádě se začaly prosazovat online služby nabízející automatický mastering hudebních nahrávek (nežřídka s využitím tzv. "umělé inteligence"), což nutí k zamyšlení nejen nad jejich technickým provedením, ale i nad povahou masteringu samotného.

Abstract

The past decade has seen an emergence of online services offering automated music mastering services (frequently advertising the use of "artificial intelligence"). This leads to questions not only about the specific use of AI techniques, but about music mastering as a whole.

Úvod.....	1
1. Mastering obecně	2
1.1. Vývoj zvukových médií	4
1.1.1 Analogová éra.....	4
1.1.2. Digitální éra.....	11
1.2. Z technika tvůrcem	13
1.3. Mastering alba vs mastering singlů.....	13
1.4. Tradiční vs stem vs NGA mastering.....	13
1.5. Tvůrčí vklad	14
1.6. Objektivní a subjektivní parametry, nástroje a úkony masteringu	15
1.7. Žánry	16
1.8. MIR.....	16
2. Umělá inteligence – AI.....	19
2.1. Definice	19
2.1.1. Obecné definice	19
2.1.2. Definice psychologie.....	19
2.1.3. Definice AI výzkumníků.....	20
2.1.4. Souhrnná definice inteligence.....	20
2.2. Definice AI, Tesslerův teorém.....	20
2.3. AI techniky.....	21
2.4. Historie neuronových sítí (NS).....	21
2.5. Neuronové sítě.....	23
2.6. Tradiční strojové učení a neuronové sítě.....	23
2.7. Tréninkové strategie neuronových sítí.....	24
2.7.1. Učení se bez učitele.....	24
2.7.2. Zpětnovazební učení	24
2.7.3. Učení s učitelem	25
2.8. Algoritmus zpětného šíření chyby v učení NS	25
2.9. Další druhy neuronových sítí	27
2.9.1. Konvoluční neuronové sítě (CNN).....	27
2.9.2. Rekurentní neuronové sítě (RNN).....	28
2.9.3. Long-term short-term sítě (LSTM).....	28
2.9.4. Transformer.....	29
2.10. Přetrénování (<i>overfitting</i>).....	30
2.11. Příprava dat pro zpracování neuronovou sítí.....	31
2.11.1. Mel-frekvenční keprstrální koeficient – MFCC	31
2.12. Deep learning a zvuk.....	32
3. Instantní mastering, Intelligent music production.....	33
3.1. Míra kontroly a varianty inteligentních prvků IMP systémů	34
3.1.1. Poznatkový	34
3.1.2. Návrhový	36
3.1.3. Nezávislý.....	37
3.1.4. Automatický	38
3.2. Reprezentace znalostí v IMP systémech.....	38
3.2.1. Zakotvená (podložená) teorie	38
3.2.2. Expertní systémy založené na reprezentaci znalostí	39
3.2.3. Přístupy založené na zpracování dat.....	40
3.3. Způsob úpravy zvuku	41
3.4. Současnost instantního masteringu	42
3.4.1. Obfuskace vnitřního fungování.....	42
3.4.2. Platformizace služeb.....	43
3.4.3. Cenová dostupnost.....	43
3.4.4. Je to vůbec mastering?	43
3.5. Lidský versus instantní mastering.....	44
3.6. Subjektivní dojmy z použití.....	44
3.6.1. Easter	45
3.6.2. Easter – dynamika	45
3.6.3. Easter – tonalita a další	46
3.6.4. Nahrávka #2.....	46
3.6.5. Nahrávka #2 - dynamika	46

3.6.6. Nahrávka #2 – tonalita a další.....	47
4. Související témata	48
4.1. Předpojatost – bias v AI a perpetuace stereotypů.....	48
4.2. Černé skříňky, Goldbergův aparát, neuronové sítě a kernel machines.....	49
4.3. Glass box.....	49
4.4. Interpretabilita neuronových sítí, projekce a haiku.....	50
4.5. Tah 37 a adversární útoky.....	51
4.6. Důvěra v automatické systémy, nahraditelnost a afektivní práce.....	52
4.7. Afektivní práce (affective labour).....	53
4.8. Originalita, smrt autora a Melancholičtí sloni.....	53
4.9. Další postřehy.....	54
Závěr	55
Bibliografie	56
Seznam obrázků	61

Seznam příloh

- Příloha 1 – „Easter“ – original mix – Jakub Trš, 2021
- Příloha 2 – „Easter“ – LANDR master – Jakub Trš, 2023
- Příloha 3 – „Easter“ – Ozone master – Jakub Trš, 2023
- Příloha 4 – „Nahrávka #2“ – original mix – Jakub Trš, 2023
- Příloha 5 – „Nahrávka #2“ – LANDR master – Jakub Trš, 2023
- Příloha 6 – „Nahrávka #2“ – Ozone master – Jakub Trš, 2023

Úvod

„Každá dostatečně vyspělá technologie je k nerozeznání od magie.“

- Arthur C. Clarke

Zkratka „AI“ („artificial intelligence“, neboli „umělá inteligence“) je dnes všudypřítomná, a to jak v podobě marketingových frází, tak i ve zcela konkrétních případech stále sofistikovanějších nástrojů čím dál běžnějších i ve zvukové tvorbě. Jedním z důsledků tohoto vývoje je vznik řady on-line služeb nabízejících levný a téměř instantní mastering hudebních nahrávek – činnost, která tradičně bývala doménou výsostně lidskou.

Tato práce si klade za cíl jednak podhalit závoj zaklínadla umělé inteligence jako takové (dále „AI“), ale také navrhnout myšlenkový rámec k uvažování o AI ve vztahu k naší profesi a prozkoumat vliv moderních sofistikovaných nástrojů na (nejen) zvukovou tvorbu na konkrétním příkladu masteringu hudebních nahrávek.

1. Mastering obecně

Většina lidí si dokáže představit, co v rámci procesu hudební produkce dělají profese jako studiový zvukař, producent či skladatel. Oproti tomu hudební mastering má auru určitého „tajemného umění“ prováděného „zasvěcenými mistry“ specialisty, k čemu přispívají nejen historické proměny profese jako takové, ale i práce probíhající nezřídka v izolaci či potřeba velmi přesného a jakostního poslechového řetězce, ať už z hlediska akustiky místnosti, nebo technického vybavení.

Technicky vzato je mastering formou zvukové postprodukce s cílem tvorby tzv. *masteru* čili nosiče finální podoby zvukové nahrávky, od které se odvozují všechny ostatní kopie. Činnost dnes obvykle považována za *mastering* je technicky vzato *premasteringem*¹ protože ke vzniku opravdového *masteru* (alespoň v případě výroby fyzických nosičů) dochází až v lisovně, nicméně termín je již všeobecně zažitý.

Slovy inženýrů samotných², mastering je posledním tvůrčím krokem v procesu celé zvukové produkce, představuje most mezi mixáží a distribucí, jde o jakousi „oponenturu mixu“³. Představuje poslední šanci něco vylepšit (či opravit) v poslechové místnosti specificky navržené a vybavené za tímto účelem.

Masteringový inženýr přináší perspektivu zkušeného a informovaného posluchače s vytříbeným vkusem, odstupem od projektu a současně expertními zvukařskými dovednostmi čili „informované druhé uši“. Může se stát, že si mix poslechne a zhodnotí, že není potřeba žádná úprava, jindy může nalézt řešení něčeho, co se nepodařilo vyřešit při mixáži.

Tradičně neodmyslitelnou rolí masteringu je také zpracování souboru nahrávek jednotlivých písní tak, aby tvořily jednolitý celek v rámci hudebního alba; ať se jedná o jednotu barevnou, hlasitostní, či temporytmickou funkčnost danou pořadím skladeb, případně jejich rozestupy. Dostupnost široké škály online platforem usnadňujících distribuci hudebních nahrávek způsobila, že z hlediska udržení pozornosti a kontaktu s fanouškovskou základnou je efektivnější častější vydávání singlů – i ty ovšem zpravidla procházejí nějakou formou masteringu.

Čtenáři neznalí tématu si mastering mohou představit jako jakýsi zvukový ekvivalent tiskařské sazby, která určuje finální formu výtisku poté, co projde editorskými úpravami a je prohlášen za hodný vydání.⁴

¹ KATZ, Robert A. *Mastering audio: the art and the science*. s.18. Oxford: Boston : Focal Press, 2002.

² OWSINSKI, Bobby a Sally ENGLEFRIED. *The mastering engineer's handbook: the audio mastering handbook*. Boston, MA : Thomson Course Technology PTR, 2008.

³ JEŽEK, Ondřej. *Rozhovor o AI & Masteringu* [Archiv autora]. 2022.

⁴ STERNE, Jonathan a Elena RAZLOGOVA. Machine Learning in Context, or Learning from LANDR: Artificial Intelligence and the Platformization of Music Mastering. *Social Media + Society*. 2019, roč. 5, č. 2. DOI: 10.1177/2056305119847525

Souhrnem se dá říci, že *gró* masteringu spočívá v optimalizaci zvuku nahrávky pro její vyznění na široké škále poslechových prostředků, s cílem zajistit co nejlepší přenos tvůrčího záměru⁵ a tím pádem konzistentní posluchačský zážitek.

⁵ Jonathan Wyner, *AES Mastering Academy pt.2.* 2022.

1.1. Vývoj zvukových médií

Vznik masteringu jako svébytné profese je úzce spjat s vývojem záznamových a distribučních médií, v následující části si tedy připomeneme několik milníků.

1.1.1 Analogová éra

1857 – tiskař a knihkupec **Édouard-Léon Scott de Martinville** vynalezl v Paříži **fonautograf**. Zařízení bylo koncipováno jako pomůcka k vizuálnímu studiu zvukových vln, ovšem bez možnosti jejich reprodukce. Zvukové vlny byly jehlou vyrývány do směsi jemného uhelného prášku a oleje na papíře nebo na skle. Až v roce 2008 skupina amerických vědeckých pracovníků z First Sounds Collective⁶ převedla nahrávku *Au Clair de la Lune*, kterou de Martinville pořídil 9. dubna 1860, do digitální podoby, což z ní činí nejstarší rozpoznatelný záznam lidského hlasu a nejstarší rozpoznatelný záznam hudby.

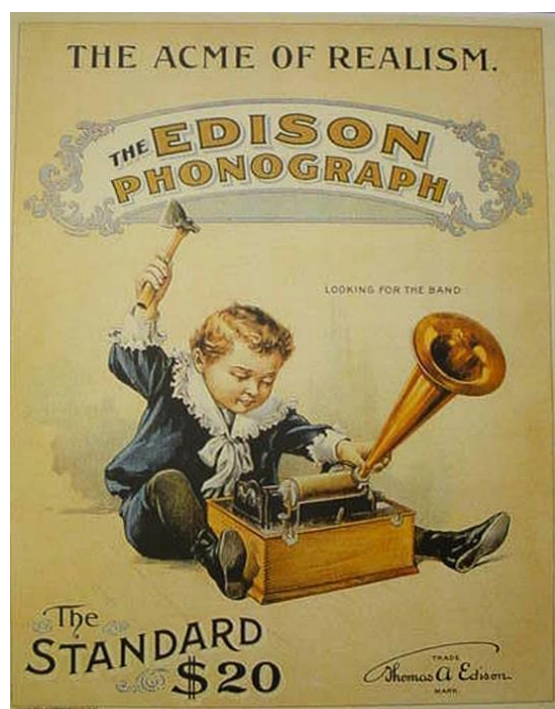


Obr. 1 - Fonautografická nahrávka

⁶ Publications :: FirstSounds.ORG. In: [cit. 21.08.2022]. Dostupné z: <http://www.firstsounds.org/research/scott.php>

1877 - Thomas Edison ohlásil vynález **fonografu**, zařízení určeného pro nahrávání a přehrávání zvuku s pomocí **válečků** pokrytých **cínovou fólií** jako vedlejší produkt práce na telegrafu a telefonu. Původním záměrem bylo sestrojít jakéhosi předchůdce záznamníku pro telegrafní zprávy, nicméně při práci si uvědomil, že zvukové vlny řeči by se možná daly zaznamenat stejně jako telegrafní tečky a čárky.

Válečky nebylo možné odstranit bez poškození zařízení – v té době ještě nikoho nenapadlo, že by se tyto prvky daly oddělit, a že by se dalo obchodovat pouze s nahrávkami samotnými.⁷



Obr. 2 - "Vrchol realismu" - dobová reklama na fonograf

1880 - Charles Sumner Tainter a Chichester Bell z Bellových laboratoří dále zdokonalovali záznamové médium, dokud neuvedli na trh **grafofon**, systém založený na Edisonově fonografu, ovšem nahrávající na válečky pokryté **voskovou vrstvou**, které bylo možné vyměňovat.

1887 - Emile Berliner vynalezl **gramofon**. Odrazil se přitom od principů popsanych Martinvillem a jeho fonautografem, ovšem jako záznamové médium zvolil místo válečku **kovový disk**. Jehla vyrývá horizontální drážku do vrstvy krycí pasty. Po skončení nahrávání je drážka do disku „nadobro“ vyleptaná s pomocí kyseliny v místech, ve kterých jehla odstranila krycí pastu. Takto vzniklý „master“ je možno dále množit např. galvanoplasticky (formou pokovení materiálu s využitím elektrolýzy).



Obr. 3 - Emile Berliner se svým vynálezem

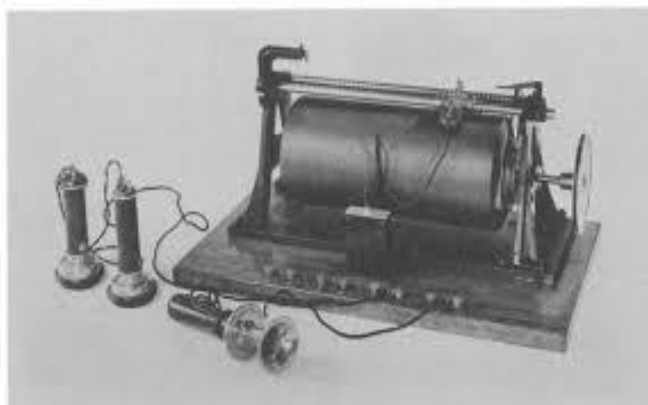
⁷ AGNEW, J.I. *From The Cylinder To The Disk Record...And Back? – PS Audio* [online] [cit. 24.08.2022]. Dostupné z: <https://www.psaudio.com/copper/article/from-the-cylinder-to-the-disk-record-and-back/>

Future generations will be able to condense into the brief space of twenty minutes the tone pictures of a lifetime—five minutes of childish prattle, five of boyish exultation, five of the man's mature reflections, ending with five moments embalming the last feeble utterances from the death-bed. Will this not seem like holding veritable communion with immortality?

From what has been stated it will be seen that the BERLINER GRAMOPHONE is to the voice what photography is to the features—*i. e.*, a simple, practical medium for securing accurate and lasting records.

Obr. 4 - Berlinerův propagační materiál.

1898 - Valdemar Poulsen předvedl princip **magnetického záznamu zvuku** čili záznam vycházející z konstantního pohybu magnetizovatelného materiálu kolem magnetické zápisové hlavy. Jeho **telegrafon**⁸, zamýšlený zejména k nahrávání telefonních hovorů, používal nejprve kovové dráty, posléze přešel na magnetické pásky.



Obr. 5 - Poulsenův telegrafon

Na Světové výstavě v Paříži v roce

1900 nahrál Poulsen hlas císaře Františka Josefa I., což se považuje za nejstarší dochovanou magnetickou nahrávku.

1902 - Edison Gold Moulded Records⁹ - až do této chvíle představovala masová produkce voskových válečků problém, protože každý váleček byl originál zaznamenaný v reálném čase, takže interpreti, jejichž výkony byly zrovna populární, museli zpívat a hrát pořad dokola před jedním či vícero fonografy, dokud nebylo vyrobeno dostatečné množství válečků.



Obr. 6 - Edisonův Gold Moulded váleček

⁸ CLARK, Mark a Henry NIELSEN. Crossed Wires and Missing Connections: Valdemar Poulsen, The American Telephonograph Company, and the Failure to Commercialize Magnetic Recording. *Business History Review*. 1995, roč. 69, č. 1. DOI: 10.2307/3117119

⁹ Brown Wax Cylinders. In: [cit. 22.08.2022]. Dostupné z: https://www.cylinder.de/guide_brown-wax-cylinders.html

V roce 1890 začal Edison s množením nahrávek ranou formou odlévání, ale zejména pantografickým kopírováním, což limitovalo množství možných kopií z jednoho „master“ válečku, protože s každou generací kopií docházelo k jeho mechanickému opotřebení, což se podepisovalo na snižující se zvukové kvalitě a když „mastery“ došly, bylo potřeba znovu zaplatit interpreta k vytvoření nových. V roce 1902 Edison podstatně zdokonalil proces odlévání a začal s výrobou *playback only* válečků z tvrdého vosku pod značkou *Edison Gold Moulded Records*, čímž odstartoval opravdovou masovou produkci těchto nosičů.

1928 - Fritz Pfeleumer

v Německu vynalezl magnetické pásky pro záznam zvuku, které se v následující dekádě staly dominantní technologií. Pfeleumer použil papírový pásek s přilakovaným práškem oxidu železitého. V roce 1928 získal patent na magnetofon, v roce 1932 prodal práva na pásku firmě AEG.¹⁰



Obr. 7 - jediná dochovaná fotografie Fritze Pfeleumera s magnetofonem

1929 - Edison, poslední zastánce válečků, přestává s jejich výrobou poté, co byly překonány kotoučovými médii.¹¹

1931 - Alan Blumlein vynalezl stereofonní záznam (tehdy pojmenován *binaurální*), když si uvědomil, že s pomocí dvou mikrofonů by bylo možné zaznamenat zvukové pole se směrovými informacemi, přenášející k posluchači dojem šířky. Přišel také na dodnes používaný způsob záznamu signálu pro oba kanály do jedné drážky desky, kdy oba kanály korespondují s diagonálními pohyby přenosky.¹²

¹⁰ DANIEL, Eric D., C. Denis MEE a Mark CLARK. *Magnetic Recording: The First 100 Years*.

¹¹ Bill Klinger; Association for Recorded Sound Collections. .

¹² Alan Blumlein and the invention of Stereo | EMI Archive Trust. In: . Dostupné z: <https://www.emiarchivetrust.org/alan-blumlein-and-the-invention-of-stereo/>

1948 - Columbia Records uvádí dlouhohrající vinylovou desku – LP. Nižší rychlost 33-1/3 otáček za minutu a miniaturizace drážky zápisu přinesly 22 minut záznamu na stranu – posun oproti 4-5 minutám předchozí generace desek se 78 otáčkami za minutu.¹³ Vinylové desky se zrodily kvůli nedostatku šelaku v průběhu 2. světové války, ovšem brzy se zjistilo, že tento nový materiál přináší vyšší mechanickou odolnost i zvukovou kvalitu. Mezi „taháky“ patřilo také to, že záznam symfonie se vešel na jednu desku oproti nutnosti kupovat vícero starších 78rpm desek, takže kulturní zážitek vyšel levněji.

45 MINUTES OF MUSIC FROM A SINGLE RECORD
... ANOTHER "FIRST" BY COLUMBIA RECORDS

COLUMBIA
LP
LONG PLAYING MICROGROOVE RECORD

Finer tone quality! So lifelike you'll hardly believe you're listening to a record. Low notes, high notes are heard without distortion. And practically no surface noise!

Uninterrupted music! Major works are recorded either on 2 sides or 1 side of a single LP record. At last—no more annoying "breaks."

More than twice as much music for your money! Columbia LP Records save you up to 60% per selection over conventional Vinylite records. Think how much farther your music budget goes... how much faster you'll build a fine record collection.

Nonbreakable Vinylite! Makes broken records practically a thing of the past—another source of savings. And super-smooth Vinylite means finer tone.

Saves storage space! Every inch of shelf space holds 3 hours of music!

Over 600 selections already in catalog! Symphonies, concertos, musical comedies, jazz, opera, children's stories—157 different records! A wonderful collection of entertainment by your favorite artists who record exclusively for Columbia. Many new releases every month.

THIS COLUMBIA LP PLAYER ATTACHMENT
plays LP records through your present radio or phonograph

You only need to add a slow-speed player attachment to your present set in order to play LP Records. The handsome Columbia Player shown here is quickly installed, attractively priced, and precision-designed for flawless reproduction of Columbia LP Records. It modernizes your present set to play both LP and your regular records. Has amazing featherweight tone arm weighing only 1/4 of an ounce! Your savings on a few LP Records pay for it. See your dealer today!

"Columbia," "Masterworks" and "LP" Trade Marks Reg. U. S. Pat. Off. Record Registrations © Trade Mark

Obr. 8 - reklamní materiál, LP

1948 - Ampex představuje svůj první komerční magnetofon, což vede k prosazení magnetické pásky jako nahrávacího média, čímž se odděluje nahrávací médium od distribučního a *master* již nevzniká živým nahráváním v reálném čase. Protože je ale vinyl stále distribučním médiem, pro dodání do lisovny musel být proveden převod (*transfer*) na vinylový master; proto se zrodila první inkarnace **masteringového inženýra**, tehdy zvaného „*transfer engineer*“.¹⁴

Presenting the new

Ampex MAGNETIC TAPE RECORDER
TRUE-TO-LIFE FIDELITY

THE GREAT NEW UNIT THAT PUT THE CROSBY SHOW ON TAPE

Obr. 9 - prezentace magnetofonu Ampex

¹³ JENKINS, Amanda. Inside the Archival Box: The First Long-Playing Disc | Now See Hear! In: *The Library of Congress* [online]. 13. 4. 2019 [cit. 28.05.2023]. Dostupné z: //blogs.loc.gov/now-see-hear/2019/04/inside-the-archival-box-the-first-long-playing-disc

¹⁴ OWSINSKI, Bobby a Sally ENGLEFRIED. *The mastering engineer's handbook*.

