AKADEMIE MÚZICKÝCH UMĚNÍ V PRAZE

FILMOVÁ A TELEVIZNÍ FAKULTA

Filmové, televizní a fotografické umění a nová média

Obor zvuková tvorba

BAKALÁŘSKÁ PRÁCE

Pitch-shifting

a možnosti jeho využití v AV díle

Matěj Lindner

Vedoucí práce: MgA. Mgr. Petr Neubauer

Oponent práce: Doc. MgA. Pavel Kopecký

Datum obhajoby: 15.9.2022

Přidělovaný akademický titul: BcA.

Praha, 2022

ACADEMY OF PERFORMING ARTS IN PRAGUE

FILM AND TELEVISION FACULTY

Film, Television and Photographic Art and New Media

Department of sound

BACHELOR'S THESIS

Pitch-shifting

and possibilities of its use in audiovision

Matěj Lindner

Supervisor of the thesis: MgA. Mgr. Petr Neubauer

Opponent of the thesis: Doc. MgA. Pavel Kopecký

Date of thesis defence: 15.9.2022

Assigned academic degree: BcA.

Praha, 2022

P r o h l á š e n í

Prohlašuji, že jsem bakalářskou práci na téma

Pitch-shifting a možnosti jeho využití v AV díle

vypracoval samostatně pod odborným vedením vedoucího práce a s použitím uvedené literatury a pramenů.

Praha, dne ....
podpis diplomanta

Upozornění

Využití a společenské uplatnění výsledků diplomové práce, nebo jakékoliv nakládání s nimi je možné pouze na základě licenční smlouvy tj. souhlasu autora a AMU v Praze.

Evidenční list

Uživatel stvrzuje svým podpisem, že tuto práci použil pouze ke studijním účelům a prohlašuje, že jí vždy řádně uvede mezi použitými prameny.

Jméno	Instituce	Datum	Podpis

Abstrakt

Bakalářská práce s názvem *Pitch-shifting a možnosti jeho využití v AV díle* se zevrubně věnuje tomuto způsobu práce se zvukovým signálem. Nejprve stručně popíše základní nutné souvislosti úzce spojené s tématem, definuje samotný termín *pitch-shifting* a dále podrobně vysvětluje jeho technickou stránku. Druhá část práce je praktičtějšího charakteru, zkoumá historický vývoj *pitch-shiftingu* a následně mapuje a hledá současné možnosti jeho využití – především ve filmu, zároveň však i v dalších odvětvích audiovizuální tvorby, s čímž je spojeno i využití v hudbě. Prostřednictvím několika ukázek tyto různé možnosti využití pitch-shiftingu názorně demonstruje.

Abstract

The bachelor's thesis entitled *Pitch-shifting and its possibilities of use in audiovision* thoroughly describes this method of sound processing. Firstly, it briefly describes the essential context, then defines the term *pitch-shifting* itself and explains its technical side in detail. The second part of the thesis contains more of a practical character. It dives into the historical development of pitch-shifting and finally searches for the current possibilities of its use – especially in film, but as well in other branches of audiovision and music. In this concluding part, the thesis demonstrates various pitch-shifting possibilities on several examples.

Poděkování

Tímto bych chtěl poděkovat mému vedoucímu práce Petru Neubauerovi za jeho odborné vedení, podnětné připomínky a lidský přístup. Dále bych rád poděkoval mé rodině a mým blízkým za trpělivost a podporu při studiu.

OBSAH

1) Úvod

Jako téma své bakalářské práce jsem si zvolil Pitch-shifting a možnosti jeho využití v audiovizuální tvorbě zejména pro můj dlouhodobý zájem o úpravy zvukového signálu. Cílem práce je nejprve se přehledně zorientovat v samotné technologické problematice pitch-shiftingu, principiálně popsat jednotlivé metody jeho provedení, ale především následně prozkoumat možnosti jeho aplikovaného využití v audiovizuální tvorbě. Z toho důvodu práce nebude až tolik do podrobna matematicky rozebírat fungování algoritmů a procesů probíhajících v útrobách výpočetních zařízení, spíše se pokusí vystihnout jejich podstatu a tu následně ukázat na konkrétních příkladech, vypovídajících jak je možno tuto metodu úpravy zvuku použít v praxi.

