



## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

<b>Název:</b>	Zpracování a úpravy výstupů automatické transkripce hudby
<b>Student:</b>	Zuzana Fílová
<b>Vedoucí:</b>	doc. RNDr. Pavel Surynek, Ph.D.
<b>Studijní program:</b>	Informatika
<b>Studijní obor:</b>	Znalostní inženýrství
<b>Katedra:</b>	Katedra aplikované matematiky
<b>Platnost zadání:</b>	Do konce letního semestru 2019/20

### Pokyny pro vypracování

Cílem práce je navrhnout systém pro vylepšování notového záznamu (například ve formátu MIDI) získaného automatickým převodem z digitálního vlnového zvukového záznamu (například ve formátu .wav). Výsledkem automatického převodu zvukového záznamu často bývá pro člověka těžko čitelný notový záznam, protože neobsahuje informace o tónině, metru, tempu atd., které jsou pro následnou reprodukci skladby podstatné a klasický notový záznam je obsahuje. Práce se tedy bude zaměřovat na vylepšení automaticky získaného notového záznamu vzhledem k jeho čitelnosti s využitím poznatků z hudební teorie. Jako možné výpočetní nástroje se nabízejí techniky z teorie jazyků a zpracování řetězců (automaty a gramatiky) nebo strojové učení.

Očekávaným výsledkem práce je teoretický koncept (pravidla či algoritmus) pro vylepšování notového záznamu. Navržený koncept řešitel dále implementuje a experimentálně vyhodnotí jeho úspěšnost.

### Seznam odborné literatury

- [1] Valentin Emiya, Roland Badeau, Bertrand David: Automatic transcription of piano music based on HMM tracking of jointly-estimated pitches. Proceedings of the 16th European Signal Processing Conference (EUSIPCO 2008), pp. 1-5, IEEE Computer Society, 2008.
- [2] Graham Grindlay, Daniel P. W. Ellis: Transcribing Multi-Instrument Polyphonic Music With Hierarchical Eigeninstruments. J. Sel. Topics Signal Processing, Volume 5(6), pp. 1159-1169, IEEE Computer Society, 2011.
- [3] Benoit Fuentes, Roland Badeau, Gaël Richard: Harmonic Adaptive Latent Component Analysis of Audio and Application to Music Transcription. IEEE Trans. Audio, Speech & Language Processing, Volume 21(9), pp. 1854-1866, IEEE Computer Society, 2011.

Ing. Karel Klouda, Ph.D.  
vedoucí katedry

doc. RNDr. Ing. Marcel Jiřina, Ph.D.  
děkan

V Praze dne 19. listopadu 2018





**FAKULTA  
INFORMAČNÍCH  
TECHNOLGIÍ  
ČVUT V PRAZE**

Bakalářská práce

## **Zpracování a úpravy výstupů automatické transkripce hudby**

*Bc. Zuzana Fílová*

Katedra aplikované matematiky

Vedoucí práce: doc. RNDr. Pavel SURYNEK, Ph.D.

13. května 2019



---

# Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval(a) samostatně a že jsem uvedl(a) veškeré použité informační zdroje v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů. V souladu s ust. § 46 odst. 6 tohoto zákona tímto uděluji nevýhradní oprávnění (licenci) k užití této mojí práce, a to včetně všech počítačových programů, jež jsou její součástí či přílohou, a veškeré jejich dokumentace (dále souhrnně jen „Dílo“), a to všem osobám, které si přejí Dílo užít. Tyto osoby jsou oprávněny Dílo užít jakýmkoli způsobem, který nesnižuje hodnotu Díla, a za jakýmkoli účelem (včetně užití k výdělečným účelům). Toto oprávnění je časově, teritoriálně i množstevně neomezené. Každá osoba, která využije výše uvedenou licenci, se však zavazuje udělit ke každému dílu, které vznikne (byť jen zčásti) na základě Díla, úpravou Díla, spojením Díla s jiným dílem, zařazením Díla do díla souborného či zpracováním Díla (včetně překladu), licenci alespoň ve výše uvedeném rozsahu a zároveň zpřístupnit zdrojový kód takového díla alespoň srovnatelným způsobem a ve srovnatelném rozsahu, jako je zpřístupněn zdrojový kód Díla.

V Praze dne 13. května 2019

.....

České vysoké učení technické v Praze  
Fakulta informačních technologií

© 2019 Zuzana Fílová. Všechna práva vyhrazena.

*Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení na předchozí straně, je nezbytný souhlas autora.*

### **Odkaz na tuto práci**

Fílová, Zuzana. *Zpracování a úpravy výstupů automatické transkripce hudby*. Bakalářská práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2019.

---

# Abstrakt

Předmětem této bakalářské práce je problematika automatické transkripce hudby (AMT). V teoretické části práce byly shrnuty základní informace o zmíněné problematice, byl zde představen aktuální stav řešení AMT a nastíněny problémy, se kterými se tato výzkumná oblast potýká. Dále byly popsány různé možnosti zhodnocení výsledků systémů AMT. Bylo zjištěno, že automatická transkripce hudby v současnosti nedosahuje uspokojivých výsledků. Výsledná reprezentace hudby (MIDI soubor nebo notový zápis) obsahuje mnoho chyb, které znemožňují praktické využití této technologie.

V praktické části práce proto byla provedena analýza nejčastějších chyb vznikajících při automatické transkripci vlnového zvukového záznamu do formátu MIDI. Na základě této analýzy byla navržena metoda pro automatické odstranění těchto chyb a úpravu získaného MIDI souboru. Cílem těchto úprav bylo vylepšení úspěšnosti systému automatické transkripce hudby. Úspěšnost navržené metody byla hodnocena pomocí F-míry a editační vzdálenosti hudebních řetězců.

Experimenty v závěrečné části práce ukázaly, že navržená metoda úprav zvětšuje podobnost výsledných notových záznamů s originálem a zásadním způsobem přispívá k lepší čitelnosti výsledných notových záznamů.

**Klíčová slova** automatická transkripce hudby, automatická oprava transkripčních chyb, hodnocení transkripčních systémů, převod zvuku do MIDI, převod zvuku do not, zpracování hudebních signálů, notový zápis

# Abstract

The subject of this bachelor thesis is focused on the automatic music transcription (AMT). The basic information related to this problematic, the state of the art in the solution of AMT and the problems tackled in this research domain are summarized in the theoretical part of this work. The approaches to the evaluation of results obtained by different AMT systems are also given. It has been found that unsatisfactory results are reached by AMT till now. Resulting music representation (MIDI file or sheet music) contains many errors that make it impossible to use this technology in practice.

Therefore, in the practical part of the thesis, an analysis of the most common errors generated by the automatic transcription of a wave audio recording into a format MIDI were analyzed. Based on this analysis, a method for automatic removal of these errors and modification of acquired MIDI file is proposed. The goal of these modifications was to improve the success of the automatic music transcription system. The success of the proposed method was evaluated by F-measure and edit distance of the musical strings.

The experiments presented in the final part of the thesis showed that the proposed method of adjustment increases the similarity of the resulting sheet music with the original and contributes significantly to the better readability of the resulting sheet music.

**Keywords** automatic music transcription, automatic correction of transcription errors, evaluation of music transcription systems, music signal processing, audio to MIDI, audio to music notation, music score, music information retrieval



---

# Obsah

Úvod	1
1 Cíl práce	3
I Teoretická část	5
2 Automatická transkripce hudby	7
2.1 Využití a motivace	8
2.2 Představení problému AMT	9
3 Reprezentace hudby v procesu transkripce	15
3.1 Zvuk	15
3.2 Parametrická reprezentace	18
3.3 Notový zápis	20
4 Současný stav řešení	29
4.1 Existující software	32
5 Hodnocení výsledků systémů AMT	37
5.1 Parametrické systémy AMT	38
5.2 Úplné systémy AMT	41
5.3 Metody porovnávání hudebních sekvencí	46
II Praktická část	51
6 Popis problému	53
6.1 Postup úplné transkripce	53
6.2 Představení problému	54

<b>7</b>	<b>Analýza chyb automatické transkripce</b>	<b>57</b>
7.1	Z monofonie se stává polyfonie . . . . .	60
7.2	Fragmentace dlouhých not . . . . .	61
7.3	Chyby v detekci výšky tónu . . . . .	61
7.4	Nepřesnosti . . . . .	62
<b>8</b>	<b>Návrh metody pro automatickou úpravu transkripce</b>	<b>63</b>
8.1	Redukce polyfonie na monofonii . . . . .	64
8.2	Implementace . . . . .	67
<b>9</b>	<b>Experimentální vyhodnocení</b>	<b>69</b>
9.1	Dataset . . . . .	69
9.2	Použité metriky . . . . .	69
9.3	Experimenty . . . . .	71
9.4	Shrnutí výsledků . . . . .	82
	<b>Závěr</b>	<b>83</b>
	<b>Literatura</b>	<b>85</b>
	<b>A Seznam použitých zkratk</b>	<b>93</b>
	<b>B Obsah příloženého CD</b>	<b>95</b>

---

## Seznam obrázků

2.1	Separace jednotlivých zdrojů zvuku z nahrávky [5] . . . . .	8
2.2	Schéma systému AMT pro transkripci se znalostí originálních not (score-informed transcription) [7] . . . . .	9
2.3	Architektura úplného systému AMT [10] . . . . .	11
2.4	Převod MIDI do notového zápisu . . . . .	11
2.5	Návrh obecné architektury systému AMT . . . . .	12
3.1	Časový průběh zvukového signálu [19] . . . . .	16
3.2	Prvních dvacet harmonických frekvencí tónu $A_1$ [20] . . . . .	16
3.3	Spektrum tónu $C\sharp$ . . . . .	17
3.4	Spektrogram . . . . .	18
3.5	Reprezentace hudby ve formě piano-roll [27] . . . . .	20
3.6	Znázornění enharmonických tónů na klaviatuře [28] . . . . .	20
3.7	Vztah mezi frekvencemi tónů, jejich názvy a odpovídajícími MIDI čísly na části klaviatury [29] . . . . .	21
3.8	Ukázka notového zápisu ve standardní hudební notaci . . . . .	23
3.9	Přehled not a pomlk . . . . .	24
3.10	Části noty . . . . .	24
3.11	Ukázka dělení a prodlužování not a pomlk [34] . . . . .	25
3.12	Přehled předznamenání durových a mollových tónin [35] . . . . .	27
3.13	Ukázka zápisu hudební skladby ve formátu abc. . . . .	28
5.1	Výsledky hodnocení MIREX detekce not v polyfonní nahrávce s více nástroji [5] . . . . .	38
5.2	Srovnání notových záznamů . . . . .	41
5.3	Rozdílná čitelnost různých notových zápisů . . . . .	43
5.4	Ukázka prvního kroku při hodnocení úspěšnosti transkripce – zrovnání [3] . . . . .	45
5.5	Přehled prací, které přistupují k notám jako k řetězcům [62] . . . . .	47

7.1	Porovnání not po automatické transkripci s originálem u písně Drowsy Maggie . . . . .	58
7.2	Porovnání not po automatické transkripci s originálem u písně Josephine's Waltz . . . . .	59
7.3	Ukázka chyb ve výstupu z AmazingMIDI . . . . .	60
8.1	Ukázka monofonní a různých polyfonních hudebních sekvencí. . . . .	64
9.1	Úspěšnost systému pro různé hodnoty parametrů. . . . .	73
9.2	Úspěšnost systému pro různé hodnoty parametrů. . . . .	74
9.3	Výsledky F-míry, editačních vzdáleností a správnosti pro jednotlivé položky z datasetu. . . . .	75
9.4	Porovnání výsledků systému AMT bez úprav / s navrženými úpravami . . . . .	76
9.5	Výsledky F-míry pro jednotlivé položky z datasetu pro systém bez úprav / s navrženými úpravami. . . . .	77
9.6	Ukázka výsledků navržené metody. . . . .	78
9.7	Ukázka výsledků navržené metody. . . . .	79
9.8	Ukázka výsledků navržené metody. . . . .	80
9.9	Porovnání různých způsobů redukce polyfonie na monofonii . . . . .	81

---

## Seznam tabulek

8.1	Třída Note (seznam atributů) . . . . .	68
9.1	Porovnání různých variant funkce redukce polyfonie . . . . .	82



---

# Úvod

Problematika automatické transkripce hudby je velmi zajímavá a komplexní. Navzdory několika desetiletím aktivního výzkumu je tato oblast stále nevyřešeným a otevřeným problémem, ve kterém lidé dosahují znatelně lepších výsledků než stroje.

Automatická transkripce hudby má širokou škálu praktických využití od prostého získání notového materiálu ke skladbám, u kterých není dostupný originál (jazz, pop, lidové písně, ...), až po různé aplikace v muzikologii či interaktivní výuce hudby.

Existují problémy, které by se svou povahou či náročností daly k automatické transkripci hudby připodobnit, jako příklad lze uvést automatické rozpoznávání řeči nebo rozpoznávání objektů na fotografiích. Je proto zajímavé, že v těchto oblastech došlo v průběhu posledních let k mnohem většímu pokroku. Automatická hudební transkripce na dobré výsledky a objevy stále ještě čeká.

Důvodů k jisté upozaděnosti této problematiky je hned několik. Jedním z nich je fakt, že není ve velkém komerčním zájmu. [1] Výzkumy v této oblasti jsou spíše akademického rázu a nebyvají financovány ze soukromého sektoru. Naproti tomu, podproblémů, které se musí při automatické transkripci vyřešit, je opravdu velké množství a žádný z nich není triviální. Ohromná rozmanitost možných hudebních signálů se spojuje do mnoha různých kombinací, což dělá z automatické transkripce hudby mimořádně obtížné výzkumné téma.

Dalším důvodem je skutečnost, že se jedná o mezioborovou disciplínu, která spojuje celou řadu oborů jako je informatika, muzikologie, matematika a fyzika. Muzikologie, neboli hudební věda, je přitom již sama o sobě vědním oborem na rozhraní humanitních a exaktních věd. Řešení mezioborových problémů s sebou přináší jisté komplikace. Hudebníci a muzikologové většinou nerozumí technické stránce věci. Někteří dokonce moderní technologie v hudbě spíše odmítají a uznávají pouze tradiční metody. Stejně tak informatici nebo matematici zpravidla nemají dostatečné znalosti v oblasti hudební

teorie a hudba není ve středu jejich osobního ani profesního zájmu. Umělecká a humanitní stránka problému je pro jejich technický a analytický způsob myšlení jistým způsobem neuchopitelná.

Na druhou stranu tato mezioborovost poskytuje zajímavé možnosti nahlížet na problém z mnoha různých pohledů a využívat různé techniky a přístupy. Setkáváme se s přístupy zaměřenými z velké části na využití znalostí z hudební teorie (pokročilé znalosti harmonie atd.), se systémy, které se snaží napodobit postup práce člověka. Existují přístupy využívající metody strojového učení, umělou inteligenci nebo naopak matematické přístupy založené například na statistice a pravděpodobnosti. Najdou zde také uplatnění přístupy vycházející z bioinformatiky, které využívají algoritmy používané například k analýze či porovnávání řetězců DNA. Zajímavých výsledků dosahují metody, které různé pohledy na hudbu a její transkripci kombinují.

Téma práce jsem si vybrala z toho důvodu, že problém automatické transkripce hudby není dosud uspokojivě vyřešen, přináší s sebou mnoho nových výzev a nabízí prostor k objevování zcela nových přístupů a náhledů na tuto komplikovanou problematiku. Dalším důvodem pro výběr tohoto tématu byla skutečnost, že systém, provádějící spolehlivě automatickou transkripci hudby, bych také prakticky využila.

Teoretická část práce může být užitečná pro kohokoliv, kdo chce získat přehled v problematice automatické hudební transkripce. Výsledek praktické části lze využít k vylepšování a opravování chyb v notovém zápisu, který byl získán pomocí automatické transkripce. To může být užitečné například pro hudebníky, kteří následně chtějí skladby interpretovat.

## Struktura práce

V teoretické části práce je nejprve definován pojem automatické transkripce hudby. Jsou zde vysvětleny základní pojmy související s problematikou a popsány tři hlavní druhy reprezentace hudby v průběhu procesu transkripce – zvuk, parametrická reprezentace a notový zápis. Dále je představen současný stav řešení problému a existující software. Poslední část je poté věnována důležitému podtématu a to hodnocení výsledků systémů AMT. Zde jsou představeny různé možnosti, jak k tomuto hodnocení přistupovat.

Praktická část práce se nejprve zabývá analýzou nejčastějších problémů a chyb vznikajících při automatické hudební transkripci a následně obsahuje návrh metody na automatické opravování, případně vylepšování, výstupů automatické transkripce hudby. Poslední část se věnuje experimentům a vyhodnocení úspěšnosti navržené metody.



---

## Cíl práce

Cílem rešeršní části práce je shrnutí problematiky automatické transkripce hudby, vysvětlení základních pojmů souvisejících s touto problematikou a zmapování současného stavu řešení problému včetně existujícího softwaru.

Cílem praktické části práce je analýza nejčastějších chyb vznikajících při automatické transkripci vlnového zvukového záznamu do formátu MIDI a následný návrh systému pro automatické odstranění těchto chyb a úpravu získaného MIDI souboru.

Výsledkem této části práce by měl být teoretický koncept (pravidla či algoritmus) pro vylepšování notového záznamu. Cílem práce je dále navržený koncept implementovat a experimentálně vyhodnotit jeho úspěšnost.



Část I  
Teoretická část



## Automatická transkripce hudby

Automatická transkripce hudby (AMT) označuje proces převodu zvukového hudebního signálu do nějaké formy hudební notace pomocí počítače.

Přesná definice pojmu není ustálená. Jednotlivé verze definice se liší ve specifikaci formy vzniklé hudební notace. Někteří autoři definují AMT jako proces převodu zvukové nahrávky do parametrické reprezentace hudby, zatímco jiní autoři tento pojem definují jako převod zvukové nahrávky do standardní hudební notace. [1]

Společné pro tyto reprezentace je to, že popisují z hudebního hlediska smysluplné parametry a mohou být využity k interpretaci hudebníkem nebo syntetizaci původní skladby. Z tohoto úhlu pohledu můžeme automatickou transkripci hudby vnímat jako formu reverzního inženýrství a hledání „zdrojového kódu“ původního zvukového signálu. [2]

Systémy AMT můžeme rozdělit do dvou kategorií podle formy reprezentace hudby, která převodem vzniká, na:

- parametrické,
- úplné.

Výsledkem práce parametrického systému AMT je, jak již vyplývá z názvu, parametrická reprezentace hudby (viz 3.2). Výsledkem práce úplného systému AMT je zápis ve standardní hudební notaci (viz 3.3).

Velká část systémů pro automatickou transkripci hudby se zabývá pouze přepisem do parametrické reprezentace hudby, kterou můžeme brát jako mezystupeň mezi vlnovým zvukem a standardním notovým zápisem. Tento způsob reprezentace dokáže plně reprezentovat původní nahrávku a je pro některé aplikace dostatečný.

V mnoha případech je ale tento způsob reprezentace hudby nedostatečný, neboť trpí absencí některých důležitých hudebních informací, které se v klasickém notovém zápisu objevují a jsou nezbytné například pro následnou interpretaci člověkem, počítačovou hudební analýzu, výuku hudby pomocí počítače

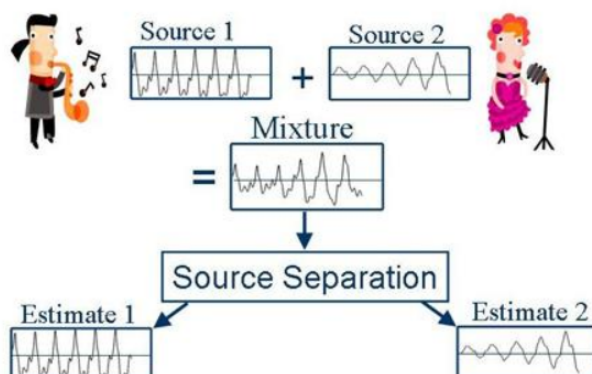
a další. Parametrická reprezentace nijak nezohledňuje použitou tóninu, tempo, neřeší enharmonické záměny ( $A\flat$  versus  $G\sharp$ ) a tak dále.

Existují také systémy, které se zabývají pouze převodem parametrického zápisu hudby do standardní hudební notace. [3]

## 2.1 Využití a motivace

Automatická transkripce hudby v sobě spojuje oblast zpracování hudebního signálu se zpracováním symbolického zápisu hudby. AMT jako taková je technologie se značným potenciálem ovlivnit hospodářskou i společenskou sféru a velice úzce souvisí s dalšími úlohami zpracování hudebních signálů. Příkladem je identifikace hudebních nástrojů nebo separace jednotlivých zdrojů zvuku z nahrávky (viz obrázek 2.1).

AMT je také užitečná pro mnoho navazujících úkolů z oblasti získávání znalostí z hudby (Music Information Retrieval – MIR) jako je detekce cover verzí nebo hodnocení hudební podobnosti. Tyto úkoly jsou snadněji řešitelné v případě, že k daným skladbám známe noty. [4]



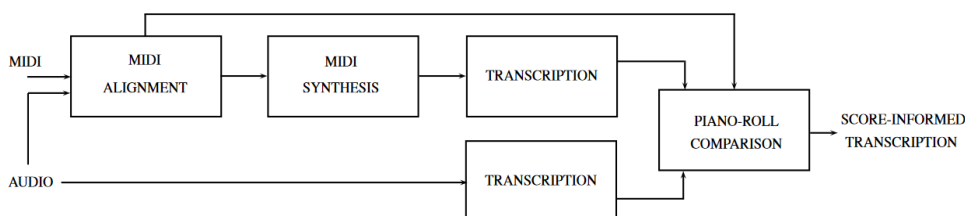
Obrázek 2.1: Separace jednotlivých zdrojů zvuku z nahrávky [5]

Úspěšný systém automatické transkripce hudby by umožňoval široké možnosti využití zahrnující získání notových záznamů ze zvukových nahrávek, pomoc při tvorbě hudby, zápis improvizovaných hudebních nápadů, automatický hudební doprovod nebo vizualizaci hudebního obsahu. [6]

Mezi další možná využití systémů AMT patří vyhledávání v hudebních záznamech (indexování a doporučování hudby podle melodie, basové linky, rytmu nebo akordů) a v neposlední řadě také využití v oblasti muzikologie (analýza jazzových improvizací, etnické a další hudby, ke které zpravidla není k dispozici notový záznam).

Dalším příkladem je rozmanitá škála využití ve výuce hudby pomocí počítače. Jedná se například o systémy pro automatickou výuku hry na hudební

nástroje, kdy student zahraje skladbu a počítač vyhodnotí, jaké chyby student udělal. Při využití AMT k těmto edukačním účelům nastává specifická situace, neboť se většinou snažíme o automatickou transkripci nahrávky, ke které zároveň známe noty. Ve většině případů existují nějaké rozdíly mezi notovým zápisem a danou zvukovou nahrávkou, které mohou být způsobeny osobitou interpretací studenta nebo jeho chybami. Systém musí tyto chyby a nesrovnalosti detekovat a studenta na ně upozornit. Informace z předem známých notových zápisů lze využít ke zjednodušení transkripčního procesu. Takovou transkripci poté nazýváme anglickým termínem *score-informed music transcription*. Schéma systému pro tento druh transkripce je zobrazeno na obrázku 2.2.



Obrázek 2.2: Schéma systému AMT pro transkripci se znalostí originálních not (score-informed transcription) [7]

Ačkoliv již proběhly první pokusy o vytvoření systémů pro tento druh hudební transkripce (např. pro klavír [8]), problém je stále poměrně neprozkoumaný. Stejně tak neprozkoumaný je i související a náročnější problém transkripce, při které neznáme noty pro všechny hlasy znějící v nahrávce, ale pouze pro vedoucí hlas. [4] V současné době relativně malá přesnost standardních metod AMT způsobuje situaci, kdy počet chyb, které udělá student, je zpravidla mnohokrát menší než počet chyb, které jsou způsobeny automatickou transkripcí. [8]

Metody získávání klasického notového zápisu z parametrické reprezentace hudby jsou poté nepostradatelné pro muzikologickou analýzu skladeb, velké množství aplikací, které zpracovávají MIDI vstupy (notační programy, interaktivní hudební systémy) a další. [9]

## 2.2 Představení problému AMT

Schopnost transkripce hudební nahrávky do notového zápisu je příkladem fascinující lidské inteligence. Člověk musí pro tuto činnost disponovat několika nepostradatelnými schopnostmi, které zahrnují percepci (sluchová analýza hudebního výstupu), kognici (rozpoznání odlišnosti zvuku jednotlivých nástrojů a schopnost rozeznání jednotlivých melodicko-rytmických struktur), znalost

hudební teorie a schopnost zhodnocení správnosti vytvořeného hudebního zápisu.

Automatická transkripce hudby je poté náročným úkolem v oblasti zpracování signálů a umělé inteligence. Vzhledem k počtu dílčích úkolů a širokému rozsahu využití je AMT považována za základní problém v oblasti zpracování hudebních signálů a MIR. [4]

Dle [1], [4] a [9] AMT zahrnuje několik dílčích úkolů:

- detekce výšek tónů (multi-pitch estimation);
- detekce začátků a konců znění tónů (onset and offset detection);
- rozpoznávání zvuku jednotlivých hudebních nástrojů;
- separaci jednotlivých hlasů;
- kvantizaci délek not;
- analýzu metra;
- rytmickou analýzu;
- identifikaci dynamiky a hudebního výrazu;
- harmonickou analýzu (detekci akordů, tóniny, ...);
- vytváření notového zápisu.