1955 - Ampex zkonstruoval pro Lese Paula první magnetofon umožňující vícestopé nahrávání, čímž osvobozuje interprety a zvukové mistry od nutnosti destruktivně nahrávat nové linky do stále stejné stopy (technika známá jako „*sound on sound*“). Vedoucí projektu Alex Snyder předpokládal, že takový výstřelek technologie zaujme nanejvýš pár nadšenců – kromě toho, že tato technika navždy proměnila nahrávací průmysl, se také prohloubil rozdíl mezi nahrávacím a masteringovým inženýrem.¹⁵

1957 - RCA Victor, Westrex, London představily **stereo vinylovou desku**, čímž se zvukové možnosti nahrávek dostaly na dnešní úroveň.¹⁶ Tvůrčí úloha masteringového inženýra tímto okamžikem získala nový rozměr, protože k přípravě na lisování masteru se začaly běžně využívat dynamické a frekvenční úpravy, po nichž výsledná nahrávka (v ideálním případě) zněla ještě lépe, než když opustila mixážní studio.

1963 - Phillips představuje **audiokazetu**. Přestože byla původně vyvinuta pro potřeby diktafonů, postupný kvalitativní vývoj a uživatelská přívětivost ji dovedly k dominanci v 70. a 80. letech.

1963 - RCA Dynagroove – první využití analogového počítače pro automatické korekce při výrobě vinylového masteru. Byť se systém (velmi zjednodušeně přirovnatelný ke komandéru) nesetkal s úspěchem, přinesl několik zcela nadčasových inovací, jako například první inkarnaci vícepásmového kompresoru (zde „*dynamic spectrum equalizer*“), nebo detekce signálových špiček pro prevenci přebuzení.

AES CONVENTION PROGRAM

Friday, October 11

1:30 p.m. TECHNICAL SESSION: STEREO PROBLEMS.

H. E. Rays, RCA Victor Record Division, Indianapolis, Ind., Chairman.

ARTIFICIAL STEREOPHONY USING SINGLE INPUT.

Manfred R. Schroeder, Bell Telephone Laboratories, Murray Hill, N. Y.

STEREOPHONIC SOUND WITH TWO TRACKS, THREE CHANNELS BY MEANS OF A PHANTOM CIRCUIT OR 2PH3 STEREO.

Paul W. Klipsch, Klipsch & Associates, Hope, Arkansas.

STEREOPHONIC SOUND REPRODUCTION IN THE HOME.

Harry F. Olson, RCA Laboratories, Princeton, N. J.

THE WESTREX STEREO DISK SYSTEM.

C. C. Davis, Westrex Corp., Hollywood, Calif.

THREE CHANNEL STEREO AND ITS SPECIFIC PURPOSE.

W. H. Miltenburg, RCA Victor Division, New York, N. Y.

PANEL:

William S. Bachman, Columbia Records, New York, N. Y.

William H. Miltenburg, RCA Division, New York, N. Y.

Obr. 10 - Program AES Convention v New Yorku, 1957, představující stereo vinylovou desku

¹⁵ Ibid.

¹⁶ Ibid.

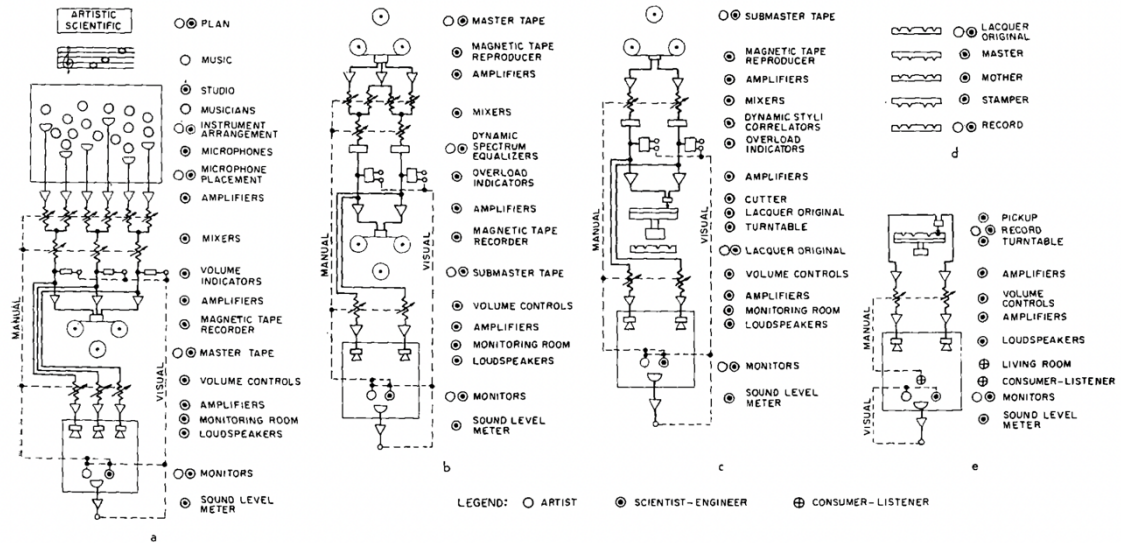


FIG. 1. DYNAGROOVE disk sound system. *a.* System for producing the master tape. *b.* System for producing the submaster tape. *c.* System for producing the lacquer original disk. *d.* Processes in the production of the disk record from the lacquer original. *e.* System for reproducing the disk record in the home.

Obr. 11 - Schéma Dynagroove systému

Dynagroove se dá považovat za předzvěst instantních masteringových služeb v tom smyslu, že šlo o automatickou kompenzaci aspektů zvuku považovaných za důležité k lepšímu přijetí nahrávky posluchači.

1.1.2. Digitální éra¹⁷

1980 - **Red Book** standard definoval podobu budoucích CD.

1982 - příchod **kompaktního disku** alias CD uspil příchod digitální éry, zásadně posunul mechanickou odolnost i zvukový potenciál nahrávek a v polovině 90. let proměnil vinyl z dominantního nosiče v nadšeneckou záležitost. Ale i když bylo cílové médium digitální, produkční řetězec předcházející lisování CD byl z většiny stále analogový.

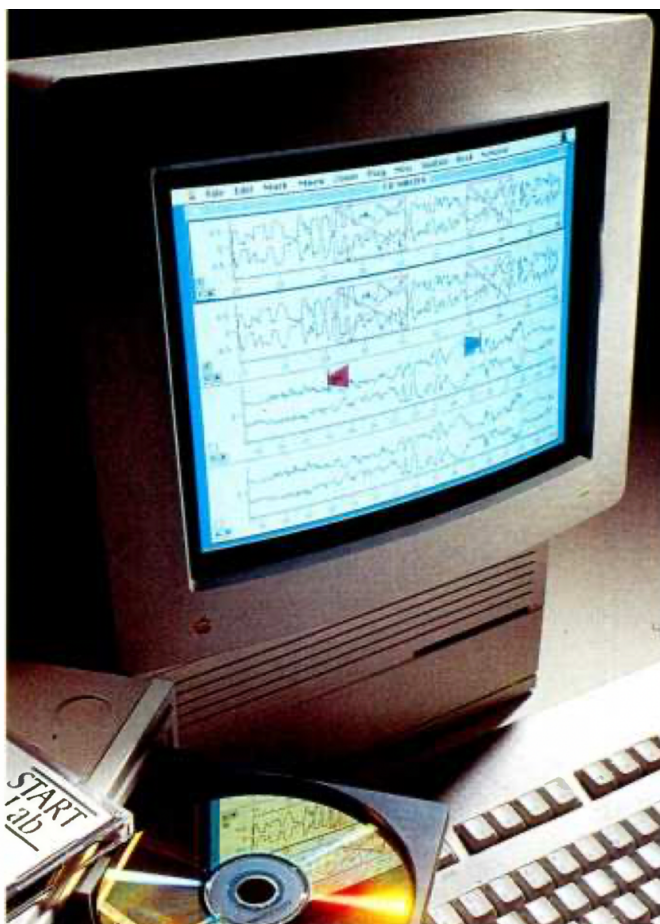


Obr. 12 - Analogová nahrávka valčíků Frederica Chopina (Philips 400 025) Claudia Arraua z března 1979 se stává prvním klasickým kompaktním diskem, který byl kdy komerčně vyroben.

1989 - **Sonic Solutions** v rámci výzkumu možností využití tehdejších počítačů pro audio produkci v reálném čase uvádí **Sonic Studio**, softwarové řešení tvorby Red Book masterů pro publikaci CD.

Sonic Solutions přineslo mnoho inovací, které dnes považujeme za samozřejmost, jako:

- NoNoise: první systém pro redukci šumu (1987);
- The Sonic System: first CD Production System, první počítačový CD authoring (1989);
- Sonic Studio HD: první plně 24 bit 96 / 192 kHz DAW (1997);



Obr. 13 - Sonic Studio

¹⁷ Ibid.

- DSD.1: první DSD Production System (1999);
- Amarra: první hi-res spotřebitelský přehrávač (2008).

1995 - internetem se začíná šířit hudba ve formátu MPEG-1 Audio Layer 3, běžně označovaném jako **.mp3**. Jeho malá velikost usnadňující šíření po tehdejším internetu způsobila revoluci v hudebním průmyslu, jejíž následky (včetně ustálení ztrátových kodeků jako distribučního standardu) pocítujeme dodnes.

2000 - současnost – CD prošlapala cestu diskům s ještě vyšší hustotou záznamu a tudíž kapacitou, jako Super Audio CD, DVD-A, Blu-ray, nebo HD DVD, nicméně stále dostupnější internetové připojení a s tím spojené sdílení souborů vedlo k tomu, že se většina zmíněných nosičů nikdy příliš neprosadila. Odpovědí na vlnu pirátství se stal až nástup dostupných streamingových služeb.

1.2. Z technika tvůrcem

Mastering tedy byla zprvu technická činnost s minimem tvůrčího vkladu spočívající v zajištění optimálního přenosu finálního mixu mezi různými médii, tedy nejčastěji z magnetického pásu na vinyl, aniž by došlo k poškození nahrávky. Práce masteringových inženýrů byla chápána jako práce dopravního strážníka nebo překladatele. Postupem času ovšem inženýři začali proces zdokonalovat s využitím technik dříve běžných spíše při mixáži – ekvalizace (zde původně ve smyslu preemfáze a deemfáze dle RIAA standardu¹⁸), či komprese dynamického rozsahu, aby se minimalizovalo riziko vyskočení přenosky z drážky při dynamických pasážích hudby. Odtud vedl už pouze krok ke kreativnímu využívání těchto nástrojů – na jedné straně zůstal stěžejním technický aspekt práce, tedy co nejtransparentnější přenos nahrávky mezi různými médii s ohledem na jejich specifika, ovšem ekvalizér, kompresor a postupem času i další nástroje dovolily masteringovým inženýrům provádět i zásahy tvůrčího charakteru, díky kterým výsledná nahrávka zněla lépe, než kdyby tímto procesem neprošla. Tento vývoj postupně vedl ke specializaci a vyhledávanosti inženýrů, jimiž zpracované nahrávky zněly nejlépe, čímž se tato profese stala tím, čím je dnes.

1.3. Mastering alba vs mastering singlů

Mastering byl tradičně spojován s prací v kontextu hudebního alba, kde bylo klíčové, aby skladby vytvářely jednotný zvukový celek, ve kterém žádná z nich nevybočovala hlasitostí, barvou nebo jinak nenarušovala jednotu zvukového charakteru.

Běžnou praxí je ovšem provádět mastering i na hudebních singlech, kde je stále benefitem poslech a případná úprava „druhým párem uší“; pro méně zkušené producenty může také jít o jakousi pečeť kvality, či externí validaci jejich zvuku.

1.4. Tradiční vs stem vs NGA mastering

Tradiční mastering probíhá na hotovém stereo mixu, kde všechna tvůrčí rozhodnutí o poměrech, barvách, prostorovosti a dynamice jednotlivých prvků již proběhla a množství stop konstituujících mix je smícháno do dvou.

S příchodem nástrojů na bázi strojového učení¹⁹ umožňujících měnit poměry jednotlivých nástrojů v již hotovém mixu se hranice mezi mixáží a masteringem dále smývá – na druhé straně ovšem dále krystalizuje fakt, že velká část masteringu je o přínosu nové perspektivy, kterou může inženýr k dílu přispět.

¹⁸ Ibid.

¹⁹ IZOTOPE. Ozone 10 Help. In: [cit. 28.05.2023]. Dostupné z: <https://docs.izotope.com/ozone10/en/master-rebalance/index.html>

Stem mastering oproti tomu pracuje místo hotového mixu s tzv. *stemy* – ty představují zpravidla submixy jednotlivých nástrojových skupin, jejichž smícháním (či součtem) docílíme tvaru, který je identický s finálním mixem.

Další okolností je fakt, že ačkoliv standardním distribučním hudebním formátem je již několik desetiletí dvoukanálové stereo, s příchodem *next-generation audio* formátů (NGA)²⁰ dojde, zdá se, konečně k masovému rozšíření zvukového média, kde počet kanálů koncového zařízení nebude nutně limitovat počet kanálů masteru nahrávky a zároveň díky konceptu objektů dojde k určitému stupni interaktivity posluchače s nahrávkou.

Vícekanálové mixy zejména populární hudby navíc svou granularitou vedou ke způsobu masteringu, který se v určitých případech může blížit spíše práci s hudebními stemy – pokud například masteringový inženýr zhodnotí, že obdržený stereo mix má příliš tichý zpěv a nelze se již vrátit do míchací fáze, obvyklá řešení mívají poměrně invazivní charakter, kdežto v případě vícekanálového mixu není výjimkou, že se zpěv nachází ve vlastním centrálním kanále a manipulace s ním je tedy nepoměrně snazší.

V době vzniku této práce stále dochází k vývoji těchto formátů, takže o budoucnosti se dá pouze spekulovat, faktem ovšem je, že po filmové postprodukci a hudební produkci se na práci s NGA formáty začínají adaptovat i masteringová studia²¹.

To, zda není stem mastering spíše mixáží, bývá otázkou přístupu masteringového inženýra – většina inženýrů zpravidla preferuje „holistický“ přístup, který podmiňuje práce na hotovém mixu, a ke stemům tak přistupuje jako k tradičnímu masteringu s tím rozdílem, že v případě potřeby může upravovat pouze určitou složku mixu²².

V této práci se budeme zabývat primárně masteringem stereo mixů, protože služby nabízející instantní mastering zatím nejsou kompatibilní s formáty s větším počtem kanálů.

1.5. Tvůrčí vklad

Pokud autor nahrávky považuje její existenci za svébytné umělecké dílo (oproti pouhé dokumentaci reality), pak se můžeme domnívat, že její subjektivní vyznění ve smyslu vnímané dynamiky a tonality a z toho plynoucí emocionální efekt jsou součástí uměleckého sdělení podobně jako samotný hudební obsah (byť ten je vždy stěžejním). Zvuková manipulace nahrávky se tudíž stává dalším prostředkem tvůrčího sdělení a přestává být pouhou technickou disciplínou naplňující exaktní kritéria stanovená obvyklými praktikami či technickými požadavky recipientů.

²⁰ EBU. 10 things you need to know about Next Generation Audio. .

²¹ *Professional Mixing and Mastering Services* [online] [cit. 28.05.2023]. Dostupné z: <https://www.audioanimals.co.uk/shop/studio-services/dolby-atmos-mastering>

²² Luca Pretolesi, *AES Mastering Academy pt.2*.

V kontextu poslechu zvuku se často setkáváme se synestetickými pojmy popisujícími subjektivní vjem, jako jsou například významové *páry tupý - ostrý, studený - teplý, temný - jasný* a další. Tyto pojmy mají svůj základ v subjektivních vlastnostech zvuku jako hlasitost nebo barva; ty jsou zase úzce provázané s jeho objektivními vlastnostmi jako je amplituda či spektrální složení.²³

Protože různé zvukové procesory jsou zpravidla navrženy tak, aby svou činností modifikovaly objektivní vlastnosti (jinými slovy soudobý ekvalizér zatím neumí číst myšlenky laického posluchače), mezi úkoly masteringového inženýra patří znalost toho, které objektivní veličiny ovlivňují které subjektivní vlastnosti, a tlumočení mezi nimi - například pokud je cílem učinit zvuk „údernějším“, sáhne pravděpodobně po nějaké formě manipulace obálky zvuku, nebo pokud je zvuk příliš „studený“, může využít například ekvalizaci.

Masteringový inženýr je tedy mediátorem, který zároveň může přispívat novými nápady a názory.

1.6. Objektivní a subjektivní parametry, nástroje a úkony masteringu

Je k diskusi, za jakých okolností je tato role nahraditelná automatickými systémy – nicméně, než diskusi otevřeme, potřebujeme poznat ještě několik dalších hledisek.

K tomu, aby automatický systém mohl konat, potřebujeme kvantifikovat jaké úkony masteringový inženýr zpravidla provádí, proto si nyní pojmenujme ty nejčastější:

- **Dynamika** – od makrodynamiky ve smyslu vzájemné úpravy hlasitosti písní v rámci alba, tak ve smyslu hlasitostí jednotlivých písňových pasáží, v mikrodynamickém smyslu pak úprava signálových špiček a práce s crest faktorem (např. kompresor, limiter, expander, transient shaper);
- **Barva** – opět jak práce s tonalitou na úrovni písní, tak zásahy chirurgičtějšího charakteru, ať už jde o potlačování nežádoucích rezonancí nebo kompenzace celkové spektrální rovnováhy, např. když mixu chybí basy, protože je mixérův poslech příliš zvýrazňoval. Mezi nástroje patří např. ekvalizér (statický, dynamický), vícepásmový kompresor, harmonická saturace apod.;
- **Prostorovost** – dojem prostorovosti do velké míry vyplývá z míry interaurální korelace čili zjednodušeně „podobnosti“ obsahu levého a pravého kanálu. Mezi nástroje patří manipulatory stereobáze, dozvukové procesory, apod.

Kromě toho provádějí masteringoví inženýři řadu dalších úkonů, mezi které patří:

- začištění začátků a konců skladeb, fade in/outy;

²³ SYROVÝ, Václav. *Hudební akustika*. Akademie múzických umění v Praze, 2013.

- v případě potřeby odstraňování nežádoucích defektů jako šum, praskání, které prošlo předešlými fázemi produkce bez povšimnutí;
- oddělení skladeb v rámci alba na CD a rozestupy mezi nimi s pomocí PQ kódů;²⁴
- zajištění ISRC kódů;²⁵
- konverze vzorkovacích frekvencí a bitové hloubky, dithering, export do nižších a ztrátových formátů pro streamovací služby apod.

1.7. Žánry

Z hlediska masteringu je podstatné, že žánry obvykle mají charakteristický „zvuk“ představující kulturně zakořeněnou konvenci, která je při zpracování nahrávky brána v potaz – ať už jejím dodržením nebo naopak cílenou ignorací.

Hudební žánr je „soubor hudebních událostí (skutečných nebo možných), jejichž průběh je řízen určitým souborem společensky přijatých pravidel.“²⁶

Jde o systematickou klasifikaci a kategorizaci hudby na základě jejich obsahových a formálních rysů; tyto žánrové kategorie slouží k identifikaci hudebních forem sdílejících společnou tradici či soubor konvencí. Žánry nám umožňují lépe porozumět a popsat různé druhy hudby, což usnadňuje komunikaci a orientaci v rozmanitosti hudebního světa.

Žánry se odlišují od *hudebních stylů* (byť v praxi často dochází k záměně těchto termínů) tím, že jde o zastřešující termín – např. *metal* je žánr, který zahrnuje styly jako *death/thrash/black metal, grindcore, apod...*)²⁷.

Představují arbitrární dělení a pro svou subjektivní povahu, stejně jako pro četné stylistické překryvy mnohých hudebních děl, může být jedna skladba zařazena mezi vícero žánrů nebo stylů.

1.8. MIR

Zatímco žánry vycházejí z kulturních konvencí a stejně jako předměty jejich vymezení procházejí v součinnosti s kulturním vývojem neustálými proměnami, pro jejich klasifikaci automatickými systémy potřebujeme rysy jednotlivých žánrů objektivně kvantifikovat.

Music information retrieval (dále MIR)²⁸ je mezidisciplinárním oborem zaměřujícím se na extrakci vhodně zvolených parametrů z hudby, na jejichž základě lze odvodit mimo jiné:

- detekci tónů, dob a tempa;

²⁴ KATZ, Robert A. *Mastering audio*, 2002.

²⁵ Ibid.

²⁶ FABRI, Franco. *What kind of Music?* . 2023.

²⁷ New Music Curriculum Guidance. In: . Dostupné z: <https://sites.google.com/site/newmusiccurriculumguidance/key-musical-concepts/1-style-genre-and-tradition>

²⁸ IŠTVÁNEK, Matěj. *Analýza interpretace hudby metodami číslicového zpracování signálu*. .

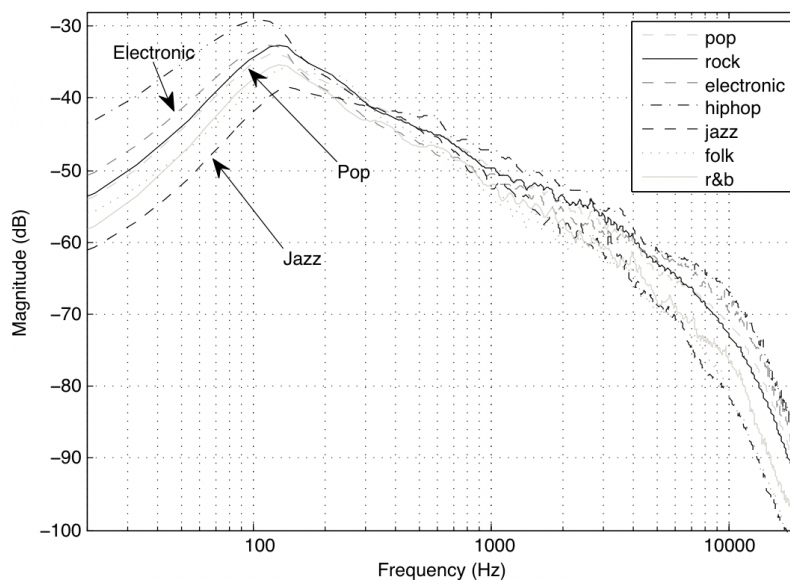
- určení žánru, akordů, nálady, skladatele, hudební epochy;
- extrakci melodie, harmonie;
- identifikaci nástrojů, separaci hudebního zdroje;
- automatickou hudební transkripci, hudební synchronizaci;
- generování hudby, partů, segmentaci.