Metodologicky práce čerpá z mnohých knižních či internetových zdrojů, z nichž některé jsou dnes již archivního charakteru. Přínos práce by měl spočívat v podání uceleného komplexního přehledu dané problematiky a v průzkumu možností aplikace tohoto nástroje v prostředí audiovize v současnosti.

Pitch-shifting bývá ve dnešní době zmiňován převážně v souvislosti s hudební produkcí a mnoho publikací či veřejných internetových blogů a diskuzí na něj nahlíží právě z této perspektivy. Tato bakalářská práce však na něj nahlíží spíše z perspektivy filmového prostředí zvukové postprodukce, čemuž bude odpovídat i zaměření ukázek v závěrečné kapitole.

2) Definice pojmu a vhled do podstatných zákonitostí

Pojem pitch-shifting představuje metody zpracování zvukového signálu, jejichž cílem je – jak již uvádí doslovný překlad tohoto termínu – posun výšky zpracovávaného zvuku. Obecně řečeno můžeme za pitch-shifting považovat jakoukoliv umělou změnu výšky zaznamenaného zvuku, je však nutné brát v potaz i ostatní vlastnosti zvuku jako časový průběh či jeho barva. Z hlediska DSP[1] se jedná o komplexní soubor různých operací, jež je nezbytné provést k dosažení uspokojivého výsledku. Tímto výsledkem pak je ve výšce změněný zvuk vycházející z charakteru původního zvukového signálu.

2.1 Zvuk, výška zvuku a jeho barva

Za zvuk se považuje každý kmitavý pohyb hmoty v různých skupenstvích, jenž v konečné podobě vyvolává sluchový vjem [1. str. 11]. Pro účel této práce však lépe poslouží definice zvuku jakožto podélného mechanického vlnění šířícího se vzduchem (či jiným médiem). Toto vlnění je vyvoláno právě kmitavým pohybem nějakého objektu – oscilátoru, jako jsou třeba hlasivky, struna či reproduktor. Tento kmitavý pohyb dále skrze vazebné síly mezi částicemi způsobuje vychylování nejbližších okolních částic, čímž vzniká zvuková vlna šířící se prostředím. Ta se dá následně charakterizovat pomocí různých fyzikálních veličin.

Pokud se kmitavý pohyb vyvolávající zvukovou vlnu opakuje v krátkých pravidelných intervalech, vzniká tzv. periodické vlnění, přičemž perioda (T) je doba tohoto jednoho opakování. S periodou je úzce spjatá další veličina – frekvence (f). Ta vypovídá o počtu těchto periodických opakování za jednotku času, konkrétně za jednu sekundu. Vztah mezi frekvencí a periodou je tedy $f=1/T$. Jednotkou frekvence je Hz (Hertz) a lidské sluchové ústrojí je schopné vnímat zvuky o frekvenci 16 až 20 000 Hz [1, str. 53]. A právě frekvence zvuku je pro pitch-

[1] Digital Signal Processing, překl. digitální zpracování signálu – matematické operace umožňující manipulaci s digitalizovaným signálem [31, str. 1]

shifting určující fyzikální veličinou – od ní je odvozena subjektivní veličina výška zvuku. Vzhledem k faktu, že již samotný název pitch-shifting v sobě obsahuje termín tzv. pitche neboli výšky zvuku, je nezbytné si jej podrobně vysvětlit.

Výška zvuku je subjektivní vjem lidského sluchového ústrojí, který úzce souvisí s frekvenčním spektrem zvukového signálu. Jednotkou výšky zvuku je 1 mel, přičemž pro frekvenci 1000 Hz odpovídá při hlasitosti 40 fonů[2] výška o hodnotě 1000 mel. [1, str. 61]