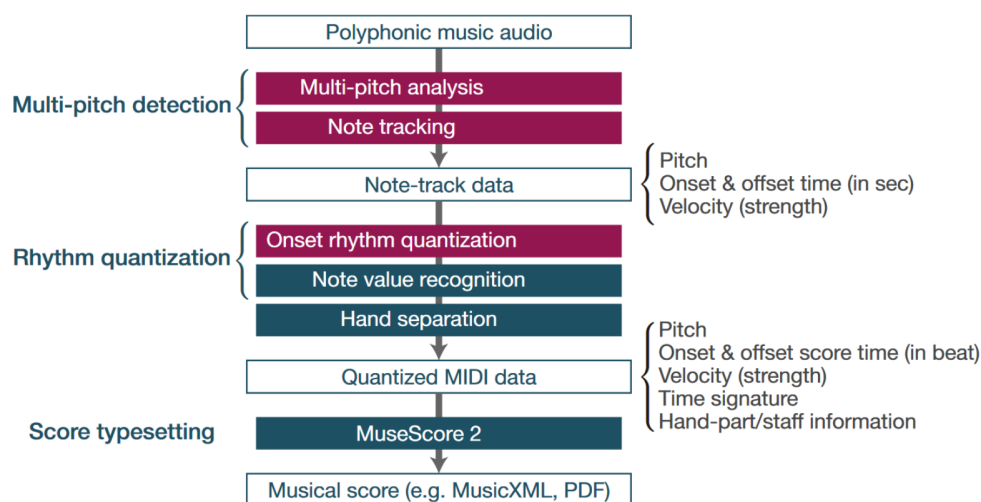
AMT zahrnuje samozřejmě také integraci všech těchto podúkolů do jednoho systému, která je také velice složitá. [10]

### 2.2.1 Systémy AMT

Od počátku vývoje systémů AMT se touto oblastí zabývaly desítky vědců a bylo navrženo mnoho různých parametrických systémů AMT. Jednotliví autoři využívají různé algoritmy a technologie, vstupní a výstupní formáty hudby atd. Rozsáhlý přehled navržených parametrických systémů mezi lety 1975 a 2010 včetně informací o reprezentaci hudby v průběhu transkripčního procesu, použitého předzpracování, evaluaci výsledků apod. lze najít v práci [11]. Další přehled parametrických systémů AMT lze najít v práci [12].

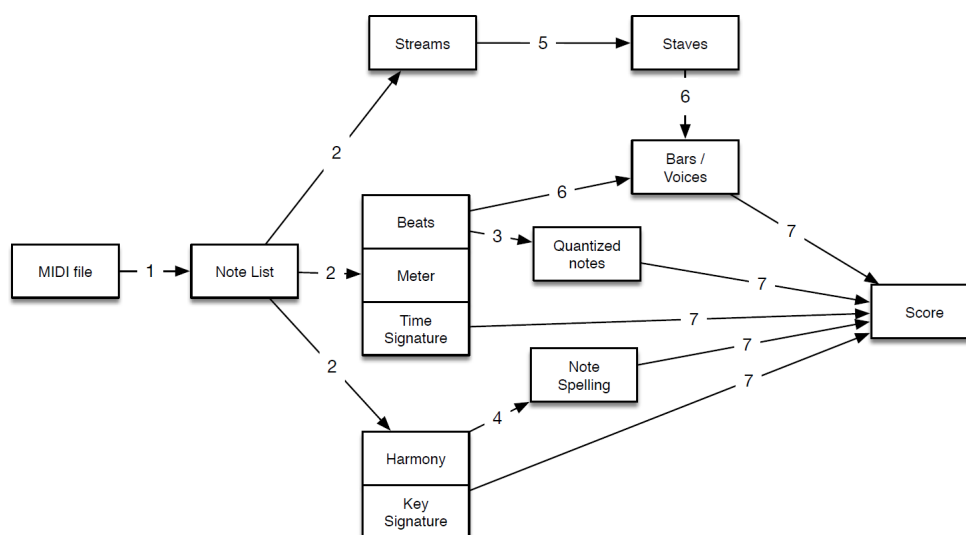
V případě úplných systémů AMT je situace odlišná. Benetos et al. [1] v roce 2013 a Cogliati et al. [6] v roce 2016 uvádí, že neexistuje žádný systém AMT pro úplnou transkripci. Nakamura et al. [10] v roce 2018 přichází s návrhem úplného systému AMT. Navržený systém integruje dosavadní metody pro zpracování polyfonního zvuku, rytmickou analýzu a odstraňuje chyby vzniklé při transkripci. Na obrázku 2.3 můžeme vidět schéma navrženého systému.





Obrázek 2.3: Architektura úplného systému AMT [10]

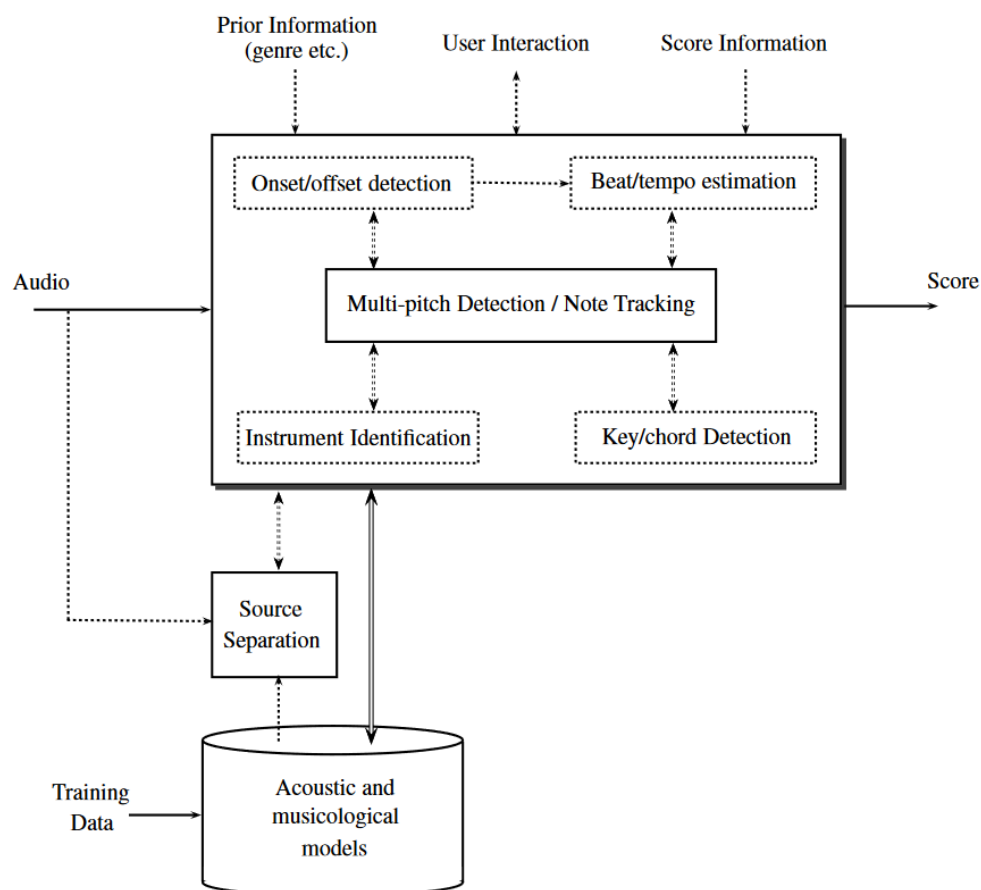
Dále existuje několik systémů AMT, které se zabývají převodem MIDI (získaného živou nahrávkou na digitální nástroje) do notového zápisu – [6], [9] a [13]. Na obrázku 2.4 je znázorněno schéma systému pro převod MIDI do notového zápisu.



Obrázek 2.4: Schéma převodu MIDI do notového zápisu. Šipky znázorňují vztahy mezi jednotlivými entitami, čísla ukazují pořadí jednotlivých kroků. [6]

### 2.2.1.1 Obecné schéma parametrického systému AMT

Přestože se jednotlivé systémy AMT od sebe značně liší, můžeme mezi nimi vysledovat určitou podobnost. Benetos et al. [1] popsali obecné schéma architektury parametrického systému AMT. Schéma je znázorněno na obrázku 2.5.



Obrázek 2.5: Návrh obecné architektury systému AMT. Podsystemy a algoritmy, které v systému nemusí být přítomny jsou znázorněny přerušovanými čarami. [1]

Jádrem transkripčního systému jsou algoritmy pro detekci znějících tónů z nahrávky (multi-pitch detection) a identifikaci not (note tracking). Další čtyři podúkoly transkripčního procesu – identifikace nástrojů, harmonická analýza (detekce tóniny a akordů), detekce začátků a konců not a rytmická analýza – jsou znázorněny v tečkovaných obdélnících a mohou, ale nemusí, být v transkripčním procesu přítomny. Separace jednotlivých zdrojů zvuku z nahrávky představuje nezávislý, ale velmi úzce související, problém. Separace jednotlivých zdrojů zvuku je řešena jako samostatná jednotka, která komuni-

kuje se systémem AMT (konkrétně s částí pro identifikaci nástrojů).

Volitelně mohou být do systému také integrovány informace o přepisované hudbě (např. žánr, nástrojové obsazení). Tyto informace mohou být získány interakcí s uživatelem nebo z částečně správného (případně neúplného) předem známého notového zápisu. Trénovací data lze využít k vytvoření akustických a muzikologických modelů, které následně mohou interagovat se systémem a dodávat do něj další užitečné informace.

### 2.2.2 Podobné problémy

Problém automatické transkripce hudby je v mnoha ohledech srovnatelný s automatickým rozpoznáním řeči, ačkoliv je samozřejmě nutné zmínit, že rozpoznávání řeči je studováno déle a je ve větším akademickém a komerčním zájmu. Společná věc charakteristická pro hudbu i pro řeč je povaha jejich vzniku, omezený počet diskretních prvků je kombinován do rozsáhlejších a složitějších struktur. V řeči se fonémy používají ke konstrukci slov a vět a v hudbě jsou jednotlivé zvuky kombinovány pro vytváření melodií, rytmů a skladeb.

Důležitým rozdílem mezi nimi je naopak to, že řeč je v podstatě monofonní, zatímco hudba je obvykle polyfonní. Při automatické hudební transkripci se ve velké míře potýkáme s problémy vyplývajícími z kombinatoriky: zvuky různých nástrojů se vyskytují v mnoha různých kombinacích a jsou skládány do rozličných hudebních skladeb. Na druhou stranu dynamická variabilita a komplexnost jedné hudební zvukové události není tak vysoká jako variabilita a komplexnost zvuku řeči.

Výhodou proto je, že při rozvoji a učení hudebního transkriberu mohou být do jisté míry využity synteticky generované hudební signály. [14]

Ve srovnání se zpracováním řeči, výzkumnou oblastí s dlouholetou tradicí, je zpracování hudby stále relativně mladou disciplínou. Zájem o ní v posledních letech však rychle roste. [15] Automatická transkripce zpěvu, pokud uvažujeme melodii i zpívaný text, v sobě tyto dva obory dokonce spojuje. [16]

Oproti rozpoznávání řeči se automatická transkripce hudby potýká s problémem získávání dat. V oblasti rozpoznávání řeči existují velké korpusy anotovaného zvuku, což umožňuje použití statistických metod a metod strojového učení (např. skryté Markovovy modely – hidden Markov model). V hudbě žádné podobné velké korpusy dat neexistují, použití těchto metod je z tohoto důvodu komplikované. [17]

Některé oblasti AMT jsou také připodobňovány ke zpracování obrazu a počítačovému vidění. [4]



---

# Reprezentace hudby v procesu transkripce

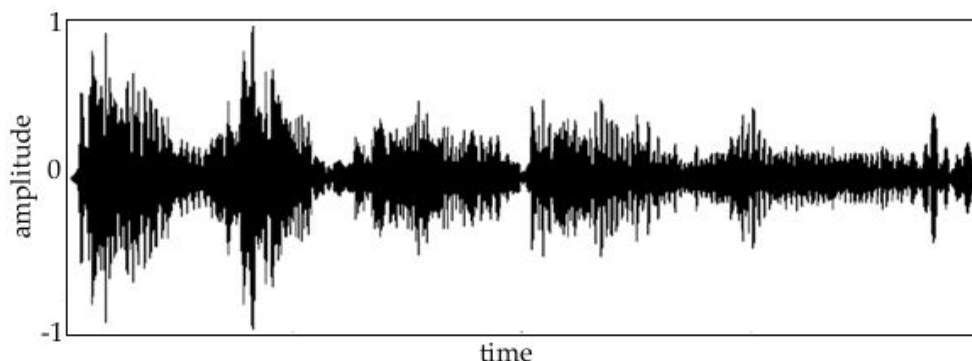
Automatická transkripce hudby, jak již napovídá její název, zahrnuje transkripci neboli převod mezi různými reprezentacemi hudby. Při návrhu systémů AMT (a obecně při zpracování hudebních signálů a další práci s hudbou) je velice důležité důkladné pochopení jednotlivých forem reprezentace hudby včetně jejich specifik. V následující kapitole budou popsány tři hlavní druhy reprezentace hudby v procesu AMT – zvuk, parametrická reprezentace, notový zápis – s ohledem na informace, které jsou důležité při AMT. Zvláštní pozornost je věnována notovému zápisu a popisu hudby z hlediska hudební vědy. Nahlížení na hudbu z tohoto hlediska je značně odlišné od „fyzikálního“ pohledu a má za následek to, že transkripce do notového zápisu je obzvláště obtížná.

## 3.1 Zvuk

Prvotní reprezentace hudby, kterou se při automatické transkripci hudby zabýváme, je zvukový signál. Zvuk vzniká chvěním hmoty, jakéhokoliv pevné látky, kapaliny nebo plynu. Chvění zdroje zvuku může být pravidelné nebo nepravidelné. Pravidelným chvěním vznikají tóny, v případě nepravidelného chvění hluky. Kmitání zdroje nejjednoduššího tónu můžeme znázornit pomocí sinusoidy. Sinusové tóny jsou poměrně vzácné, častěji se setkáváme s tóny složenými. Příkladem složených tónů mohou být všechny tóny produkované na hudební nástroje. [18]

Složený tón vnímáme jako jediný tón i přesto, že se skládá z několika sinusových tónů různých výšek. Tyto tóny nazýváme částkové nebo také alikvótní. Jednotlivé alikvótní tóny jsou uspořádány a tvoří takzvanou harmonickou řadu. Harmonická řada alikvótních tónů je pro hudbu základním akustickým jevem.

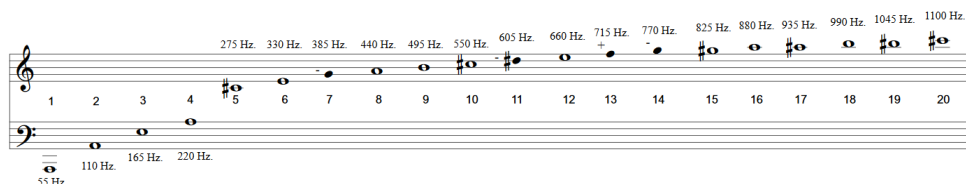
### 3. REPREZENTACE HUDBY V PROCESU TRANSKRIPCE



Obrázek 3.1: Časový průběh zvukového signálu [19]

Frekvence jednotlivých tónů v harmonické řadě můžeme vyjádřit pomocí aritmetické posloupnosti  $a_{n+1} = a_n + d$ , kde  $a_1 = f_0$  a  $d = f_0$ .  $f_0$  označuje frekvenci základního tónu.

Z hudebního hlediska harmonická řada postupuje vždy ve stejných hudebních intervalech – oktáva, kvinta, kvarta, atd. Jednotlivé tóny harmonické řady můžeme zobrazit pomocí not na odpovídajícím místě notové osnovy. Tento tradiční způsob záznamu harmonických řad je méně přesný než údaje o frekvencích tónů. Na obrázku 3.2 je zobrazeno prvních dvacet částkových tónů k základnímu tónu  $A_1$ . Symboly  $-$  a  $+$  označují, že frekvence daného tónu v notách je významně nižší či vyšší než frekvence odpovídajícího tónu hraného na piano.

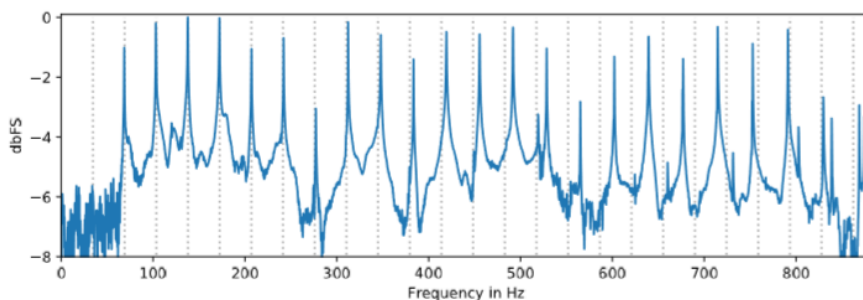


Obrázek 3.2: Prvních dvacet harmonických frekvencí tónu  $A_1$  [20]

Sluchem vnímáme výšku složeného tónu vždy jako výšku nejspodnějšího sinusového tónu v harmonické řadě. Při produkci tónů reálnými zdroji zvuku zní v harmonické řadě jednotlivé alikvótní tóny různě silně a některé mohou i chybět. Poměr jednotlivých alikvótních tónů určuje barvu tónu, díky níž jsme například sluchem schopni rozpoznat jednotlivé druhy hudebních nástrojů.

Mimo alikvótních tónů se ve složeném tónu mohou objevit i tóny neharmonické, které vznikají nepravidelnostmi ve chvění zvukového zdroje a do uvedené harmonické řady nepatří. Soubor všech harmonických i neharmonických částkových tónů nazýváme spektrum tónu. Ukázka spektra tónu  $C\sharp$  je

na obrázku 3.3. V případě, že ve spektru tónu převládají harmonické složky, jedná se o tón, v opačném případě jde o hluk. Některé zvuky mohou být na hranici mezi tóny a hluky neboli šumy. [18]



Obrázek 3.3: Spektrum tónu  $C\sharp$  hraného na piano. Tuhost strun způsobuje posun frekvencí částkových tónů od ideálních harmonických frekvencí. (znázorněny vertikálními tečkovanými čarami) Můžeme si všimnout, že 23. částkový tón je již na pozici, kde bychom předpokládali 24. částkový tón. [4]

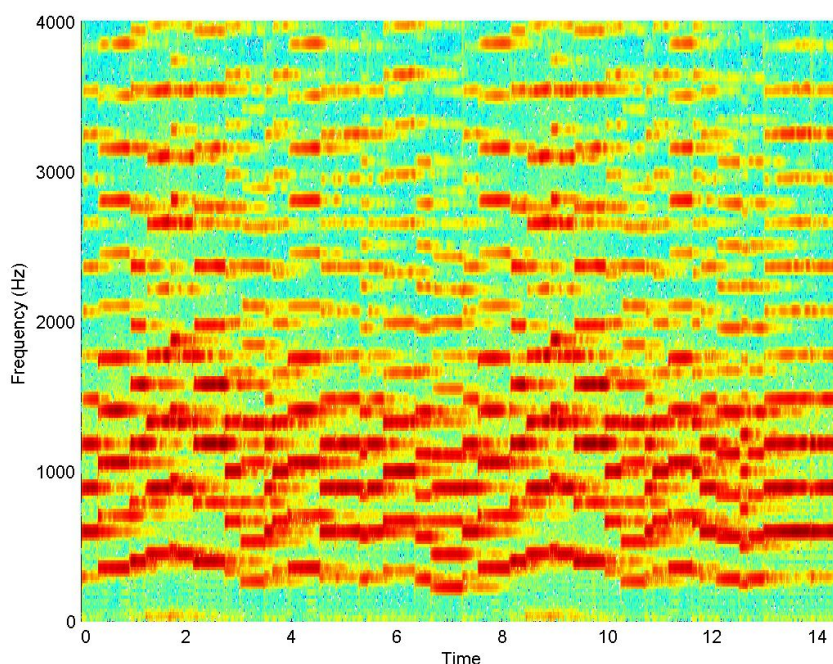
Z hlediska počtu současně znějících tónů můžeme hudební zvukové nahrávky rozdělit na polyfonní a monofonní. V případě monofonní hudby zní v každém okamžiku pouze jeden základní tón. V polyfonní hudbě může v jednom okamžiku znít neomezené množství tónů.

Extrakce výšek tónů z nahrávky zvukového signálu není jednoduchá právě kvůli přítomným harmonickým částkovým tónům. Obzvláště komplikovaná je poté extrakce výšek tónů z polyfonních nahrávek. Představíme-li si, že pokud slyšíme čtyři tóny znějící naráz, ve zvukovém signálu je přítomno ještě nezanedbatelné množství (třeba několik desítek) dalších částkových tónů, které se navzájem prolínají a které sluchem běžně nevnímáme. [21]

Vezmeme-li si například jednoduchý akord C dur složený ze tří tónů C, E a G. Procento částkových tónů, které se překrývají s ostatními je pro jednotlivé tóny C, E a G 46,7 %, 33,3 % a 60 %. [4]

Zatímco spektrogram<sup>1</sup> hudebního signálu (viz obrázek 3.4) může být vypočítán přímočaře například pomocí krátkodobé Fourierovy transformace (short-time Fourier transform), získání hudební reprezentace ve formě piano-roll (3.2.2) je netriviální a systémy, které se o to pokoušejí, bývají velmi složité. Převodem zvuku do piano-roll (nebo jiné parametrické reprezentace hudby) se zabývají parametrické systémy AMT. [21]

<sup>1</sup>Spektrogram je vizuální reprezentace zobrazující intenzitu jednotlivých frekvenčních složek signálu v závislosti na čase.



Obrázek 3.4: Spektrogram znázorňující prvních 14 sekund písně Greensleeves hrané na hoboj [22]

## 3.2 Parametrická reprezentace

Parametrická reprezentace hudby může být (případě parametrických systémů AMT) konečným výstupem transkripce. V případě úplných systémů AMT je mezistupněm při převodu do standardní hudební notace. V některých případech může být tento mezistupeň při převodu úplně vynechán. Takových systémů je však velmi málo a zaměřují se pouze na dílčí úkoly transkripce. (Příkladem může být detekce dob přímo z audio nahrávky.)

Parametrická reprezentace hudby je taková symbolická reprezentace hudby, která popisuje hudbu pomocí fyzikálních veličin. Hudba je v této reprezentaci popsána nejčastěji pomocí frekvence znějícího tónu a času (čas začátku/konce znění tónu nebo délka noty v milisekundách). Příkladem parametrické reprezentace hudby je formát MIDI. [6]

### 3.2.1 MIDI

Music Instrument Digital Interface (MIDI) je průmyslový standardizovaný protokol, který byl navržen pro komunikaci v reálném čase mezi elektronickými hudebními nástroji, počítači, sequencery, ovladači a další audiotechnikou



a videotechnikou. Součástí standardu je definice způsobu přenosu dat mezi zařízeními a způsob uložení dat v takzvaném Standardním MIDI souboru (SMF). [23]

Neexistuje žádný oficiálně přijímaný standardní formát souboru určený k elektronické reprezentaci hudby. Nicméně MIDI se v průběhu let stalo nejoblíbenějším a nejpoužívanějším formátem. Formát standardního MIDI souboru však nebyl nikdy určen k použití jako formát pro kompletní hudební notaci, proto není překvapující, že není pro tento účel nejvhodnější. [24]

MIDI nepřenáší zvukový signál, ale jedná se o proud tzv. událostí. Nejpoužívanějšími událostmi jsou začátek znění tónu (Note On) a konec znění tónu (Note Off). Událost začátku nebo konce znění tónu obsahuje údaje o času události, výšce tónu a hodnotu velocity (hlasitost, rychlost nástupu noty).

Délku noty můžeme získat odečtením času začátku od času konce dané noty. Výška tónu je v MIDI souboru reprezentována celými čísly v rozmezí od 0 do 127. Rozdíl 1 odpovídá intervalu jednoho půltónu. Nota C4 má v MIDI číslování standardně číslo 60.

MIDI číslo odpovídající výšce tónu můžeme ke každému tónu vypočítat z jeho základní frekvence  $f_0$  podle vzorce:

$$\text{MIDI note number} = 69 + 12 * \log_2\left(\frac{f_0}{440 \text{ Hz}}\right)$$

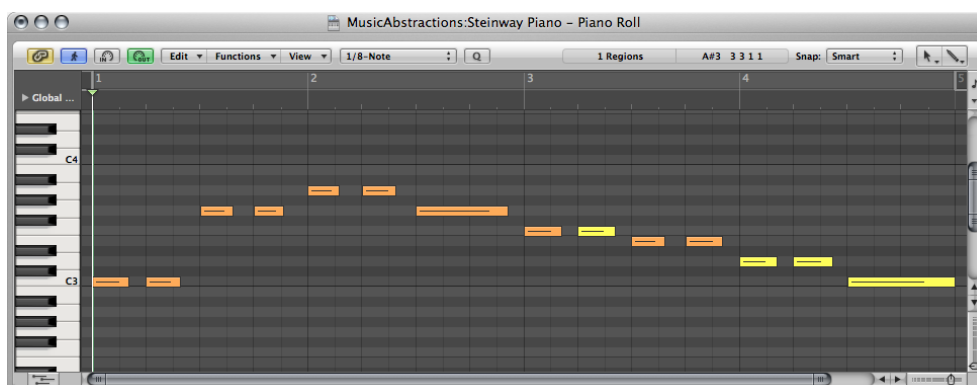
Časy začátků a konců not jsou v MIDI souboru nekvantované a nabývají tedy velkého množství hodnot. Naproti tomu ve standardní hudební notaci může délka noty nabývat jen několika málo hodnot a časy začátků tónů musí přímo odpovídat hudební „časové mřížce“, která je určena pomocí metra a rytmu. (viz kapitola 3.3)

Standardní MIDI soubor dále přímo neobsahuje určení rytmu, tóniny, rozdělení na hlasy a další hudební informace. Tyto informace nelze z MIDI souboru jednoduchým způsobem získat. Je k tomu potřeba využití pokročilejších metod a algoritmů. Proces získání notového zápisu z MIDI je netriviální a pro převod do standardní hudební notace je potřeba vyřešit ještě celou řadu úkolů. [6], [25], [26]

### 3.2.2 Piano-roll

Jelikož je MIDI binární formát, pro editaci či jinou práci se soubory ve formátu MIDI využíváme zobrazení ve formě seznamu událostí nebo jako tzv. piano-roll (3.5). V kontextu AMT je piano-roll druh reprezentace hudby, která zobrazuje výšku znějících tónů v závislosti na čase. Piano-roll můžeme znázornit jako graf, kde vertikální osa  $y$  představuje výšku tónu (klávesy na klavíru) a horizontální osa  $x$  zobrazuje čas. Reprezentace hudby ve formě piano-roll se používá i nezávisle na MIDI formátu.