Mezi MIR parametry relevantní z hlediska strojového učení patří mimo jiné:

- **MFCC** – viz kapitolu 2.11. o přípravě dat pro neuronovou síť;
- **Poměr energie pásem** – *band energy ratio*, vypočítá se jako energie určitého frekvenčního pásma vzhledem k celkové energii všech frekvenčních pásem; využití nalézá např. při oddělení hudby od řeči, či při klasifikaci nálady a žánru;
- **Spektrální centroid** – odhaluje těžiště, či dominantní složku frekvenčního spektra, kde je koncentrováno nejvíce energie, metrika je úzce svázána s percepčním termínem „jasnosti“, vnímáním barvy zvuku;
- **Šířka pásma** – *spectral spread*, odvozena od spektrálního centroidu; pokud je energie distribuována rovnoměrně, šířka pásma je větší, pokud je energie rozložena spíše v užších frekvenčních „pásmech“, šířka pásma je menší.

Genre	Spectral Centroid (Hz)	Spectral Crest	Spectral Slope	Peak Magnitude (dB)
Pop	868	0.0158	-0.9433	-30.58
Rock	858	0.0153	-0.9793	-30.66
Elect.	845	0.0194	-0.7461	-27.70
Hip-hop	662	0.0265	-0.8141	-22.52
Jazz	785	0.0141	-1.2929	-35.58
Folk	603	0.0191	-1.1824	-32.54
Disco	963	0.0148	-0.8042	-30.31
R&B	811	0.0149	-1.0336	-33.87
Soul	760	0.0157	-1.0303	-32.94

Obr. 14 - příklady naměřených hodnot MIR parametrů napříč žánry



Obr. 15 - zprůměrovaná frekvenční spektra napříč žánry

Pro naše účely je podstatné, že každý žánr má z časového i spektrálního hlediska nějaké charakteristické rysy, které jsou díky MIR metodám extrahovatelné do počítačově zpracovatelné formy.

2. Umělá inteligence – AI

2.1. Definice²⁹

Pro definici „umělé inteligence“ se jako logické východisko nabízí nejprve definovat inteligenci jako takovou – zde ovšem rychle narazíme na problém, protože v současnosti žádná obecně přijímaná definice neexistuje; ta se liší v závislosti na oboru, ale i na filozofickém východisku definovatele. Po zbytek práce budeme používat zkratku „AI“ pro její všeobecnou zažitost.

2.1.1. Obecné definice

„Schopnost učit se a aplikovat fakta a dovednosti, zejména když je tato schopnost široce rozvinuta...“

- Encarta World English Dictionary, 2006

„Schopnost učit se a porozumět problémům, nebo je řešit.“

- Word Central Student Dictionary, 2006

2.1.2. Definice psychologie

„Schopnost disponovat věděním.“

- V. A. C. Henmon, 1921

„Intelligence je to, co se měří testy inteligence.“ *

- E. Boring, 1923

„Intelligence není jedinou, jednolitou schopností, ale spíše souborem několika funkcí. Termín popisuje kombinaci těchto schopností nezbytných pro přežití a pokrok v určité kultuře.“

- A. Anastasi, 1992

„Intelligence je schopnost učit se, konat na základě úsudku, a využívat představivost.“

- J. Huarte

„Schopnost myslet abstraktně.“

- L. M. Terman

„Schopnost nabývat schopností.“

- H. Woodrow

„[Intelligence je] vlastnost mysli, díky které je schopna vnímat řád v chaosu.“

- R. W. Young

²⁹ LEGG, Shane a Marcus HUTTER. *A Collection of Definitions of Intelligence* [online]. arXiv, 2007 [cit. 01.08.2022]. Dostupné z: <http://arxiv.org/abs/0706.3639>

* *patří se podotknout, že tato zdánlivě duchaprostá definice vychází z deziluze nad neuchopitelností inteligence jako jediné měřitelné metriky poté, co mu byl svěřen masivní program testování vojsk v první světové válce*

2.1.3. Definice AI výzkumníků

„Dosahování cílů v komplexních prostředích.“

- B. Goertzel

„Intelligence je schopnost optimálně využít omezené zdroje – včetně času - za účelem dosažení cílů.“

- R. Kurzweil

„[...] schopnost řešit složité problémy.“

- M. Minsky

„Intelligence je schopnost infromatického systému přizpůsobit se svému prostředí s nedostatečnými znalostmi a zdroji.“

- P. Wang

2.1.4. Souhrnná definice inteligence

V současnosti žádná obecně přijímaná definice neexistuje; ta se liší v závislosti na oboru, ale i na filozofickém východisku definovatele. Výše citované definice pocházejí ze sbírky³⁰ vytvořené s cílem nalezení společných rysů týkajících se inteligence, kterými jsou:

- vlastnost individuálního agenta (entita schopna konání) vyplývající z interakce s jeho okolím;
- vlastnost spojená s agentovou schopností uspět ve vztahu k nějakému cíli;
- inteligence záleží na míře schopnosti adaptace na různé cíle a prostředí.

Syntézou těchto znaků je následující definice:

„Intelligence zkoumá agentovu schopnost dosahovat cílů napříč různorodými prostředími.“

- S. Legg and M. Hutter³¹

2.2. Definice AI, Tesslerův teorém

„Umělá inteligence“ je infromatickou disciplínou zabývající se studiem a tvorbou programů či strojů, které by se daly označit za „inteligentní“³² – takové, které napodobují kognitivní funkce obvykle spojované s lidskou myslí, jako např. schopnost *řešit problémy a učit se*. Tyto stroje lze označovat také jako „*racionální agenty*“.

V 70. letech formuloval Larry Tessler myšlenku, která se později stala známou jako „Tesslerův teorém“, nebo „AI efekt“: „AI je to, co ještě nebylo naprogramováno.“

Tessler tím naráží na okolnosti, že jakmile se podaří s pomocí programu nasimulovat nějakou schopnost dříve vyhrazenou pouze lidskému intelektu, lidé ji brzy přestanou považovat za

³⁰ LEGG, Shane a Marcus HUTTER. *A Collection of Definitions of Intelligence*.

³¹ Ibid.

³² Machine Learning Glossary. In: *Google for Developers* [online] [cit. 28.05.2023]. Dostupné z: <https://developers.google.com/machine-learning/glossary>

nezbytnou složku *skutečného myšlení*. Opravdové jádro inteligence se vždy nachází v *té další věci*, která zatím nebyla naprogramována.³³

2.3. AI techniky

Na poli umělé inteligence figuruje množství technik³⁴:

- **evoluční algoritmy** - využívají modelů evolučních procesů k nalezení řešení náročných a rozsáhlých úloh;
- **expertní systémy** - řešící problémy na základě pevně daných pravidel a znalostní báze;
- **A*** - algoritmus hledání cesty používaný k určení nejlepší cesty v daném prostředí;
- **strojové učení** - (*machine learning, ML*) - sada metod, které počítače používají k vytváření a vylepšování předpovědí nebo chování na základě dat; systémy na této bázi řeší problémy bez explicitních instrukcí a mohou se učit z dat.

Aktuálně nejrozšířenější metodou strojového učení je takzvaný „**deep learning**“ (hluboké učení), což je hybridní termín pro „machine learning“ využívající „deep neural networks“, čili neuronové sítě s množstvím skrytých vrstev neuronů mezi vstupní a výstupní vrstvou.

2.4. Historie neuronových sítí (NS)

V roce 1957 navrhl Frank Rosenblatt z Cornellovy univerzity *perceptron*, primitivního jednovrstvého předchůdce NS se schopností analyzovat obraz s rozlišením 20x20 a kapacitou rozeznat jednoduché tvary.

Počáteční vlna rychlého rozvoje neuronových sítí (a dalších technik na poli umělé inteligence) postupně začala narážet na limity jednak konceptuální, jednak na limity tehdy dostupného výpočetního výkonu, což vedlo k období stagnace označovanému jako „AI zima“.³⁵

V roce 1986 byl zájem znovu oživen zdokonalením učební metody zpětným šířením chyby s využitím částečných derivací (*backpropagation algorithm*), díky čemuž začalo být možné efektivně trénovat hluboké neuronové sítě až s desítkami vrstev. Neuronové sítě se tedy opět staly relevantními, nicméně stále byly výpočetně náročné a zdálo se, že hloubka přesahující pět vrstev již přináší pouze zanedbatelná zlepšení.³⁶

³³ HOFSTADTER, Douglas R. *Gödel, Escher, Bach: an eternal golden braid*. New York: Basic Books, 1999.

³⁴ VELARDO, Valerio. 2- AI, machine learning and deep learning - YouTube. In: [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/>

³⁵ RUSSELL, Stuart J. a Peter NORVIG. *Artificial intelligence: a modern approach*. Hoboken: Pearson, 2021.

³⁶ RUMELHART, David E., Geoffrey E. HINTON a Ronald J. WILLIAMS. Learning representations by back-propagating errors. *Nature*. 1986, roč. 323, č. 6088. DOI: 10.1038/323533a0

Zlom přišel v roce 2012 – neuronová síť s přezdívkou AlexNet³⁷ přesvědčivě vyhrála soutěž v rozpoznávání obrázků ImageNet, jejímž zadáním bylo roztřídit cca 1 milion popsaných obrázků do cca 1000 definovaných kategorií. Mechanismy, na kterých postavil AlexNet svůj úspěch, byly známy již relativně dlouho (jde o konvoluční neuronovou síť), ovšem klíčovou se stala jejich souhra – kombinace obrovského množství tréninkových dat společně s dostupným výpočetním výkonem.

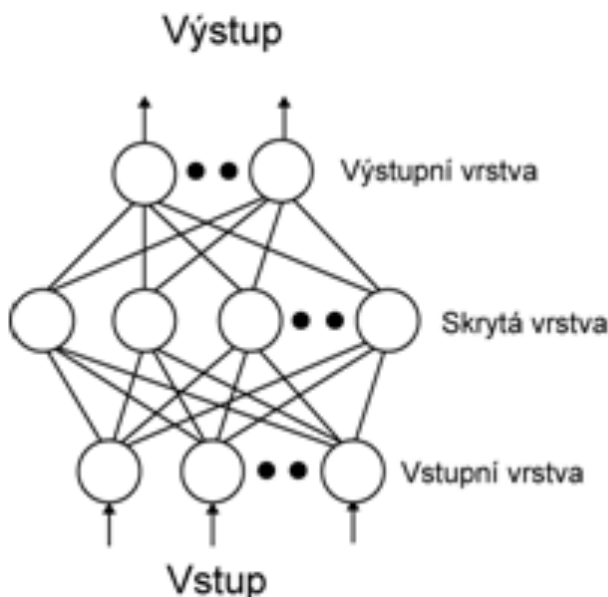
Protože je většina služeb a nástrojů pro automatické zpracování zvuku v současnosti založena na strojovém učení, konkrétně na neuronových sítích, budeme se v této práci věnovat právě jim.

³⁷ KRIZHEVSKY, Alex, Ilya SUTSKEVER a Geoffrey E HINTON. *ImageNet Classification with Deep Convolutional Neural Networks* [online]. Curran Associates, Inc., 2012 [cit. 28.05.2023]. Dostupné z: https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

2.5. Neuronové sítě

„...počítačový systém tvořen množstvím jednoduchých vzájemně propojených prvků zpracovávajících informace dynamickými změnami vlastního stavu v reakci na vnější podněty.“

- Dr Robert Hecht-Nielsen³⁸



Obr. 16 - zjednodušené schéma vrstev neuronů v NS

Umělé neuronové sítě jsou výpočetní modely volně inspirované nervovou strukturou mozku. Oproti svým biologickým předobrazům jsou menší – zatímco mozek obsahuje miliardy nervových buněk, umělé neuronové sítě se pohybují v „pouhých“ desítkách až stovkách tisíc neuronů.³⁹

2.6. Tradiční strojové učení a neuronové sítě

Jak jsme již naznačili v odstavci o technikách strojového učení, zatímco ostatní techniky AI provádějí úkony na datech na základě předem naprogramovaných pravidel, systémy strojového učení jsou postaveny před množstvím tréninkových dat a jejich úkolem je vyvodit pravidla právě na jejich základě. Jde vlastně o aproximátor funkcí proto, že trénink vychází z předpokladu existence zákonitostí určujících souvislost mezi výchozími daty a správnými výsledky.

Tradiční ML techniky extrahují parametry z relativně malých datasetů, jsou výpočetně relativně nenáročné a dají se nejlépe aplikovat na problémy s přímočarým řešením, zatímco deep

³⁸ CAUDILL, Maureen. Neural Network Primer: Part I. *AI Expert*. 1989.

³⁹ A Basic Introduction To Neural Networks. In: [cit. 28.05.2023]. Dostupné z: <https://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html>

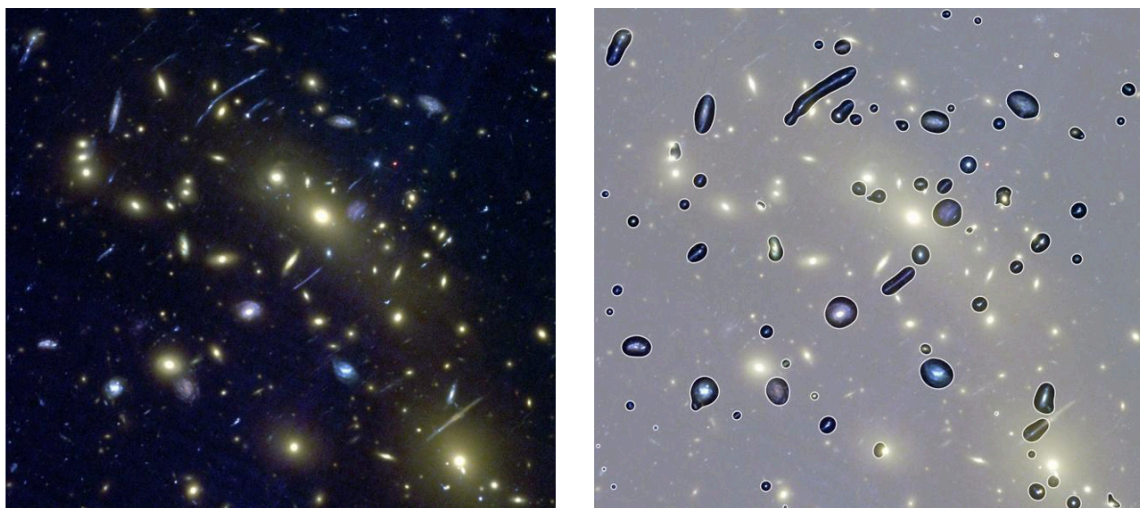
learning vyžaduje velké množství tréninkových dat, výpočetního výkonu a nejlépe se hodí k řešení komplexních problémů.

Nastiňme si základy vnitřního fungování deep learningu. S ohledem na zkoumání tématu z pohledu tvůrčí profese se nebudeme do hloubky zabývat matematickou stránkou věci.

2.7. Tréninkové strategie neuronových sítí

2.7.1. Učení se bez učitele

V tomto případě probíhá trénování modelu na základě učebních dat zcela prostých požadovaných výstupních informací – to dává smysl v případě algoritmů navržených s cílem hledat vzorce či anomálie v množině informací, jako jsou dosud neobjevené galaxie na snímcích pořízených teleskopy⁴⁰.



Obr. 17 - automatické pátrání po dosud neobjevených galaxiích

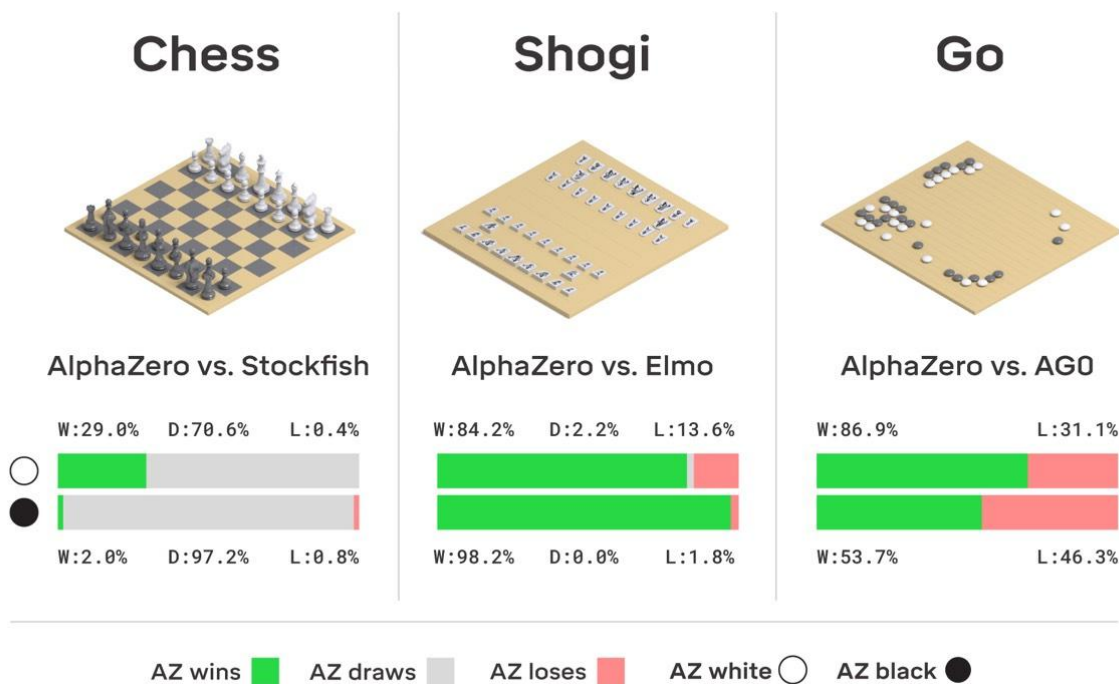
2.7.2. Zpětnovazební učení

Trénink probíhá ve virtuálním prostředí prostřednictvím trestů a odměn. Podobně jako lze domestikovaná zvířata cvičit jejich odměňováním za správně splněné úkoly, agenta lze odměnit jedničkami a nulami a posunout jeho učení kýženým směrem. Zpětnovazební učení je výborně aplikovatelné v případech, kdy lze problém „gamifikovat“, tedy popsat formou hry

⁴⁰ STAFF, News. Astronomers Teach AI to ‘See’ Astronomical Images | Sci.News. In: *Sci.News: Breaking Science News* [online]. 8. 7. 2015 [cit. 28.05.2023]. Dostupné z: <https://www.sci.news/astronomy/science-ai-unsupervised-learning-algorithm-02998.html>

s přesně formulovanými výherními podmínkami a s jasnými pravidly a parametry, s pomocí kterých lze konat.

Příkladem může být AlphaZero, program na bázi zpětnovazebního učení, hrající hry jako šachy, go nebo šógi na vysoké úrovni bez explicitně naprogramovaných pravidel či strategií. AlphaZero opakovaně hraje proti sobě samému, postupně se učí a zdokonaluje své strategie prostřednictvím zpětnovazebního učení.



Obr. 18 – výsledky učení AlphaZero

2.7.3. Učení s učitelem

V tomto případě je model vytvářen na základě dat, která obsahují jak vstupní informace, tak požadované výsledky. Pokud bychom například chtěli vytvořit model rozpoznávající, zda je na obrázku žába, učební data by obsahovala jak obrázky s tímto živočichem, tak obrázky bez něj a každý z nich by zároveň obsahoval informaci o jeho přítomnosti, či absenci. V případě kombinace učení s učitelem i bez něj (*semi-supervised learning*) tyto informace obsahuje pouze část učebních dat.

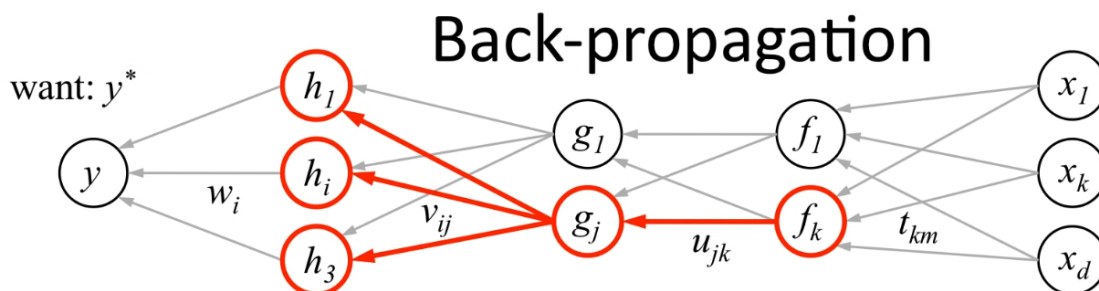
2.8. Algoritmus zpětného šíření chyby v učení NS

Neuronové sítě se skládají z vrstev uzlů neboli neuronů, které představují nejmenší výpočetní jednotku. Rozlišujeme vrstvu vstupní (neurony přijímající data), jednu nebo více skrytých vrstev (termín *hluboké učení* je odvozen od „hloubky“ neuronové sítě, dané právě množstvím těchto vrstev) a výstupní vrstvu, jejíž prostřednictvím se dozvídáme výsledek výpočtu. Neurony v rámci každé vrstvy nabývají určitou hodnotu a jsou propojeny se všemi neurony následující

vrstvy. Tato spojení mají různou váhu, která ovlivňuje jak moc se podepíšou na výsledné hodnotě aktivace následujícího neuronu. Ta je určena součtem předchozích aktivací násobených váhami, které slouží jako vstupní hodnota pro aktivační funkci následující buňky. Pro zhodnocení správnosti výkonu sítě se používá tzv. **chybová funkce**, která v podstatě vyjadřuje rozdíl mezi výslednými a ideálními hodnotami – čím je její hodnota vyšší, tím horší výkon síť podala.