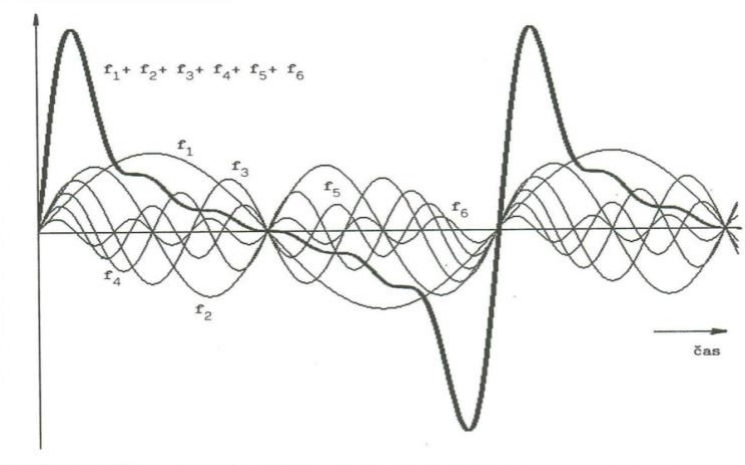
„Sluchový vjem výšky zvuku odráží periodické chování jeho časového průběhu. Pokud je tento průběh opravdu periodický, pak vjem výšky přímo odpovídá frekvenci vnímaného tónu. Avšak i u zvuků netónové povahy se hovoří o jejich výšce, která obdobně odráží náznaky periodicity časového průběhu, resp. výskyt odpovídajícího lokálního maxima či maxim ve frekvenčním spektru tohoto průběhu." [1, str. 60]

Je tedy důležité si uvědomit, že nejen periodické tóny, ale všechny druhy zvuků v sobě zahrnují frekvenční složky, a tedy i námi vnímanou výšku – ať už jde o rozezněnou strunu kytary, hluk motoru nastartovaného auta či třeba dopad kamene vhozeného do vody nebo prostě jen zvuk vydávaný při chůzi po schodech. A co je podstatné: lidská řeč je taktéž souborem znělých tonálních složek (samohlásky) a různých netonálních částic (souhlásky), což bude pro později popsané posouvání výšky řeči poměrně zásadní fakt (více v kapitole 2.2). Měli bychom tedy uvažovat, že u pitch-shiftingu se jedná o transformaci zvuku komplexně jako celku napříč jeho frekvenčním spektrem, nikoliv jen o posunutí výšky hudebních tónů (ačkoliv termíny „tón" a „zvuk" bývají v běžné řeči často zaměňovány). Vztah mezi subjektivní výškou zvuku a objektivní fyzikální veličinou frekvencí zvuku je přibližně logaritmický – rozdíl ve výšce mezi tóny o frekvencích 100 Hz a 160 Hz vnímáme jako stejně velký mezi tóny o frekvencích 1000 Hz a 1600 Hz. Záleží tedy na poměru frekvencí. Výška je lineární, zatímco frekvence exponenciální.

[2] Jednotka subjektivní hlasitosti zvuku odpovídající hladině akustického tlaku v závislosti na frekvenci

Ačkoliv výšku zvuku vnímáme převážně podle jeho jedné základní frekvence, zvuk jako celek se téměř vždy sestává i z dalších frekvencí – vyšších harmonických a neharmonických složek. Pokud uvažujeme hudební tóny, harmonické složky jsou násobkem jejich základní frekvence, tedy například pro tón G2 o základní frekvenci 196 Hz jsou harmonické frekvence 392 Hz, 588 Hz, 784 Hz a tak dále.

„Z hudebního hlediska představují tyto složky resp. tóny sled konkrétních intervalů. Mezi 1. a 2. harmonickou je to oktáva, 2. a 3. harmonickou kvinta, mezi 3. a 4. harmonickou kvarta, následuje velká tercie, malá tercie atd. Vhodnou volbou velikostí amplitud a fází jednotlivých harmonických kmitů je potom možno složit libovolný periodický průběh kmitání (obr. 2.2)." [1, str. 24]



Obr. 2.1: Součet řady harmonických vln [1, str. 24]

Sled těchto harmonických složek je pak matematicky vždy přesně dán násobkem fundamentální frekvence. Odychylky od těchto přesných násobků jsou pak takzvané neharmonické složky spektra. Jejich příkladem mohou být zvuky vydávané perkusivními nástroji, avšak i hudební nástroje, jež běžně vnímáme jako naprosto periodické, zahrnují neharmonické složky, které vznikají zejména nelineárním chováním oscilátorů v podobě strun, membrán, jazýčků apod. [1, str. 111]. Hlasitostní poměr mezi jednotlivými (ne)harmonickými složkami pak udává tzv. barvu zvuku – v případě lidské řeči se jedná o barvu hlasu.