### 3. REPREZENTACE HUDBY V PROCESU TRANSKRIPCE

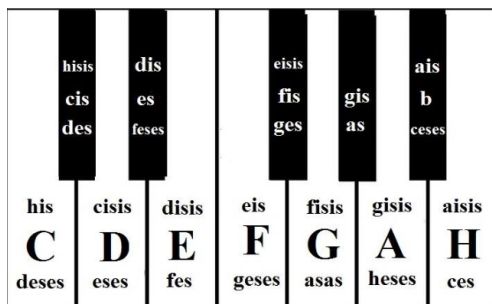


Obrázek 3.5: Reprezentace hudby ve formě piano-roll [27]

### 3.3 Notový zápis

Následující sekce obsahuje popis a vysvětlení základních pojmů využívaných při náhledu na hudbu z hlediska hudební vědy.

Popis všech symbolů a možností zápisu hudby ve standardní hudební notaci v této práci není zdaleka vyčerpávající. V hudebním zápise se objevuje mnoho dalších symbolů či informací důležitých při interpretaci skladeb. Jedná se o artikulační znaménka, označení dynamiky, melodické ozdoby, označení nálady a výrazu přednesu a mnohé další. Stejně tak úvod do hudební teorie je zde z důvodu rozsahu práce pouze velmi stručný.



Obrázek 3.6: Znázornění enharmonických tónů na klaviatuře [28]

#### 3.3.1 Tóny a tónová soustava

U tónů rozlišujeme čtyři základní vlastnosti výšku, délku, barvu a sílu. V akustice se místo síly používá označení hlasitost.

Tónová soustava je uspořádání všech tónů podle jejich výšek. Výška odpovídá frekvenci tónu. Základem tónové soustavy je sedm základních tónů c, d, e, f, g, a, h. Těchto sedm tónů se v tónové soustavě několikrát opakuje

v různých výškových polohách. Od výchozího tónu c k dalšímu opakovanému c je osm stupňů, proto se vzdálenost mezi těmito tóny nazývá oktáva. Tónová soustava obsahuje celkem devět oktáv.

Každý základní tón můžeme jednou nebo dvakrát zvýšit či snížit. Takto vzniklé tóny označujeme jako tóny odvozené. Názvy těchto tónů obsahují základní tón a příponu -is, -isis pro zvýšené tóny a -es, -eses pro tóny snižené.

Základních a odvozených tónů je v jedné oktávě celkem 35, přestože tónových výšek rozlišujeme pouze 12. Z toho vyplývá, že tóny stejné výšky se mohou jmenovat různě. Tóny, které mají stejnou výšku, ale jiný název, se nazývají enharmonické. Pro jejich záměnu (např. tón označíme jako fis místo ges) využíváme pojem enharmonická záměna. Na obrázku 3.6 je nákres části klaviatury, kde jsou na jednotlivých klávesách vypsány názvy všech odpovídajících základních i odvozených tónů. [18]

Vztah mezi frekvencemi, MIDI čísly a názvy tónů můžeme vidět na obrázku 3.7.

<u>MIDI number</u>	<u>Note name</u>	<u>Keyboard</u>	<u>Frequency Hz</u>		<u>Period ms</u>	
21	A0		27.500		36.36	
23	B0		30.868	29.135	32.40	34.32
24	C1		32.703		30.58	
26	D1		36.708	34.648	27.24	28.86
28	E1		41.203	38.891	24.27	25.71
29	F1		43.654		22.91	
31	G1		48.999	46.249	20.41	21.62
33	A1		55.000	51.913	18.18	19.26
35	B1		61.735	58.270	16.20	17.16
36	C2		65.406		15.29	
38	D2		73.416	69.296	13.62	14.29
40	E2		82.407	77.782	12.13	12.86
41	F2		87.307		11.45	
43	G2		97.999	92.499	10.20	10.81
45	A2		110.00	103.83	9.091	9.631
47	B2		123.47	116.54	8.099	8.581
48	C3		130.81		7.645	
50	D3		146.83	138.59	6.811	7.216
52	E3		164.81	155.56	6.068	6.428
53	F3		174.61		5.727	
55	G3		196.00	185.00	5.102	5.405
57	A3		220.00	207.65	4.545	4.816
59	B3		246.94	233.08	4.050	4.290
<b>60</b>	<b>C4</b>		<b>261.63</b>		<b>3.822</b>	
62	D4		293.67	277.18	3.405	3.608
64	E4		329.63	311.13	3.034	3.214

Obrázek 3.7: Vztah mezi frekvencemi tónů, jejich názvy a odpovídajícími MIDI čísly na části klaviatury [29]

#### 3.3.2 Intervaly

Slovo interval má v hudbě odlišný význam než v matematice. Interval mezi dvěma notami lze považovat za vzdálenost mezi jejich výškami. Pro ladění klavíru se v dnešní době používá rovnoměrně temperované ladění, což znamená, že interval mezi každými dvěma sousedními klávesami (bílou a černou nebo dvěma bílými) je vždy stejný. Tento hudební interval se nazývá půltón nebo také malá sekunda a je to nejmenší běžně užívaná vzdálenost mezi výškami dvou tónů. Interval dvou půltónů se nazývá tón nebo také velká sekunda. Oktáva se skládá z dvanácti půltónů. [30]

Název intervalu je složen ze dvou slov, číslovky, která určuje přibližnou velikost intervalu a přídavného jména, které velikost upřesňuje. Pro označení intervalu se používají latinské číslovky prima, sekunda, tercie, kvarta, kvinta, sexta, septima a oktáva. Přesnou velikost intervalů určujeme pomocí přídavných jmen malý, velký, čistý, zvětšený, zmenšený, dvojnásobně zvětšený, a dvojnásobně zmenšený. Uděláme-li u jednoho nebo obou tónů enharmonickou záměnu, dostaneme tzv. enharmonický interval. Enharmonické intervaly stejně znějí, ale odlišně se jmenují. (Obdobně jako u enharmonických tónů.) Enharmonické intervaly mívají v hudbě jinou výrazovou platnost. Příkladem může být zvětšená sekunda, která se ve skladbě vyznačuje silným napětím a vyžaduje pokračování a naproti tomu jí enharmonická malá tercie, která je obvykle intervalem klidovým.

Harmonické intervaly (tj. takové intervaly, kde oba tóny zní současně) můžeme rozdělit na konsonance (libozvučné) a disonance (nelibozvučné). [18] Zjednodušeně lze říci, že dva tóny znějí libozvučně, pokud můžeme poměr frekvencí tónů vyjádřit jako poměr  $n : m$ , kde  $n$  a  $m$  jsou malá kladná celá čísla. Čím menší jsou tato čísla, tím dostáváme libozvučnější interval. Po primě (dva stejné tóny) je nejlibozvučnějším intervalem oktáva s poměrem  $1 : 2$ . [30] Toto rozdělení se uplatňuje především při poslechu jednotlivých intervalů mimo hudební souvislosti. Při poslechu hudební skladby se ale intervaly často jeví jinak v závislosti na celkovém hudebním kontextu.

Pokud se zaměříme na tuto problematiku podrobněji, zdůvodnění rozdílů mezi konsonancemi a disonancemi není jednoduché. Musíme brát v úvahu hledisko akustické, fyziologické (proces slyšení) i psychologické a při posuzování hudební souvislosti i estetické a historické. Při vnímání hudby získáváme určité návyky, které uplatňujeme i při poslechu osamocených intervalů. Rozdělení intervalů na libozvučné a nelibozvučné tedy není jednoznačné a nelze ani jednoduše říci, že konsonance působí na člověka příjemným dojmem a disonance naopak nepříjemným. [18]

#### 3.3.3 Notopis

Hudební notace označuje způsob záznamu skladby do formy hudebního písma. Hudební notace se v průběhu historie měnila a vyvíjela. I v současnosti existuje



Obrázek 3.8: Ukázka zápisu ve standardní hudební notaci. Úryvek ze skladby Fantasia od Johanna Sebastiana Bacha. [31]

tují různé druhy hudební notace. V této práci popíšeme a následně budeme pracovat s aktuálně nejpoužívanější formou hudební notace, která bývá také označována jako standardní hudební notace, moderní hudební notace nebo také západní hudební notace. Na obrázku 3.8 je ukázka notového záznamu ve standardní hudební notaci.













Ve standardní hudební notaci tóny zapisujeme pomocí symbolů - not. Tvar noty nám udává délku tónu. Výšku tónu rozlišujeme pomocí umístění noty v notové osnově. Síla tónu, rychlost hry a způsob přednesu jsou v notovém zápisu označeny symboly nebo krátkým textem. Barva tónu je dána nástrojovým či pěveckým obsazením. Kromě tónů mají v hudbě velký význam i přestávky, odmlčení. Označují se speciálními symboly – pomlkami a mají stejné hodnoty jako noty. Na obrázku 3.9 je uveden přehled základních hodnot a tvarů not a pomlek včetně jejich délky trvání.

Následuje-li za sebou více osminových (nebo kratších) not, můžeme jednotlivé noty spojit trémcem. Tomuto spojení dáváme přednost z důvodu snazší čitelnosti. Na obrázku 3.9 jsou vedle jednotlivých not vidět osminové, šestnáctinové a dvaatřicetinové noty spojené trémcem do různě početných skupin. Spojení trémcem můžeme využívat i u not nestejných délek.

Noty píšeme do notových osnov. Notová osnova má pět linek a čtyři mezery. Jména not v osnově určuje znaménko zvané klíč, který udává umístění jedné noty a podle ní se řídí pojmenování ostatních. (Např. houslový neboli G klíč udává umístění tónu g na druhou linku odspodu.)

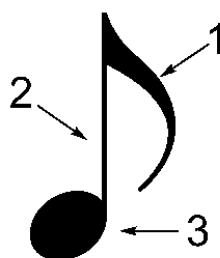
Odvozené tóny se v notovém zápise značí posuvkami umístěnými před notou. Zvýšený tón označuje křížek  $\sharp$ , snížený tón bé  $\flat$ . Existují také dvojkřížky a dvojbéčka. Platnost posuvek ruší taktové čáry nebo odrážky  $\downarrow$ .

### 3. REPREZENTACE HUDBY V PROCESU TRANSKRIPCE

celá	4 doby		
půlová	2 doby		
čtvrt'ová	1 doba		
osminová	půl doby		
šestnáctinová	čtvrt doby		
dvaatřicetinová	osmina doby		

Obrázek 3.9: Přehled základních hodnot a tvarů not a pomlk včetně jejich délky trvání [32]

Sazba notového zápisu se řídí mnoha dalšími pravidly a má také určitý hudební význam. Pravidla stanovují například způsob psaní nožek not (části noty viz obrázek 3.10) v závislosti na výšce noty, rozložení hlasů a mnoha dalších faktorech. Noty patřící do stejného hlasu by měly mít nožky směřující stejným směrem a rozdělení do trámčů by mělo odpovídat metru (3.3.6) dané hudební pasáží. Dodržování těchto pravidel ovlivňuje zásadním způsobem čitelnost notového zápisu. Při sazbě notového zápisu dochází také k mnoha výjimkám, které jsou obvykle dělány právě za účelem ještě většího zlepšení čitelnosti zápisu. Podrobný popis těchto pravidel není předmětem této práce. O jejich dodržování se starají notační programy. Při automatické transkripci hudby je však potřeba mít tato pravidla také na paměti, neboť pro jejich dodržování je potřeba využívat znalosti získané při transkripci.



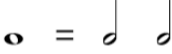
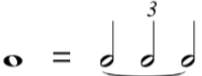
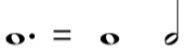
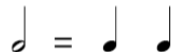
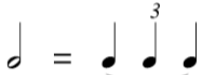


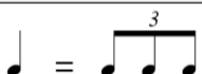




Obrázek 3.10: Části noty: 1 – praporec, 2 – nožka, 3 – hlavička [33]

#### 3.3.4 Dělení a prodlužování not a pomlk

Základní hodnota not a pomlk je uvedena v jejich názvech, z nich je patrné, že hlavní a nejčastější způsob dělení not je dělení na dvě poloviny. Celá nota se

dělí na dvě noty půlové atd. Kromě toho známe v hudbě ještě také dělení na tři stejné části. Skupině not, která vznikla tímto dělením říkáme triola. Různé druhy dělení not jsou vidět na obrázku 3.11. Mezi méně časté až velmi vzácné dělení patří dělení na pět, šest, sedm nebo i více částí.

Délku trvání noty může prodloužit pomocí ligatury nebo tečkami uváděnými za notou. Ligatura je oblouček spojující dvě nebo více not stejné výšky. Uplatňuje se především při spojování not přes taktové čáry. Tečka za notou prodlužuje trvání noty o polovinu původní délky. [18] Druhá tečka za notou prodlužuje dobu trvání noty o další čtvrtinu původní délky a analogicky můžeme spočítat délku trvání noty i s více než dvěma tečkami. Více než dvě tečky za notou se však v notovém zápisu vyskytují velice zřídka. [30]

DVOUDÍLNÉ DĚLENÍ:	TRIOLY:	NOTY S TEČKOU:
		
		
		
		

Obrázek 3.11: Ukázka dělení a prodlužování not a pomlk [34]

### 3.3.5 Tempo

Tempo znamená rychlost, jakou se skladba hraje. Tempové označení se uvádí na začátku skladby, případně i v průběhu skladby v případě změny. Existují různá tempová označení. Tempo může být zadáno přesně pomocí hodnoty BPM, která udává, kolik čtvrtových not se vejde do jedné minuty. Někdy je však tempo udáváno pouze přibližně pomocí ustálených italských pojmenování např. *allegro* – rychle, *adagio* – pomalu.

Při interpretaci skladby se tempo nedodrhuje úplně přesně, protože přednes skladby vyžaduje drobné odchylky od strojově přesného a stále stejného tempa. Tyto odchylky od tempa se nazývají agogika a interpret ji při interpretaci provádí podle svého citu a vkusu. Dále v notovém zápisu můžeme nalézt značky pro zrychlení či zpomalení.

### 3.3.6 Metrum a rytmus

Notový zápis dělíme na krátké úseky trvající určitý počet dob. Tyto úseky nazýváme takty a oddělujeme je taktovými čarami. Doba je v hudbě základní jednotkou metra a rytmu a představuje v závislosti na tempu skladby pevně daný časový interval. Taktové předznamenání je uvedeno v notovém zápisu za symbolem klíče a skládá se ze dvou celých čísel  $n$  a  $r$ , kde  $n \in \mathbb{Z}^+$  a  $r$  je mocnina čísla 2. Tato čísla píšeme za klíč ve tvaru  $\frac{n}{r}$ . Písmeno  $n$  vyjadřuje počet dob v taktu a  $r$  časovou hodnotu noty, která odpovídá jedné době. (Například  $\frac{3}{4}$  takt – tříčtvrtový takt.) První doba v taktu má vždy větší přízvuk, což znamená, že první nota z taktu se hraje silněji než ostatní noty. Základní takty jsou takt dvoudobý a třídobý.

Metrum označuje v hudbě střídání přízvučných a nepřízvučných dob a většinou odpovídá taktům. Rytmus označuje střídání tónů různých délek a je vždy plně vypsán v notách. Naproti tomu metrum tvoří pro rytmus pravidelné pozadí, které je v hudbě pro člověka dobře patrné, ale nemusí se vždy projevit v notaci. Metrum a rytmus se doplňují a často se zkoumají dohromady jako metro-rytmické vztahy v hudbě. [18], [30]

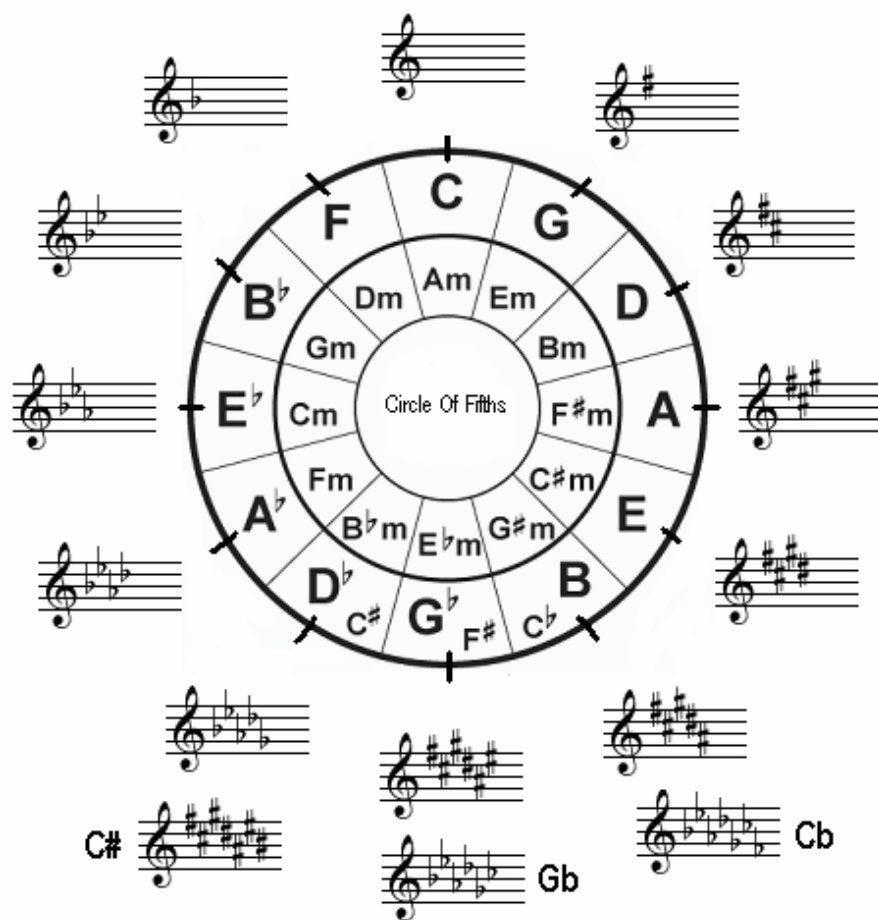
### 3.3.7 Tónina

Každá hudební skladba je psána v nějaké tónině. Tónina je v podstatě podmnožina tónů vybraná z množiny všech možných tónů podle určitých pravidel. Jeden tón z tóniny je hlavní a nazývá se tónika. Od tohoto tónu vychází pojmenování tóniny. Každá tónina má tzv. předznamenání, které určuje, které tóny budou v tónině zvýšené či snižené. Předznamenání je v notovém zápise uvedeno na začátku každého řádku pomocí křížků nebo béček na příslušném místě notové osnovy. Přehled existujících předznamenání je zobrazen a na obrázku 3.12. Při poslechu skladby tóninu určujeme sluchem pomocí tonálního cítění. Melodie často končí tónikou. Existují různé druhy tónin – durové, mollové, dórské a další. Tónina se může v průběhu skladby také měnit, toto označujeme termínem modulace.

### 3.3.8 Melodie

Melodie je posloupnost not, které jsou v hudební skladbě nejvýraznější a slouží k definování a charakterizaci hudebního díla. V populární písni může být příkladem melodie posloupnost tónů, které zpívá sólový zpěvák, zatímco jiné noty jsou hrány jinými nástroji jako doprovod. V hudbě hraje melodii často (ale není to pravidlem) nejvyšší nástroj nebo nástrojová sekce. Melodie však není definována a rozpoznávána pouze podle výšek jednotlivých not, ale také pomocí rytmu a celkového hudebního kontextu. [30]





Obrázek 3.12: Přehled předznamenání durových a mollových tónin [35]

### 3.3.9 Harmonie

Nauka o harmonii postihuje ten aspekt hudby, ve kterém zní dva a více tónů současně. Základním stavebním prvkem harmonie je akord. Akord je souzvuk nejméně tří tónů různé výšky. Ve vícehlasých skladbách rozlišujeme akordy a průběžné harmonie. Průběžné harmonie jsou jakékoliv souzvučky vznikající v hudební skladbě hraním ozdobných a melodických tónů, případně v polyfonních skladbách samostatným vedením hlasů. Základním harmonickým pojmem je harmonická funkce akordu, která označuje význam akordu v tonální harmonické větě. Nauka o harmonii se dělí na tři velké oddíly: diatoniku, chromatiku a moderní harmonii. [18]

### 3.3.10 Repräsentace notace v počítači

Otázka, jakým způsobem lze reprezentovat notové zápisy v počítači, je aktuální od samotného vzniku elektronických hudebních nástrojů a začátku využívání počítačů pro tvorbu a přehrávání hudby. Existuje široká paleta různých formátů, které se o počítačovou reprezentaci notace pokoušejí, od vysoce prosperujících formátů až po neúspěšné, velice složité a málo známé formáty.

Elektronická reprezentace notových zápisů je běžně využívána například v notačních programech. Notační programy umožňují tvorbu notových záznamů, jejich editaci, konverzi do audia nebo do formy obrázků či PDF a mnoho dalšího.

V současné době existuje několik běžně využívaných formátů, které dokáží plně reprezentovat notové záznamy ve standardní hudební notaci. Jedním ze způsobů reprezentace, který v současnosti patří mezi nejrozšířenější a nejvíce podporované, je formát MusicXML. [24], [36]

Dalším z využívaných formátů je formát abc. Abc je textový formát, který využívá pro reprezentaci hudební notace znaky ASCII. Ve formátu abc je možno vyjádřit složité hudební struktury. Abc obsahuje širokou škálu symbolů pro noty, délky not, trioly, akordy, ozdoby a další. Ukázka hudební skladby zapsané ve formátu abc je na obrázku 3.13. Píseň zapsaná v jazyce abc může být konvertována do MIDI. [37]

```
X: 1
T: Farewell Marian
Z: bunyip
S: https://thesession.org/tunes/1427#setting1427
R: waltz
M: 3/4
L: 1/8
K:F major
AdA =B^cd|e2 A4|fgf ede|^c2 A4|AdA GAF|BdB ABG|FA F2E2|D6|
AdA =B^cd|e2 A4|fgf ede|^c2 A4|AdA GAF|BdB ABG|FA F2E2|D6|
|A2 c3A|G2 A4|A2 c3d|e2 A4|f2 g3g|f2e2d2|ce c2=B2|A6|
AdA =B^cd|e2 A4|fgf ede|^c2 A4|AdA GAF|BdB ABG|FA F2E2|D6|
```

Obrázek 3.13: Ukázka zápisu hudební skladby ve formátu abc.

---

## Současný stav řešení

Výkonnost nejlepších akademických systémů AMT je stále jednoznačně pod úrovní transkripce hudebním odborníkem. A to i navzdory skutečnosti, že ani sami odborníci nejsou schopni vždy produkovat dokonalé výsledky, přestože využívají předchozí znalosti a zkušenosti a jsou schopni komplexního pohledu na celou problematiku. [1]

Ve srovnání s jinými problémy v oblasti zpracování signálů existuje několik faktorů, které činí AMT zvláště náročnou:

- V polyfonní hudbě zní najednou mnoho tónů (nebo obecně zvuků) z více rozdílných zdrojů (různé hudební nástroje nebo zpěv) s rozdílnou výškou, hlasitostí a barvou. Každý ze zdrojů zvuku ještě produkuje jeden nebo více hudebních hlasů. Získání některých informací (například výšky tónů) z takového smíšeného hudebního signálu je extrémně obtížné a ve většině případů neexistuje jednoznačné řešení tohoto problému.
- Současně znějící tóny vzájemně interagují a frekvence jejich alikvótních tónů se do určité míry překrývají, což dělá separaci jednotlivých hlasů ještě složitější.
- Jednotlivé hlasy jsou v hudbě řízeny pravidelnou časovou strukturou zvanou metrum. Hudebníci při interpretaci skladeb ale věnují speciální pozornost vzájemné synchronizaci různých hlasů. (Příkladem může být situace, kdy se druhý hlas podřizuje prvnímu hlasu a nedodrží přesně svůj notový záznam.) Toto narušuje předpoklad nezávislosti jednotlivých zdrojů zvuku a komplikuje jejich oddělení.
- Anotace notových záznamů je velmi časově náročná a vyžaduje vysokou odbornost. Nedostatek anotovaných notových záznamů limituje využití technik supervizovaného strojového učení pouze na specifické podproblémy AMT. Příkladem může být transkripce hudby hrané na klavír,

u které umíme proces anotace notových záznamů do jisté míry automatizovat díky existenci digitálních pian. Dalším problémem je, že ani anotovaný notový záznam není časově závislý na audio signálu a obvykle neposkytuje přesnou reprezentaci hudební nahrávky. Využití notových záznamů v technikách strojového učení pro proces AMT je tedy obecně komplikované. [4]

Problém automatické transkripce hudby není doposud uspokojivě vyřešen, složitost problému a výše uvedené faktory vedou k velkému množství chyb ve výstupech systémů AMT.

I při využití nejpokrokovějších metod je chybovost při rozpoznávání tónů z nahrávky většinou mezi 20-40 %, nebo dokonce vyšší v případech polyfonní hudby. Chybovost transkripce se může snížit omezením podmínek nebo zahrnutím dalších informací od uživatele do procesu transkripce. Přesnost transkripce v obecném případě však stále zabraňuje využití technologie AMT v praktických aplikacích. [8]

Pokud se pokusíme blíže specifikovat jednotlivé podúkoly, které zvládají systémy AMT, podle [5] lze říci, že:

Současné systémy AMT umí (s určitou mírou úspěšnosti):

- detekovat výšku, začátky a konce znění tónů z nahrávky;
- identifikovat jednotlivé hudební nástroje v polyfonní hudbě;
- přiřadit detekované noty jednotlivým hudebním nástrojům.