Naším cílem je hodnotu funkce minimalizovat, což matematicky představuje hledání jejího minima. Výpočtem gradientu této funkce v každém uzlu sítě lze určit, které váhy je potřeba upravit, aby se výsledná hodnota snížila (a síť tím pádem podala lepší výkon).

Na základě chybové funkce pak probíhá **algoritmus zpětného šíření chyby**, který s využitím částečných derivací upravuje váhy pro minimalizaci hodnoty chybové funkce. Výsledná konfigurace sítě je pak ona pověstná „černá skříňka“, která může být použita pro reálné nasazení.



1. receive new observation $\mathbf{x} = [x_1 \dots x_d]$ and target y^*
2. **feed forward:** for each unit g_j in each layer $1 \dots L$ compute g_j based on units f_k from previous layer: $g_j = \sigma \left(u_{j0} + \sum_k u_{jk} f_k \right)$
3. get prediction y and error $(y - y^*)$
4. **back-propagate error:** for each unit g_j in each layer $L \dots 1$

<p>(a) compute error on g_j</p> $\frac{\partial E}{\partial g_j} = \sum_i \underbrace{\sigma'(h_i)}_{\substack{\text{should } g_j \\ \text{be higher} \\ \text{or lower?}}} \underbrace{v_{ij}}_{\substack{\text{how } h_i \text{ will} \\ \text{change as} \\ g_j \text{ changes}}} \underbrace{\frac{\partial E}{\partial h_i}}_{\substack{\text{was } h_i \text{ too} \\ \text{high or} \\ \text{too low?}}}$	<p>(b) for each u_{jk} that affects g_j</p> <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>(i) compute error on u_{jk}</p> $\frac{\partial E}{\partial u_{jk}} = \frac{\partial E}{\partial g_j} \underbrace{\sigma'(g_j)}_{\substack{\text{do we want } g_j \text{ to} \\ \text{be higher/lower}}} \underbrace{f_k}_{\substack{\text{how } g_j \text{ will change} \\ \text{if } u_{jk} \text{ is higher/lower}}}$ </div> <div style="width: 45%;"> <p>(ii) update the weight</p> $u_{jk} \leftarrow u_{jk} - \eta \frac{\partial E}{\partial u_{jk}}$ </div> </div>
---	---

Copyright © 2014 Victor Lavrenko

Obr. 19 - algoritmus zpětného šíření chyby v NS

Síť, kterou jsme popsali, je plně propojená a dopředná (*feedforward*), což z ní činí tzv. vícevrstvý perceptron (*MLP*).

2.9. Další druhy neuronových sítí

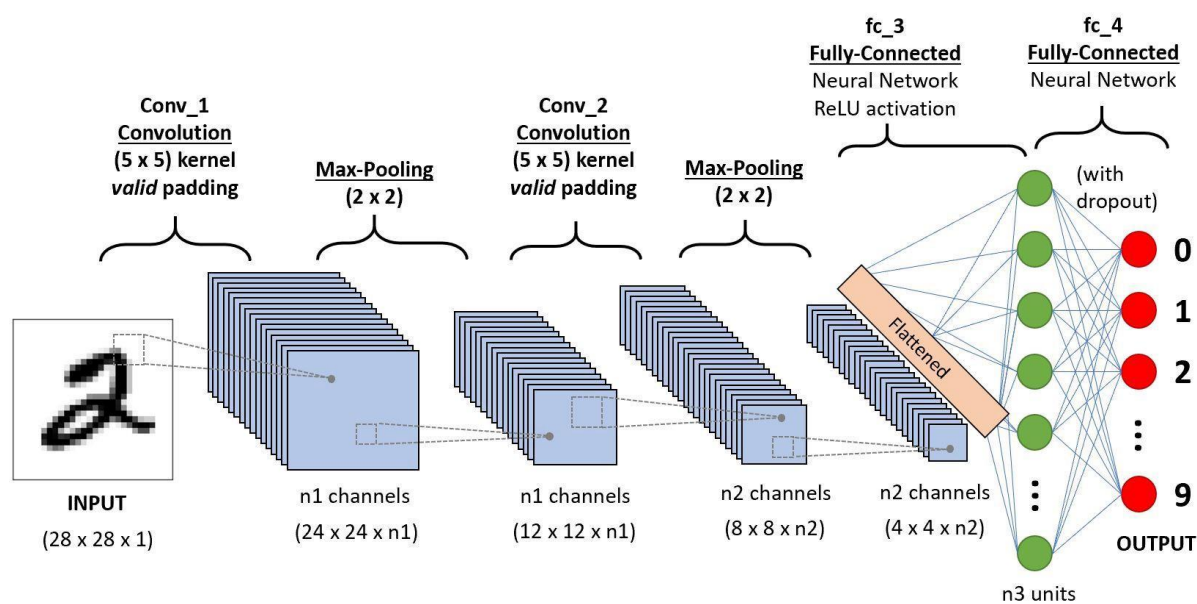
2.9.1. Konvoluční neuronové síť (*CNN*)

Tato architektura funguje v určitých ohledech podobně jako centrum lidského mozku zpracovávající obrazový vjem. Provádí konvoluční operace, které snižují dimenzionalitu přicházejících dat a zároveň extrahují vlastnosti důležité pro klasifikaci. Tato architektura dokáže zachytit prostorové a časové závislosti ve vstupních datech.

CNN se ukázaly jako velmi účinné při práci s grafickými daty (lepší než vícevrstvé perceptrony), protože obvykle obsahují struktury, které síť dokáže detekovat, navíc mají obvykle méně parametrů než MLP vrstvy.

CNN za sebe skládá série konvolučních a sdružujících vrstev - síť nejprve extrahuje jednoduché vlastnosti jako jsou hrany apod., a každá následující vrstva rozpoznává čím dál obecnější aspekty – například „toto je žába“.

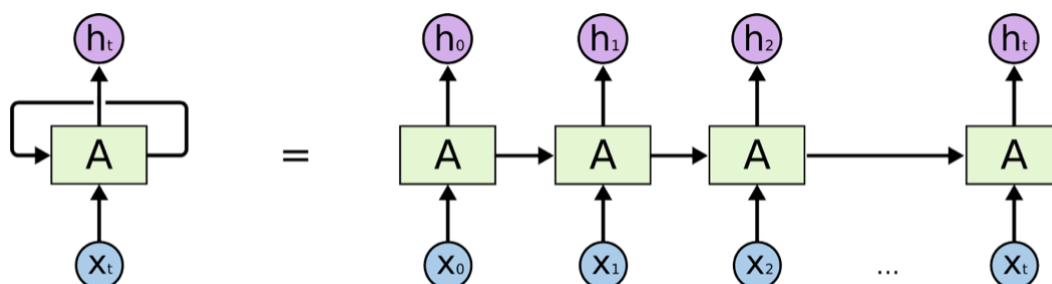
Jak ovšem neuronová síť vhodná pro zprostředkování strojového vidění souvisí se zvukem? Jednoduše – protože MFCC (mel-frekvenční koeficienty, viz kapitola 2.11.1) jsou ve své podstatě grafickými daty, mohou být interpretovány jako multidimenzionální matice ideální pro vstupní vrstvu neuronové sítě.



Obr. 20 - architektura konvoluční NS

2.9.2. Rekurentní neuronové síť (RNN)⁴¹

Lidské myšlení nezačíná každou vteřinu od nuly. Při četbě tohoto textu chápete každé slovo v kontextu těch předcházejících; vaše myšlenkové pochody jsou do jisté míry persistentní – narozdíl od tradičních MLP. V RNN se snažíme tento problém obejít využitím smyčky na úrovni neuronů, díky které má právě vypočítaná hodnota aktivace vliv i na budoucí aktivaci stejného neuronu.



Obr. 21 - paměťová buňka v RNN

Tyto sítě jsou vhodnější pro práci se sekvenciálními daty, kde má jejich pořadí svůj význam. Jsou navrženy tak, aby zpracovávaly vstupní sekvence proměnlivé délky, dokáží lépe respektovat dlouhodobé závislosti v datech, což je vhodné pro úlohy, které vyžadují kontext nebo paměť, jako je zpracování a generování sekvencí dat, kde aktuální výstup závisí na předchozích vstupech a aktuálním vstupu – typicky zpracování přirozeného jazyka (NLP – *natural language processing*), rozpoznávání řeči, generování hudby a predikce časových řad.

Achillovou patou RNN je ovšem relativně krátká paměť a náchylnost k problému **mizejícího gradientu**, což je situace, kdy se gradient (sklon křivky chybové funkce) při zpětné propagaci postupně zmenšuje a přibližuje se k nule. Když to nastane, váhy se přestávají aktualizovat a síť se přestává učit – pokud se podařilo nalézt globální minimum, síť je v neefektivnější konfiguraci, ovšem pokud se proces zastavil v jednom z lokálních minim, efektivita tréninku je pouze částečná.

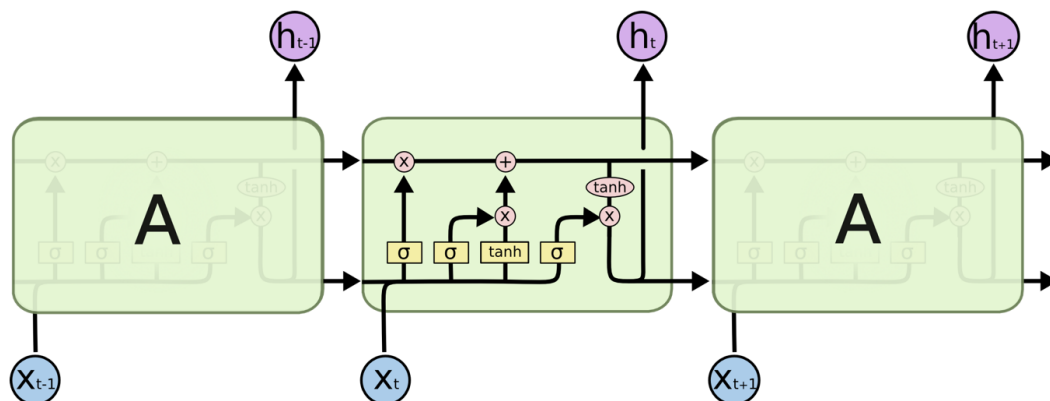
2.9.3. Long-term short-term síť (LSTM)⁴²

Tyto sítě jsou typem RNN, které jsou navrženy mimo jiné s ohledem na problém mizejícího gradientu a jsou efektivnější ve zpracování sekvencí s dlouhodobými závislostmi. LSTM mají oproti RNN navíc paměťovou buňku ukládající informace po relativně dlouhou dobu, která umožňuje síti selektivně zapomínat nebo si ukládat informace na základě vstupních dat.

⁴¹ 17- *Recurrent Neural Networks Explained Easily* [online]. 2020 [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/watch?v=DY82Goknf0s>

⁴² 18- *Long Short Term Memory (LSTM) Networks Explained Easily* [online]. 2020 [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/watch?v=eCvz-kB4yko>

LSTM jsou komplexnější než tradiční RNN, a tudíž bývají náročnější na trénink z hlediska výpočetního výkonu.

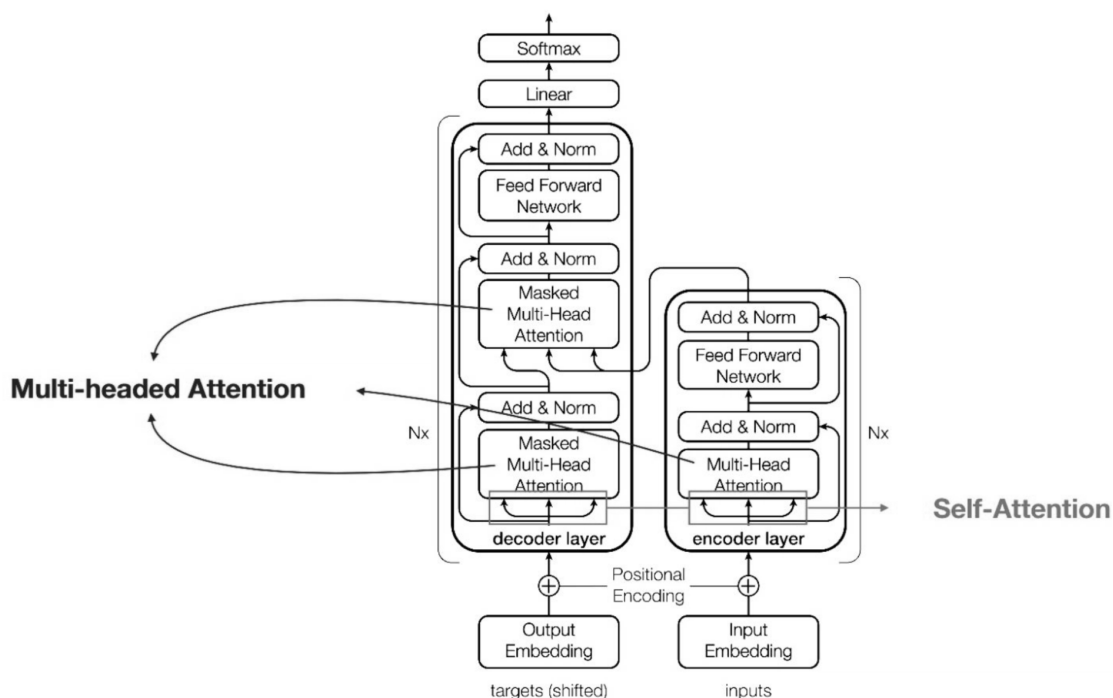


Obr. 22 - interagující vrstvy v rámci paměťové buňky LSTM

2.9.4. Transformer⁴³

Transformer je typ neuronové sítě používaný pro zpracování sekvenčních dat, ovšem narozdíl od tradičních rekurentních neuronových sítí nepoužívá rekurentní architekturu, ale místo toho se spoléhá na mechanismus **pozornosti**, který umožňuje modelu přidělit větší váhu specifickým vstupům v závislosti na jejich významu pro výstup. Díky tomu transformer pracuje s dlouhodobými závislostmi v datech ještě lépe než LSTM.

⁴³ VASWANI, Ashish et al. *Attention Is All You Need* [online]. arXiv, 2017 [cit. 28.05.2023]. Dostupné z: <http://arxiv.org/abs/1706.03762>



Obr. 23 - Schéma transformeru

Na rozdíl od MLP, který je typickou dopřednou neuronovou sítí, je transformer založený na tzv. enkodér-dekodér architektuře. Enkodér slouží k převodu vstupní sekvence na vektor, zatímco dekodér slouží k dekodování výsledného vektoru na výstupní sekvenci. Tento model se používá například při strojovém překladu mezi jazyky, kde se přeložený text generuje postupně pomocí dekodéru. Transformer je v oblasti zpracování přirozeného jazyka a strojového vidění efektivnější (zejména díky lepší možnosti paralelizace tréninku) než dřívější RNN, které od svého popsání v roce 2017 postupně vytlačuje; je mj. základem pro GPT a ostatní velké jazykové modely (LLM).

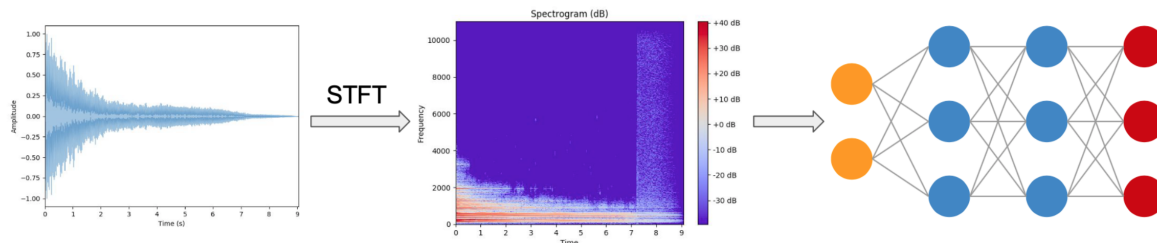
2.10. Přetrénování (*overfitting*)⁴⁴

Takto označujeme situaci, kdy je model příliš dobře vycvičen na tréninkových datech, takže ztrácí schopnost generalizovat nová, nikdy neviděná data, což popírá smysl jeho existence. Overfittingu lze během trénování předejít různými technikami, jako je regularizace, brzké zastavení (*early stopping*) a rozšíření tréninkových dat. Vědět, kdy použít kterou techniku, je součástí „alchymie“ práce s NS.

⁴⁴ 14- SOLVING OVERFITTING in neural networks - YouTube. In: [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/>

2.11. Příprava dat pro zpracování neuronovou sítí⁴⁵

Pro zpracování neuronovou sítí musí být každý signál zpracován do specifického formátu dovolujícím jeho použití v roli vstupních dat.



Obr. 24 - extrakce zvukových parametrů pro NS

V tradičních technikách strojového učení bývalo zapotřebí provádět velké množství extrakce zpracovatelných parametrů. Ovšem protože princip neuronových sítí spočívá v tom, že si důležité parametry sítě vyvodí samy v průběhu učebního procesu, komplexita a nutnost jejich předběžné extrakce je tím značně snížena.

Spektrogram stále není vhodnou formou pro zpracování signálu v rámci neuronové sítě, k tomu je zapotřebí ještě několik kroků. Lidský sluch nefunguje lineárně: mezi tóny A1 (55 Hz) a A2 (110 Hz) vnímáme stejný rozdíl ve výšce jako mezi A4 (440 Hz) a A5 (880 Hz). Oba tyto intervaly představují oktávy, ale rozdíl mezi tóny té první je 55 Hz, zatímco pro druhou oktávu je to 440 Hz. To řeší *mel-stupnice* definující frekvenční pásma, která jsou rovnoměrně rozložena vzhledem k vnímaným frekvencím, tudíž banka filtrů založená na této stupnici poskytuje oproti standardní logaritmické stupnici vyšší rozlišení na nižších frekvencích a nižší rozlišení na vyšších frekvencích podobně, jako v případě lidských uší.

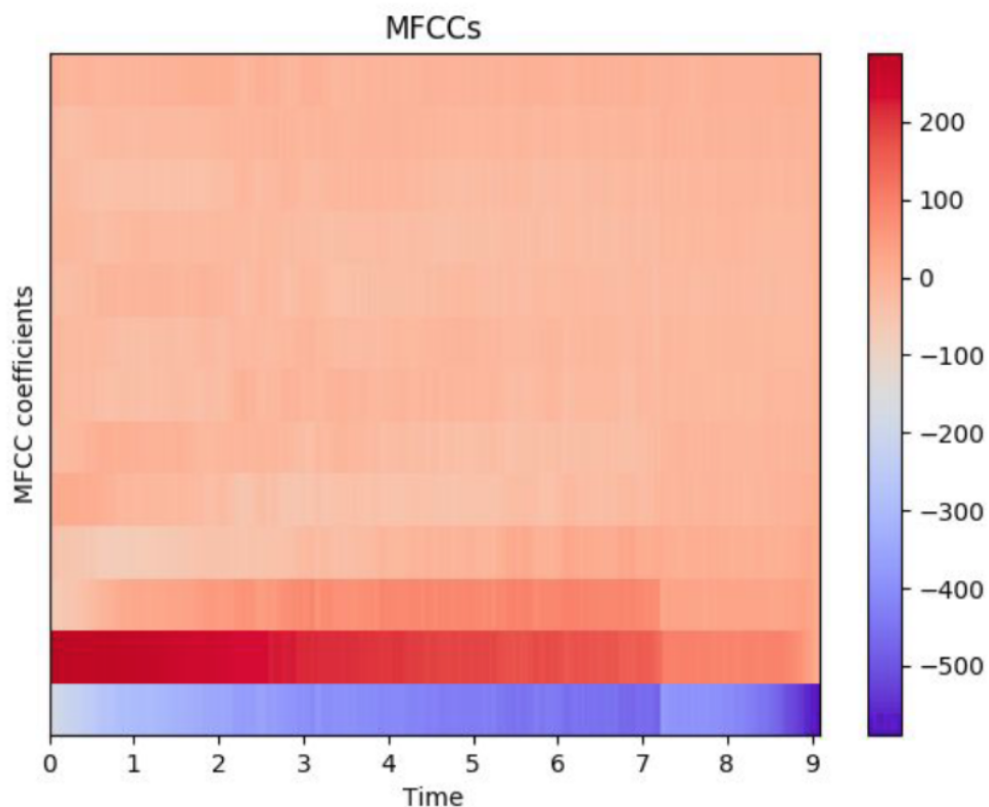
2.11.1. Mel-frekvenční keprální koeficient – MFCC^{46,47}

MFCC je reprezentací vlastností zvukového signálu beroucí v úvahu nelinearitu a maskovací vlastnosti lidského sluchu a současně skrze spektrální rozložení energie hudebního signálu dobře popisuje timbrální či texturální rysy zvuku formou praktickou pro počítačové zpracování. MFCC jsou již dlouhou dobu používány při zpracování lidské řeči, avšak své využití najdou i při zpracování hudebního signálu.

⁴⁵ How to Extract Audio Features - YouTube. In: [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/watch?v=8A-W1xk7qs8&t=500s>

⁴⁶ IŠTVÁNEK, Matěj. *Analýza interpretace hudby metodami číslicového zpracování signálu.*

⁴⁷ Mel-Frequency Cepstral Coefficients Explained Easily - YouTube. In: [cit. 28.05.2023]. Dostupné z: https://www.youtube.com/watch?v=4_SH2nfbQZ8



Obr. 25 - MFCC

Jejich původní využití bylo v oblasti rozpoznávání řeči, ale svou užitečnost prokázaly také v klasifikaci hudebních žánrů a nástrojů. MFCC jsou užitečnými deskriptory větších spektrálních struktur, které jsou snadno čitelné a „interpretovatelné“ neuronovými sítěmi.

2.12. Deep learning a zvuk⁴⁸

Umělé neuronové sítě nabízejí řadu využití souvisejících se zvukem:

- rozpoznávání, detekce řeči;
- rozpoznávání emocí na základě hlasu;
- rozpoznávání hluku, nežádoucích složek signálu;
- klasifikace hudebních žánrů, nástrojů, nálady;
- generování zvuku, hudby;
 - ať už VAE (*variational autoencoder*) s predikcí následujícího samplu (DADABOTS, OpenAI Jukebox), difúzní modely tvořící spektrogramy (Riffusion) nebo ještě lépe jejich kombinace s LLM (Google MusicLM).

⁴⁸ VELARDO, Valerio. 2- AI, machine learning and deep learning - YouTube.

3. Instantní mastering, Intelligent music production

Instantním masteringem míníme „automatický algoritmičtý mastering založený na strojovém rozpoznávání *patternů*, volbě adekvátních processingových řetězců a jejich automatické aplikaci“.⁴⁹

V posledních několika letech přišlo na trh několik služeb nabízejících instantní mastering, mezi nimi:

- LANDR (2014)⁵⁰
- Emastered (2015)⁵¹
- CloudBounce (2016)⁵²
- MajorDecibel (2016)⁵³
- BandLab Mastering (2016)⁵⁴
- MasteringBox (2015)⁵⁵
- a řada dalších.

Jak tyto služby skutečně fungují (nebo jakým způsobem konkrétně využívají strojové učení) není jasné, protože jde o předmět obchodního tajemství; nicméně jak jsme již nastínili v předcházejících kapitolách, řetězec masteringových inženýrů se obvykle skládá z variant tradičních procesorů – v nějaké formě téměř vždy dojde k ekvalizaci, kompresi a limitaci. Práce s nimi pak znamená nastavování jejich parametrů na hodnoty vedoucí ke kýženému výsledku, podmíněnému inženýrovým úsudkem založeným na snaze najít shodu mezi objektivními a cílovými subjektivními parametry zvuku.

V uplynulých letech si lze všimnout šíření nástrojů pro hudební produkci explicitně využívajících AI techniky, potažmo techniky strojového učení - za průkopníka lze v tomto

⁴⁹ ELLIOTT, Mitchell a Song Hui CHON. A Comparative Study of Music Mastered by Human Engineers and Automated Services. *Journal of the Audio Engineering Society*. 2022, roč. 70, č. 9. DOI: 10.17743/jaes.2022.0050

⁵⁰ LANDR: Creative Tools for Musicians. In: *LANDR* [online]. 20. 3. 2023 [cit. 28.05.2023]. Dostupné z: <https://www.landr.com/>

⁵¹ Online Audio Mastering by Grammy Winning Engineers | eMastered. In: [cit. 28.05.2023]. Dostupné z: <https://eMastered.com>

⁵² CloudBounce - Instant Audio Mastering. Just hear it. In: [cit. 28.05.2023]. Dostupné z: <https://www.cloudbounce.com/>

⁵³ MajorDecibel | Instant Online Audio Mastering Service. In: [cit. 28.05.2023]. Dostupné z: <https://majordecibel.com/>

⁵⁴ BandLab | Unlimited FREE Instant Online Audio Mastering. In: *BandLab* [online] [cit. 28.05.2023]. Dostupné z: <https://www.bandlab.com/mastering>

⁵⁵ Free on-line Mastering. In: *MasteringBOX* [online] [cit. 28.05.2023]. Dostupné z: <https://www.masteringbox.com/>

ohledu považovat společnost iZotope⁵⁶ a její produktové řady Neutron⁵⁷, Ozone⁵⁸ a RX⁵⁹ pro usnadnění mixáže, masteringu a restaurování zvuku; dalším významným hráčem je pak společnost Sonible⁶⁰ se svou řadou tradičních procesorů obohacených o „inteligentní“ prvky, jako Smart EQ 3⁶¹ a další, které mají za cíl právě usnadnění hledání shody mezi objektivními a subjektivními parametry.

V možnostech, parametrech a kritériích automatických procesů v hudební produkci nám může pomoci vznikající výzkumný obor „Intelligent music production“, (dále IMP)⁶². Ačkoliv je jeho primárním zaměřením zejména disciplína hudební mixáže, pro účely práce je IMP relevantní z důvodu fundamentálně stejných stavebních bloků processingových řetězců v kontextu mixáže i masteringu, díky čemuž nám může poskytnout potřebný myšlenkový rámec spojující jak tradiční masteringové postupy, tak možnosti automatizace vyplývající z deep learningu a také podklady k odbornému odhadu principů fungování instantních masteringových služeb.

IMP systémy lze dělit z hlediska míry kontroly nad úpravami zvuku, reprezentace znalostí nebo způsobu manipulace se zvukem. Může jít o asistivní systémy, jako jsou právě výše jmenované produkty od iZotope a Sonible, které na základě analýzy zvuku podávají uživateli návrhy na úpravy parametrů procesorů, nebo na opačném spektru o plně automatické systémy s minimem uživatelské kontroly, jako jsou videoherní zvukové mixy – na pomezí těchto extrémů nalezneme instantní mastering.

3.1. Míra kontroly a varianty inteligentních prvků IMP systémů

Podoba inteligentních prvků IMP systémů závisí na míře kontroly, která má systému být svěřena; podle toho lze dle Moffata a Sandlera⁶³ rozlišit několik variant.

3.1.1. Poznatkový

Přístup založený na poznacích poskytuje uživateli největší úroveň kontroly, přínos IMP systému zde spočívá v poskytování nové perspektivy nebo návrhu dalšího postupu. V této situaci má systém nejnižší míru kontroly nad situací; obvykle neprovádí zásahy do zvuku samotného, ale cílem je zvýšení informovanosti uživatele skrze informační prvky, jako jsou

⁵⁶ iZotope and Assistive Audio Technology. In: *iZotope* [online] [cit. 28.05.2023]. Dostupné z: <https://www.izotope.com/en/learn/izotope-and-assistive-audio-technology.html>

⁵⁷ IZOTOPE. Neutron 4 Help Documentation. In: [cit. 28.05.2023]. Dostupné z: <https://s3.amazonaws.com/izotopedownloads/docs/neutron4/en/index.html>

⁵⁸ IZOTOPE. *Ozone 10 Help*.

⁵⁹ IZOTOPE. RX 10 Help. In: [cit. 28.05.2023]. Dostupné z: <https://docs.izotope.com/rx10/en/dialogue-isolate/index.html>

⁶⁰ *sonible plug-ins: Artificial Intelligence inside* [online] [cit. 28.05.2023]. Dostupné z: <https://www.sonible.com/blog/plugins-with-ai/>

⁶¹ *smart:EQ 3 - the intelligent equalizer by sonible - Get your 30-day trial now* [online] [cit. 28.05.2023]. Dostupné z: <https://www.sonible.com/smarteq3/>

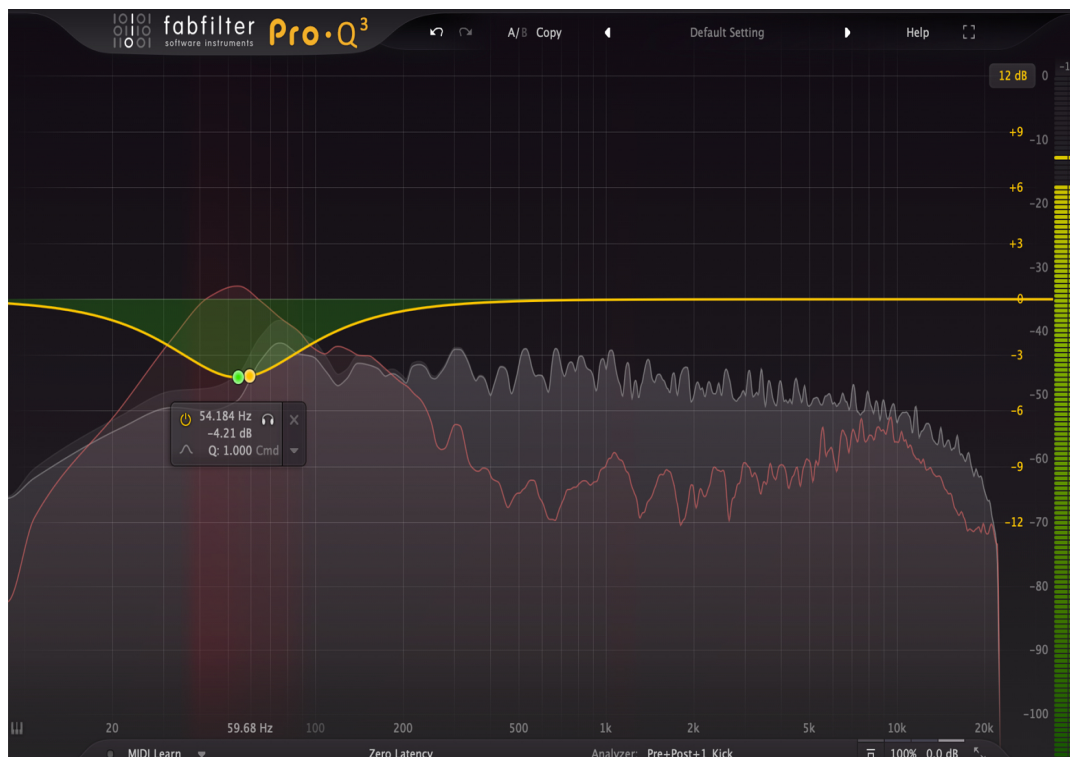
⁶² BRECHT DEMAN, RYAN STABLES, a JOSHUA D. REISS. Intelligent Music Production. .

⁶³ MOFFAT, David a Mark B. SANDLER. Approaches in Intelligent Music Production. *Arts*. 2019, roč. 8, č. 4. DOI: 10.3390/arts8040125

různé druhy vizualizace signálu apod. Do této kategorie bychom mohli zařadit například „Metric AB” od ADAPTR Audio - plugin nabízející širokou škálu vizualizací a metrik vstupního signálu společně s jejich srovnáváním s řadou uživatelsky volitelných referencí; nebo například ukazatele frekvenčního maskování v současných ekvalizérech, jejichž instance jsou schopny vzájemné komunikace napříč stopami.



Obr. 26 - Metric AB od ADAPTR Audio



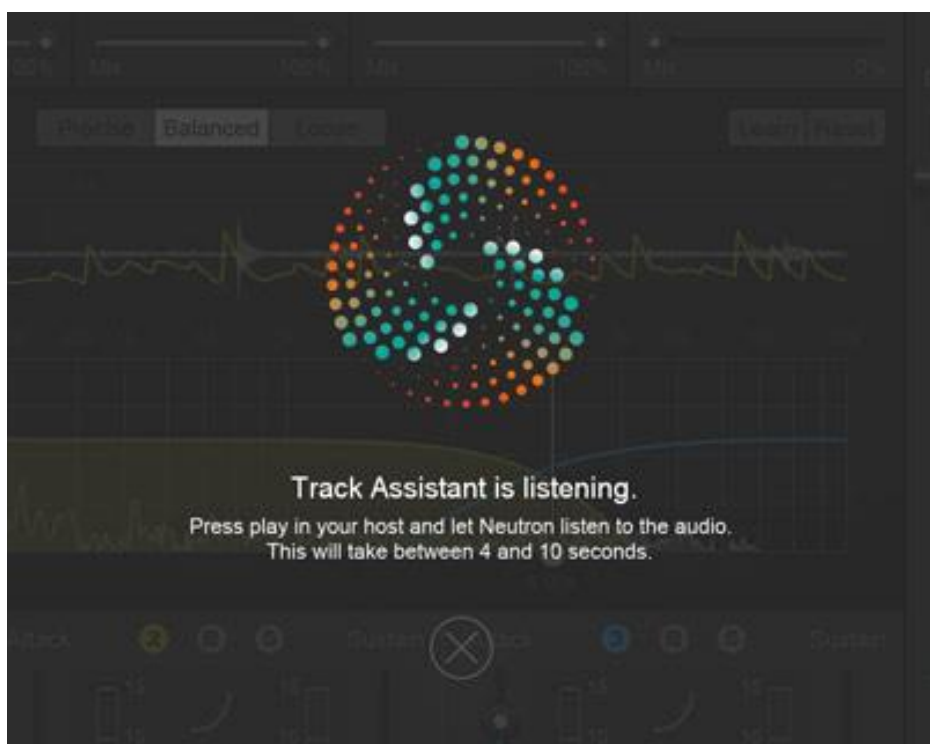
Obr. 27 - Znárodnění frekvenčního maskování ve FabFilter Pro-Q 3

3.1.2. Návrhový

O krok dál jde návrhový systém, který uživateli umožňuje požádat o analýzu a interpretaci mixu a na jejím základě navrhnout (a případně rovnou aplikovat) nastavení parametrů řetězce, ať už na úrovni jednotlivých stop nebo submixů s ohledem na jejich vzájemné vztahy (např. s cílem minimalizace frekvenčního maskování).

Příkladem jsou například výše zmíněné nástroje od iZotope a od Sonible, které obsahují uživatelsky spustitelné asistivní prvky, po jejichž aktivaci nástroj chvíli „poslouchá“ vstupní signál a na základě jeho analýzy a několika uživatelsky volených parametrů pak navrhuje nastavení parametrů procesorů, jako jsou ekvalizér, kompresor a dalších; uživatel pak tento návrh může přijmout či odmítnout.

Při práci se zvukovým materiálem si zpravidla můžeme všimnout existence jakéhosi „optimálního výchozího stavu“ pro jeho další zpracování v kontextu jiných zvuků – pokud touto kvalitou neoplývá, je obvyklé nejprve strávit čas technickou prací se zvukem samotným, ať je to formou hledání a odstraňování rezonancí, nežádoucích hluků a šumů, či nepatřičného dynamického rozsahu; až poté je nasnadě přejít k takzvané „tvůrčí“ práci vycházející z estetického vjemu.



Obr. 28 - iZotope Neutron, analýza signálu



Obr. 29 - Prostředí Sonible Smart EQ, bílá křivka vychází z automatické analýzy

Tato technická a do určité míry kvantifikovatelná část práce je nezdávka svěřena asistentským pozicím, aby s nimi hlavní tvůrčí entita nemusela ztrácet čas, a právě tuto část práce se uvedené návrhové systémy snaží automatizovat.

Dle spoluzakladatele Sonible Alexe Wankhammera⁶⁴ díky tomu tyto systémy nalézají využití zejména v prostředí živého vysílání (kde je často k dispozici naprosté minimum času na hledání adekvátní barvy mluveného slova a automatický systém, který je schopen toto zastoupit během několika okamžiků, je vítanou pomocí); mezi začátečníky, kterým mohou návrhy posloužit jako vítaná studijní pomůcka; ale také mezi profesionály, kterým smršňující se rozpočty nedovolují zaměstnávat asistenty.

3.1.3. Nezávislý

Nezávislý systém jde z hlediska uživatelské kontroly opačným směrem než návrhový systém – v tomto případě jsou systému přiděleny úkoly, které plní do velké míry autonomně a uživatel zde plní roli supervizora.

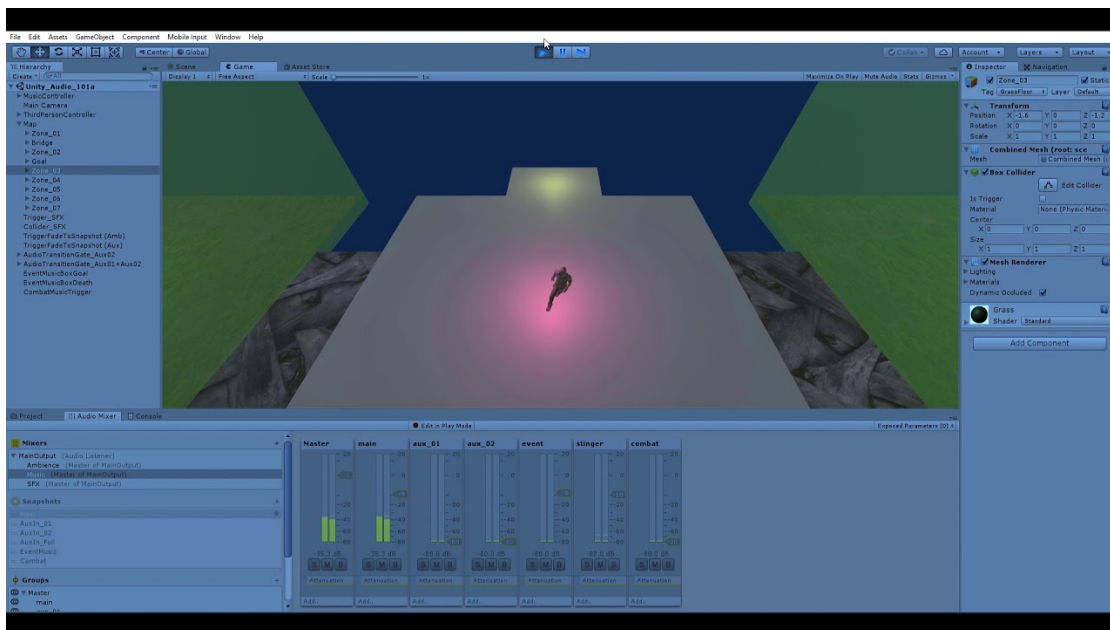
Nejde ještě o zcela automatický systém bez možnosti uživatelského zásahu, ale zredukované processingové parametry, omezená uživatelská kontrola a zpracování mimo reálný čas řadí instantní masteringové služby jako LANDR, Cloudbounce a další právě do této kategorie.

⁶⁴ *Should we be afraid of A.I!?!?* [online]. 2022 [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/watch?v=ZGetnk222YU>

3.1.4. Automatický

V plně automatickém systému je veškerá kontrola přenechána systému, který plně automatizuje parametry za účelem dosažení předem daného cíle – ten může být navržený manuálně nebo s využitím strojového učení⁶⁵.

Plně automatický systém může být užitečný například pro amatéry hledající co nejjednodušší způsob zpracování svých nahrávek, ale i v případech, kdy je potřeba zvuk zpracovávat v reálném čase bez možnosti lidské intervence, například v uměleckých instalacích nebo v kontextu videoherního zvukového mixu⁶⁶.



Obr. 30 - Prostředí herního enginu Unity, který lze použít pro vytváření automatických mixů v videoherním kontextu.

3.2. Reprezentace znalostí v IMP systémech

Způsob reprezentace znalostí (čili zpracování informací z reálného světa do počítačově zpracovatelné podoby) v rámci inteligentního systému je určujícím faktorem pro jeho schopnost s nimi pracovat, a také pro možnosti uživatelské interakce se systémem. IMP rozlišuje několik přístupů.

3.2.1. Zakotvená (podložená) teorie

Tato výzkumná metoda se často používá ve společenských vědách, kde jsou teorie systematicky formulovány na základě sběru a analýzy dat. Cílem je formalizovat proces práce se zvukem skrze empirická pozorování práce reálných zvukařů a na jejich základě modelovat sadu pravidel popisujících tento proces. Tento přístup je zásadní pro porozumění lidskému prvku ve zvukové tvorbě – ten však současně představuje nástrahu, protože spousta tvůrčích

⁶⁵ RAMÍREZ, Marco A Martínez, Emmanouil BENETOS a Joshua D REISS. A GENERAL-PURPOSE DEEP LEARNING APPROACH TO MODEL TIME-VARYING AUDIO EFFECTS. .

⁶⁶ STEVENS, Richard a Dave RAYBOULD. *The Game Audio Tutorial*. Routledge, 2013. DOI: 10.4324/9780240817279

rozhodnutí vychází z vědomého či nevědomého porušování pravidel; při jejich modelování je tedy nutné počítat s určitou flexibilitou.

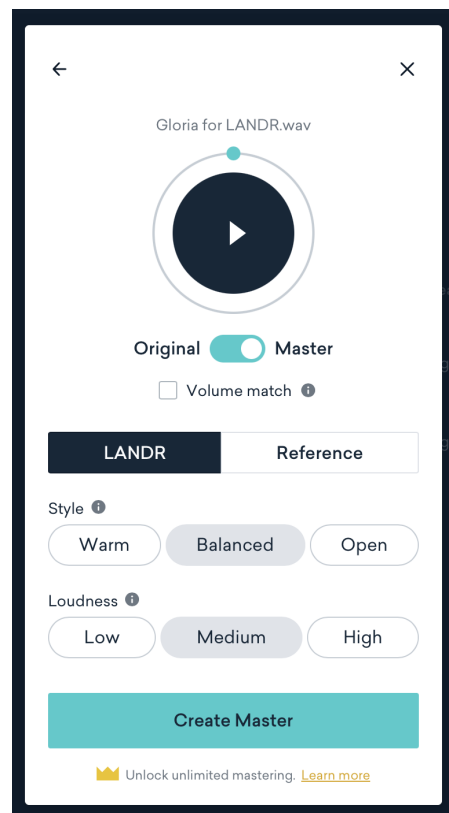
3.2.2. Expertní systémy založené na reprezentaci znalostí

Zvuková tvorba (potažmo tvorba obecně) není matematickou rovnicí s jediným správným řešením; těch může být celá řada a jejich podoba je závislá na vkusu uživatele v kombinaci s kulturními či žánrovými konvencemi. Jak jsme již nastínili dříve, součástí práce se zvukem je úprava objektivních parametrů na základě pojmenovávání těch subjektivních, čili například pokud je cílem učinit zvuk „hřejivějším“, inženýr může s pomocí ekvalizéru ubrat vyšší frekvence. Jak může ale systém vědět, že subjektivní popis „hřejivý“ může znamenat právě takovýto zásah?