Některé systémy AMT zvládají (s menší úspěšností) také:

- detekovat a integrovat do systému rytmické informace;
- detekovat melodii;
- vytvořit prstoklady (pro určité nástroje);
- určit hlasitost jednotlivých detekovaných not;
- kvantizovat délky not, zarovnat noty do metrické mřížky.

Z výše uvedených bodů si lze všimnout, že k vytvoření úspěšného úplného systému AMT vede v současnosti ještě velice dlouhá cesta.

Proces získání klasického notového zápisu z parametrické reprezentace hudby (např. nekvantizovaného MIDI) je netriviální a v současné době neexistuje systém, který by tento problém uspokojivě řešil, implementoval a byl popsán a publikován v odborné literatuře. Tato část transkripce je zvláště obtížná také kvůli velkému přesahu do muzikologie. [6]

Určitého pokroku v oblasti úplné transkripce bylo dosaženo v roce 2018 [10]. Sami autoři však hodnotí výsledky svého navrženého systému AMT takto: Získané notové záznamy, s výjimkou hudebně a zvukově jednoduchých

---

případů, obsahují nesprávné či nehratelné noty a nejsou stále ani zdaleka uspokojivé.

Jedním z dalších aktivních problémů v oblasti zpracování hudebních signálů je detekce a klasifikace zvuků, u kterých nelze určit jejich výšku. Ve většině případů hovoříme o problému transkripce bicích nástrojů, které obsahuje většina současné hudby. Úkolem transkripčního systému je zachytit a klasifikovat zvuk jednotlivých bubnů, činelů a dalších částí bicí soupravy. Automatická transkripce nahrávek s přítomností bicí soupravy je komplikovaná z toho důvodu, že obsahuje velké množství harmonických i neharmonických zvuků a také jsou zde kladeny větší požadavky na rytmickou přesnost. Systémy určené k transkripci bicích nástrojů jsou vyvíjeny nezávisle na běžných systémech AMT. Vytvoření robustního systému pro přepis hudby obsahující skladby hrané s doprovodem bicí soupravy zůstává stále otevřeným problémem. [4]

Dalším problémem, který nebyl dosud vyřešen a který je klíčový pro rozvoj v oblasti úplných transkripčních systémů a sazbu výsledných notových zápisů je správný odhad metra nahrávky.

Je nutno zmínit, že pokroky ve výzkumu AMT v dnešní době pocházejí především od informatiků, zejména těch, kteří se specializují na strojové učení. Příspěvek do výzkumu ze strany počítačové muzikologie, hudební psychologie, akustiky a dalších příbuzných oborů je minimální.

K dalšímu vývoji AMT by bylo potřeba integrovat znalosti z těchto oblastí do systémů AMT a více zapojit odborníky z těchto oblastí jako spolupracovníky do výzkumu. [1]

#### 4.0.1 Polo-automatické přístupy

Skutečnost, že ani nejmodernější systémy AMT nedosahují úrovně transkripce lidmi, vyvolává otázku, zda a jak by mohl člověk napomoci procesu transkripce. Určité úkoly, které jsou pro člověka relativně snadné (identifikace nástrojů, detekce začátků not), jsou při algoritmičtém řešení naopak velice obtížné. Kombinace lidského vnímání a znalostí s algoritmičtými přístupy by mohla vést k výsledkům, které jsou přesnější než plně automatická transkripce a které jsou zároveň získány v kratším čase než transkripce prováděná pouze člověkem.

Je potřeba, aby vstup do systému AMT nevyžadoval od uživatele příliš mnoho času, úsilí a hudební odbornosti. Pro systémy AMT může být v zásadě užitečná jakákoliv informace týkající se přímo přepisované nahrávky nebo výsledného notového zápisu např. tónina, tempo, taktové označení, počet a druh hudebních nástrojů v nahrávce. Uživatel může být dokonce vyzván k přiřazení několika not nebo akordů ke každému nástroji.

Tvorba polo-automatických systémů AMT je v současnosti jedním z nevyřešených výzkumných problémů. Hlavním úkolem je identifikace oblastí, ve

kterých může být vstup od uživatele přínosem pro AMT, a integrace získaných informací do nízkourovňové analýzy signálů.

Kromě praktického využití mohou polo-automatizované systémy také přispět k vývoji robustnějších plně automatizovaných systémů, neboť umožňují zaměřit se algoritmicky pouze na podmnožinu požadovaných úkolů, zatímco mají k dispozici potřebné a správné informace z jiných dílčích úkolů transkripce. [1]

### 4.1 Existující software

Navzdory určitému pokroku ve výzkumu stále neexistuje žádná aplikace pro koncové uživatele, která by dokázala provádět přesně a spolehlivě automatickou transkripci hudby neohledně na použítou kombinaci nástrojů a žánr hudby. Většina existujícího softwaru se zaměřuje pouze na parametrickou transkripci a i v ní je pozadu oproti nejmodernějším akademickým systémům AMT. [1]

Je také důležité zmínit, že zde existuje určitá oddělenost akademického a komerčního světa. Většina vědeckých prací o AMT se zdržuje explicitního srovnávání výsledků s komerčně dostupným softwarem pro hudební transkripci. Důvodem může být jiné zaměření akademických a komerčních nástrojů. Komerční software je přizpůsobován odlišným požadavkům ze strany cílových uživatelů. [4]

Existuje celá řada softwarových produktů, které se pokoušejí o automatickou transkripci hudby nebo alespoň o detekci znějících frekvencí v hudebním záznamu. V případě komplexnějších programů bývá nějaká forma automatické transkripce pouze jednou z mnoha funkcí programu. V naprosté většině případů se jedná o parametrické systémy AMT a výsledkem programu je MIDI soubor.

Vzhledem ke stavu řešení problému AMT je software v této oblasti koncipován ve většině případů spíše jako pomocný nástroj při transkripci člověkem. Program provede automatickou transkripci a dále umožňuje uživateli editovat vzniklou reprezentaci hudby. (V tomto případě tedy nejčastěji MIDI soubor.) Vzhledem k chybovosti transkripce se předpokládá, že uživatel bude chtít vzniklý soubor dále upravovat, manuálně odstranit chybné noty, doplnit noty chybějící, upravovat délky not, přehrávat si určité úseky nahrávky zpomaleně spolu s výsledným MIDI apod. Programy poskytují uživateli pouze určitou představu o přepisované nahrávce a ne její přesný hudební obsah.

Nelze říci, který program je v současnosti při automatické transkripci hudby nejúspěšnější. Otázka zhodnocení výsledků systémů AMT je sama o sobě poměrně komplikovaná (viz kapitola 5). Jednotlivé programy umožňující automatickou transkripci hudby nebývají nijak objektivně hodnoceny ani srovnávány navzájem mezi sebou. Každý program také cílí na jiné uživatele a zaměřuje se na různé části transkripce nebo jiný druh hudby, tudíž nelze obecně určit, který existující software je v současnosti nejlepší.

Dle [4], [5], [38] a [39] patří mezi nejznámější dostupný software na AMT následující programy.

### **Akoff Music Composer**

Akoff Music Composer je zaměřen převážně na aranžování a komponování skladeb. Uživatel zapíše nebo zazpívá melodii skladby a program ji poté převede do MIDI. Uživatel může výsledné MIDI přehrát a upravit ho. Další funkcí programu je doplnění akordového postupu k získané melodii. [40]

### **AmazingMIDI**

AmazingMIDI je program pro automatickou transkripci hudby, který převádí soubory ve formátu WAV do formátu MIDI. Program je zaměřen na transkripci polyfonní hudby s jedním nástrojem, dokáže analyzovat akordy. Není však určen na transkripci bicích nástrojů. Tento program byl použit v rámci praktické části této práce. [41]

### **AnthemScore**

AnthemScore provádí úplnou transkripci hudby. Dovoluje převod souborů ve formátu MP3 nebo WAV do notového zápisu. Převod dvouminutové skladby trvá cca 3-5 minut. Po převodu je uživateli zobrazena vizualizace přepisované skladby ve formě spektrogramu s detekovanými notami a je mu umožněno provádět úpravy (mazání chybných not apod.). Výsledný notový zápis je možné uložit jako PDF, MIDI nebo MusicXML. [42]

### **AudioScore**

AudioScore je program pro úplnou transkripci hudby a je vyvíjen souběžně s jedním z nejznámějších notačních programů – Sibelius. Umožňuje převod zvuku do notového záznamu (včetně transkripce bicích nástrojů), výstupy transkripce lze importovat do programu Sibelius, který poskytuje širokou škálu možností editace notových záznamů. [43]

### **IntelliScore**

Program IntelliScore je zaměřen na parametrickou transkripci hudby. Umožňuje převod polyfonní hudby s více nástroji do MIDI. Dovoluje i transkripci bicích nástrojů, ale pouze v případě sólové hry (v nahrávce nejsou přítomny jiné nástroje). V průběhu transkripčního procesu využívá program informace o nahrávce získané od uživatele (počet a druh hrajících nástrojů, klepání do rytmu apod.). [44]

### ScoreCloud

Program ScoreCloud umožňuje převod zvuku do notového zápisu. Předpokladem je, že nahrávka bude nahrána přes připojené MIDI klávesy nebo pomocí mikrofону. Program se zaměřuje na přepis nahrávek hraných na piano nebo kytaru, případě s doprovodem zpěvu. Po nahrání a převedení do MIDI program uživateli umožňuje manuální editaci MIDI souboru. Poté je MIDI převedeno do notového zápisu. Přestože program umožňuje také import a převod jakýchkoliv audio nahrávek, na webových stránkách je uvedeno, že v případě např. popových písní, nahrávek s bicí soupravou nebo více nástroji, nedává program uspokojivé výsledky. Analýza nahrávky trvá cca 1,5 násobek jejího času. V současnosti se pracuje na dalším zlepšení. [45]

### Solo Explorer

Solo Explorer je program umožňující konverzi zvukového signálu do formátu MIDI. Je určen pouze pro monofonní hudební nahrávky. Program nabízí přesnou detekci výšek tónů a může být použit také pro kontrolu správné intonace (což je užitečné například při ladění nástrojů nebo výuce hudby). [46]

### Transcribe!

Program Transcribe! se zaměřuje především na poskytování podpory při transkripci hudby uživatelem. Umožňuje zpomalené přehrávání prepisovaných nahrávek a další úpravy vstupního zvukového signálu (např. transpozici – posun celé nahrávky o určitý hudební interval), dále obsahuje analyzátor not a akordů, který zobrazuje spektrum znějících frekvencí a na jejich základě navrhuje název noty, případně akordu. [47]

### WIDI Recognition System

WIDI Recognition System je program pro parametrickou transkripci hudby, který umožňuje převod zvuku do formátu MIDI. Zaměřuje se i na transkripci bicích nástrojů. Program dále poskytuje uživateli zpomalené přehrávání nahrávek a široké možnosti editace MIDI souborů. Obsahuje hudební editor TrueTone, který umožňuje vizualizaci hudby a opravu chyb vzniklých při transkripci. Rozdělení do taktů a metrum je doplněno manuálně uživatelem. [48]

### Programy pro práci s hudbou

Dále můžeme jmenovat ještě Logic pro X (Audio to score) [49], Autotune [50] a Melodyne [51], které patří mezi rozsáhlé programy pro práci s hudbou. Nabízí velké množství funkcí od nahrávání a úpravy zvuku, přes editaci MIDI souborů až po práci s notovými zápisy. Některé funkce, které programy nabízí, se dají zařadit do oblasti AMT – detekce tóniny, převod MIDI do not atd.



## Notační programy

Existují také některé volně dostupné i komerční notační programy, které se nezaměřují na automatickou transkripci hudby, ale dovedou mimo jiné převést importovaný MIDI soubor do standardní hudební notace.

Příkladem těchto programů je Finale [52], Sibelius [53] či MuseScore [54]. Tyto programy zpravidla neprovádí žádnou pokročilejší analýzu, která je při tomto procesu v rámci AMT nutná. Při převodu není počítáno s přítomností chyb a velkých nepřesností. Výsledky převodu proto nejsou dostatečné. V případě neanotovaných MIDI souborů převod také typicky není plně automatizovaný a potřebuje doplnění informací od uživatele. (Uživatel musí ručně zadat taktové označení, tóninu, tempo apod.)

Ze tří výše uvedených programů tvoří určitou výjimku program MuseScore, který má pokročilejší způsob převodu MIDI do standardní hudební notace, s možností volby převodu takového MIDI, které bylo získané z živé nahrávky na MIDI klávesy. MuseScore provádí kvantizaci MIDI, rytmickou analýzu a dovoluje volit různá nastavení. V případě MuseScore je proces převodu plně automatizovaný. Detaily procesu převodu nejsou zdokumentovány na webových stránkách programu a zatím nebyly ani publikovány v odborných člancích. [6] Program MuseScore je využit v praktické části této práce.



---

## Hodnocení výsledků systémů AMT

Přístupů k řešení problému automatické transkripce hudby je celá řada a velmi se od sebe liší, přístupy k evaluaci těchto transkripčních systémů jsou stejně tak variabilní. [55]

Mezi nejvýznamnější výzkumné počiny obecně v oblasti evaluace algoritmů a metod používaných v počítačovém zpracování hudby patří MIREX (The Music Information Retrieval Evaluation eXchange). MIREX představuje komunitní evaluační rozhraní pro hodnocení algoritmů a technik týkajících se MIR, digitálních hudebních knihoven (music digital libraries) a výpočetní muzikologie (computational musicology). Je koordinován a řízen specializovanou mezinárodní laboratoří International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL) z Univerzity Illinois v Urbana Champaign. MIREX využívá standardizované datasey, metriky a metodologie, z nichž některé lze využít i při hodnocení jednotlivých podúkolů řešených v systémech AMT.

Existuje mnoho výzev, kterým musí výzkum v oblasti MIR čelit. Rychlý výzkum v této oblasti brzdí také otázky ohledně duševního vlastnictví hudby. Například MIREX z důvodu autorských práv využívané datasey nezveřejňuje. Dalším velkým problémem je obecně neexistence dostatečného množství datasetů, anotovaných hudebních nahrávek, notových záznamů, atd., které by se daly při výzkumu a následné evaluaci výsledků využít.

Datové sady, které jsou v současnosti běžně používány pro hodnocení, jsou malé. Běžně se například používá malý dataset z hudební databáze RWC<sup>2</sup> čítající pouze 12 skladeb (i přesto, že databáze RWC obsahuje celkově mnohem více záznamů). Datasey takového rozsahu nelze považovat za reprezentativní, nebezpečí přeučení modelů nebo nadhodnocení výkonu systémů AMT je vysoké. [56]

---

<sup>2</sup><https://staff.aist.go.jp/m.goto/RWC-MDB/>

## 5. HODNOCENÍ VÝSLEDKŮ SYSTÉMŮ AMT

MIREX vypisuje každoročně témata výzkumu a poté poskytuje jedenkrát za rok rozhraní k otestování výsledků. Tímto dává odborníkům možnost zhodnotit své výzkumné výsledky a porovnat je s výsledky ostatních týmů. Vypsány výzkumnými tématy na rok 2019 jsou například detekce tóniny z nahrávky, transkripce bicích nástrojů, odhad tempa z nahrávky, detekce rytmu nebo identifikace cover verzí. [57], [58]

Na obrázku 5.1 jsou pro ukázkou zobrazeny výsledky hodnocení MIREX pro výzkumný problém detekce not v polyfonní nahrávce s více hudebními nástroji. Můžeme si všimnout, že výsledky celkově nejsou příliš uspokojivé a v průběhu let dosáhl pouze jeden výzkumný tým úspěšnosti vyšší než 0,5. Pro vyhodnocení byla využita F-míra (viz 5.1.1).

### Results for Task 2 (onset/offset-based F-measure)

Teams	2009	2010	2011	2012	2013	2014
Yeh and Roebel	0.31	0.33	0.35	-	-	-
Dressler	-	-	-	0.45	-	0.44
Benetos and Dixon/Weyde	-	-	0.21	0.23	0.33	0.36
Duan, Han and Pardo	0.22	0.19	-	-	-	-
Fuentes et al.	-	-	-	0.39	-	-
Elowsson and Friberg	-	-	-	-	-	0.58
Cheng et al.	-	-	-	-	0.29	-
Su and Yang	-	-	-	-	-	0.29
Böck	-	-	-	0.09	-	0.14
Dessein et al.	-	0.24	-	-	-	-
Duan and Temperley	-	-	-	-	-	0.28

Obrázek 5.1: Výsledky hodnocení MIREX detekce not v polyfonní nahrávce s více nástroji [5]

### 5.1 Parametrické systémy AMT

Existuje celá řada metod, které se používají ke zhodnocení výsledků parametrických systémů AMT. Hodnocení těchto systémů se většinou zaměřuje na ohodnocení dvou dílčích úkolů patřících do této části AMT, a to na detekci výšek tónů z nahrávky (multi-pitch estimation) a detekci začátků a konců znění tónů (onset and offset detection).

Nejpoužívanější metody hodnocení jsou založeny na vyhledání chyb a jejich následném kvantitativním vyhodnocení. Nepanuje však jednotný názor na to, co přesně by mělo být při transkripci již považováno za chybu (typicky jsou tolerovány určité časové nebo frekvenční odchylky) a také na to, které aspekty výsledné reprezentace do hodnocení zahrnout (např. zda pouze začátky znění tónů nebo také jejich konce, které je obtížnější detekovat). [59] Nejrozšířenější mírou hodnocení je v této oblasti F-míra (5.1.1). Je však nutno říci, že i když je F-míra pro měření kvality parametrických systémů AMT používaná, není pro úplné ohodnocení transkripčního systému dostatečná, protože nebere v úvahu aspekty jako sekvenčnost, akordy, harmonické nebo tonální vztahy atd. [55]

Vytvoření více vypovídající míry, která by zohledňovala váhy jednotlivých chyb a pozitivně korelovala s hodnocením lidmi, zůstává otevřeným problémem. [4]

Přístupy k evaluaci parametrických systémů AMT můžeme rozdělit podle působu porovnávání originálu s výslednou parametrickou reprezentací hudby na dva typy: frame-based a note-based evaluace.

### 5.1.1 Frame-based evaluace

Frame-based způsob evaluace je založen na rovnoměrném rozdělení skladby na krátké časové úseky. MIREX [57] například využívá rozdělení skladby na úseky dlouhé 10 ms. Tato délka se může v různých pracích lišit. Odpovídající časové úseky v originále a výstupu se navzájem jednotlivě porovnávají. V rámci každého časového úseku jsou zkoumány detekované základní frekvence znějících tónů.

Pro zhodnocení výsledků [59] definuje přesnost (precision), úplnost (recall) jako:

$$\text{Precision} = \frac{\sum_{t=1}^T TP(t)}{\sum_{t=1}^T TP(t) + FP(t)}$$

$$\text{Recall} = \frac{\sum_{t=1}^T TP(t)}{\sum_{t=1}^T TP(t) + FN(t)}$$

Počet souběžně znějících tónů v jednotlivých časových úsecích originálu se mění. Pro každý časový úsek (index časového úseku označujeme  $t$ ) jsou určeny hodnoty:  $TP(t)$  (true positives) počet tónů, které jsou v daném úseku určeny správně,  $FP(t)$  (false positives) počet tónů, které jsou v daném úseku ve výsledku navíc a  $FN(t)$  počet tónů, které v daném úseku ve výsledku chybí. Tyto hodnoty jsou při výpočtu přesnosti a úplnosti sečteny přes všechny časové úseky ( $T$ ).

Výsledná úspěšnost transkripčního systému se poté spočítá pomocí F-míry (F-measure), která je kompromisem mezi přesností (precision) a úplností (recall).

$$\text{F-measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

Můžeme také spočítat míru správnosti (accuracy) jako:

$$\text{Accuracy} = \frac{\sum_{t=1}^T TP(t)}{\sum_{t=1}^T TP(t) + FN(t) + FP(t)}$$

Míra správnosti nabývá hodnot z intervalu  $[0, 1]$ , kde 1 odpovídá dokonalé transkripci.

### 5.1.2 Note-based evaluace

Note-based způsob evaluace vychází z rozdělení skladby na jednotlivé tóny – noty. Na skladbu v tomto případě nahlížíme jako na posloupnost tónů, u každého sledujeme tři vlastnosti: výšku (frekvenci), čas začátku znění tónu a čas ukončení tónu. Každý tón poté zvlášť porovnáváme s odpovídajícím tónem v originálu.

Typicky existuje určitá tolerance, v rámci níž tóny ještě považujeme za stejné. Bay et al. [59] tolerují odchylku do 50 ms v času začátku znění tónu a čtvrttón ve výšce tónu, Dixon [60] 70 ms v začátku znění tónu (časy konců tónů nejsou při hodnocení brány v úvahu) a naproti tomu Ewert et al. [8] tolerují odchylku až do 250 ms v čase začátku a konce znění tónu.

Dále postupujeme obdobně jako při frame-based evaluaci. Porovnáním s originálem rozdělíme výsledné tóny do tří kategorií, správně klasifikované tóny (true positives – TP), tóny, které jsou ve výsledku navíc (false positives – FP) a tóny, které ve výsledku chybí (false negatives – FN).

Z počtu tónů ve výše uvedených kategoriích spočítáme přesnost a úplnost pomocí vzorce:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Z přesnosti a úplnosti můžeme stejně jako v případě frame-based evaluace vypočítat F-míru.

$$\text{F-measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

Souhrnnou míru správnosti (accuracy) vypočteme vzorcem:

$$\text{Accuracy} = \frac{TP}{TP + FN + FP}$$

K celkovému hodnocení systému AMT je poté nejčastěji využívána právě F-míra.

## 5.2 Úplné systémy AMT

Výsledkem úplného systému AMT je notový záznam ve standardní hudební notaci. Hodnocení takového systému je složitější a výzkum v této oblasti je oproti parametrickým systémům značně pozadu. Při zhodnocení výstupů úplného transkripčního systému se nejčastěji využívá nějaké formy porovnání vzniklého notového zápisu s originálním notovým zápisem. Čím podobnější si tyto notové záznamy jsou, tím můžeme transkripční systém označit za úspěšnější.

Porovnávání notových záznamů je aktuální výzkumný problém, jež nebyl doposud uspokojivě vyřešen. Jeho řešení má využití nejen v automatické transkripci hudby, ale také v mnoha dalších oblastech jako je například muzikologická analýza hudebních děl nebo vyhledávání v databázích notových záznamů. Je zřejmé, že čím více se bude výzkum v oblasti AMT posouvat, tím nezbytnější bude existence standardizované funkce, která dokáže určit podobnost notových záznamů. Tento úkol je pro další vývoj klíčový. [61]

V současnosti jsou pokusy o vytvoření plnohodnotného transkripčního systému bržděny mimo jiné také absencí objektivní metriky, která by dokázala zhodnotit výsledky navrženého systému. Hodnocení výsledků transkripčních systémů bývá nejčastěji prováděno odborníky v oblasti hudební teorie. Hodnocení lidmi je ale časově i finančně náročné a nedává dostatečnou zpětnou vazbu k dalšímu vylepšování systémů AMT, proto zde existuje snaha toho hodnocení automatizovat.



Obrázek 5.2: Srovnání notových záznamů vzniklých různými systémy AMT. Notový záznam ve spodní části představuje originál. [10]

Ukázku notových záznamů vznikajících při automatické transkripci mů-

žeme vidět na obrázku 5.2. V horní části obrázku jsou tři notové zápisy vzniklé různými systémy AMT, ve spodní části je poté zobrazen originální notový zápis.

Vytvoření metriky pro hodnocení úplných systémů AMT není jednoduché vzhledem k inherentní složitosti hudebního zápisu. Vhodná metrika by musela brát v úvahu všechny aspekty vyšších hudebních struktur přítomných v notaci. [3] Vytvoření takové metriky zůstává stále otevřeným problémem. [4]

### 5.2.1 Přístupy k hodnocení

Vzhledem k neexistenci jednotné metriky pro evaluaci úplných transkripčních systémů se většinou přistupuje k následujícím zjednodušením:

- míra úspěšnosti je navržena přímo pro konkrétní transkripční systém;
- úspěšnost jednotlivých podúkolů při transkripci se vyhodnocuje zvlášť;
- některé aspekty úplného notového zápisu jsou ignorovány (zápis je zredukován na posloupnost not a pomlk, vyšší aspekty notového zápisu, jako je metrum, rozdělení do taktů a tak dále, nejsou brány v úvahu a nejsou hodnoceny).

Tato zjednodušení se často kombinují. Všechna zjednodušení však mají určité nevýhody.

Pokud je míra úspěšnosti navržena na míru pro konkrétní transkripční systém, může být ovlivněna autorovým přístupem k řešení problému transkripce a můžeme tedy spekulovat o objektivnosti takového hodnocení. Další nevýhodou je, že úspěšnost jednotlivých systémů mezi sebou nelze dobře porovnávat. Vzhledem k tomu, že v současnosti si navržené transkripční systémy zpravidla nekladou za cíl použitelnost pro jakýkoliv žánr hudby, druhy hudebních nástrojů atd., ale naopak se cíleně vymezují pouze na určitou oblast, porovnávání úspěšnosti jednotlivých systémů mezi sebou je obecně složité.

Z důvodu zjednodušení komplexního problému se jednotlivé podúkoly při AMT často řeší samostatně. (Dílními úkoly myslíme například separaci hlasů, analýzu metra, harmonickou analýzu, ...) I v případě, že systém AMT řeší více těchto podúkolů najednou, hodnocení úspěšnosti se provádí odděleně pro každý podúkol zvlášť bez ohledu na další části systému. Pro relevantní ohodnocení kvality transkripce je ale nedostatečné pouze zprůměrovat jednotlivá dílní hodnocení.

Dalším problémem tohoto zjednodušení je, že chyby vzniklé při řešení některého podúkolu se mohou různým způsobem řetězit a také se mohou promítnout i do řešení dalších podúkolů. Například chyby vzniklé při detekci not se započítají znovu jako chyby v rytmické analýze, i přesto, že ta již byla provedena správně.