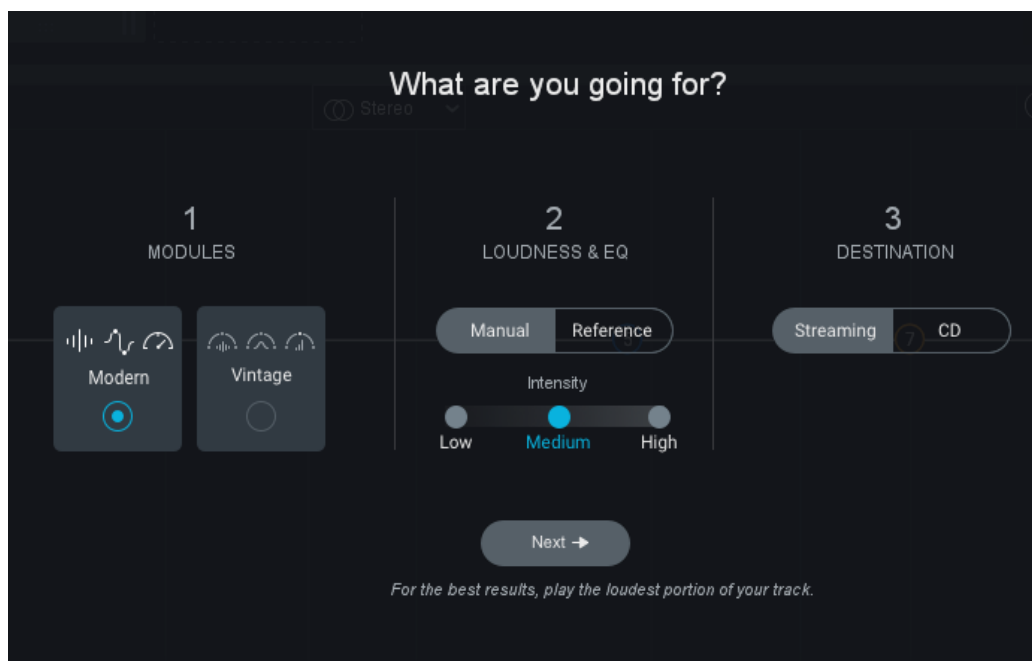
Právě odvození adekvátních pravidel a jejich reprezentace ve strojově interpretovatelné formě je hlavní výzvou při návrhu expertních systémů.

Prvky tohoto přístupu lze vidět ve zjednodušení ovládacích prvků, kdy jsou např. dynamické úpravy nahrávky ve své komplexitě zredukovány na pojem „intenzita“ nebo „hlasitost“, nebo celková tonalita na pojmy jako „hřejivý“, „vyvážený“, „moderní“ nebo „vintage“.

Dá se předpokládat, že s rozvojem velkých jazykových modelů a objevování nových možností jejich implementace se brzy dočkáme systému, který bude schopen přijímat jako vstupní podnět běžné hovorové vyjadřování, na jehož základě bude schopen volit další postup.



Obr. 31 - uživatelsky přístupné parametry LANDR



Obr. 32 - prostředí iZotope Ozone, dialog asistivní části

3.2.3. Přístupy založené na zpracování dat

De facto se jedná o přístupy založené na strojovém učení – jak víme z předcházejících kapitol, tyto systémy se neřídí předem určenými pravidly, ale pravidla samy vyvozují na základě tréninku na velkém objemu dat obsahujících specifické aspekty zvuku, na jejichž zpracování tvůrce systému cílí.

Například Smart EQ od Sonible obsahuje sérii tzv „profilů“, což jsou v podstatě natrénované modely se specifickým určením – nalezneme zde profily pro různé druhy hudebních nástrojů nebo hlasů.

Trénování těchto modelů probíhá formou učení s učitelem, kdy vývojáři nejprve nashromáždí velké množství dobře znějících vzorků daného nástroje – k tréninku je potřeba mnoho hodin materiálu, který musí projít lidským výběrem. Těmto vzorkům poté různými způsoby sníží kvalitu (přičemž cílem je napodobovat situace, jejichž výsledkem může nedostatečná kvalita být, čili vzorky se zkreslují, konvolvují impulsními odezvami a dále překrucují), systému je následně předložen znehodnocený vzorek společně s tím původním a systém se snaží úpravami parametrů co nejvíce přiblížit výchozí vzorek tomu cílovému⁶⁷.

iZotope veřejně takto detailně svůj postup nerozebírá, ale vzhledem k principiální podobnosti fungování a četným zmínkám o využití strojového učení v asistivních částech procesorů se lze domnívat, že jsou založeny na podobném principu.

Achillovou patou tohoto přístupu je dostupnost kvalitních tréninkových dat, protože k efektivnímu tréninku je jich potřeba velké množství.

Aktuální generace inteligentních nástrojů v sobě obsahuje prvky všech tří přístupů.

⁶⁷ *Should we be afraid of A.I.?*

3.3. Způsob úpravy zvuku

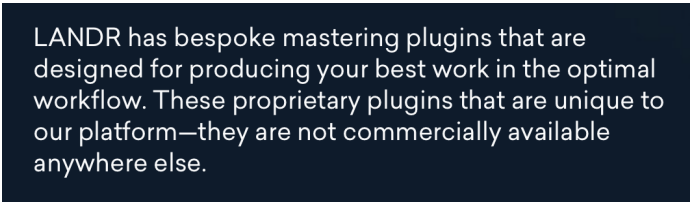
IMP systémy lze dále dělit z pohledu toho, jakým způsobem do zvuku zasahují. Dosud popsané druhy systémů zpravidla pracují s parametry tradičních procesorů, ovšem možné jsou také systémy upravující zvuk na úrovni datového toku resyntézou nového audia na základě zadaných parametrů, což by mohlo vést ke zcela novým zvukovým možnostem přesahujícím dnešní zvukové mantinely vycházející z používání tradičních procesorů.

3.4. Současnost instantního masteringu

3.4.1. Obfuskace vnitřního fungování

I když by mohlo jít o plně automatický systém, který by na základě analýzy a určení cílového zvuku mohl audio zcela resyntetizovat, není to pravděpodobné, protože odborná literatura ani patenty společností stojícími za těmito službami v tomto směru nenaznačují žádné radikální pokroky^{68,69}. Je mnohem pravděpodobnější, že v rámci analýzy služba využije strojového učení k identifikaci žánru nahrávky a stanovení odchylky od idealizovaného cílového zvuku – jak jsme již viděli v kapitole o MIR, každý žánr má nějaký charakteristický zvuk, který lze s pomocí MIR metod kvantifikovat do počítačem zpracovatelné podoby. Služba posléze na základě odchylky vstupní nahrávky od cíle navrhne processingový řetězec, který lze uživatelsky ovlivnit pouze velmi obecným způsobem (viz kapitola 2.3).

Většina služeb instantního masteringu tedy nějakým způsobem „mlží“ - materiály LANDRu odkazují na využití AI, eMastered a CloudBounce explicitně zmiňují klasifikační analýzu využívající strojové učení, na jejímž základě sestavují řetězce tradičních zvukových procesorů na míru poskytnuté skladbě. Zato některé další služby jako MajorDecibel a MasteringBox nezmiňují ani jedno ani druhé a místo toho používají vágní termíny jako „masteringový engine“ nebo „komplexní algoritmus“; výjimkou je pak instantní masteringová služba spadající pod BandLab, která inzeruje „artist-driven mastering“ a oproti AI technikám se ve svých materiálech přímo vymezuje.



LANDR has bespoke mastering plugins that are designed for producing your best work in the optimal workflow. These proprietary plugins that are unique to our platform—they are not commercially available anywhere else.

Obr. 33 - Propagační materiál LANDRu o vlastním vývoji zvukových procesorů

Všechny služby spojují zmínky o „vlastnoručně vyvinutých nástrojích a procesorech“, což má pravděpodobně zvýšit vnímanou přidanou hodnotu služeb skrze exkluzivitu.

Některé navíc nabízejí zpracování nahrávky vůči poskytnuté referenci, kdy uživatel může poskytnout další nahrávky, jejímuž zvuku se ve finálním masteru chce přiblížit.

Zajímavostí je, že z výše jmenovaných pouze MajorDecibel ve svých materiálech uvádí, že se nedá považovat za plnohodnotnou náhradu lidských masteringových inženýrů, ale explicitně cílí zejména na ty, kteří by na jejich služby buď neměli rozpočet, nebo chtějí vylepšit zvuk

⁶⁸ STERNE, Jonathan a Elena RAZLOGOVA. *Machine Learning in Context, or Learning from LANDR*.

⁶⁹ ELLIOTT, Mitchell a Song Hui CHON. *A Comparative Study of Music Mastered by Human Engineers and Automated Services*.

svého dema, které by pro profesionálního masteringového inženýra nebylo dostatečně atraktivním.

3.4.2. Platformizace služeb

Zatímco masteringoví inženýři se obvykle soustřeďují pouze na mastering, služby instantního masteringu někdy nabízejí další služby a stávají se spíše platformami, v rámci kterých je mastering pouze částí celkové nabídky služeb. Např. LANDR nabízí kromě masteringu i hudební distribuci na streamovací služby, pluginy, samplý nebo nástroje pro vzdálenou spolupráci mezi hudebníky hostované na své platformě. BandLab také mastering nabízí v rámci širšího rámce svého ekosystému postaveného kolem on-line DAWu BandLab Studio.⁷⁰

3.4.3. Cenová dostupnost

Instantní masteringové služby účtují buď za hotovou píseň nebo fungují na základě předplatného. V případě zpracování jediné nahrávky si služby účtují kolem 10 USD, ale všechny mají tendenci nabízet spíše předplatné, které stojí obvykle mezi 10-30 USD měsíčně. Mastering s legendárním inženýrem Bobem Katzem začíná na 200 USD za hodinu, online mastering lidským inženýrem v Abbey Road Studios začíná na 100 GBP za nahrávku, ceník studia 3 Bees uvádí cenu masteringu od 1500 Kč za hodinu.

Cenová dostupnost instantních služeb je tedy řádově nižší.

Our engine will make your track louder, more crisp and full of life

Obr. 34 - z materiálů LANDR, podobná vyjádření nalezneme napříč všemi službami

3.4.4. Je to vůbec mastering?

Jak jsme viděli v předcházejících kapitolách, mastering se z ryze technické činnosti stal činností s potenciálem výrazného tvůrčího vkladu a odpověď na tuto otázku bude záviset na tom, komu je položena, respektive do jaké míry vidí dotýčný benefit v možnosti vést tvůrčí dialog s jinou lidskou bytostí. Služby instantního masteringu vlastně nabízejí svou vlastní definici (takovou, která je založená na algoritmicky podložené externí validaci zvuku) a je na uživateli, zda na ni přistoupí či nikoliv.

Aby byla jakákoliv ML implementace úspěšná ve tvůrčím snažení, musí její stvořitelé vymezit podmínky úspěchu: „hotová“ nahrávka je estetické rozhodnutí a současně pohyblivý cíl.

⁷⁰ STERNE, Jonathan a Elena RAZLOGOVA. Tuning sound for infrastructures: artificial intelligence, automation, and the cultural politics of audio mastering. *Cultural Studies*. 2021, roč. 35, č. 4–5. DOI: 10.1080/09502386.2021.1895247

3.5. Lidský versus instantní mastering

Výzkumníci z Belmont University v Nashville⁷¹ provedli srovnání lidských a instantních masterů s cílem prozkoumat platnost obecně rozšířeného názoru, že instantní mastering je inherentně méně kvalitní než lidská práce. Jako výchozí vzorky posloužily části dvou žánrově odlišných nahrávek, alternativního rocku a jazz-fusion.

V první části testu se zjišťovala schopnost účastníků správně identifikovat lidský a instantní master, ve druhé části pak zda je možné vysledovat obecnou preferenci účastníků jedním, či druhým směrem.

Správně identifikovat lidské a instantní mastery dokázalo pouze 48 % účastníků, preference směrem k lidským masterům se pak pohybovaly na hranici statistické zanedbatelnosti.

Z toho lze vyvodit několik možných závěrů – pokud se v kontextu masteringu dá mluvit o objektivních měřítkách kvality, je možné, že instantní masteringové služby jsou již na poměrně vysoké úrovni; je ale k diskusi, jak by si služby vedly na širším spektru nahrávek, které by zároveň byly méně konformní s předpokládaným složením tréninkových dat těchto služeb.

Je také možné, že účastníci prostě nebyli zvyklí poslouchat instantní mastery a s dalším tréninkem by jejich schopnost rozpoznat nuance instantního a lidského masteru vzrostla.

V neposlední řadě hraje roli také fakt, že důležitou součástí masteringu je pořád ještě sjednocování nahrávek v rámci alba a lidské mastery by tudíž zněly jinak, kdyby měly fungovat v kontextu jiných písní.

3.6. Subjektivní dojmy z použití

Autor se rozhodl použít dvě nahrávky ze svého archivu a podrobit je jednak masteringu služby LANDR (s výchozím nastavením zpracování tonality a dynamiky), jednak automatickému processingu iZotope Ozone 10.

První nahrávka s názvem *Easter* je experimentální skladbou s metalovými prvky na hranici písně a zvukové koláže s recitací a výraznými rozdíly mezi jednotlivými pasážemi a ostrými přechody mezi nimi. Skladba vznikla jako soutěžní a v zadání bylo explicitně uvedeno, že výsledek má znít co nejhotověji, aniž by prošel masteringovými zásahy, zejména pak excesivní limitací signálových špiček.

Druhá nahrávka, taktéž soutěžní (a tentokrát bezejmenná) je zvukově mnohem uchopitelnější – jedná se o tříaktovou akční filmovou hudbu s výraznými orchestrálními prvky. Tato nahrávka rovněž neprošla žádnými excesivními masteringovými zásahy; přítomen byl pouze limiter zachytávající jen ty nejvýraznější signálové špičky.

Pro subjektivní posouzení byly všechny verze následně normalizovány na hlasitost –23LUFS.

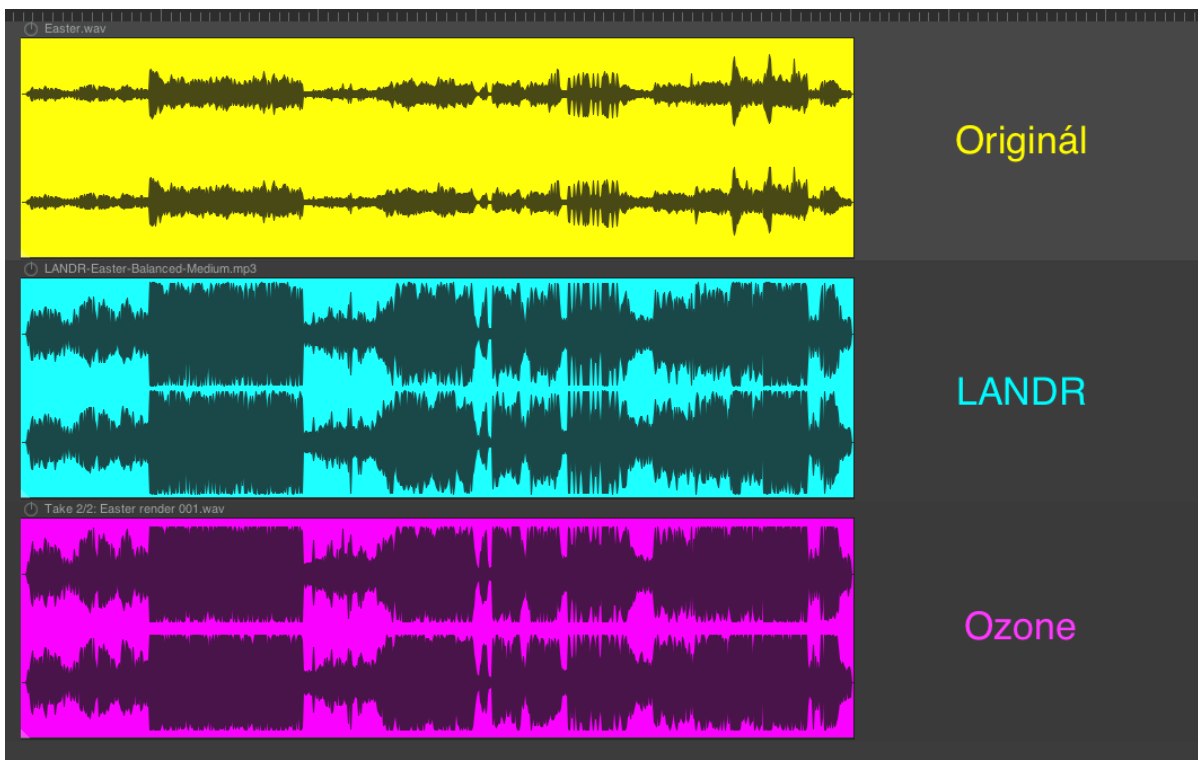
⁷¹ ELLIOTT, Mitchell a Song Hui CHON. *A Comparative Study of Music Mastered by Human Engineers and Automated Services*.

3.6.1. Easter

V souladu s původním zadáním nechává výchozí mix Easter velký dynamický headroom pro potenciálního masteringového inženýra, tudíž výrazné zvýšení hlasitosti je první nejmarkantnější změnou.

#	Track	Item	Integrated	Range	True peak	Maximum short-term	Maximum moment...
1	SOURCE	Easter.wav	-23.0 LUFS	12.0 LU	-2.5 dBTP	-17.6 LUFS	-12.9 LUFS
2	LANDR	LANDR-Easter-Balanced-Medium.mp3	-11.8 LUFS	9.4 LU	-0.2 dBTP	-8.2 LUFS	-5.9 LUFS
3	Ozone	Easter render 001.wav	-11.2 LUFS	8.3 LU	-1.0 dBTP	-8.8 LUFS	-7.3 LUFS

Obr. 35 - naměřená hlasitost všech verzí Easter



Obr. 36 - waveformy všech verzí Easter

3.6.2. Easter – dynamika

Tento mix je s trochou nadsázky vymyšlený tak, aby se nedal klasickým způsobem masterovat – vokální prvky, které nejsou zpěvem, ale spíše recitací, jsou ze zvukově dramaturgického hlediska myšlené jako voiceover a podle toho jsou v mixu mnohem výše, než by u normálního zpěvu bylo obvyklé. Kompresi částí s nimi má pak poměrně znatelný efekt na instrumentální složku mixu.

Dalším prvkem vzpírajícím se kompresi jsou basové a subbasové efekty - na obrázku jsou viditelné na waveformě původního mixu jako ty největší špičky. Záměrem jejich použití je dramatický efekt, který je obvyklý spíše při sledování filmů než při poslechu současné hudby s jejím často malým dynamickým rozsahem, což ale není možné bez adekvátního headroomu. Oba algoritmičké mastery proto trpí znatelným „dýchnutím“ v okamžiku sweepu (ale i v momentě cca ve třetí čtvrtině skladby, kdy výrazný basový buben hraje současně s delikátním mluveným slovem), kdy dynamická komprese drží celkovou energii mixu pod

určitým prahem, a přitom se dočasně sníží hlasitost celého mixu. LANDR si se situací poradil lépe než Ozone, který navrhl agresivně rychlou limitaci vedoucí k výraznému zkreslení – ovšem narozdíl od LANDRu, tento neduh lze s velkou přesností řešit úpravou parametrů jednotlivých procesorů v rámci Ozone.

3.6.3. Easter – tonalita a další

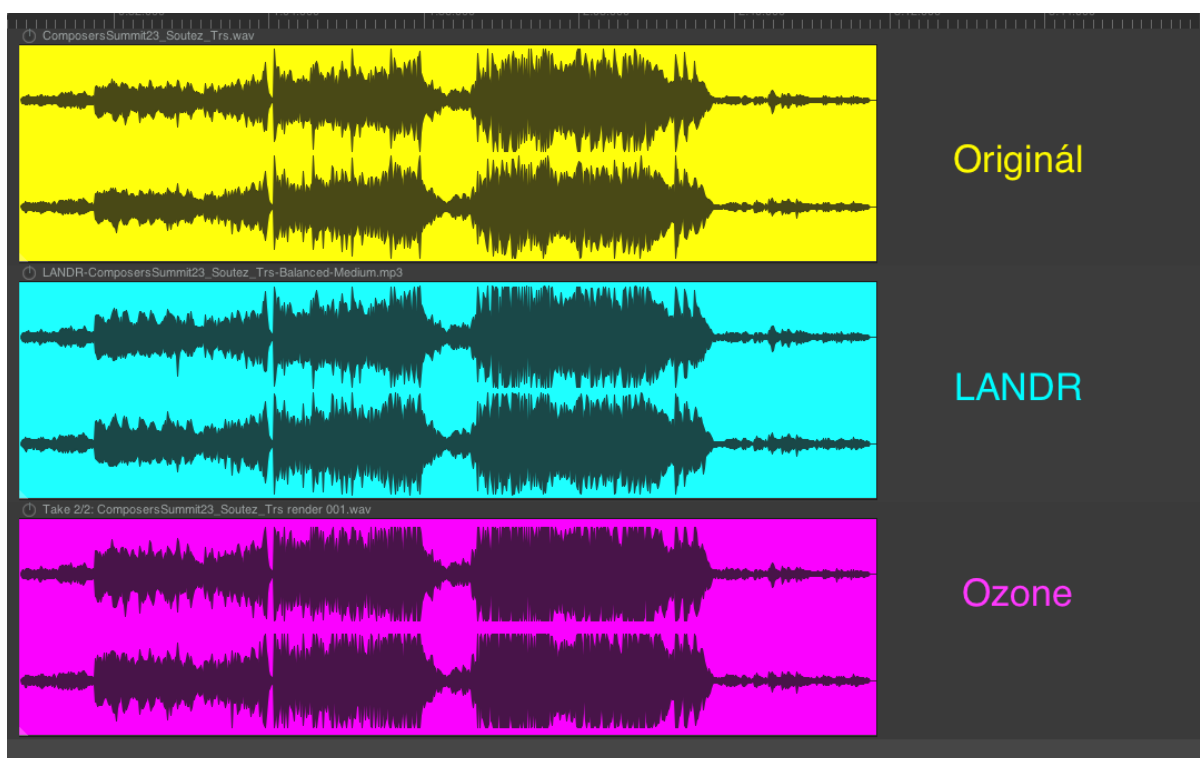
Originál a LANDR verze se na poslech příliš neliší – LANDR verze zní o něco mohutněji na basech, ale ne příliš. Ozone oproti tomu nabízí zdvih na středech a výškách od 10kHz výše – míra vhodnosti tohoto zásahu je spíše subjektivní.