Pokud ignorujeme některé aspekty notového zápisu, dochází zde ke ztrátě mnoha informací a takováto evaluace systému může být poté zredukována



(podle míry ignorovaných aspektů) na porovnávání hudebních sekvencí nebo až na hodnocení systému, který poskytuje pouze parametrickou reprezentaci hudby. [61]

### 5.2.2 Kvalita transkripce

Pokud odhlédneme od celkového hodnocení úspěšnosti úplných systémů AMT a zaměříme se pouze na posouzení kvality transkripce jednoho konkrétního hudebního úseku (případně celé skladby), zjistíme, že ani to není příliš jednoduchým úkolem. Přesné posouzení kvality transkripce není triviálním úkolem ani pro zkušeného muzikologa. Vyjádřit kvalitu transkripce automaticky souhrnně jedním nebo několika čísly je o to složitější.

Na notový zápis ve standardní hudební notaci může být nahlíženo jako na posloupnost hudebních objektů (klíče, předznamenání, noty a pomlky) s připuštěním výskytu více objektů současně jako jsou překrývající se noty nebo akordy. Chyby vzniklé při transkripci zahrnují špatné udání taktu, přebývající či chybějící noty a pomlky, chyby v délkách trvání not a pomlky, enharmonické záměny, špatný směr nožek not a špatné seskupení not do trámců a další. Všechny tyto chyby přispívají v různé míře ke snížení kvality výsledné transkripce.



(a) Ground truth



(b) Transcription with a wrong pickup measure



(c) Transcription off by a 16th note

Obrázek 5.3: Porovnání dvou transkripcí stejné skladby. Transkripce obsahují podobné chyby, ale jejich čitelnost se značně liší. [10]

Na obrázku 5.3 jsou vidět dvě různé transkripce téže skladby. Obě transkripce obsahují stejnou chybu – špatnou detekci metra. První uvedená transkripce se ale vyznačuje výrazně lepší čitelností a je tedy lidmi hodnocena jako kvalitnější.

Dalším faktem je, že podobné chyby jsou lidmi vnímány jako různě závažné v závislosti na širokém hudebním kontextu dané skladby. Jako příklad můžeme uvést špatně určenou notu, která se výškou nehodí do harmonie skladby a notu, která je posunutá o oktávu nebo jiný libozvučný interval oproti originálu. Nicméně konkrétní vliv různých chyb či kategorií chyb na kvalitu transkripce nebyl doposud dostatečně dobře prozkoumán. (Výzkumem v této oblasti se zabývali například v práci [55], ale problém stále není uspokojivě vyřešen.)

Při posuzování kvality transkripce je důležité mít na paměti, že výstupy transkripčního systému (notové záznamy) budou využívány přímo lidmi. Kvůli výše zmíněným skutečnostem je důležitým faktorem při výběru vhodné metriky na ohodnocení kvality transkripce také pozitivní korelace výsledků se subjektivním hodnocením lidmi. [3]

### 5.2.3 Aktuální návrhy metrik

V nedávné době byly učiněny pokusy o vytvoření komplexních metrik pro evaluaci výsledků úplných systémů AMT. Jedná se o první návrhy a výzkum tímto směrem bude jistě ještě dále pokračovat. Níže uvádíme dva příklady.

#### 5.2.3.1 McLeod a Steedman, 2018

McLeod a Steedman [61] navrhuje metriku MV2H pro evaluaci parametrických i úplných systémů AMT. Metrika může být použita i pro systémy, které některý z podúkolů automatické transkripce hudby neřeší. Například systémy, které provádí detekci tónů z nahrávky a analýzu rytmu, ale nic dalšího. Navržená metrika vychází, v případech, kde je to možné, z již využívaných metrik pro dílčí podúkoly, ale zaměřuje se na to, aby vzniklé chyby nepenalizovala vícekrát v různých částech systému.

Výpočet se skládá z pěti částí. Ohodnocení detekce tónů, separace hlasů, rytmické analýzy, detekce délek not a ohodnocení harmonické analýzy. Výsledné hodnocení kvality transkripce je poté aritmetickým průměrem těchto pěti mezivýsledků.

Autoři uvádí, že i přesto, že se jejich navržená metrika pravděpodobně nestane konečnou standardizovanou metrikou pro úkol hudební transkripce, měla by se stát součástí diskuze a zásady, které stály za jejím vytvořením, by měly být brány v úvahu ve všech budoucích navrhovaných metrikách.

## 5.2.3.2 Cogliati a Duan, 2017

Cogliati a Duan [3] navrhují metriku pro hodnocení přesnosti automatické hudební transkripce pomocí porovnání výsledného notového záznamu s originálem. Evaluace je zaměřena také na čitelnost výsledného notového zápisu. Metrika je modifikací editační vzdálenosti a je založena na podobných metrikách užívaných v bioinformatice a lingvistice. Její výpočet probíhá ve dvou krocích. V prvním kroku jsou notové záznamy navzájem zarovnány podle začátků a výšky jednotlivých not (viz 5.4). V druhém kroku jsou vyčísleny rozdíly mezi notovými zápisy.



Obrázek 5.4: Ukázka prvního kroku při hodnocení úspěšnosti transkripce – zarovnání [3]

Při výpočtu je bráno v úvahu dvanáct aspektů notového zápisu:

- taktové čáry,
- předznamenání,

- taktové označení,
- klíče,
- noty,
- názvy not,
- délky not,
- shlukování not pomocí trámců,
- směr nožek not,
- pomlky,
- délky pomlk,
- rozdělení not do notových osnov.

Výsledky byly porovnávány s hodnocením odborníky. Experimenty ukázaly jasnou korelaci mezi vypočtenou hodnotou a průměrným hodnocením člověka, i když korelace nebyla dokonalá. Navržená metrika byla testována na skladbách získaných z živé nahrávky na MIDI klávesy. Skladby tedy neobsahovaly typické chyby vzniklé při automatické transkripci ze zvukového záznamu. Pro využití této metriky na notovém materiálu obsahujícím chyby je potřeba modifikací a dalšího výzkumu.

### 5.3 Metody porovnávání hudebních sekvencí

Další přístup k hodnocení úplných systémů AMT nabízí metody porovnávání hudebních sekvencí. Metody porovnávání hudebních sekvencí se v současnosti častěji používají například pro vyhledávání podobných melodií ve velkých databázích než přímo v oblasti evaluace transkripčních systémů, je však možné je s úspěchem využít i pro tento účel. [55]

(Pojmem hudební sekvence v tomto kontextu myslíme obecné pojmenování pro posloupnost hudebních objektů či událostí, může být vyjádřena pomocí notového zápisu či jinou reprezentací.)

#### 5.3.1 Notové zápisy jako řetězce

Existuje velké množství algoritmů pro práci s řetězci, které se obvykle používají na textových nebo biologických řetězcích (např. DNA nebo řetězce proteinů). Při porovnávání hudebních sekvencí se často vychází z předpokladu, že notový zápis může být považován za řetězec hudebních entit, jako jsou noty, akordy atd. Pokud se pokusíme aplikovat algoritmy pro práci s řetězci na

### 5.3. Metody porovnávání hudebních sekvencí

hudební sekvence, vyvstává zde však mnoho otázek ohledně reprezentace hudební struktury. Měla by být hudební sekvence reprezentována jednoduše jako jediný řetězec uspořádaných n-tic (například uspořádaných dvojic – výška, délka) nebo by jednotlivé parametry měly být považovány za oddělené řetězce? Bereme-li v úvahu melodii, měla by být udána absolutní výškou, relativní výškou vzhledem k tonálnímu centru, pomocí intervalů nebo jiným složitějším způsobem? Rytmus by měl být udáván v délce trvání not nebo vyjádřen ve zlomcích jako poměr délky noty a délky taktu? Jak zohlednit, že některé noty jsou v hudební sekvenci důležitější v závislosti na délce trvání, harmonickém kontextu nebo metru? Tyto a mnoho dalších otázek je třeba zodpovědět, přičemž odpovědi zdaleka nejsou jednoznačné. [62]

Stručný přehled prací, které přistupují k notám jako k řetězcům je uveden na obrázku 5.5.

	pitch representation	rhythm representation	other structural factors	pre-segmentation required	type of pattern processing	type of matching	string matching algorithm
Mongeau & Sankoff, 1990	pitch/degree difference from tonal centre	durations	weights based on degree of consonance		comparison	approximate	dynamic programming
Stammen & Pennycook 1993	intervals in semitones	duration ratios	no	yes	recognition	approximate	dynamic programming
Smith, McNab Witten, 1997	intervals in semitones & contour	durations	no	no	recognition	exact & approximate	dynamic programming
Rowe, 1995 Rowe & Li 1995	intervals in semitones	durations	no	yes	recognition & induction	approximate	dynamic programming
Rolland 1996a,b 1998	absolute pitch & intervals	durations & ratios	contribution weights (e.g. long dur.)	no	induction	approximate	dynamic programming
Bakhmutova, Gusev, Titkova 1997	scale-step intervals		metric position	no	induction	approximate	dynamic programming
Cope 1990, 1991	intervals in semitones	durations	elimination of very short notes	no	m-length pattern induction	near-exact	brute-force algorithm
McGettrick 1997	abs. pitch, intervals in semitones	duration ratios	accented notes	no	recognition	exact	Boyer-Moore algorithm
Coyle & Shmulevich 1998	intervals in semitones (+error)	duration ratios (+error)	key-finding algorithm		comparison	exact (+ error absolute and perceptual)	equal length comparison
Hsu, Liu & Chen, 1998	absolute pitch	durations	elementary chords and metre	no	induction	exact	dynamic programming (only exact)
Hiraga 1997	interv: sem, scale-steps, step-leap, contour	durations: exact, log-ratio, shorter-longer-equal	reduction of surface	yes (tentative)	induction	exact (emphasis in immediate repetition)	not described
Cambouropoulos 1998	interv: sem, scale-steps, step-leap, contour	durations: exact, ratio, shorter-longer-equal	reduction of surface	no	induction	exact	Chrochemore (1981)

Obrázek 5.5: Přehled prací, které přistupují k notám jako k řetězcům [62]

### 5.3.2 Levenshteinova vzdálenost

Jedním z nejvyužívanějších způsobů porovnávání hudebních sekvencí reprezentovaných posloupností not je Levenshteinova vzdálenost a její různé modifikace. Tato vzdálenost je velmi známá a využívána při zpracování textu a řetězců.

Levenshteinova vzdálenost je metrika, která udává vzdálenost mezi dvěma řetězci jako minimální počet editačních operací potřebných ke konverzi jednoho řetězce na druhý. Mezi editační operace patří v případě Levenshteinovy vzdálenosti tyto tři operace:

- smazání znaku,
- přidání znaku,
- záměna znaku za jiný znak. [63]

Existují různé možnosti výpočtu Levenshteinovy vzdálenosti. Jedním z nich je využití Wagner–Fischerova algoritmu [64], který využívá techniky dynamického programování. Pseudokód algoritmu je uveden zde 1. Časová složitost výpočtu je  $\mathcal{O}(|A| * |B|)$ , kde  $|A|$  a  $|B|$  jsou délky porovnávaných řetězců.

**Vstup :** řetězce  $A$  a  $B$

**Výstup:** tabulka  $D$ ,  $D[|A|, |B|]$  obsahuje Levenshteinovu vzdálenost mezi řetězci  $A$  a  $B$

```

D[0, 0] ← 0;
for i ← 1 to |A| do
  | D[i, 0] ← D[i - 1, 0] + 1
end
for j ← 1 to |B| do
  | D[0, j] ← D[0, j - 1] + 1
end
for i ← 1 to |A| do
  | for j ← 1 to |B| do
    |   m1 ← D[i - 1, j - 1] + 1;
    |   m2 ← D[i - 1, 0] + 1;
    |   if Ai == Bj then cost ← 0 else cost ← 1;
    |   m3 ← D[0, j - 1] + cost;
    |   D[i, j] ← min(m1, m2, m3);
  | end
end

```

**Algoritmus 1:** Wagner–Fischerův alg. pro výpočet Levenshteinovy vzdálenosti

Mongeau a Sankoff [65] adaptovali tuto vzdálenost na porovnávání monofonních hudebních sekvencí. Na notový záznam je nahlíženo jako na posloupnost uspořádaných dvojic, kde první položka z dvojice vyjadřuje výšku noty

a druhá její délku. Levenshteinova vzdálenost je rozšířena o započtení vah jednotlivých operací – vložení, smazání nebo záměna noty. Váha operace je závislá jak na výškách, tak i délkách daných not.

Využití této vzdálenosti bylo později rozšířeno na polyfonii, avšak pouze s velkým omezením. Přípustné jsou pouze sekvence akordů, kdy současně znějící noty mají vždy stejný čas začátku znění tónu a délku. Takováto kritéria splňuje pouze malý zlomek polyfonní hudby. [55]

Různé modifikace Levenshteinovy vzdálenosti používá pro porovnávání hudebních sekvencí mnoho dalších autorů (např. [66], [67], [68], [69] a [70]).

Další přístup k výpočtu podobnosti mezi hudebními sekvencemi bere noty jako body nebo úsečky v multidimenzionálním prostoru. Vzdálenost mezi hudebními sekvencemi je poté vzdáleností těchto geometrických útvarů. Například autoři [71] reprezentují noty jako body v dvoudimenzionálním prostoru výšky a času. Každému bodu je poté přiřazena jeho váha podle délky dané noty. Dále je v daném prostoru definována vzdálenost mezi dvěma body - Eukleidovská vzdálenost. Celková vzdálenost dvou hudebních sekvencí je poté spočtena pomocí metriky Earth Movers Distance (EMD) nebo Proportional Transportation Distance (PTD). [71]

Výše uvedené metody jsou pouze malou částí z velkého množství různých existujících metod. Notový zápis můžeme v určitých případech reprezentovat také pomocí vektorů či grafů a metody výpočtu podobnosti se poté odvíjí od zvolené reprezentace (příklad pro reprezentaci ve formě grafů [72] a stromů [73]).

Vzhledem k rozdílné povaze rytmu a melodie bývá rytmická a melodická podobnost v některých případech oddělována. Další velká část autorů a prací se tedy zaměřuje pouze na porovnávání rytmu ([74], [75], [76], [77], [78], [79] a [80]) nebo porovnávání melodie ([71], [81], [82] a [83]).





Část II

**Praktická část**



---

## Popis problému

Problém automatické transkripce hudby v současné době není uspokojivě vyřešen. Špatná kvalita provedení transkripcí a velké množství chyb ve výstupu znemožňuje praktické využití této technologie. Při současném stavu řešení problému AMT v dostupném softwaru lze dokonce říci, že manuální oprava výstupů do prakticky využitelných not mnohdy zabere uživateli s průměrným hudebním sluchem více času a je pro něj komplikovanější než přímý přepis ze zvukové nahrávky.

Cílem praktické části této práce je návrh metody, která výstupy AMT upraví a vylepší automaticky.

Při řešení tohoto problému bohužel nemůžeme navázat na žádnou předchozí výzkumnou práci. Aby tedy bylo možné výstupy AMT opravit a vylepšit, je nutné nejprve analyzovat, jaké chyby při automatické transkripci vznikají, a následně vymyslet způsob pro jejich odstranění a vylepšení notového zápisu.

### 6.1 Postup úplné transkripce

Naprostá většina výzkumných prací se zabývá pouze parametrickými systémy AMT, jinak je tomu i v případě softwaru. V době psaní této práce neexistuje program pro úplnou transkripci, který by bylo možné využít.

Jelikož však existují programy pro převod zvuku do MIDI a jiné programy, které umí převádět MIDI do notového zápisu, je možné je postupně použít a získat v konečném důsledku notový zápis ve standardní hudební notaci ze zvukové nahrávky.

#### 6.1.1 Výběr programu pro převod zvuku do MIDI

Převod zvukového signálu do MIDI byl v našem případě realizován pomocí programu AmazingMIDI. Existuje více programů, které se pokoušejí o automatickou transkripci hudby, žádný z nich bohužel nedosahuje uspokojivých

výsledků (více viz 4.1). Mezi kritéria pro výběr vhodného programu patřilo především:

dostupnost programu (zdarma a volně dostupný program),

plně automatický převod (některé programy vyžadují určitý vstup uživatele do procesu transkripce např. klepání do rytmu),

možnost spuštění ve Windows 10 nebo online,

kvalita provedené transkripce.

Z existujících programů byl vybrán právě program AmazingMIDI ([41]), který se pro naše potřeby ukázal jako nejvhodnější.

Navrženou metodu lze samozřejmě uplatnit i na výstupech AMT získaných pomocí transkripce jinými programy, nemusí však dávat srovnatelné výsledky, neboť analýza chyb a následný návrh metody pro jejich odstranění a vylepšení MIDI souboru byly prováděny na výstupech z programu AmazingMIDI.

### 6.1.2 Výběr programu pro převod MIDI do notového zápisu

V procesu AMT je převod MIDI (nebo jiné parametrické reprezentace hudby) do notového zápisu (a to i ve srovnání s první částí AMT – převodu zvuku do MIDI) značně komplikovaný a neprozkoumaný podproblém. Některé notační programy však dovedou převést MIDI soubor do notového zápisu. Více detailů o této problematice je možno si přečíst v kapitole 4.1.

Pro převod MIDI do notového zápisu byl vybrán program Musescore ([54]). Tento program je volně a zdarma dostupný a oproti jiným notačním programům má pokročilé funkce pro převod MIDI do notového zápisu.

## 6.2 Představení problému

V ideálním případě by systém AMT fungoval tímto způsobem. Vstupem do systému by byla zvuková nahrávka a jeho výstupem poté správný notový zápis (se všemi hudebními prvky, které má notový zápis obsahovat) odpovídající vstupní nahrávce. V současném stavu řešení problému AMT je tento ideální stav zatím nedosažitelný a naším cílem je se tomuto stavu s dostupnými prostředky co nejvíce přiblížit.

Vzhledem k šíři a komplikovanosti problému jsme se omezili na monofonní hudbu, tedy takovou hudbu, kdy v jednom okamžiku zní maximálně jeden tón.

V první fázi transkripce je zvuková nahrávka převedena programem AmazingMIDI do formátu MIDI.

Automatická transkripce hudby není v současnosti na takové úrovni, aby dokázala tento převod provést bezchybně. (Jedná-li se o zvukový záznam s vysokým stupněm složitosti a polyfonie, nabízí se otázka, zda je převod vůbec teoreticky možný.)

V případě, že MIDI získáváme právě pomocí AMT, chyby vzniklé převodem zvuku do MIDI se ještě připočítávají k nepřesnostem, které jsou v případě využití živé nahrávky již přítomny v lidské interpretaci hudby. Výsledný MIDI soubor tedy navíc obsahuje různé nezanedbatelné množství chyb.

Proces získání notového zápisu z MIDI je netriviální i v případě, že využíváme MIDI získané z živé nahrávky na MIDI klávesy tj. takové MIDI, které je také nekvantizované a neanotované, ale nebylo získáno převodem z vlnového zvukového záznamu. Pro úplné dokončení transkripce je potřeba vyřešit ještě celou řadu úkolů. [6]

Pokud chceme tento MIDI soubor dále zpracovávat a převést jej do standardní hudební notace, přítomnost chyb tento úkol zásadním způsobem ztěžuje. Metody či algoritmy, které nejsou v celkovém výsledku moc úspěšné ani pro MIDI neobsahující chyby, někdy zcela selhávají při zpracování MIDI souborů s chybami. [10]

Úpravy výstupů a odstranění chyb je z tohoto důvodu vhodné dělat právě ve fázi, ve které je hudba reprezentována pomocí parametrické reprezentace – v našem případě MIDI. Navržená metoda upravuje vzniklý MIDI soubor a odstraňuje chyby z první fáze transkripce.

Následně je pro vizualizaci výsledků a případné další využití tento MIDI soubor převeden pomocí programu Muscore do notového záznamu. Množství chyb vzniklých v první části transkripce má za následek to, že Muscore není schopen smysluplně zobrazit původní neopravený MIDI soubor (myslíme tím smysluplně z hlediska čitelnosti a hudební teorie a praxe) a podobnost výsledného notového zápisu s originálem je jen velmi malá. Naším cílem je tento MIDI soubor upravit tak, abychom po jeho převodu do notového zápisu získali čitelný notový zápis co nejvíce odpovídající originálnímu notovému zápisu.



---

## Analýza chyb automatické transkripce

Ve výstupech AMT se vyskytuje značné množství chyb. Toto množství závisí na způsobu a kvalitě provedené transkripce a také na kvalitě a složitosti původní zvukové nahrávky. Lze si všimnout, že chyby zpravidla nevznikají náhodně, ale většinou vychází z některých vlastností a specifík povahy zvuku, které systémy AMT neumí v současnosti dostatečně dobře postihnout. Z tohoto důvodu také plyne skutečnost, že druhy chyb, které se ve výstupech různých systémů AMT objevují, jsou si navzájem podobné.

Dle literatury [3], [4] a [55] patří mezi typické chyby, vznikající při automatickém převodu zvuku na formát MIDI:

- alikvótní tóny,
- oktávové chyby,
- časové nepřesnosti,
- chybějící noty,
- přebývající noty,
- a mnohé další.

Nejčastější chyby vznikající konkrétně při transkripci pomocí programu AmazingMIDI identifikujeme porovnáním not vzniklých touto transkripcí s jejich notovým originálem. Při porovnávání notových zápisů a analýze chyb je třeba uplatnit znalosti hudební teorie a čtení složitějšího notového zápisu.

Následně v práci navrhne systém, který bere při převodu automaticky získaného MIDI do notového zápisu tyto chyby v úvahu a zároveň je odstraňuje. Výstupem transkripce programu AmazingMIDI je pouze parametrický zápis hudby - MIDI, pro účely analýzy nejčastějších vzniklých chyb je vhodné

## 7. ANALÝZA CHYB AUTOMATICKÉ TRANSKRIPCE

MIDI zobrazit ve standardní hudební notaci, což provedeme pomocí volně dostupného programu MuseScore.

Pro účely srovnání s originálem bereme notový zápis jako posloupnost objektů – not – s následujícími vlastnostmi:

- výška,
- délka,
- relativní pozice v taktu.

Relativní pozicí v taktu máme na mysli údaj o tom, na kterou dobu nota začíná. Všechny tyto vlastnosti jsou pro každou notu z notového zápisu jednoznačně určitelné. Pokud se u dvou porovnávaných notových zápisů neshoduje taktové označení (např. 3/4 a 4/4 takt), relativní pozice not v taktu nelze mezi sebou smysluplně porovnávat.

The image shows two musical staves for the song 'Drowsy Maggie'. The top staff, labeled (a), is the original notation in 4/4 time, featuring a melody of eighth and sixteenth notes. The bottom staff, labeled (b), is an automatic transcription of the same piece. It uses color-coding to highlight differences: black notes indicate perfect matches with the original, while purple notes indicate errors in pitch. The transcription also includes triplets and rests, which are marked with '3' and 'z' respectively.

(a) Část písně Drowsy Magiie - notový originál

(b) Část písně Drowsy Magiie po automatické transkripci

Obrázek 7.1: Porovnání not po automatické transkripci s originálem u písně Drowsy Maggie

Na obrázku 7.1 lze vidět srovnání notového originálu části písně Drowsy Magiie s odpovídající částí téže písně po automatické transkripci. V tomto případě program MuseScore správně určil tóninu i taktové označení. Pro větší přehlednost jsou noty po automatické transkripci obarveny podle způsobu, jakým se liší od originálu:

**černé noty:** shodují se s originálem ve všech vlastnostech;

**fialové noty:** s originálem se shodují v délce a pozici v taktu, ale byla u nich špatně detekována výška;



---

**oranžové noty:** od originálu se liší rytmicky, mají jinou délku nebo jsou špatně umístěné v taktu;

**zelené noty:** jsou v zápise navíc a neodpovídají žádným původním notám;

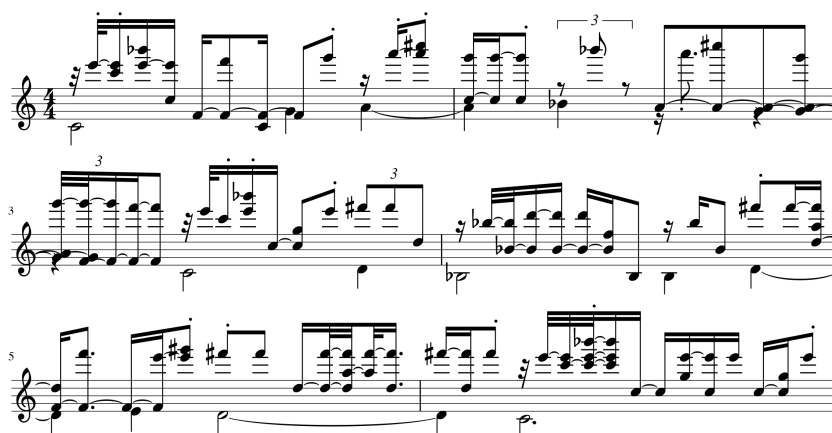
**červené noty:** odpovídají alikvótním tónům jiné noty, která v dané chvíli zní.

Na dalším obrázku 7.2 je zobrazeno srovnání notového originálu části písně Josephine's Waltz s odpovídající částí téže písně po automatické transkripci. Můžeme si všimnout, že v tomto případě nebyl ani program MuseScore schopen správně určit taktové označení a tóninu, pravděpodobně kvůli příliš velkému množství chyb vzniklých při transkripci.

V tomto případě je notový zápis po transkripci od původního tak odlišný, že se notové zápisy mezi sebou porovnávají jen s velkými obtížemi. Opět si zde ale můžeme všimnout stejných chyb jako v předchozím případě. Navíc je zde dobře patrné (např. v první době ve třetím taktu), že program AmazingMIDI má potíže se zpracováním dlouhých tónů, které rozděluje na více kratších not.