3.6.4. Nahrávka #2

Hlasitostní měření prozrazují, že výchozí mix je sám o sobě poměrně hlasitý – algoritmické zásahy vedly ke zvýšení integrované hlasitosti o necelé 2LU.

#	Track	Item	Integrated	Range	True peak	Maximum short-term	Maximum moment...
1	SOURCE	ComposersSummit23_Soutez_Trns.wav	-14.6 LUFS	20.7 LU	0.0 dBTP	-10.7 LUFS	-8.2 LUFS
2	LANDR	LANDR-ComposersSummit23_Soutez_Trns-Balanced-Medium.mp3	-13.8 LUFS	19.7 LU	-0.2 dBTP	-10.1 LUFS	-8.2 LUFS
3	Ozone	ComposersSummit23_Soutez_Trns render 001.wav	-12.2 LUFS	20.4 LU	-1.0 dBTP	-8.5 LUFS	-7.0 LUFS

Obr. 37 - naměřená hlasitost všech verzí Nahrávky #2



Obr. 38 - waveformy všech verzí Nahrávky #2

3.6.5. Nahrávka #2 - dynamika

V tomto případě není možné mix příliš komprimovat, aniž by dramatické momenty neztratily na síle, proto zůstala hlasitost všech verzí na velmi podobné úrovni – pocitově jsou ovšem tiché pasáže LANDR verze výraznější než ty originální při zachování podobné hlasitosti forte

pasáží, je tedy možné, že LANDR aplikuje nějakou formu dynamických úprav na úrovni celých pasáží.

3.6.6. Nahrávka #2 – tonalita a další

Všechny verze znějí velice podobně, snad jen LANDR má o něco výraznější basy (což vede k tomu, že v závěrečné tiché pasáži je najednou zcela jasně slyšet pedál piana) a Ozone navrhl zejména mírné přidání vyšších středů.

4. Související témata

4.1. Předpojatost – bias v AI a perpetuace stereotypů

Hluboké učení stojí a padá s kvalitou tréninkových dat s ohledem na plánované využití modelu.

Dobře to ilustruje několik kontroverzí uplynulých let:

- model označující lidi tmavé pleti jako gorily⁷²;
- model detekující spolehlivě maligní kožní nálezy díky přítomnosti pravitka v grafice⁷³;
- model určující doktory výhradně jako muže a sestřičky výhradně jako ženy⁷⁴.

Tyto zdánlivě nesouvisející incidenty jsou pro nás naprosto relevantní, protože pro ML model jsou to všechno „pouze“ data. Pokud je tedy model stojící za službou typu LANDR naučen pouze na vzorcích mainstreamové hudby, tak uživatelé tvořící hudbu konformní aktuálně módním stylům/žánrům, u kterých existuje obecně sdílená představa, jak mají znít „správně“, budou pravděpodobně spokojeni, zatímco tvůrci experimentální hudby zkoumající hudební obzory budou automatickou službou pravděpodobně zklamáni.

Ten, kdo ovládá tréninková data, ovládá výsledný model – dá se tedy polemizovat, zda konzervativnost při správě datasetu ze strany korporátní entity nemůže perpetuovat zvukové stereotypy a brzdí rozvoj nových žánrů.

Tato problematika jde za hranice masteringu a AI. Laptop, na kterém autor tuto práci píše, má s ohledem na své fyzické dispozice poměrně senzační zvuk – což se ovšem projeví pouze u některých žánrů, zejména u populární hudby. V momentě, kdy má přehrát nahrávku např. smíšeného sboru, zvuk se začne velmi rychle rozpadat, dochází ke slyšitelnému zkreslení již při nižších hlasitostech. Je na zvážení, nakolik technologie může favorizovat některé žánry více než jiné.

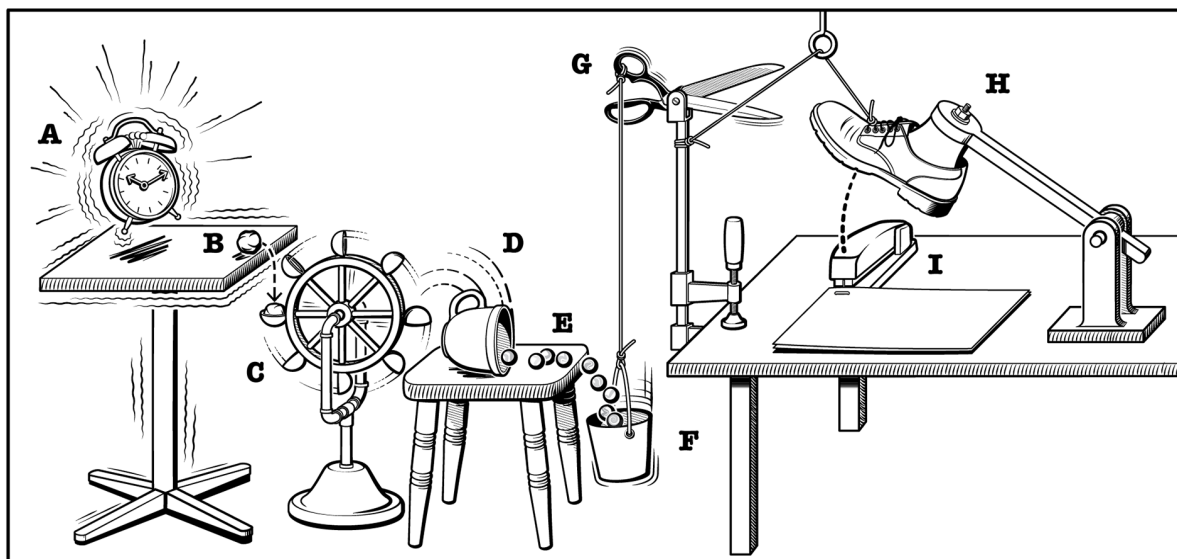
Výše popsané problémy jsou jedním z důvodů, proč se na systémy strojového učení často nahlíží jako na záhadné stroje, o jejichž vnitřním fungování se lze pouze dohadovat.

⁷² ZHANG, Maggie. Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software. In: *Forbes* [online] [cit. 28.05.2023]. Dostupné z: <https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/>

⁷³ *When AI flags the ruler, not the tumor — and other arguments for abolishing the black box (VB Live)* [online]. 2021 [cit. 28.05.2023]. Dostupné z: <https://venturebeat.com/business/when-ai-flags-the-ruler-not-the-tumor-and-other-arguments-for-abolishing-the-black-box-vb-live/>

⁷⁴ BUONOCORE, Tommaso. Man is to Doctor as Woman is to Nurse: the Dangerous Bias of Word Embeddings. In: *Medium* [online]. 3. 3. 2020 [cit. 28.05.2023]. Dostupné z: <https://towardsdatascience.com/gender-bias-word-embeddings-76d9806a0e17>

4.2. Černé skříňky, Goldbergův aparát⁷⁵, neuronové sítě a kernel machines



Obr. 39 - Schéma Goldbergova aparátu

Goldbergův aparát (pojmenovaný po svém „vynálezci“, karikaturistovi Rube Goldbergovi) je stroj provádějící jednoduchý úkol záměrně složitým způsobem, obvykle zahrnujícím nějaký druh řetězové reakce.

Není pochyb o tom, že neuronové sítě dosahují leckdy uhrančivých výsledků v mnoha vědeckých a technických úlohách, ovšem empirické úspěchy předstihly naše základní znalosti o jejich struktuře a funkci, což jim dodává mystickou auru magických (občas až Goldbergovských) černých skříněk. V současnosti probíhá výzkum hledající ekvivalence mezi neuronovými sítěmi a již popsanými matematickými entitami, jako jsou např. *kernel machines*, což by umožňovalo využít již existující teorii a demystifikovat vnitřní fungování NS.^{76,77}

4.3. Glass box⁷⁸

„Černá skříňka“ je cokoliv, u čeho uživatel má přístup ke vstupním a výstupním datům, ale nemá přístup k vnitřnímu fungování – typicky proprietární software apod. Nicméně v případě neuronových sítí je nasnadě spíše přirovnání ke „skleněné skříňce“ - vnitřní struktura je nabíledni, kdykoliv připravena ke zkoumání a jednotlivé prvky v izolaci není těžké definovat. Nic nám nebrání ukázat si a pojmenovat vstupní, výstupní či skrytou vrstvu, neuron, či hodnotu té které váhy nebo aktivace. Problém se začíná rýsovat v momentě, kdy se snažíme definovat

⁷⁵ Rube Goldbergův stroj. In: *Necyklopedie* [online]. 20. 6. 2013 [cit. 02.08.2022]. Dostupné z: [//necyklopedie.org/wiki/Rube_Goldberg%C5%AFv_stroj](http://necyklopedie.org/wiki/Rube_Goldberg%C5%AFv_stroj)

⁷⁶ JACOT, Arthur, Franck GABRIEL a Clément HONGLER. *Neural Tangent Kernel: Convergence and Generalization in Neural Networks* [online]. arXiv, 2020 [cit. 02.08.2022]. Dostupné z: <http://arxiv.org/abs/1806.07572>

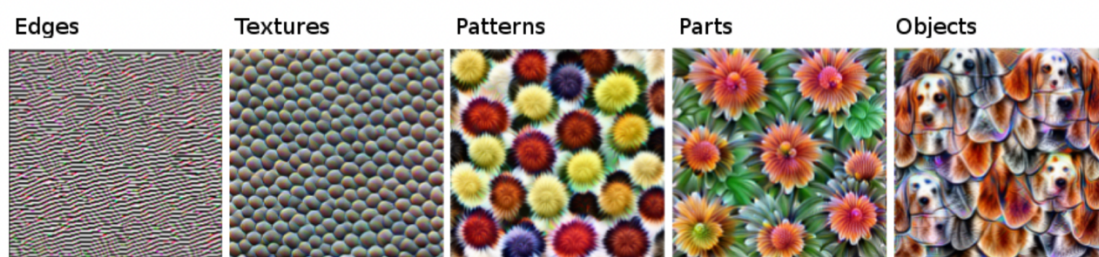
⁷⁷ KHARE, Yana. *OpenAI's New Tool Explains Behavior of Language Model At Every Neuron Level* [online]. 2023 [cit. 26.05.2023]. Dostupné z: <https://www.analyticsvidhya.com/blog/2023/05/openai-new-tool-explains-behavior-of-language-model-on-every-neuron-level/>

⁷⁸ RAI, Arun. Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science*. 2020, roč. 48, č. 1. DOI: 10.1007/s11747-019-00710-5

a přiřadit jejich součinnost k nějakému konkrétnímu výstupu sítě – pokud například síť na nějakém obrázku identifikuje žabu⁷⁹, máme tendenci začít zkoumat, který shluk neuronů stojí za „úvahou“ vedoucí k tomuto výsledku. Shodou okolností právě vizuální doména a její zpracování konvoluční sítí jsou dobrým základem pro úvahy o interpretabilitě neuronových sítí.

4.4. Interpretabilita neuronových sítí, projekce a haiku

Podívejme se následující vizualizaci zpracování dat v konvoluční neuronové síti:



Obr. 40 - vizualizace detekovaných rysů v jednotlivých vrstvách CNN

Hluboké neuronové sítě se učí vysokoúrovňové rysy (*high level features*) ve skrytých vrstvách – to je jedna z jejich nejsilnějších stránek, která snižuje potřebu předběžné extrakce parametrů. Jak obraz prochází dalšími a dalšími konvolučními vrstvami, síť se učí nové a stále složitější rysy. Transformovaná obrazová informace poté prochází plně propojenými vrstvami, aby vyústila v klasifikaci nebo predikci.

Vizualizace rysů mohou vyvolat iluzi, že rozumíme tomu, co neuronová síť dělá, ale rozumíme skutečně tomu, co se v neuronové síti děje? I když se podíváme na stovky nebo tisíce vizualizací funkcí, neuronové sítě neporozumíme. Kanály spolu komplexně interagují, pozitivní a negativní aktivace spolu nesouvisí, více neuronů se může naučit velmi podobné rysy a pro mnoho rysů nemáme odpovídající lidské pojmy. Nesmíme se nachytat a věřit, že rozumíme neuronovým sítím jen proto, že jsme viděli, že neuron 859 ve 4. vrstvě je aktivován kopretinami. Analýza sítí ukázala, že architektury jako ResNet nebo Inception mají jednotky, které reagují na určité pojmy. Ale přesnost není tak velká a často mnoho jednotek reaguje na stejný koncept a některé na žádný koncept. Kanály jsou spolu vždy nějakým způsobem propojené a nelze je interpretovat izolovaně.⁸⁰

Projekce je psychologický mechanismus, kterým se jedinec snaží přenášet své myšlenky a impulsy na jiné osoby nebo objekty. Inspirován počítačem generovanými haiku, které slyšel

⁷⁹ Pepe the Frog. In: *Wikipedia* [online]. 2023 [cit. 28.05.2023]. Dostupné z: https://en.wikipedia.org/w/index.php?title=Pepe_the_Frog&oldid=1156293322

⁸⁰ MOLNAR, Christoph. *Interpretable machine learning: a guide for making black box models explainable (2nd ed.)*. 2022.

v rádiu, Douglas Hofstadter se koncem 70. let rozhodl napsat vlastní program pro generování gramaticky koherentního textu.⁸¹

A male pencil who must laugh clumsily would quack. Must program not always crunch girl at memory? The decimal bug which spits clumsily might tumble. Cake who does sure take an unexpected man within relationship might always dump card.

Program ought run cheerfully.

The worthy machine ought not always paste the astronomer.

Oh, program who ought really run off of the girl writes musician for theater.

The businesslike relationship quacks.

The lucky girl which can always quack will never sure quack.

The game quacks. Professor will write pickle. A bug tumbles. Man takes the box who slips.

Obr. 41 - výstup Hofstadterova programu

Po několika iteracích usoudil, že hlavním prvkem dodávajícím jeho výstupům záblesky smyslu a koherence je hlavně fakt, že při čtení textu mají lidé přirozenou tendenci přisuzovat slovům jejich plný význam, jako kdyby vyplýval ze shluku písmen, která tato slova tvoří.

Stejnou optikou se lze dívat na interpretaci fungování neuronových sítí. Ptát se, jak neuronová síť *dospěla* k rozhodnutí, je kategoricky špatná otázka vycházející z představy, že síť k predikci *dospěla* po nějaké *zralé úvaze*. Z pohledu sítě je její predikce ovšem pouze systémovou reakcí na stimulus. Neuronová síť je především výpočetní model operující s čísly, a jeho úspěšnost je závislá na tom, do jaké míry uživatel dokáže svůj úmysl vyjádřit v řeči čísel. Význam je výsledkům přisouzen v mysli operátora. Implikace pro lidskou inteligenci v této práci nebudeme rozebírat.

Shakespeare je jedním z těch, kdo zastával myšlenku, že „krása tkví v oku pozorovatele“. Autor se domnívá, že totéž se dá říci o rysech inteligence.

4.5. Tah 37 a adversární útoky^{82,83}

Hra go byla ve své „komplexní jednoduchosti“ tradičně považována za baštu lidské inteligence a intuice, čímž představovala nezdolný úkol pro výzkumníky na poli umělé inteligence. V roce 2016 program AlphaGo s využitím kombinace zpětnovazebního učení a tradičních heuristických metod přesvědčivě porazil světového šampiona v Go Lee Sedola (dnes již legendárním „tahem 37“), díky čemuž byl označen za „génia s nadlidskou schopností hrát go“. Začátkem roku 2023 vytvořili výzkumníci program specificky hledající chyby ve stylu hry Kata-Go, dalšího pokročilého programu na hraní go, a podařilo se jim najít způsob, jak jej spolehlivě porazit díky zneužití situace, která nebyla součástí tréninkového datasetu.

⁸¹ HOFSTADTER, Douglas R. *Gödel, Escher, Bach*.

⁸² Adversarial Policies in Go - Game Viewer. In: [cit. 28.05.2023]. Dostupné z: https://goattack.far.ai/pdfs/go_attack_paper.pdf?uuid=WgOkT29w1fP7rG7p1682

⁸³ Move 37: Artificial Intelligence, Randomness, and Creativity. In: *John Menick* [online]. 17. 10. 2016 [cit. 06.08.2022]. Dostupné z: <http://localhost:4000/writing/move-37-alpha-go-deep-mind.html>

Důležité je zmínit, že adversární program není v go příliš zběhlý – ve standardních hrách jej jednoduše porázejí i amatéři. Jde o typický příklad adversárního útoku, který podtrhuje, že zdánlivě nadlidsky inteligentní programy a algoritmy jsou plné blataných děr, které navíc vzhledem k jejich netransparentní povaze nemusí být ihned patrné. Při interakci se systémy založenými na strojovém učení je vždy nutné mít na paměti, že ten, kdo má kontrolu nad tréninkovými daty, má také kontrolu nad výsledným chováním modelu, a také to, že jakýkoliv výstup tohoto systému je třeba brát s rezervou a používat vlastní úsudek, či v našem případě uši.

4.6. Důvěra v automatické systémy, nahraditelnost a afektivní práce^{84,85}

Rozhodujícím faktorem úspěchu spolupráce člověka se strojem je důvěra, kterou má lidský člen týmu ke svému strojovému kolegovi. Automatizace a autonomní systémy nabízejí větší konzistenci a přesnost, než které by mohl dosáhnout člověk, a to nezřídka v úlohách, které jsou pro člověka příliš repetitivní, rychlé nebo nebezpečné. S těmito novými spoluhráči přichází příležitost k větší týmové efektivitě – pokud lidský operátor důvěřuje autonomnímu spoluhráči, může se soustředit na své vlastní úkoly a využívat své silné stránky. Nejistota a zranitelnost, kterou s sebou přináší nutnost spolehnout se na „strojové kolegy“, naplňují klasickou definici důvěry jako „[...] přesvědčení, že agent pomůže dosáhnout cílů člověka v situaci, která se vyznačuje nejistotou a zranitelností“.⁸⁶

Důvěra v automatické systémy je polem zkoumaným zejména v průmyslovém kontextu, nicméně na práci se zvukem (či tvůrčí práci obecně) lze také nahlížet jako na součást zábavního průmyslu.

Studium nových sofistikovaných nástrojů na poli zvukové tvorby se většinou nevyhne úvahám o tom, zda mají stroje potenciál nahradit lidi ve tvůrčích profesích. Nabízí se paralela s digitální fotografií – dnes je pro většinu lidí naprosto přirozené, že pro pořízení fotografie digitálním fotoaparátem stačí namířit na fotografovaný objekt a fotoaparát sám nastaví parametry, jako expozice, vyvážení bílé apod., zatímco k tomu, co se obecně považuje za „hotový zvuk“, je stále potřeba množství manuální práce ve formě processingu apod. Na jedné straně je to podmíněné odlišností technologií, na druhé straně zase kulturně determinovanou představou o zamýšleném výsledku.

Soudě podle dosavadního vývoje, každá zásadní technologická změna s sebou přinesla eliminaci určitých (nezřídka technických či repetitivních) aspektů dotčené činnosti – například v době DAW nemusíme ručně a destruktivně stříhat magnetické pásky, což nám uvolňuje čas

⁸⁴ KOHN, Spencer C. et al. Measurement of Trust in Automation: A Narrative Review and Reference Guide. *Frontiers in Psychology* [online]. 2021, roč. 12 [cit. 02.05.2023]. Dostupné z: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.604977>

⁸⁵ MORAY, Neville, John LEE a Bonnie M. MUIR. *Trust and human intervention in automated systems*. 1994.

⁸⁶ KOHN, Spencer C. et al. *Measurement of Trust in Automation*.

a myšlenkovou kapacitu na jiná rozhodnutí. Autor si je vědom dvousečnosti tohoto fenoménu, ať už jde o nutnost přemýšlet nad tím, co děláme, pokud máme omezený počet pokusů, nebo obecná očekávání týkající se času na práci ve vztahu k výsledné kvalitě produktu. Nicméně každá větší automatizace s sebou většinou kromě znehodnocení repetitivní práce přináší také nové tvůrčí možnosti a zpravidla také nepodmíněně přehodnocení toho, v čem tkví jádro lidského přínosu v oboru procházejícím automatizací.

4.7. Afektivní práce (affective labour)⁸⁷

Označuje práci, jejímž cílem je vyvolat nebo změnit emocionální prožitky lidí. Lidské masteringové inženýry lze považovat za *affective labourers*, protože jsou svými klienty považováni za tvůrčí partnery, se kterými je možné vést dialog a společně hledat pochopení a naplnění tvůrčího záměru, většinou podmíněného nějakou formou duševního vkladu klienta v díle.

Dá se předpokládat, že dokud bude tvůrčí dialog v kurzu, lidští tvůrci nevyumřou – což ovšem neznamená, že by se neměli učit s novými nástroji a ovládnout jejich silné i slabé stránky, pokud chtějí zůstat relevantními.

4.8. Originalita, smrt autora a Melancholičtí sloni

„Umělci si po staletí namlouvali, že tvoří. Ve skutečnosti nic takového nedělají. Objevují. V povaze reality je obsažena konečná řada kombinací hudebních tónů, které bude lidská nervová soustava vnímat jako libozvučné. Po tisíciletí je objevujeme, jsou implicitně obsaženy ve vesmíru – a namlouváme si, že jsme je "stvořili".

Vytvořit implikuje nekonečné množství možností, objevit znamená konečné množství. Myslím, že jako živočišný druh se s myšlenkou, že jsme místo tvůrců objevitelé, vyrovnávat špatně.”⁸⁸

- *Melancholy of Elephants, Spider Robinson*

„Nyní víme, že text není sérií slov nesoucí jediný "teologický" význam ("poselství" autora-Boha), ale vícerozměrný prostor, v němž se mísí, střetávají a proměňují různé texty, z nichž žádný není původní. Text je směsí citátů čerpaných z nespočetných kulturních východisek.”⁸⁹

- *Roland Barthes*

Dá se spekulovat, že nápodoba a začarovaný kruh uniformizace hudebního zvuku skrze automatické systémy je pouhým logickým pokračováním jevu, který je příčinou existence střídání dominantních žánrových a stylových proudů v průběhu staletí a dekad – umělci se vzájemně inspirují, opisují od sebe, nebo přímo kradou. Většina tvorby je vědomým či nevědomým amalgamem již existujících děl, vycházejících z předchozí zkušenosti autorů;

⁸⁷ BIRCHNELL, Thomas. Listening without ears: Artificial intelligence in audio mastering. *Big Data & Society*. 2018, roč. 5, č. 2. DOI: 10.1177/2053951718808553

⁸⁸ ROBINSON, Spider. *Melancholy elephants*. New York, N.Y.: Tom Doherty Associates, 1985.

⁸⁹ BARTHES, Roland a Stephen HEATH. *Image, music, text: essays*. London: Fontana, 1977.

i v případě masteringu se často používají jako reference již hotové nahrávky. Styl se recykluje, dokud není obecný vkus saturován (*dle Hofstadtera: nuda se nedostaví s vyčerpáním repertoáru chování věci, ale když se nám podaří zmapovat limity prostoru, v rámci kterého chování té věci probíhá*), ve kterém momentě se upne k dalšímu stylu či žánru, který nemá v nedávné paměti - ať už je zcela originální, nebo pouze zapomenutý a znovuobjevený. Automatické systémy tento proces pouze urychlují a je na zvážení, kam povede zvyšující se míra eklekticismu vyplývající z toho, že nové a nové žánry a styly budou objevovány a znovuobjevovány, aniž by se na ně stačilo stejnou měrou také zapomínat. Autor si dovoluje spekulovat, že obecný vkus se buď dostane do bodu jakési postmoderní singularity, kde je všechno současně *in* i *out*, nebo limity lidského vnímání pod náporom nových a nových stylů a žánrů povedou k exponenciálnímu znásobení subkultur a kulturních „*niche bublin*“, ve kterých bude kulturní vývoj probíhat fyziologicky únosným (*ustíhatelným*) tempem. Obě varianty můžeme do jisté míry pozorovat již dnes.

4.9. Další postřehy

Na věc se dá nahlížet tak, že stroje se neučí nové věci, ale lidé konečně nacházejí nové možnosti enkódování informací, které v daném jsovcu byly vždy přítomny. Počítače rozumí pouze číslům a pokroky ve strojovém učení znamenají, že jsme objevili nový způsob interpretace částí světa kolem nás při zachování relevantních logických vztahů.

Alarmisté ve tvůrčích oborech zřejmě leckdy vycházejí z předpokladu, že dostupnost sofistikovaných nástrojů znamená, že je všichni začnou okamžitě používat. Z vlastní zkušenosti ale autor ví, že dostupnost nástrojů se možná mění, ale reálně jde o potenciál lidí konat – autor si dovoluje spekulovat, že ten se zřejmě nebude měnit tak závratným tempem. Autor se mnohokrát obklopil novými pluginy či nástroji s cílem pravidelně je používat a stvořit spoustu děl, jen aby na ně během týdne zapomněl a začal se věnovat se jiným starostem.

Možná, že to, co v automatických systémech (ve tvůrčím kontextu) doopravdy hledáme, není inteligence, ale *duše*. Umělá inteligence je v tomto ohledu přesně tím, čím je nazývána – umělou. Ostatně, již dnes se označení něčeho jako „vygenerovaného umělého inteligencí“ v kontextu tvůrčích aktivit používá nezdědka jako nadávka – *tento scénář je zcela o ničem, vypadá skoro jako kdyby jej vygenerovala AI*.

Závěr

„Ve zlaté horečce prodávej lopaty.”

- Kalifornské přísloví

Když lidé platí jiné lidi, aby pro ně něco dělali, tak proto, že jim chybí buď čas, nebo patřičná dovednost, popř. obojí. Rozšiřování sofistikovaných automatických systémů snižuje bariéry k tomu, aby lidé dělali dříve specializované věci sami, a někteří toho jistě využijí (všichni víme, co se během posledních několika dekád stalo s množstvím nahrávacích studií). Ale když zkusíme používat například některé z posledních generativních modelů, rychle přijdeme na to, že jedna bariéra byla nahrazena druhou – rok před vznikem této práce „*prompt engineering*” jako profese ještě neexistovala.

Je pravděpodobné, že AI mastering zvýší obecný zvukový standard amatérské produkce, protože skrze strojové učení je relativně jednoduché automaticky kompenzovat ty největší zvukové nedostatky, ovšem AI zřejmě masteringové inženýry jen tak nenahradí, protože ti se mezitím stačili vyvinout z techniků v „*affective labourers*” - kreativní partnery lidí, kteří je oslovují.

Je pravděpodobné, že automatizované systémy nás nahradí pouze tehdy, pokud jim to dovolíme, tj. pokud je odmítneme vzít v potaz a využít je ve svůj prospěch.

Je pravděpodobné, že dojde k dalšímu přehodnocení toho, v čem mastering vlastně spočívá a dojde k dalšímu rozmazávání jeho definice v rámci hierarchie zvukové tvorby, nicméně autor této práce její psaní ukončuje s větší jistotou než dříve, že lidský vklad ještě neřekl poslední slovo.

Bibliografie

AGNEW, J.I. *From The Cylinder To The Disk Record...And Back? – PS Audio* [online] [cit. 24.08.2022]. Dostupné z: <https://www.psaudio.com/copper/article/from-the-cylinder-to-the-disk-record-and-back/>

BARTHES, Roland a Stephen HEATH. *Image, music, text: essays*. 13. [Dr.]. vyd. London: Fontana, 1977. ISBN 978-0-00-686135-5.

BIRCHNELL, Thomas. Listening without ears: Artificial intelligence in audio mastering. *Big Data & Society*. 2018, roč. 5, č. 2, s. 205395171880855. ISSN 2053-9517, 2053-9517. DOI: 10.1177/2053951718808553

BRECHT DEMAN, RYAN STABLES, a JOSHUA D. REISS. *Intelligent Music Production*. , s. 219.

BUONOCORE, Tommaso. Man is to Doctor as Woman is to Nurse: the Dangerous Bias of Word Embeddings. In: *Medium* [online]. 3. 3. 2020 [cit. 28.05.2023]. Dostupné z: <https://towardsdatascience.com/gender-bias-word-embeddings-76d9806a0e17>

CAUDILL, Maureen. *Neural Network Primer: Part I. AI Expert*. 1989

CLARK, Mark a Henry NIELSEN. Crossed Wires and Missing Connections: Valdemar Poulsen, The American Telegraph Company, and the Failure to Commercialize Magnetic Recording. *Business History Review*. 1995, roč. 69, č. 1, s. 1–41. ISSN 0007-6805, 2044-768X. DOI: 10.2307/3117119

DANIEL, Eric D., C. Denis MEE a Mark CLARK. *Magnetic Recording: The First 100 Years*. ISBN 0-7803-4709-9.

EBU. 10 things you need to know about Next Generation Audio.

ELLIOTT, Mitchell a Song Hui CHON. A Comparative Study of Music Mastered by Human Engineers and Automated Services. *Journal of the Audio Engineering Society*. 2022, roč. 70, č. 9, s. 764–776. ISSN 15494950. DOI: 10.17743/jaes.2022.0050

FABBRI, Franco. *What kind of Music?* . 2023

HOFSTADTER, Douglas R. *Gödel, Escher, Bach: an eternal golden braid*. 20th anniversary ed. vyd. New York: Basic Books, 1999. ISBN 978-0-394-75682-0.

IŠTVÁNEK, Matěj. *Analýza interpretace hudby metodami číslicového zpracování signálu*.

IZOTOPE. Neutron 4 Help Documentation. In: [cit. 28.05.2023]. Dostupné z: <https://s3.amazonaws.com/izotopedownloads/docs/neutron4/en/index.html>

IZOTOPE. Ozone 10 Help. In: [cit. 28.05.2023]. Dostupné z: <https://docs.izotope.com/ozone10/en/master-rebalance/index.html>

IZOTOPE. RX 10 Help. In: [cit. 28.05.2023]. Dostupné z: <https://docs.izotope.com/rx10/en/dialogue-isolate/index.html>

JACOT, Arthur, Franck GABRIEL a Clément HONGLER. *Neural Tangent Kernel: Convergence and Generalization in Neural Networks* [online]. arXiv, 2020 [cit. 02.08.2022]. Dostupné z: <http://arxiv.org/abs/1806.07572>

JENKINS, Amanda. Inside the Archival Box: The First Long-Playing Disc | Now See Hear! In: *The Library of Congress* [online]. 13. 4. 2019 [cit. 28.05.2023]. Dostupné z: [//blogs.loc.gov/now-see-hear/2019/04/inside-the-archival-box-the-first-long-playing-disc](https://blogs.loc.gov/now-see-hear/2019/04/inside-the-archival-box-the-first-long-playing-disc)

JEŽEK, Ondřej. *Rozhovor o AI & Masteringu* [Archiv autora]. 2022.

KATZ, Robert A. *Mastering audio: the art and the science*. Oxford: Boston : Focal Press, 2002. ISBN 978-0-240-80545-0.

KHARE, Yana. *OpenAI's New Tool Explains Behavior of Language Model At Every Neuron Level* [online]. 2023 [cit. 26.05.2023]. Dostupné z: <https://www.analyticsvidhya.com/blog/2023/05/openai-new-tool-explains-behavior-of-language-model-on-every-neuron-level/>

KOHN, Spencer C. et al. Measurement of Trust in Automation: A Narrative Review and Reference Guide. *Frontiers in Psychology* [online]. 2021, roč. 12 [cit. 02.05.2023]. ISSN 1664-1078. Dostupné z: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.604977>

KRIZHEVSKY, Alex, Ilya SUTSKEVER a Geoffrey E HINTON. *ImageNet Classification with Deep Convolutional Neural Networks* [online]. Curran Associates, Inc., 2012 [cit. 28.05.2023]. Dostupné z: https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

LEGG, Shane a Marcus HUTTER. *A Collection of Definitions of Intelligence* [online]. arXiv, 2007 [cit. 01.08.2022]. Dostupné z: <http://arxiv.org/abs/0706.3639>

MOFFAT, David a Mark B. SANDLER. Approaches in Intelligent Music Production. *Arts*. 2019, roč. 8, č. 4, s. 125. ISSN 2076-0752. DOI: 10.3390/arts8040125

MOLNAR, Christoph. *Interpretable machine learning: a guide for making black box models explainable (2nd ed.)*. 2022. ISBN 9798411463330.

MORAY, Neville, John LEE a Bonnie M. MUIR. *Trust and human intervention in automated systems*. 1994.

OWSINSKI, Bobby a Sally ENGLEFRIED. *The mastering engineer's handbook: the audio mastering handbook*. 2nd ed. vyd. Boston, MA : Thomson Course Technology PTR, 2008. ISBN 978-1-59863-449-5.

RAI, Arun. Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science*. 2020, roč. 48, č. 1, s. 137–141. ISSN 0092-0703, 1552-7824. DOI: 10.1007/s11747-019-00710-5

RAMÍREZ, Marco A Martínez, Emmanouil BENETOS a Joshua D REISS. A GENERAL-PURPOSE DEEP LEARNING APPROACH TO MODEL TIME-VARYING AUDIO EFFECTS.

ROBINSON, Spider. *Melancholy elephants*. New York, N.Y: Tom Doherty Associates, 1985. ISBN 978-0-8125-5231-7.

RUMELHART, David E., Geoffrey E. HINTON a Ronald J. WILLIAMS. Learning representations by back-propagating errors. *Nature*. 1986, roč. 323, č. 6088, s. 533–536. ISSN 0028-0836, 1476-4687. DOI: 10.1038/323533a0

RUSSELL, Stuart J. a Peter NORVIG. *Artificial intelligence: a modern approach*. Fourth edition. vyd. Hoboken: Pearson, 2021. Pearson series in artificial intelligence. ISBN 978-0-13-461099-3.

- STAFF, News. Astronomers Teach AI to 'See' Astronomical Images | Sci.News. In: *Sci.News: Breaking Science News* [online]. 8. 7. 2015 [cit. 28.05.2023]. Dostupné z: <https://www.sci.news/astronomy/science-ai-unsupervised-learning-algorithm-02998.html>
- STERNE, Jonathan a Elena RAZLOGOVA. Machine Learning in Context, or Learning from LANDR: Artificial Intelligence and the Platformization of Music Mastering. *Social Media + Society*. 2019, roč. 5, č. 2, s. 205630511984752. ISSN 2056-3051, 2056-3051. DOI: 10.1177/2056305119847525
- STERNE, Jonathan a Elena RAZLOGOVA. Tuning sound for infrastructures: artificial intelligence, automation, and the cultural politics of audio mastering. *Cultural Studies*. 2021, roč. 35, č. 4–5, s. 750–770. ISSN 0950-2386, 1466-4348. DOI: 10.1080/09502386.2021.1895247
- STEVENS, Richard a Dave RAYBOULD. *The Game Audio Tutorial*. 0. vyd. Routledge, 2013. ISBN 978-1-136-12702-1. DOI: 10.4324/9780240817279
- SYROVÝ, Václav. *Hudební akustika*. Akademie múzických umění v Praze, 2013. ISBN 978-80-7331-297-8.
- VASWANI, Ashish et al. *Attention Is All You Need* [online]. arXiv, 2017 [cit. 28.05.2023]. Dostupné z: <http://arxiv.org/abs/1706.03762>
- VELARDO, Valerio. 2- AI, machine learning and deep learning - YouTube. In: [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/>
- ZHANG, Maggie. Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software. In: *Forbes* [online] [cit. 28.05.2023]. Dostupné z: <https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/>
- 14- SOLVING OVERFITTING in neural networks - YouTube. In: [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/>
- 17- *Recurrent Neural Networks Explained Easily* [online]. 2020 [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/watch?v=DY82Goknf0s>
- 18- *Long Short Term Memory (LSTM) Networks Explained Easily* [online]. 2020 [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/watch?v=eCvz-kB4yko>
- A Basic Introduction To Neural Networks. In: [cit. 28.05.2023]. Dostupné z: <https://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html>
- Adversarial Policies in Go - Game Viewer. In: [cit. 28.05.2023]. Dostupné z: https://goattack.far.ai/pdfs/go_attack_paper.pdf?uuid=WgOkT29w1fP7rG7p1682
- AES Mastering Academy pt.2. 2022.*
- Alan Blumlein and the invention of Stereo | EMI Archive Trust. In: . Dostupné z: <https://www.emiarchivetrust.org/alan-blumlein-and-the-invention-of-stereo/>
- BandLab | Unlimited FREE Instant Online Audio Mastering. In: *BandLab* [online] [cit. 28.05.2023]. Dostupné z: <https://www.bandlab.com/mastering>
- Bill Klinger; Association for Recorded Sound Collections. , s. 16.
- Brown Wax Cylinders. In: [cit. 22.08.2022]. Dostupné z: https://www.cylinder.de/guide_brown-wax-cylinders.html

CloudBounce - Instant Audio Mastering. Just hear it. In: [cit. 28.05.2023]. Dostupné z: <https://www.cloudbounce.com/>

Free on-line Mastering. In: *MasteringBOX* [online] [cit. 28.05.2023]. Dostupné z: <https://www.masteringbox.com/>

How to Extract Audio Features - YouTube. In: [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/watch?v=8A-W1xk7qs8&t=500s>

iZotope and Assistive Audio Technology. In: *iZotope* [online] [cit. 28.05.2023]. Dostupné z: <https://www.izotope.com/en/learn/izotope-and-assistive-audio-technology.html>

LANDR: Creative Tools for Musicians. In: *LANDR* [online]. 20. 3. 2023 [cit. 28.05.2023]. Dostupné z: <https://www.landr.com/>

Machine Learning Glossary. In: *Google for Developers* [online] [cit. 28.05.2023]. Dostupné z: <https://developers.google.com/machine-learning/glossary>

MajorDecibel | Instant Online Audio Mastering Service. In: [cit. 28.05.2023]. Dostupné z: <https://majordecibel.com/>

Mel-Frequency Cepstral Coefficients Explained Easily - YouTube. In: [cit. 28.05.2023]. Dostupné z: https://www.youtube.com/watch?v=4_SH2nfbQZ8

Move 37: Artificial Intelligence, Randomness, and Creativity. In: *John Menick* [online]. 17. 10. 2016 [cit. 06.08.2022]. Dostupné z: <http://localhost:4000/writing/move-37-alpha-go-deep-mind.html>

New Music Curriculum Guidance. In: . Dostupné z: <https://sites.google.com/site/newmusiccurriculumguidance/key-musical-concepts/1-style-genre-and-tradition>

Online Audio Mastering by Grammy Winning Engineers | eMastered. In: [cit. 28.05.2023]. Dostupné z: <https://eMastered.com>

Pepe the Frog. In: *Wikipedia* [online]. 2023 [cit. 28.05.2023]. Dostupné z: https://en.wikipedia.org/w/index.php?title=Pepe_the_Frog&oldid=1156293322

Professional Mixing and Mastering Services [online] [cit. 28.05.2023]. Dostupné z: <https://www.audioanimals.co.uk/shop/studio-services/dolby-atmos-mastering>

Publications :: FirstSounds.ORG. In: [cit. 21.08.2022]. Dostupné z: <http://www.firstsounds.org/research/scott.php>

Rube Goldbergův stroj. In: *Necyklopedie* [online]. 20. 6. 2013 [cit. 02.08.2022]. Dostupné z: http://necyklopedie.org/wiki/Rube_Goldberg%C5%AFv_stroj

Should we be afraid of A.I?!? [online]. 2022 [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/watch?v=ZGetnk222YU>

smart:EQ 3 - the intelligent equalizer by sonible - Get your 30-day trial now [online] [cit. 28.05.2023]. Dostupné z: <https://www.sonible.com/smarteq3/>

sonible plug-ins: Artificial Intelligence inside [online] [cit. 28.05.2023]. Dostupné z: <https://www.sonible.com/blog/plugins-with-ai/>

When AI flags the ruler, not the tumor — and other arguments for abolishing the black box (VB Live) [online]. 2021 [cit. 28.05.2023]. Dostupné z: <https://www.youtube.com/watch?v=ZGetnk222YU>

z: <https://venturebeat.com/business/when-ai-flags-the-ruler-not-the-tumor-and-other-arguments-for-abolishing-the-black-box-vb-live/>

Seznam obrázků

Obr. 1 - Fonautografická nahrávka	4
Obr. 2 - "Vrchol realismu" - dobová reklama na fonograf	5
Obr. 3 - Emile Berliner se svým vynálezem	5
Obr. 4 - Berlinerův propagační materiál.	6
Obr. 5 - Poulsenův telegrafon	6
Obr. 6 - Edisonův Gold Moulded váleček	6
Obr. 7 - jediná dochovaná fotografie Fritze Pflaumera s magnetofonem	7
Obr. 8 - reklamní materiál, LP	8
Obr. 9 - prezentace magnetofonu Ampex	8
Obr. 10 - Program AES Convention v New Yorku, 1957, představující stereo vinylovou desku	9
Obr. 11 - Schéma Dynagroove systému	10
Obr. 12 - Analogová nahrávka valčíků Frederica Chopina (Philips 400 025) Claudia Arraua z března 1979 se stává prvním klasickým kompaktním diskem, který byl kdy komerčně vyroben.	11
Obr. 13 - Sonic Studio	11
Obr. 14 - příklady naměřených hodnot MIR parametrů napříč žánry	17
Obr. 15 - zprůměrovaná frekvenční spektra napříč žánry	18
Obr. 16 - zjednodušené schéma vrstev neuronů v NS	23
Obr. 17 - automatické pátrání po dosud neobjevených galaxiích	24
Obr. 18 - výsledky učení AlphaZero	25
Obr. 19 - algoritmus zpětného šíření chyby v NS	26
Obr. 20 - architektura konvoluční NS	27
Obr. 21 - paměťová buňka v RNN	28
Obr. 22 - interagující vrstvy v rámci paměťové buňky LSTM	29
Obr. 23 - Schéma transformery	30
Obr. 24 - extrakce zvukových parametrů pro NS	31
Obr. 25 - MFCC	32
Obr. 26 - Metric AB od ADAPTR Audio	35
Obr. 27 - Znázornění frekvenčního maskování ve FabFilter Pro-Q 3	35
Obr. 28 - iZotope Neutron, analýza signálu	36
Obr. 29 - Prostředí Sonible Smart EQ, bílá křivka vychází z automatické analýzy	37
Obr. 30 - Prostředí herního enginu Unity, který lze použít pro vytváření automatických mixů ve videoherním kontextu.	38
Obr. 31 - uživatelsky přístupné parametry LANDR	39
Obr. 32 - prostředí iZotope Ozone, dialog asistivní části	40
Obr. 33 - Propagační materiál LANDRu o vlastním vývoji zvukových procesorů	42
Obr. 34 - z materiálů LANDR, podobná vyjádření nalezneme napříč všemi službami	43
Obr. 35 - naměřená hlasitost všech verzí Easter	45
Obr. 36 - waveformy všech verzí Easter	45
Obr. 37 - naměřená hlasitost všech verzí Nahrávky #2	46
Obr. 38 - waveformy všech verzí Nahrávky #2	46
Obr. 39 - Schéma Goldbergova aparátu	49
Obr. 40 - vizualizace detekovaných rysů v jednotlivých vrstvách CNN	50
Obr. 41 - výstup Hofstadterova programu	51