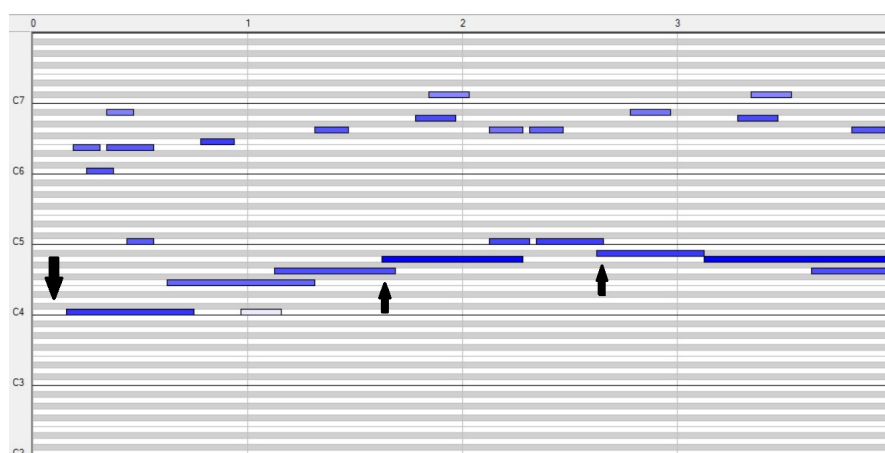
(a) Část písně Josephine's waltz - notový originál



(b) Část písně Josephine's waltz po automatické transkripci

Obrázek 7.2: Porovnání not po automatické transkripci s originálem u písně Josephine's Waltz

Jelikož program MuseScore má již zabudován pokročilejší systém převodu MIDI do standardní hudební notace, v jeho výstupu nemůžeme vidět všechny chyby, které při transkripci vznikly, protože některé z nich již dokázal odstranit. Další chyby můžeme vidět přímo na vzniklém MIDI souboru na obrázku



Obrázek 7.3: Ukázka chyb ve výstupu z AmazingMIDI

7.3. Dochází zde například k posunu začátku skladby a dalším nepřesnostem v začátcích, koncích a délkách trvání tónů. (První šipka zleva ukazuje na posunutý začátek skladby, další dvě šipky na příklady nepřesností.)

V následujících podkapitolách zjištěné chyby rozdělíme do několika kategorií a budeme se zabývat jednotlivými chybami podrobněji.

## 7.1 Z monofonie se stává polyfonie

Ačkoliv jsme se omezili na monofonii, z notových záznamů na obrázcích 7.1 a 7.2 je vidět, že po automatické transkripci zvukové nahrávky do MIDI se objevuje polyfonie. Dokonce si lze všimnout, že polyfonie - znění dvou a více tónů současně - se zde neobjevuje pouze zřídka, ale naopak vysoce převažuje nad monofonií. Tento problém je z pohledu správnosti transkripce a čitelnosti notového záznamu zásadní.

Vzhledem k tomu, že víme, že původní nahrávka byla monofonní, jakoukoliv polyfonii můžeme jednoznačně označit za chybu. Z tohoto důvodu při úpravách a vylepšování notového zápisu veškerou polyfonii odstraníme. Výsledný notový zápis bude tedy opět monofonní.

Níže jsou uvedeny dvě nejčastější příčiny vzniku polyfonie při automatické transkripci.

- Alikvótní tóny

Se základním tónem zní vždy celá řada alikvótních tónů. Lidé různou intenzitou jednotlivých alikvótních tónů vnímají jako barvu zvuku a dokáží díky nim poslechem rozeznat, který hudební nástroj tón vydává. Při automatické transkripci nejsou tyto tóny správně identifikovány jako součást již přítomného základního tónu, ale jsou naopak chybně považovány za samostatné tóny.

- Tóny nejsou včas ukončeny

Začátek tónu bývá zpravidla jasnější a ráznější než jeho konec. Při automatické transkripci se proto někdy stává, že konec tónu není správně detekován. Tón není včas ukončen a kvůli tomu vytváří polyfonii s následujícími tóny. Pokud bychom se tímto jevem zabývali podrobněji, museli bychom brát v úvahu také fakt, že ve zvukovém záznamu skladby záleží na mnoha dalších vlivech souvisejících s akustikou prostoru, ve kterém byla skladba nahrána. Jednotlivé tóny mohou znít ještě nějakou dobu po té, co byla ukončena jejich produkce. Drobné překryvy sousedních tónů mohou být také způsobeny časovými nepřesnostmi při transkripci.

## 7.2 Fragmentace dlouhých not

Při analýze chyb vzniklých automatickou transkripcí se ukázalo, že program AmazingMIDI je nerobustní vůči vibratu<sup>3</sup>, které se ve zvuku přirozeně vyskytuje. Zvuk jednotlivých druhů nástrojů se mírou vibrata liší. Velká míra vibrata, která je přítomna například ve zvuku flétny, saxofonu a dalších dechových nástrojů, způsobuje značnou chybovost transkripce. Program mylně interpretuje jeden dlouhý tón jako posloupnost více kratších tónů.

## 7.3 Chyby v detekci výšky tónu

V případě, že při transkripci dojde u konkrétního tónu k chybě v detekci výšky tónu, nelze bohužel bez návratu k původní zvukové nahrávce tuto chybu automaticky identifikovat a opravit. Nejčastěji se jedná o odchylku oproti původní výšce o půltón, tón nebo oktávu (12 půltónů).

Pokud se v nahrávce objeví šum, zvuk nehudební povahy nebo je část nahrávky nějakým způsobem poškozená, program AmazingMIDI tyto chyby interpretuje jako součást skladby. Ve výsledném MIDI se tyto chyby objeví jako hodně hluboké nebo hodně vysoké tóny. Jejich výška tedy oproti průměrné výšce tónu ve skladbě značně vybočuje směrem dolů nebo nahoru. V tomto případě lze s určitou pravděpodobností odhadnout, že tento tón do skladby nepatří a jedná se o chybu.

Program AmazingMIDI bohužel neumí identifikovat nástroje použité v nahrávce. Pokud by to uměl nebo v případě využití jiného programu, který by měl tuto funkci implementovanou, chyby tohoto typu bychom mohli snáze identifikovat a následně odstranit s využitím znalosti výškového rozsahu jednotlivých nástrojů. Například rozsah většiny dechových hudebních nástrojů je omezen na dvě až čtyři oktávy, což je již značné omezení oproti celému rozsahu výšek tónů.

<sup>3</sup>Termín vibrato označuje pravidelné přirozené kolísání frekvence a amplitudy tónu při zpěvu nebo nástrojové hře.

### 7.4 Nepřesnosti

Ačkoliv standardní notový zápis je ve své podstatě jednoznačný a přesný (např. všechny čtvrtové noty ve skladbě by měly trvat v závislosti na tempu skladby stejný počet milisekund), ve skutečnosti živě zahraná hudba není nikdy strojově přesná. Tento fakt je dán jednak neschopností člověka předepsanou přesnost dodržet, ale zároveň je určitá míra nepřesnosti také jako umělecký záměr součástí interpretace skladby.

Při automatické transkripci se k těmto nepřesnostem přidávají ještě další nepřesnosti vzniklé přímo při převodu. Tyto chyby jsou patrné ve vzniklém MIDI souboru. Typickou chybou je fakt, že krátké ticho na začátku nahrávky je již započítáno jako součást skladby. Dochází tak k celkovému posunu všech tónů, což může mít zásadní vliv při identifikaci rytmu skladby. Dalším problémem je, že začátky, konce a délky tónů neodpovídají pomyslné časové mřížce. Tímto problémem se zabývá kvantizace MIDI. (Nutno však říci, že typická kvantizace MIDI se zabývá MIDI soubory získanými například z živých nahrávek na MIDI klávesy, zatímco při AMT vznikají zpravidla mnohem větší časové nepřesnosti než v tomto případě.)

---

## Návrh metody pro automatickou úpravu transkripce

Jako hlavní problém při analýze chyb se ukázal vznik polyfonie v monofonní nahrávce (samotná polyfonie vzniká z různých příčin). Při úpravách notového zápisu se tedy zaměříme především na tento problém. Cílem metody však není pouhé libovolné odstranění polyfonie (toto by bylo v principu jednoduché), ale odstranění polyfonie takovým způsobem, aby byl výstup metody monofonní a navíc co nejpodobnější originálu před provedením automatické transkripce.

U některých druhů chyb není snadné (nebo to ani není možné) identifikovat, zda se jedná o chybu, bez návratu a přezkoumání původní zvukové nahrávky. V našem případě však víme, že jakýkoliv vznik polyfonie je chybou, protože původní nahrávka byla monofonní.

Pro účely metody definujeme notu  $N$  jako uspořádanou pětici  $(s, e, p, v, l)$ , kde

$s$  je start neboli číslo tiků začátku noty,

$e$  je end neboli číslo tiků konce noty,

$p$  je pitch (výška noty),

$v$  je velocity (hlasitost),

$l$  je length (délka noty).

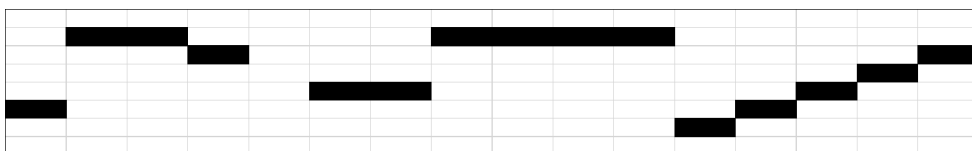
Konečná posloupnost těchto pětic představuje notový zápis. Posloupnost je seřazená podle hodnoty  $s$  – začátku znění noty.

## 8.1 Redukce polyfonie na monofonii

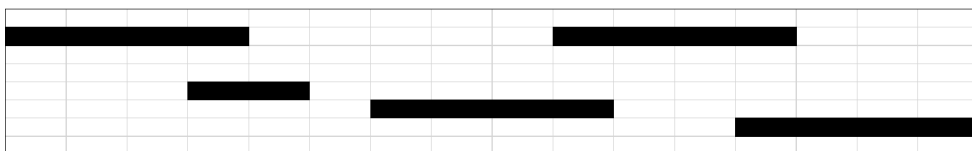
V případě monofonní sekvence pro každé dvě za sebou jdoucí noty  $N_n$  a  $N_{n+1}$ , kde  $n \in J, J \subset \mathbb{N}$ , platí, že  $e_n \leq s_{n+1}$ . (Předchozí nota skončí nejpozději ve stejnou chvíli, kdy začíná znít další nota.)

V polyfonní sekvenci poté existuje minimálně jedna dvojice za sebou jdoucích not  $N_n$  a  $N_{n+1}$ , kde  $n \in J, J \subset \mathbb{N}$ , pro které platí, že  $e_n > s_{n+1}$ . Navržená metoda všechnu polyfonii odstraní.

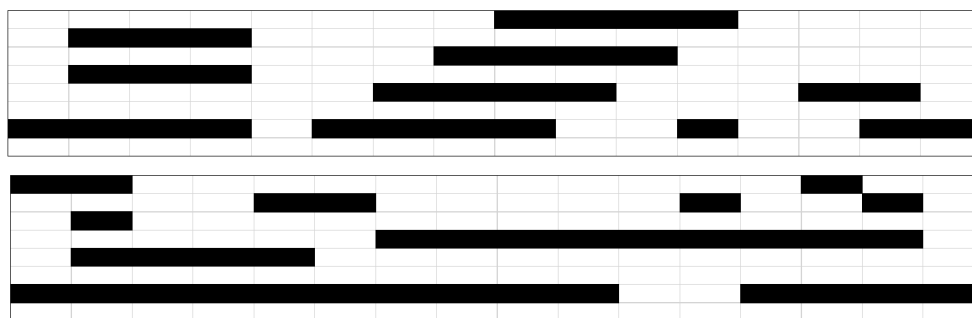
Navržená metoda postupně prochází vstupní posloupnost not získanou automatickou transkripcí a seřazenou podle hodnoty  $s$  – začátek znění noty. Všechny případy polyfonie (překryvy v trvání not) jsou postupně identifikovány a odstraněny. V polyfonní hudební sekvenci nemusí docházet pouze k jednoduchému překryvu mezi dvěma následujícími notami, může zde naráz znít i více not, noty se mohou rozmanitě překrývat atd. Na obrázku 8.1 jsou pro ukázkou znázorněny různé polyfonní hudební sekvence.



(a) Monofonní hudební sekvence



(b) Hudební sekvence s jednoduchou polyfonií



(c) Hudební sekvence se složitější polyfonií

Obrázek 8.1: Ukázka monofonní a různých polyfonních hudebních sekvencí ve formě piano-roll.

Noty jsou v jakkoliv složité polyfonii vyhodnocovány postupně po dvojicích. Způsob odstranění polyfonie pro konkrétní dvojici not je určen pomocí funkce `noteCompare`.

Výstupem metody je opět posloupnost not seřazená podle hodnoty  $s$  – začátek znění noty, tentokrát však bez připuštění polyfonie. Pro každé dvě za sebou jdoucí noty  $N_n$  a  $N_{n+1}$ , kde  $n \in J, J \subset \mathbb{N}$ , tedy platí, že  $e_n \leq s_{n+1}$ . Pseudokód metody je uveden níže.

**Vstup** : zásobník `noteStack` (posloupnost not seřazená podle začátků, dřívější na vrcholu zásobníku)

**Výstup**: pole `result` (posloupnost not seřazená podle začátků bez polyfonie)

```

while noteStack není prázdný do
  first ← noteStack.pop();
  second ← noteStack.pop();
  if nenastává překryv then
    vlož first na konec result;
    noteStack.push(second);
  else
    switch noteCompare(first, second) do
      case smazat první do
        | noteStack.push(second)
      end
      case smazat druhou do
        | noteStack.push(first)
      end
      case ořez do
        | first.end ← second.start;
        | vlož first na konec result;
        | noteStack.push(second);
      end
    end
  end
end
end

```

#### Algoritmus 2: Metoda redukce polyfonie

Důležitou částí procesu je funkce `noteCompare`. Tato funkce má hlavní vliv na výslednou podobu notového zápisu. Vstupem do funkce jsou dvě noty, u kterých dochází k překryvu v jejich znění. Výstupem je jedno ze tří možných rozhodnutí, jakým způsobem je nejhodnější notový zápis upravit.

Definujeme tři možná řešení problému polyfonie:

smažeme první notu,

smažeme druhou notu,

provedeme ořez.

Poslední způsob řešení – ořez – vychází z myšlenky, že noty nemusíme nutně brát jako neměnné objekty. U jednotlivých not můžeme měnit začátek či konec znění noty nebo její délku. V tomto případě žádnou z not nesmažeme, ale upravíme konec první noty tak, aby se s následující notou nepřekrývala a nedocházelo tak k polyfonii.

Konkrétní způsob řešení problému polyfonie je určen porovnáním not pomocí několika kritérií. Snažíme se zjistit, zda některá z not je chybná a do výsledného notového zápisu nepatří. Prvním z kritérií je rozdíl výšky porovnávaných not. Pokud vypočtený rozdíl odpovídá vzdálenosti alikvótních tónů, vyšší nota je smazána. Dalším kritériem je podíl délek porovnávaných not. Pokud je podíl menší než stanovená hranice *lenDiff*, kratší nota je smazána. Posledním kritériem je rozdíl hlasitosti porovnávaných not. Pokud je rozdíl větší než stanovená hranice *velDiff*, tišší nota je smazána.

Pokud podle žádného z předchozích kritérií není rozhodnuto o smazání jedné z not, dochází k ořezu. Předpokládáme, že žádná z porovnávaných not není úplně chybná a že vznik polyfonie byl zapříčiněn časovými nepřesnostmi, špatnou detekcí konce tónů apod.

Konec první noty je upraven tak, aby se s následující notou nepřekrývala. Nabízí se i jiné způsoby ořezu (například upravit začátek druhé noty), ty však navržená metoda neprovádí. Tento způsob ořezu byl zvolen na základě analýzy chyb a skutečnosti, že detekce konce znění tónu je obtížnější než detekce jeho začátku. V detekci konců tónů z tohoto důvodu vzniká více chyb.

Pseudokód funkce *noteCompare* je uveden níže. Při provádění experimentů můžeme tuto funkci různě upravovat a sledovat, jaký vliv mají tyto změny na výslednou podobu notového zápisu.

```
Vstup : Note first, Note second
Výstup: způsob redukce polyfonie

if rozdíl výšek odpovídá alikvótnímu tónu then
  | smaž vyšší tón
end
else if rozdíl hlasitosti not > hranice velDiff then
  | smaž tišší tón
end
else if podíl délek < hranice lenDiff then
  | smaž kratší tón
end
else
  | ořez
end

Procedura noteCompare(Note first, Note second)
```



## 8.2 Implementace

Implementace metody pro úpravu výstupů AMT byla realizována v programovacím jazyce Java. Pro práci s MIDI soubory byla použita knihovna `javax.sound.midi`.

V průběhu úprav a následných experimentech se hudba nachází v různých formátech (MIDI soubor, vnitřní reprezentace notového zápisu, formát abc), kromě implementace samotné metody pro úpravu not bylo potřeba implementovat ještě další pomocné funkce.

### 8.2.1 Realizované funkce programu

Mezi funkce, které program provádí patří:

- Načtení vstupního MIDI souboru.
- Převod jednotlivých událostí Note on a Note off do formy vnitřní reprezentace notového zápisu.
- Zarovnání not do časové mřížky, odstranění not s nulovou délkou.
- Úprava not, redukce polyfonie na monofonii.
- Převod vnitřní reprezentace hudby zpět do MIDI souboru. (Opravený MIDI soubor je výstupem programu.)

Pro účely následných experimentů provádí program také:

- Načtení abc souboru s originálním notovým zápisem a jeho převod do vnitřní reprezentace not.
- Porovnání originálního notového zápisu s výsledným notovým zápisem podle zvolené metriky. (Výpočet editační vzdálenosti, případně F-míry.)

### 8.2.2 Zarovnání not do časové mřížky

MIDI zápis vzniklý automatickou transkripcí je nekvantovaný, tóny začínají a končí v jakémkoliv čase a délky znění tónů tedy nabývají mnoha různých hodnot. Naproti tomu noty v notovém zápise nabývají pouze několika hodnot (nota celá, půlová, atd.). Abychom mohli z MIDI vytvořit posloupnost not, je potřeba provést kvantizaci. V notovém zápise existují přesně definované časy, ve kterých může nota začínat a končit, ty jsou určeny pomyslnou časovou mřížkou, která vychází z metra a rytmu skladby. Přesnost časové mřížky můžeme měnit rozhodnutím, jaké nejkratší noty v notovém zápise povolíme. V navržené metodě byla zvolena přesnost šestnáctinových not. (V notových zápisech se šestnáctinové noty běžně vyskytují. Dvaatřicetinové či kratší noty

nejsou kvůli špatné čitelnosti v praxi již tolik používány. Z tohoto důvodu byla zvolena právě tato hodnota.)

V navržené metodě jsou hodnoty začátku a konce znění noty zarovnány (zaokrouhleny) na nejbližší bod časové mřížky. Pro účely naší metody nepovolujeme jiné dělení not než na poloviny. (Nepovolujeme tedy trioly, kvartoly apod.)

Při analýze chyb bylo zjištěno, že při automatické transkripci hudby dochází často k posunu začátku celé skladby. Přestože v notových zápisech není teoreticky úplně vyloučeno, aby skladba začínala pomlčkou – tichem, ticho na začátku skladby v naší metodě považujeme vždy za chybu. Začátek skladby je posunut na čas, kdy začne znít první tón.

### 8.2.3 Vnitřní reprezentace not

Hudební sekvence je v programu představována polem not. Navržená třída Note (tabulka 8.1) uchovává a spojuje dostatek informací z rozdílných reprezentací hudby, ve kterých se nota v průběhu transkripčního procesu nachází (MIDI soubor, notový zápis – formát abc).

Struktura byla navržena s ohledem na možné budoucí rozšíření (například přidáním modulu pro převod MIDI do notového zápisu zahrnující rytmickou, harmonickou analýzu atd.) nebo změny v hodnotících kritériích pro úspěšnost transkripce. Ať již směrem k nízkoúrovňové frame-based evaluaci nebo naopak vysokoúrovňové evaluaci kompletních notových zápisů – hodnocení enharmonických záměn, rozdělení do taktů atd. Pomlka je v této reprezentaci brána jako speciální případ noty s nulovou výškou.

Pole not seřazených podle jejich začátků poté představuje notový zápis.

Typ	Název	Popis
double	length	počet šestnáctinových not trvání
int	pitch	výška udaná v MIDI čísle, nula je pomlka
string	noteName	jméno základního tónu
string	origName	původní řetězec z abc
int	barNumber	číslo taktu
long	startMIDI	číslo tiky, kdy nota začíná
long	endMIDI	číslo tiky, kdy nota končí
long	lengthMIDI	počet tiků trvání noty
int	velocity	hlasitost z MIDI souboru

Tabulka 8.1: Třída Note (seznam atributů)

---

## Experimentální vyhodnocení

V následující kapitole jsou představeny výsledky navržené metody úprav výstupů automatické transkripce hudby. Bylo provedeno několik různých experimentů zaměřených na zjištění úspěšnosti navržené metody pomocí porovnávání výsledných hudebních sekvencí s originálem. Úspěšnost transkripce a navržené metody úprav byly hodnoceny pomocí F-míry a editační vzdálenosti řetězců. Výsledky experimentů jsou vizualizovány pomocí grafů. Kapitola obsahuje také ukázky srovnání notových záznamů bez úprav a výsledných notových záznamů. Na konci kapitoly je uvedeno krátké shrnutí výsledků.

### 9.1 Dataset

Dataset využitý při experimentech se skládá z 25 monofonních notových záznamů a k nim příslušejícím MIDI souborů, které vznikly automatickou transkripcí hudby.

Notové záznamy byly získány ve formátu abc z [84] a [85]. Do datasetu byly vybrány již existující jednohlasé skladby s ohledem na různorodost jednotlivých položek. Dataset obsahuje notové zápisy ve všech nejpoužívanějších taktech (2/4, 4/4, 3/4, 6/8), v různých tóninách a různých hudebních formách (pochod, waltz, polka, reel, jig, menuet, koleda). Písně přísluší do klasické hudby a lidové hudby různých národností. Z notových záznamů byly vygenerovány syntetické MIDI nahrávky s využitím zvuku flétny. Tyto nahrávky byly převedeny do formátu wav pomocí programu MuseScore. Nahrávky byly dále převedeny pomocí programu na automatickou transkripci hudby AmazingMIDI do MIDI.

### 9.2 Použité metriky

Jak bylo zmíněno v kapitole 5, hodnocení úspěšnosti systémů AMT není nijak standardizováno a jednotlivé metody hodnocení se od sebe velmi liší. Pro účely

experimentů v této práci jsme využili dva odlišné přístupy k hodnocení.

### 9.2.1 Editační vzdálenost hudebních řetězců

Při porovnávání řetězců not využíváme Levenstheinovu vzdálenost (5.3.2). Využití Levenstheinovy vzdálenosti je omezeno pouze na porovnávání monofonních hudebních sekvencí.

Na notový záznam je v tomto případě nahlíženo jako na posloupnost uspořádaných dvojic. V tomto kontextu je postačující notu charakterizovat pouze dvěma vlastnostmi - výškou a délkou. Výška je udána MIDI číslem a délka je udána v počtu šestnáctinových not. (Šestnáctinová nota má délku jedna, osminová délku dva atd.)

Při hodnocení systému AMT nás zajímá podobnost originálního notového záznamu s výsledným notovým záznamem. Levenstheinova vzdálenost nám v tomto případě udává, jaký počet smazání, přidání nebo záměny noty/pomlky je potřebný k převodu výsledného hudebního řetězce na originální řetězec. Způsob výpočtu editační vzdálenosti mezi dvěma hudebními řetězci odpovídá výpočtu všeobecně známé Levenstheinovy vzdálenosti mezi řetězci. Algoritmus je implementován pomocí dynamického programování.

Pro účely hodnocení transkripce nám pouhý počet potřebných operací ke konverzi jednoho řetězce na druhý nedává dostatečné informace. Porovnávání řetězce jsou různě dlouhé a deset potřebných operací ve stoznakovém řetězci je zcela jistě lepší výsledek než deset v desetiznakovém řetězci. Z tohoto důvodu je editační vzdálenost vztažena k délce porovnávaných řetězců a přepočtena na míru podobnosti pomocí vzorce  $s = \frac{x-e}{x}$ , kde  $x$  označuje počet znaků delšího z porovnávaných řetězců a  $e$  vypočtenou editační vzdálenost. Podobnost nabývá hodnot z intervalu  $[0, 1]$ , kde 1 znamená dokonalou transkripci.

Při hodnocení úspěšnosti automatické transkripce používáme dvě varianty hodnocení.

#### Celková editační vzdálenost

Pro výpočet uvažujeme hudební řetězec skládající se z not a pomlky. Při porovnávání jsou noty považovány za stejné pouze v případě, že mají zároveň stejnou délku i výšku. Pomlky jsou stejné, pokud mají shodnou délku. Nota a pomlka nejsou stejné nikdy.

#### Melodická editační vzdálenost

Hudební řetězec pro účely výpočtu melodické editační vzdálenosti se skládá pouze z not, pomlky zde nejsou uvažovány. Při porovnávání jsou noty považovány za stejné v případě, že mají stejnou výšku. Délka not není brána v úvahu. Tato editační vzdálenost se tedy zaměřuje pouze na melodii a nehodnotí rytmus hudebního řetězce.

## 9.2.2 Přesnost, úplnost, F-míra a správnost

Druhá využívaná metoda hodnocení vychází ze způsobu hodnocení úspěšnosti klasifikačních problémů. Při výpočtu přesnosti, úplnosti, F-míry a správnosti používáme postup a vzorce popsané v sekci 5.1.2. Opět porovnáváme vzniklý notový zápis s originálním notovým zápisem. Výhodou této metriky (oproti výše uvedené Levenstheinovy vzdálenosti) je možnost výpočtu i pro polyfonní hudební záznamy.

Při výpočtu F-míry v hodnocení systémů AMT typicky existuje určitá tolerance, kdy tóny ještě považujeme za stejné. V našem případě je tolerance následující. V průběhu úprav notových záznamů dochází kromě redukce polyfonie také k zarovnání not do časové mřížky. Následkem tohoto zarovnání je vznik tolerance v začátku a konci znění tónu  $\frac{1}{2}$  šestnáctinové noty, což odpovídá časové toleranci 31,25 ms. Při porovnání jednotlivých not považujeme noty za stejné již pouze v případě, že mají stejnou délku, začínají na stejném místě časové mřížky a také končí na stejném místě časové mřížky. Co se týče výšky tónu, výšky tónu jsou již programem AmazingMIDI zaokrouhleny na půltóny, další tolerance ve výškách tónu tedy není žádoucí, proto není povolena.

Přesnost vyjadřuje poměr správně určených not ve výsledku ku všem notám ve výsledku. Úplnost vyjadřuje poměr správně určených not ve výsledku ku všem notám originálu. V ideálním případě je přesnost i úplnost rovna jedné. Přesnost a úplnost jsou kombinovány do souhrnné F-míry (F-measure). Pro doplnění můžeme vypočítat ještě celkovou správnost (accuracy).

Z těchto čtyř výše uvedených hodnot bývá jako rozhodující ukazatel úspěšnosti systémů AMT považována právě F-míra.

## 9.3 Experimenty

### 9.3.1 Výběr vhodných parametrů

Obecně se ve výzkumné oblasti AMT často využívá algoritmů s mnoha parametry, jejichž volba má velký vliv na úspěšnost výsledného systému AMT, ale metodika výběru parametrů není jednoznačně dána. Nejinak je tomu i v případě navržené metody pro úpravu a vylepšování výstupů automatické transkripce hudby, která je předmětem této práce.

Pokud budeme uvažovat volbu parametrů v širším kontextu, vyvstává ohledně ní mnoho otázek. Samotné sestavení funkce, která má za úkol redukovat monofonii na polyfonii, totiž můžeme považovat za parametr. Jak tuto funkci koncipovat? Které aspekty notových záznamů brát v úvahu? Jakým způsobem zarovnávat tóny do časové mřížky? Je lepší noty nejprve zarovnat a poté odstranit polyfonii nebo obráceně? Otevírá se nám velký parametrický prostor. Při návrhu metody pro zlepšování výstupů AMT bylo vyzkoušeno mnoho různých variant a postupů. Z nich jsme vybrali pomocí průběžného

testování na menším datasetu tu variantu, která se jevila jako nejlepší. Nyní odhlédneme od tohoto širokého kontextu a budeme se věnovat volbě nejlepších hodnot parametrů pro výslednou metodu představenou v předchozí části práce.

Parametry, jejichž hodnotu je potřeba určit, jsou následující:

- velDiff (hranice rozdílu hlasitosti),
- lenDiff (hranice podílu délky).

Smysluplné hodnoty parametru rozdíl hlasitosti nabývají celých čísel od 0 do 127, pro podíl délky jsou to poté reálná čísla od 0 do 1. Na obrázku 9.2 vidíme porovnání výsledků systému pro vybrané kombinace hodnot těchto parametrů. První tabulka ukazuje hodnoty získané pomocí celkové editační vzdálenosti, druhá tabulka pomocí melodické editační vzdálenosti a třetí tabulka pomocí F-míry. Jedno políčko tabulky odpovídá průměru hodnot pro všechny soubory z datasetu. Barevná škála od červené, přes bílou k modré znázorňuje hodnoty od nejnižších po nejvyšší. Tmavě červeně políčka označují nejvyšší hodnoty. Můžeme si všimnout, že rozložení výsledků je při měření pomocí editační vzdálenosti, melodické editační vzdálenosti a F-míry nápadně podobné.

Dalším parametrem, který jsme testovali, byl binární parametr, zda první alikvótní tón (1. AL) vymazávat z výsledných not nebo ne. (Důvodem k této spekulaci je skutečnost, že interval oktávy, který tomuto alikvótnímu tónu odpovídá, se v melodii písní poměrně běžně vyskytuje, mohlo by tedy být nežádoucí takové noty mazat.)

Obrázek 9.1 znázorňuje porovnání výsledků systému s odstraněním (spodní tabulka) a bez odstranění (horní tabulka) prvního alikvótního tónu. Oranžově a červeně jsou označeny nejvyšší hodnoty v dané tabulce. Je vidět, že pokud odstraňujeme i první alikvótní tón, dostáváme celkově lepší výsledky.

Pro další experimenty byly vybrány ty hodnoty parametrů, které byly při testování vyhodnoceny jako nejlepší. (velDiff = 60, lenDiff = 0,15 a 1.AL = true).

### 9.3. Experimenty

		length											
veloc		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10		0.383612	0.383612	0.383612	0.387503	0.387033	0.382693	0.381383	0.371472	0.352909	0.284115	0.263952	0.255757
20		0.485487	0.485487	0.485487	0.488284	0.488015	0.491473	0.488043	0.461509	0.420514	0.311424	0.280337	0.263355
30		0.551858	0.551858	0.551858	0.553827	0.55042	0.553255	0.545614	0.508569	0.463146	0.329809	0.284631	0.264445
40		0.54837	0.54837	0.54837	0.550644	0.547386	0.546575	0.552489	0.515299	0.473325	0.334237	0.289137	0.265472
50		0.547814	0.547814	0.547814	0.550263	0.54709	0.545872	0.551973	0.514961	0.475737	0.335709	0.289754	0.264849
60		0.549488	0.549488	0.549488	0.551935	0.548759	0.547935	0.554293	0.516309	0.477917	0.338245	0.291377	0.266476
70		0.544044	0.544044	0.544044	0.546484	0.542989	0.543824	0.550522	0.512608	0.477871	0.338245	0.291377	0.266476
80		0.539967	0.539967	0.539967	0.542225	0.538729	0.53957	0.546985	0.510138	0.475966	0.337673	0.290806	0.265905
90		0.534811	0.534811	0.534811	0.537068	0.533567	0.534655	0.54296	0.506813	0.475728	0.337673	0.290806	0.265905
100		0.533595	0.533595	0.533595	0.535852	0.532351	0.533717	0.54257	0.506543	0.475728	0.337673	0.290806	0.265905
110		0.53258	0.53258	0.53258	0.535197	0.531696	0.533406	0.542222	0.506192	0.475728	0.337673	0.290806	0.265905
120		0.532349	0.532349	0.532349	0.534965	0.531696	0.533406	0.542222	0.506192	0.475728	0.337673	0.290806	0.265905

		length											
veloc		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10		0.389269	0.389269	0.389269	0.389269	0.389269	0.384486	0.38377	0.375022	0.35311	0.284341	0.264194	0.255999
20		0.50447	0.50447	0.50447	0.50402	0.503584	0.499407	0.491987	0.46702	0.421278	0.311718	0.280632	0.263649
30		0.577317	0.577317	0.577317	0.577061	0.574374	0.567678	0.555181	0.519693	0.463464	0.329656	0.284478	0.264292
40		0.590298	0.590298	0.590298	0.590063	0.587377	0.575057	0.570116	0.533811	0.473796	0.334237	0.289137	0.265472
50		0.589241	0.589241	0.589241	0.58901	0.586323	0.573674	0.568965	0.532853	0.476207	0.335709	0.289754	0.264849
60		0.591253	0.591253	0.591025	0.588339	0.575738	0.571283	0.534194	0.478387	0.338245	0.291377	0.266476	
70		0.58418	0.58418	0.58418	0.583942	0.580931	0.57013	0.567155	0.530104	0.478341	0.338245	0.291377	0.266476
80		0.579781	0.579781	0.579781	0.579537	0.576526	0.565761	0.563622	0.527632	0.476437	0.337673	0.290806	0.265905
90		0.57446	0.57446	0.57446	0.574216	0.571205	0.560795	0.559597	0.524308	0.476199	0.337673	0.290806	0.265905
100		0.573208	0.573208	0.573208	0.572964	0.569953	0.559824	0.559207	0.524037	0.476199	0.337673	0.290806	0.265905
110		0.572087	0.572087	0.572087	0.572245	0.569234	0.559491	0.558843	0.523671	0.476199	0.337673	0.290806	0.265905
120		0.571845	0.571845	0.571845	0.572003	0.569234	0.559491	0.558843	0.523671	0.476199	0.337673	0.290806	0.265905

Obrázek 9.1: Porovnání úspěšnosti systému s mazáním (1.tabulka) a bez mazání (2.tabulka) prvního alikvótního tónu. Měřeno pomocí celkové editační vzdálenosti.

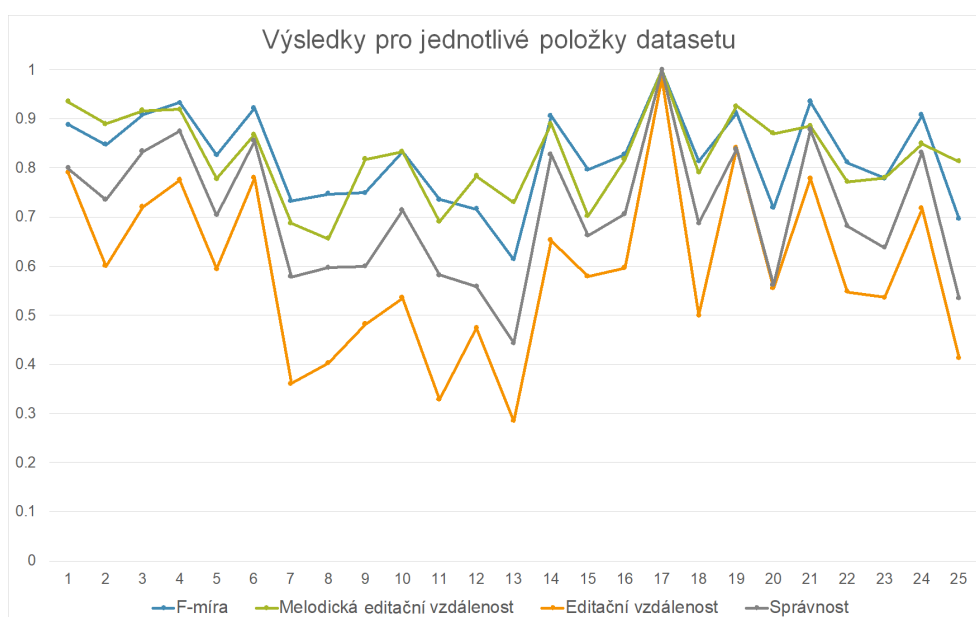






### 9.3.2 Celková úspěšnost navržené metody

Pro účely zjištění celkové úspěšnosti navržené metody bylo postupováno následovně. Pro každou položku z datasetu byl MIDI soubor získaný automatickou transkripcí hudby opraven navrženou metodou. Následně byla zkoumána podobnost nově vzniklých notových záznamů s originálními notovými záznamy. Pro každou položku byly vypočteny editační vzdálenosti a přesnost, úplnost, F-míra a správnost. Pro získání celkové úspěšnosti navržené metody byl vypočten aritmetický průměr těchto hodnot (pro každou metriku zvlášť).



Obrázek 9.3: Výsledky F-míry, editačních vzdáleností a správnosti pro jednotlivé položky z datasetu.

Navržená metoda dosáhla při měření pomocí celkové editační vzdálenosti úspěšnosti 59 % a s použitím melodické editační vzdálenosti poté 82 %.

Při použití druhého způsobu měření úspěšnosti dosáhla navržená metoda přesnosti 74 % a úplnosti 89 %. Dosažená hodnota F-míry byla 80 % a správnosti (accuracy) 69 %.

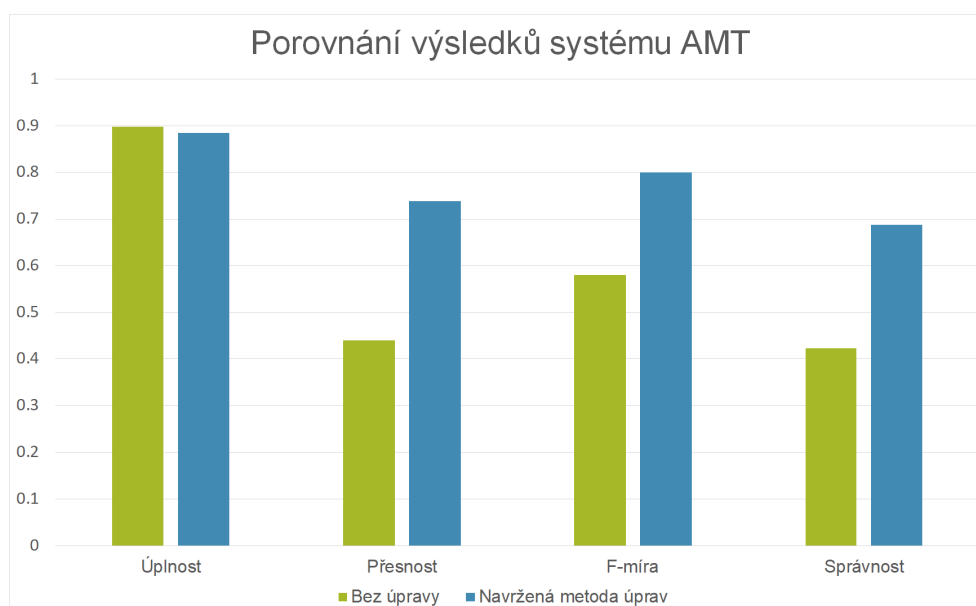
Graf na obrázku 9.3 ukazuje hodnoty F-míry, editačních vzdáleností a správnosti pro jednotlivé položky z datasetu. Lze si všimnout, že rozložení hodnot výsledků je pro jednotlivé metriky podobné. Hodnocení pomocí jednotlivých metrik vykazuje vzájemnou pozitivní korelaci.

### 9.3.2.1 Porovnání s úspěšností systému bez úprav

Důležitým ukazatelem úspěšnosti navržené metody je porovnání úspěšnosti systému bez provedených úprav s úspěšností systému s provedenými úpravami. Celková úspěšnost systému totiž samozřejmě závisí i na kvalitě transkripce pomocí programu AmazingMIDI, kterou nemůžeme následnými úpravami nijak ovlivnit.

Jelikož v tomto případě porovnáváme monofonní sekvenci s polyfonní, nelze použít editační vzdálenosti. Využijeme proto srovnání pomocí přesnosti, úplnosti, F-míry a správnosti.

I v systému „bez úprav“ odstraníme posunutý začátek skladby a zarovnáme noty do časové mřížky. Pokud bychom to neudělali, vypočtená úspěšnost transkripce by byla v tomto případě pravděpodobně velmi malá nebo dokonce nulová. (Žádné časové události by neodpovídaly originálním notám.)



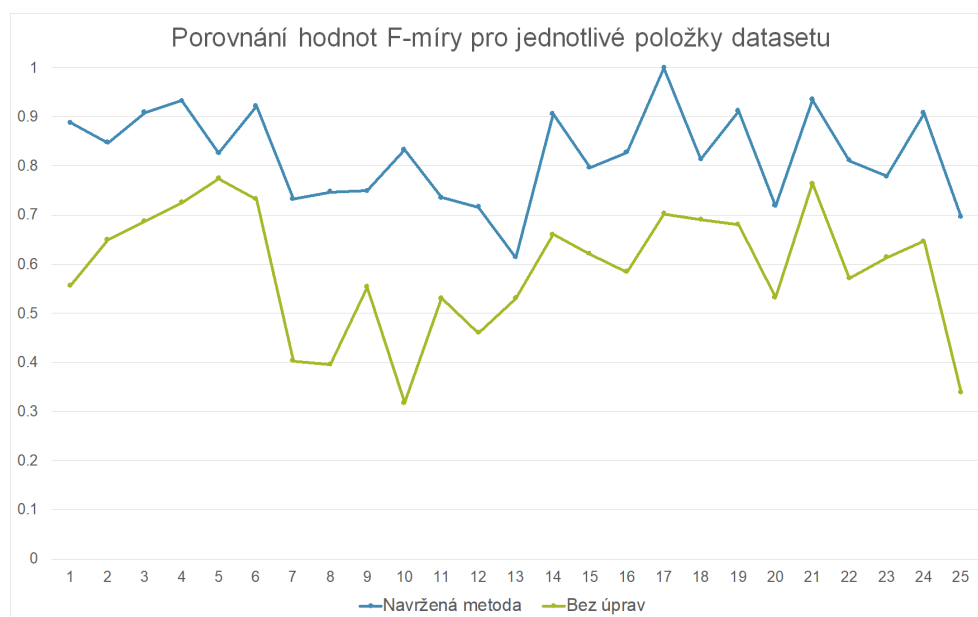
Obrázek 9.4: Porovnání výsledků systému AMT bez úprav / s navrženými úpravami

Na obrázku 9.4 vidíme graf ukazující přesnost, úplnost, F-míru a správnost systému bez úprav a po úpravě navrženou metodou. Z výsledků je zřejmé, že po provedení úprav dochází ke značnému zvýšení přesnosti (30 %), aniž by se výrazně snížila úplnost. V úplnosti dochází pouze k malému snížení (1 %). Ke zlepšení dochází i v případě F-míry (22 %) a správnosti (26 %). (Hodnoty jsou udávány jako průměr pro všechny položky datasetu.)

Tyto hodnoty nám ukazují, že metoda upravuje notový zápis způsobem, který ve většině případů ve výsledném notovém zápisu ponechává noty, které jsou správné, případně noty správně opravuje, a naopak odstraňuje chybné

noty, které do původní skladby nepatří. Navržená metoda se ukazuje jako účinná.

Na obrázku 9.5 vidíme hodnoty F-míry pro jednotlivé položky z datasetu bez úpravy a poté po úpravě navrženou metodou. Ke zlepšení došlo u všech zkoumaných položek.



Obrázek 9.5: Výsledky F-míry pro jednotlivé položky z datasetu pro systém bez úprav / s navrženými úpravami.

### 9.3.2.2 Subjektivní pohled na výsledky

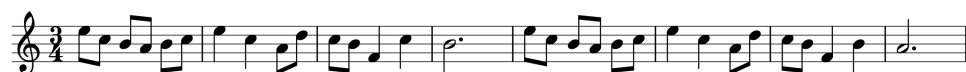
Přestože vypočtená přesnost, úplnost, F-míra a editační vzdálenosti hudebních řetězců poskytují určitou představu o úspěšnosti systému, nedokážou v celé šíři postihnout hodnocení tak komplexního problému, jako je automatická transkripce hudby. Jelikož zatím neexistují metriky pro kompletní hodnocení notových zápisů (viz kapitola 5), subjektivní hodnocení lidmi (např. tvůrci systému AMT nebo hudebními odborníky) je v současnosti nenahraditelné a je při výzkumech v této oblasti používáno [6]. Výsledné notové záznamy jsou navíc čteny a využívány opět lidmi, subjektivní pohled je tedy v tomto případě velice důležitý.

Přikládáme zde několik ukázek výsledků navržené metody úprav výstupů automatické transkripce. Na obrázcích 9.6, 9.7 a 9.8 lze vidět, že zlepšení je z lidského pohledu naprosto zásadní. Na obrázku 9.8 je ukázána část písně, u které došlo (v rámci celého datasetu) při hodnocení pomocí F-míry k nejmenšímu zlepšení (viz 9.3.2.1). I toto nejmenší pozorované zlepšení v F-míře (5 %)

## 9. EXPERIMENTÁLNÍ VYHODNOCENÍ

však odpovídá značnému subjektivnímu zlepšení čitelnosti a správnosti notového zápisu.

Navržená metoda úprav se ukázala jako účinná. Praktická využitelnost výsledných notových záznamů díky navržené metodě stoupá a co se týče čitelnosti, dochází zde také k mimořádnému zlepšení.



(a) Prvních 8 taktů originálního notového zápisu



(b) Prvních 9 taktů not po automatické transkripci



(c) Prvních 8 taktů notového zápisu po úpravách

Obrázek 9.6: Ukázka výsledků navržené metody. Viditelné rozdíly v notových záznamech odpovídají zvětšení F-míry o 16,5 %. (Průměrné zlepšení bylo 22 %.)



(a) Prvních 9 taktů originálního notového zápisu

A multi-staff musical score showing the first 6 measures of the automatically transcribed score. It includes five staves (numbered 1-5) with various musical notations, including triplets, slurs, and accidentals. The notation is significantly more complex and cluttered than the original score.

(b) Prvních 6 taktů not po automatické transkripci



(c) Prvních 9 taktů notového zápisu po úpravách

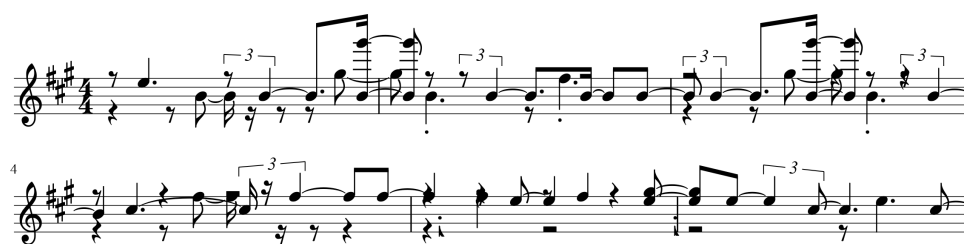
Obrázek 9.7: Ukázka výsledků navržené metody. Viditelné rozdíly v notových záznamech odpovídají zlepšení při měření pomocí F-míry o 51 % – největší zlepšení v datasetu.

## 9. EXPERIMENTÁLNÍ VYHODNOCENÍ

---



(a) Prvních 6 taktů originálního notového zápisu



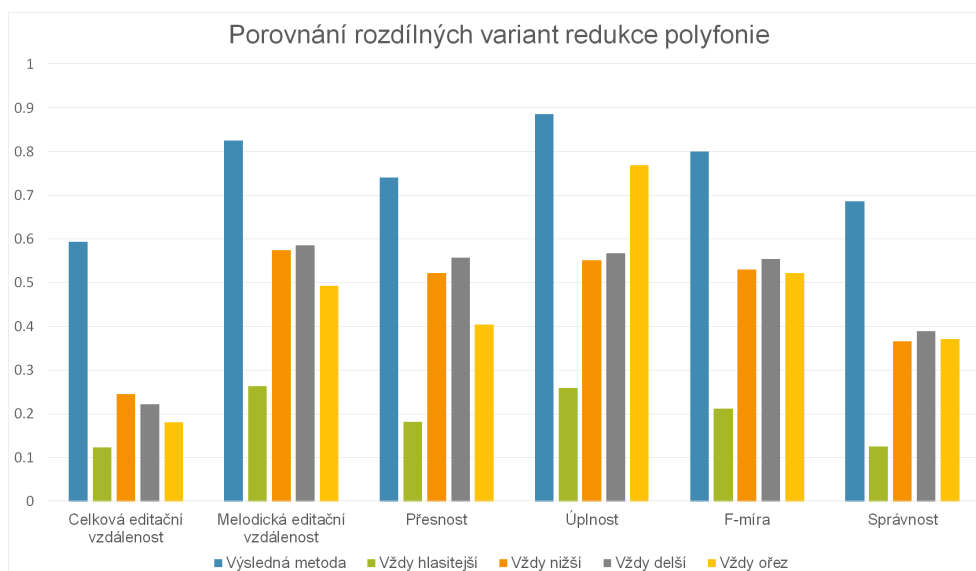
(b) Prvních 6 taktů not po automatické transkripci



(c) Prvních 6 taktů notového zápisu po úpravách

Obrázek 9.8: Ukázka výsledků navržené metody. Viditelné rozdíly v notových záznamech odpovídají zvětšení F-míry o 5 % – nejmenší zlepšení v datasetu.

### 9.3.3 Testování dalších variant úprav



Obrázek 9.9: Porovnání různých způsobů redukce polyfonie na monofonii

Další experiment vychází z otázky, zda by bylo možné funkci `noteCompare` při redukcí polyfonie na monofonii nějakým způsobem zjednodušit. Například tím, že bychom při porovnávání not brali v úvahu pouze určité kritérium. Byly proto navrženy následující čtyři zjednodušené varianty a jejich úspěšnost následně otestována a srovnána s výslednou metodou úprav.

**Rozhodování pouze podle výšky not.** Při současném znění dvou not vždy do výsledného notového zápisu vybereme tu, která je vyšší.

**Rozhodování pouze podle hlasitosti not.** Při současném znění dvou not vždy do výsledného notového zápisu vybereme tu, která je hlasitější (má větší hodnotu `velocity`).

**Rozhodování pouze podle délky not.** Při současném znění dvou not vždy do výsledného notového zápisu vybereme tu, která je delší.

**Vždy provedeme ořez.** Když zazní nová nota, předchozí notu ukončíme.

Výsledky (obrázek 9.9 a tabulka 9.1) ukazují, že navržená metoda ve všech měřených kritériích dosahuje lepších výsledků, než zjednodušené varianty.

## 9. EXPERIMENTÁLNÍ VYHODNOCENÍ

---

	Výsledná metoda	Vždy hlasitější	Vždy nižší	Vždy delší	Vždy ořez
Celková editační vzdálenost	0.59	0.12	0.25	0.22	0.18
Melodická editační vzdálenost	0.82	0.26	0.57	0.58	0.49
Přesnost	0.74	0.18	0.52	0.56	0.40
Úplnost	0.89	0.26	0.55	0.57	0.77
F-míra	0.80	0.21	0.53	0.55	0.52
Správnost	0.69	0.12	0.37	0.39	0.37

Tabulka 9.1: Porovnání různých variant funkce redukce polyfonie

### 9.4 Shrnutí výsledků

Provedené experimenty se zaměřovaly na hodnocení kvality transkripce a provedených úprav. Pomocí editačních vzdáleností a F-míry byly porovnávány vzniklé notové zápisy s originálními notovými zápisy.

Výsledky ukázaly, že navržená metoda úprav je účinná a zvětšuje podobnost vzniklých notových zápisů s originálem. Při hodnocení pomocí F-míry došlo k průměrnému zlepšení o 22 %. Dále bylo zjištěno, že navržená metoda dosahuje ve všech měřených aspektech větší úspěšnosti než její zjednodušené varianty.

Při subjektivním pohledu na výsledné notové záznamy si lze všimnout, že zlepšení, ke kterému dochází po provedení úprav navrženou metodou, je z lidského pohledu velmi zásadní.

Navržené úpravy přispívají k lepší čitelnosti a praktickému využití výsledných notových záznamů.



---

## Závěr

V rámci teoretické části práce byly shrnuty základní informace o problematice automatické transkripce hudby a popsány jednotlivé druhy reprezentace hudby, které se v průběhu transkripčního procesu vyskytují (zvuk, parametrická reprezentace – MIDI, notový zápis). Dále zde byl představen aktuální stav řešení problému AMT včetně existujícího softwaru a nastíněny problémy a výzvy, se kterými se tato výzkumná oblast potýká. V neposlední řadě zde bylo popsáno několik možností, jak lze přistupovat k hodnocení výsledků systémů AMT.

Cílem praktické části práce bylo analyzovat nejčastější chyby vznikající při automatické transkripci vlnového zvukového záznamu do formátu MIDI a na základě této analýzy navrhnout vhodnou metodu pro vylepšení a úpravu získaného MIDI souboru. Ukázalo se, že množství chyb ve výsledné reprezentaci hudby je značné a znemožňuje praktické využití technologie AMT. Mezi chyby objevující se ve výsledné reprezentaci hudby patří především přítomnost alikvótních tónů, špatná detekce konců tónu a různé časové nepřesnosti.

Byla navržena nová metoda úpravy MIDI souborů, která tyto chyby automaticky odstraňuje a provádí úpravy s cílem celkového vylepšení úspěšnosti systému automatické transkripce hudby. Tato metoda byla implementována a její úspěšnost byla poté vyhodnocena pomocí F-míry a editační vzdálenosti hudebních řetězců.

Experimenty v závěrečné části práce ukázaly, že navržená metoda úprav zvětšuje podobnost výsledných notových záznamů s originálem. Při hodnocení pomocí F-míry bylo dosaženo průměrného zlepšení o 22 %. Z ukázek notových záznamů před a po úpravách je patrné, že toto zlepšení odpovídá zásadnímu zlepšení v čitelnosti a praktické využitelnosti výsledných notových záznamů. Navržená metoda se ukázala jako účinná.

## Budoucí práce

Problematika automatické transkripce hudby je natolik rozsáhlá a komplexní, že je možné se jejímu výzkumu věnovat prakticky celý život. Zároveň není v současnosti dostatečně prozkoumaná a poskytuje tedy široké možnosti dalšího výzkumu v rozličných podoblastech, ať už se jedná o zlepšení detekce tónů z nahrávky, návrh metrik pro hodnocení úspěšnosti úplných transkripčních systémů nebo například transkripci bicích nástrojů. Zvlášť neprobádaná je ta část transkripce, která se zaměřuje na oblast převodu parametrického zápisu hudby do standardní hudební notace. Tato práce přispěla pouze malým kouskem do ohromné mozaiky. Stále zde zůstává mnoho otázek, které je potřeba zodpovědět, a mnoho problémů, které je potřeba vyřešit.

Vzhledem k tomu, že současné vědecké poznání není v oblasti hudby ještě zdaleka na svém vrcholu (například dosud není dostatečně prozkoumáno a objasněno, jak hudbu vnímá a zpracovává lidský mozek), jistý pokrok ve zlepšení automatické transkripce hudby můžeme v budoucnosti jistě očekávat souběžně s pokrokem v dalších přírodních a medicínských vědách – biologie, neurologie a neuropsychologie. Třeba se jednou dočkáme doby, kdy počítači pustíme komplikovanou polyfonní skladbu hranou na několik různých hudebních nástrojů a vzápětí si budeme moci vytisknout tuto skladbu zapsanou v notách, rozdělenou do partů. K tomu ovšem vede ještě velice dlouhá a složitá cesta.

---

## Literatura

1. BENETOS, Emmanouil; DIXON, Simon; GIANNOULIS, Dimitrios; KIRCHHOFF, Holger; KLAPURI, Anssi. Automatic music transcription: challenges and future directions. *Journal of Intelligent Information Systems*. 2013, roč. 41, č. 3, s. 407–434. Dostupné z DOI: 10.1007/s10844-013-0258-3.
2. RODRÍGUEZ, José Luis Oropeza; GUERRA, Sergio Suárez; LÓPEZ, Omar Velázquez. *Artificial Intelligence Methods for Automatic Music Transcription using Isolated Notes in Real-Time* [online]. 2018 [cit. 2019-04-13]. Dostupné z: <https://www.cicling.org/micai/2018/papers>.
3. COGLIATI, Andrea; DUAN, Zhiyao. A Metric for Music Notation Transcription Accuracy. In: *18th ISMIR Conference* [online]. 2017, s. 407–413 [cit. 2019-05-12]. Dostupné z: [https://ismir2017.smcnus.org/wp-content/uploads/2017/10/131\\_Paper.pdf](https://ismir2017.smcnus.org/wp-content/uploads/2017/10/131_Paper.pdf).
4. BENETOS, Emmanouil; DIXON, Simon; DUAN, Zhiyao; EWERT, Sebastian. Automatic Music Transcription: An Overview. *IEEE Signal Processing Magazine*. 2019, roč. 36, č. 1, s. 20–30. ISSN 1053-5888. Dostupné z DOI: 10.1109/MSP.2018.2869928.
5. DUAN, Zhiyao; BENETOS, Emmanouil. *ISMIR 2015 Tutorial - Automatic Music Transcription* [online]. 2015 [cit. 2019-04-13]. Dostupné z: <http://c4dm.eecs.qmul.ac.uk/ismir15-amt-tutorial/>.
6. COGLIATI, Andrea; TEMPERLEY, David; DUAN, Zhiyao. Transcribing Human Piano Performances into Music Notation. In: *17th ISMIR Conference* [online]. 2016, s. 758–764 [cit. 2019-04-13]. Dostupné z: <http://www2.ece.rochester.edu/projects/air/publications/>.
7. BENETOS, Emmanouil; KLAPURI, Anssi; DIXON, S. Score-informed transcription for automatic piano tutoring. In: *Signal Processing Conference (EUSIPCO)*. 2012, s. 2153–2157. ISBN 978-1-4673-1068-0.

8. EWERT, Sebastian; WANG, Siying; MÜLLER, Meinard; SANDLER, Mark B. Identifying Missing and Extra Notes in Piano Recordings Using Score-Informed Dictionary Learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2017, roč. 25, č. 10, s. 1877–1889. ISSN 2329-9290. Dostupné z DOI: 10.1109/TASLP.2017.2724203.
9. CAMBOUROPOULOS, Emilios. From MIDI to Traditional Musical Notation. In: *Proceedings of the AAAI Workshop on Artificial Intelligence and Music: Towards Formal Models for Composition, Performance and Analysis* [online]. 2000 [cit. 2019-04-13]. Dostupné z: <http://users.auth.gr/emilios/papers/aaai2000.pdf>.
10. NAKAMURA, Eita; BENETOS, Emmanouil; YOSHII, Kazuyoshi; DIXON, Simon. Towards Complete Polyphonic Music Transcription: Integrating Multi-Pitch Detection and Rhythm Quantization. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2018. Dostupné z DOI: 10.1109/icassp.2018.8461914.
11. ARGENTI, Fabrizio; NESI, Paolo; PANTALEO, Gianni. Automatic Music Transcription: From Monophonic to Polyphonic. In: 2011, s. 27–46. Dostupné z DOI: 10.1007/978-3-642-22291-7\_3.
12. KLAPURI, Anssi P. Automatic Music Transcription as We Know it Today. *Journal of New Music Research*. 2004, roč. 33, č. 3, s. 269–282. Dostupné z DOI: 10.1080/0929821042000317840.
13. TAKEDA, H.; SAITO, N.; OTSUKI, T.; NAKAI, M.; SHIMODAIRA, H.; SAGAYAMA, S. Hidden Markov model for automatic transcription of MIDI signals. In: *2002 IEEE Workshop on Multimedia Signal Processing*. 2002, s. 428–431. Dostupné z DOI: 10.1109/MMSP.2002.1203337.
14. KLAPURI, Anssi. Introduction to Music Transcription. *Signal Processing Methods for Music Transcription*. 2006, s. 3–20. Dostupné z DOI: 10.1007/0-387-32845-9\_1.
15. MUELLER, Meinard; PARDO, Bryan A; MYSORE, Gautham J; VALIMAKI, Vesa. Recent Advances in Music Signal Processing [From the Guest Editors]. *IEEE Signal Processing Magazine*. 2019, roč. 36, č. 1, s. 17–19. ISSN 1053-5888. Dostupné z DOI: 10.1109/MSP.2018.2876190.
16. WANG, Chong-kai; LYU, Ren-Yuan; CHIANG, Yuang-chin. An automatic singing transcription system with multilingual singing lyric recognizer and robust melody tracker. In: *INTERSPEECH*. 2003.
17. DIXON, Simon. On the Computer Recognition of Solo Piano Music. In: *Australasian Computer Music Conference*. 2003, s. 31–37.
18. ZENKL, Luděk. *ABC hudební nauky*. 8. vyd., v Editio Bärenreiter Praha vyd. 2. Praha: Editio Bärenreiter Praha, 2003. ISBN 80-86385-21-3.

19. *Zvukový signál* [online obrázek] [cit. 2019-04-14]. Dostupné z: [http://sites.music.columbia.edu/cmcc/MusicAndComputers/chapter1/01\\_01.php](http://sites.music.columbia.edu/cmcc/MusicAndComputers/chapter1/01_01.php).
20. BAIN, Reginald. *The Harmonic Series: A Path to understanding musical intervals, scales, tuning and timbre* [online]. 2003 [cit. 2019-04-14]. Dostupné z: <http://in.music.sc.edu/fs/bain/atmi02/hs/hs.pdf>.
21. KLAPURI, Anssi. Multiple fundamental frequency estimation by summing harmonic amplitudes. In: *ISMIR*. 2006, s. 216–221.
22. *Spectrogram* [online obrázek] [cit. 2019-04-14]. Dostupné z: <https://auditoryneuroscience.com/pitch/melodies-timbre>.
23. *The complete MIDI 1.0 detailed specification: incorporating all recommended practices*. MIDI Manufacturers Association, 2006.
24. CUNNINGHAM, Stuart. Suitability of musicxml as a format for computer music notation and interchange. In: *Proceedings of IADIS Applied Computing 2004 International Conference, Lisbon, Portugal* [online]. 2004 [cit. 2019-05-12]. Dostupné z: [https://www.researchgate.net/publication/228705496\\_Suitability\\_of\\_MusicXML\\_as\\_a\\_Format\\_for\\_Computer\\_Music\\_Notation\\_and\\_Interchange](https://www.researchgate.net/publication/228705496_Suitability_of_MusicXML_as_a_Format_for_Computer_Music_Notation_and_Interchange).
25. KLAPURI, A.; DAVY, M. *Signal Processing Methods for Music Transcription*. Springer US, 2007. ISBN 9780387328454. Dostupné také z: <https://books.google.cz/books?id=AF30yR41GIAC>.
26. HABUDOVÁ-HRUŠKOVÁ, N. Representation and Pattern Matching Techniques for Music Data. In: *Student Research Conference*. 2011, sv. 1, s. 125–131.
27. *Piano Roll and Event List Views* [online obrázek]. 2018 [cit. 2019-04-14]. Dostupné z: <http://digitalsoundandmusic.com/3-2-1-piano-roll-and-event-list-views/>.
28. *Enharmonická záměna* [online obrázek] [cit. 2019-04-14]. Dostupné z: <https://www.luciedittrichova.cz/wp-content/uploads/2016/09/1C-v%C3%BDukov%C3%A9-materi%C3%A1ly.pdf>.
29. *Midi čísla, noty, frekvence* [online obrázek] [cit. 2019-04-14]. Dostupné z: <https://newt.phys.unsw.edu.au/jw/notes.html>.
30. WRIGHT, David. *Mathematics and music*. American Mathematical Soc., 2009. ISBN 9780821848739.
31. BACH, Johann Sebastian. *Fantaisie and Fugue in A minor, BWV 904* [online] [cit. 2019-04-14]. Dostupné z: [https://imslp.org/wiki/File:PMLP08009-Fantaisie\\_and\\_Fugue\\_in\\_A\\_minor,\\_BWV\\_904.pdf](https://imslp.org/wiki/File:PMLP08009-Fantaisie_and_Fugue_in_A_minor,_BWV_904.pdf).
32. *Hudební nauka* [online obrázek] [cit. 2019-04-14]. Dostupné z: <https://www.zuszasbraslav.cz/images/hudebninauka/Prehled%20uciva%20HN%20-%20teorie.pdf>.

33. *Nota – části noty* [online obrázek] [cit. 2019-04-14]. Dostupné z: [https://cs.wikipedia.org/wiki/Nota#/media/File:Parts\\_of\\_a\\_musical\\_note.svg](https://cs.wikipedia.org/wiki/Nota#/media/File:Parts_of_a_musical_note.svg).
34. *Dělení not* [online obrázek] [cit. 2019-04-14]. Dostupné z: <https://docplayer.cz/17439653-1-houslovy-altovy-tenorovy-a-basovy-klic.html>.
35. *Tóniny* [online obrázek] [cit. 2019-04-15]. Dostupné z: <https://www.piano-keyboard-guide.com/key-signatures.html>.
36. GOOD, Michael; RECORDARE, LLC. Lessons from the Adoption of MusicXML as an Interchange Standard. In: *Proceedings of XML*. 2006, s. 5–7.
37. OLIWA, Tomasz Michal. Genetic algorithms and the abc music notation language for rock music composition. In: *Proceedings of the 10th annual conference on Genetic and evolutionary computation*. 2008, s. 1603–1610.
38. WEIHS, Claus; LIGGES, Uwe. Parameter Optimization in Automatic Transcription of Music. *From Data and Information Analysis to Knowledge Engineering Studies in Classification, Data Analysis, and Knowledge Organization*. 2006, s. 740–747. Dostupné z DOI: 10.1007/3-540-31314-1\_91.
39. LEÃO, Henrique BS; GUIMARÃES, Germano F.; RAMALHO, Geber L.; CAVALCANTE, Sérgio V. *Benchmarking Wave-to-MIDI Transcription Tools* [online] [cit. 2019-04-28]. Dostupné z: <http://gsd.ime.usp.br/~lku/site-sbcm/2000/papers/leao.pdf>.
40. AKOFF SOUND LABS. *Akoff Music Composer. Version 3.0* [software] [cit. 2019-04-29]. Dostupné z: <http://www.akoff.com>.
41. ARAKISOFTWARE. *AmazingMIDI. Version 1.70* [software]. 2003 [cit. 2019-04-29]. Dostupné z: <http://www.pluto.dti.ne.jp/~araki/amazingmidi/>.
42. LUNAVERUS. *AnthemScore. Version 4.3.1* [software]. 2019 [cit. 2019-04-29]. Dostupné z: <https://www.lunaverus.com/>.
43. NEURATRON LTD. *AudioScore 2018* [software]. 2018 [cit. 2019-04-29]. Dostupné z: <http://www.sibelius.com/products/audioscore/ultimate.html>.
44. INNOVATIVE MUSIC SYSTEMS. *IntelliScore. Version 8.1* [software]. 2017 [cit. 2019-04-29]. Dostupné z: <http://www.intelliscore.net/>.
45. DOREMIR MUSIC RESEARCH AB. *ScoreCloud. Version 4.* [software]. 2018 [cit. 2019-04-29]. Dostupné z: <https://scorecloud.com/download>.
46. GAILIUS RASKINIS. *Solo Explorer. Version 1.007* [software]. 2002 [cit. 2019-05-08]. Dostupné z: <http://www.recognisoft.com/>.

47. SEVENTH STRING SOFTWARE. *Transcribe! Version 8.73* [software]. 2008 [cit. 2019-04-29]. Dostupné z: <https://www.seventhstring.com/xscribe/overview.html>.
48. WIDISOFT. *WIDI Recognition System. Version 4.5* [software] [cit. 2019-04-29]. Dostupné z: <http://www.widisoft.com/english/download.html>.
49. APPLE INC. *Logic pro X. Version 10.4.4* [software]. 2019 [cit. 2019-04-29]. Dostupné z: <https://www.apple.com/logic-pro/>.
50. ANTARES AUDIO TECHNOLOGIES. *Autotune* [software]. 2019 [cit. 2019-04-29]. Dostupné z: <https://www.antarestech.com/>.
51. CELEMONY. *Melodyne 4* [software] [cit. 2019-04-29]. Dostupné z: <https://shop.celemony.com/cgi-bin/WebObjects/CelemonyShop>.
52. MAKEMUSIC. *Finale. Version 26.0.1.655* [software]. 2018 [cit. 2019-05-04]. Dostupné z: <https://www.finalemusic.com/>.
53. AVID TECHNOLOGY. *Sibelius* [software]. 2019 [cit. 2019-04-29]. Dostupné z: <http://www.sibelius.com/download/index.html>.
54. MUESCORE BVBA. *MuseScore. Version 2.3.2* [software]. 2018 [cit. 2019-04-29]. Dostupné z: <https://musescore.org/en/download>.
55. DANIEL, Adrien; EMIYA, Valentin; DAVID, Bertrand. Perceptually-based evaluation of the errors usually made when automatically transcribing music. In: *ISMIR 2008: Proceedings of the 9th International Conference of Music Information Retrieval*. Drexel University, 2008. ISBN 9780615248493.
56. BENETOS, Emmanouil; DIXON, Simon; GIANNOULIS, Dimitrios; KIRCHHOFF, Holger; KLAPURI, Anssi. Automatic music transcription: Breaking the glass ceiling. *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012* [online]. 2012 [cit. 2019-04-29]. Dostupné z: <https://www.eecs.qmul.ac.uk/~simond/pub/2012/Benetos-et-al-ISMIR2012-MIRrors.pdf>.
57. *MIREX* [online]. IMIRSEL [cit. 2019-04-15]. Dostupné z: [https://www.music-ir.org/mirex/wiki/2019:Main\\_Page](https://www.music-ir.org/mirex/wiki/2019:Main_Page).
58. DOWNIE, J. Stephen. The Music Information Retrieval Evaluation eXchange (MIREX). *D-Lib Magazine*. 2006, roč. 12, č. 12.
59. BAY, Mert; EHMANN, Andreas F; DOWNIE, J Stephen. Evaluation of Multiple-F0 Estimation and Tracking Systems. In: *ISMIR*. 2009, s. 315–320.
60. DIXON, S. On the computer recognition of solo piano music. In: *2000 Australasian Computer Music Conference* [online]. 2000, s. 31–37 [cit. 2019-05-10]. Dostupné z: <https://pdfs.semanticscholar.org/fa23/62d077f11db4ce62d68504cfd060414c839e.pdf>.

61. MCLEOD, Andrew; STEEDMAN, Mark. Evaluating automatic polyphonic music transcription. In: *19th ISMIR Conference* [online]. 2018 [cit. 2019-05-10]. Dostupné z: [http://ismir2018.ircam.fr/doc/pdfs/148\\_Paper.pdf](http://ismir2018.ircam.fr/doc/pdfs/148_Paper.pdf).
62. CAMBOUROPOULOS, Emilios; CRAWFORD, Tim; ILIOPOULOS, C. Pattern Processing in Melodic Sequences: Challenges, Caveats and Prospects. *Computers and the Humanities*. 2001, roč. 35, s. 9–21. Dostupné z DOI: 10.1023/A:1002646129893.
63. TOUSSAINT, Godfried T. Algorithmic, geometric, and combinatorial problems in computational music theory. *Proceedings of X Encuentros de Geometria Computacional* [online]. 2003, s. 101–107 [cit. 2019-05-10]. Dostupné z: <https://pdfs.semanticscholar.org/d90d/1d980548f5d26c2ef6e81c970c001f4803b7.pdf>.
64. WAGNER, Robert A.; FISCHER, Michael J. The String-to-String Correction Problem. *J. ACM*. 1974, roč. 21, č. 1, s. 168–173. ISSN 0004-5411. Dostupné z DOI: 10.1145/321796.321811.
65. MONGEAU, Marcel; SANKOFF, David. Comparison of musical sequences. *Computers and the Humanities*. 1990, roč. 24, č. 3, s. 161–175. Dostupné z DOI: 10.1007/bf00117340.
66. TOUSSAINT, Godfried T.; OH, Seung Man. Measuring musical rhythm similarity: Edit distance versus minimum-weight many-to-many matchings. In: *Proceedings on the International Conference on Artificial Intelligence (ICAI)*. 2016, s. 186.
67. TREMBLAY, G.; CHAMPAGNE, F. Marking musical dictations using the edit distance algorithm. *Software: Practice and Experience*. 2007, roč. 37, č. 2, s. 207–230. Dostupné z DOI: 10.1002/spe.764.
68. POST, Olaf; TOUSSAINT, Godfried. The Edit Distance as a Measure of Perceived Rhythmic Similarity. *Empirical Musicology Review*. 2011, roč. 6, č. 3, s. 164–179. Dostupné z DOI: 10.18061/1811/52811.
69. LEMSTRÖM, Kjell; UKKONEN, Esko. Including interval encoding into edit distance based music comparison and retrieval. In: *Proc. AISB*. 2000, s. 53–60.
70. KADOTA, Takashi; HIRAO, Masahiro; ISHINO, Akira; TAKEDA, Masayuki; SHINOHARA, Ayumi; MATSUO, Fumihiro. Musical sequence comparison for melodic and rhythmic similarities. In: *8th Symposium on String Processing and Information Retrieval, SPIRE 2001*. 2001. Dostupné z DOI: 10.1109/spire.2001.989744.



71. VELTKAMP, Remco; TYPKE, Rainer; GIANNOPOULOS, Panos; WIERING, F; OOSTRUM, R.W. van. Using Transportation Distances for Measuring Melodic Similarity. *ISMIR* [online]. 2003 [cit. 2019-05-12]. Dostupné z: <http://www.cs.uu.nl/research/techreps/repo/CS-2003/2003-024.pdf>.
72. ORIO, Nicola; RODA, Antonio. A Measure of Melodic Similarity based on a Graph Representation of the Music Structure. In: *10th ISMIR Conference*. 2009, s. 543–548.
73. VALERO, David Rizo; INESTA, José Manuel. Tree-structured representation of melodies for comparison and retrieval. In: *Int. Workshop on Pattern Recognition in the Information Society, PRIS 2002*. 2002, s. 140–155.
74. TOUSSAINT, Godfried T.; OH, Seung Man. Measuring Musical Rhythm Similarity: Further Experiments with the Many-to-Many Minimum-Weight Matching Distance. *Journal of Computer and Communications*. 2016, roč. 04, č. 15, s. 117–125. Dostupné z DOI: 10.4236/jcc.2016.415011.
75. TOUSSAINT, Godfried T. A Comparison of Rhythmic Similarity Measures. In: *ISMIR* [online]. 2004 [cit. 2019-05-12]. Dostupné z: <http://www-cgri.cs.mcgill.ca/~godfried/teaching/mir-reading-assignments/Comparison-of-Rhythmic-Similarity-Measures.pdf>.
76. MANZOLLI, Jônatas; GIMENES, Marcelo; MAIA JR, Adolfo; MARTINS, João M. Similarity Measures for Rhythmic Sequences. *NICS Reports*. 2013, č. 2, s. 4–13.
77. HOLZAPFEL, Andre; STYLIANOU, Yannis. A scale transform based method for rhythmic similarity of music. *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2009. Dostupné z DOI: 10.1109/icassp.2009.4959584.
78. PAULUS, Jouni; KLAPURI, Anssi. Measuring the similarity of Rhythmic Patterns. In: *ISMIR* [online]. 2002 [cit. 2019-05-12]. Dostupné z: <http://ismir2002.ircam.fr/proceedings/02-FP05-1.pdf>.
79. HOLZAPFEL, Andre; STYLIANOU, Yannis. Rhythmic similarity of music based on dynamic periodicity warping. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2008. Dostupné z DOI: 10.1109/icassp.2008.4518085.
80. MARTINS, Joao M; GIMENES, Marcelo; MANZOLLI, Jônatas; MAIA JR, Adolfo. Similarity measures for rhythmic sequences. In: *Proceedings of the 10th Brazilian Symposium on Computer Music (SBCM)*. 2005.

81. URBANO, Julián; LLORÉNS, Juan; MORATO, Jorge; CUADRADO, Sonia. Melodic Similarity through Shape Similarity. *Exploring Music Contents Lecture Notes in Computer Science*. 2011, s. 338–355. Dostupné z DOI: 10.1007/978-3-642-23126-1\_21.
82. MÜLLENSIEFEN, Daniel; FRIELER, Klaus. Optimizing Measures Of Melodic Similarity For The Exploration Of A Large Folk Song Database. In: *ISMIR* [online]. 2004 [cit. 2019-05-12]. Dostupné z: <http://www.ee.columbia.edu/~dpwe/ismir2004/CRFILES/paper178.pdf>.
83. ALOUPIS, Greg; FEVENS, Thomas; LANGERMAN, Stefan; MATSUI, Tomomi; MESA, Antonio; NUÑEZ, Yurai; RAPPAPORT, David; TOUS-SAINT, Godfried. Algorithms for Computing Geometric Measures of Melodic Similarity. *Computer Music Journal*. 2006, roč. 30, č. 3, s. 67–76. Dostupné z DOI: 10.1162/comj.2006.30.3.67.
84. WALSHAW, Chris. *Abc notation* [online] [cit. 2019-04-24]. Dostupné z: <http://abcnotation.com/>.
85. *The Session* [online] [cit. 2019-04-24]. Dostupné z: <https://thesession.org/>.

## Seznam použitých zkratk

**AMT** Automatic Music Transcription

**ASCII** American Standard Code for Information Interchange

**BPM** Beat per minute

**MIDI** Musical Instrument Digital Interface

**MIR** Music Information Retrieval

**MIREX** The Music Information Retrieval Evaluation eXchange

**SMF** Standard MIDI File



---

## Obsah přiloženého CD

readme.txt .....	stručný popis obsahu CD
src	
├─ impl .....	zdrojové kódy implementace
├─ thesis .....	zdrojová forma práce ve formátu $\LaTeX$
text .....	text práce
├─ thesis.pdf .....	text práce ve formátu PDF