Výška zvuku také do určité míry závisí nejen na jeho frekvenčním průběhu, ale i na jeho intenzitě. U frekvencí pod 1000 Hz vnímáme slaběji znějící tóny jako o něco málo vyšší (max. o 10 %) než ty hlasitější, ačkoliv jejich frekvence je stále stejná. V oblasti kolem 2000 Hz je výška na hlasitosti téměř nezávislá a v oblasti nad 2 kHz je tomu zas naopak – slabší tóny vnímáme jako hlubší než při silné intenzitě. Proto také mluvíme o výšce zvuku jako o subjektivní veličině. Za zmínku ještě v souvislosti s posunem výšky zvuku stojí uvést nejmenší sluchem rozlišitelný rozdíl mezi dvěma frekvencemi. Zde opět závisí, v jaké oblasti frekvenčního spektra se daný zvuk nachází. Nejpřesnější citlivost má lidský sluch v kolem 2 kHz, kde je schopen rozeznat rozdíl 0,1 % až 0,2 % dané frekvence, což při tomto kmitočtu odpovídá rozdílu 2 až 4 Hz. U nižších a vyšších kmitočtů je minimální rozpoznatelná změna výšky tónu vždy horší [1, str. 61].

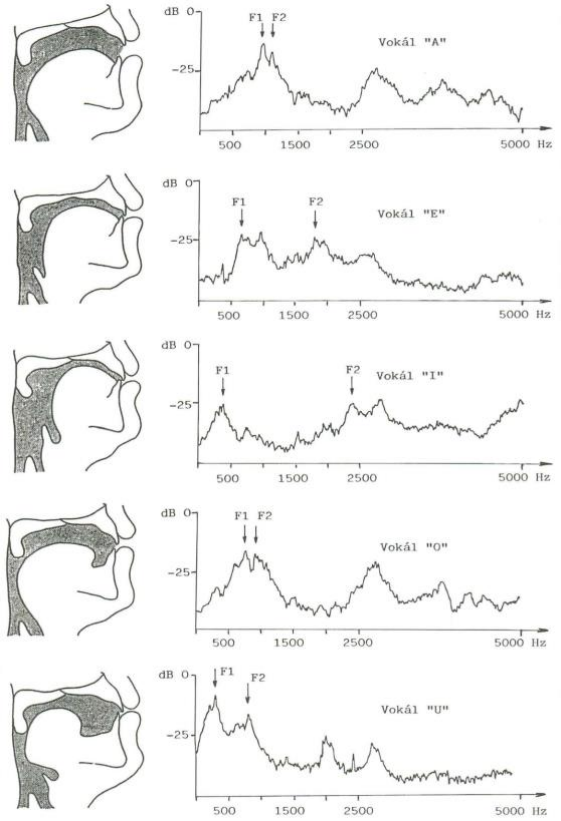
2.2 Akustika lidské řeči a hlasu

„A good understanding of human voice production is the starting point of improving voice and developing algorithms for voice and speech technology." [3, str.3] [3]

Lidská řeč je výsledkem proudění vzduchu, které vzniká pohybem dýchacích svalů. Dále se přes průdušky a průdušnice dostává ke hlasivkám, do nichž naráží, rozkmitává je, načež vzniká primární akustický kvaziperiodický signál. Ten putuje dále přes hrtan a hltan do ústní a nosní rezonanční dutiny, kde je dále tvarován a následně vypuštěn do okolního prostředí. Společně s tímto kvaziperiodickým signálem, který přísluší samohláskám – vokálům, vzniká vlivem turbulentního proudění vzduchu narážením do překážek (jazyk, zuby, rty) stochastický signál šumového charakteru. Ten označujeme jako souhlásky – konsonanty.

[3] *Překlad: Správné porozumění procesu vzniku lidského hlasu je výchozím bodem k vylepšení hlasu a vývoji algoritmů pro technologie hlasu a řeči.*

Samotný hlasivkový tón není o zcela stálé frekvenci – jemně kolísá v závislosti na emocionálním projevu řečníka. Jeho samotné frekvenční rozmezí (bez ovlivnění rezonančními dutinami) je v rozmezí 70 až 680 Hz. Hlasový šum tvořící souhlásky je pak širokopásmového charakteru s klesající úrovní 6 dB/okt. nad frekvenci 10 kHz a pod frekvenci 800 Hz. [1, str. 205]. Trubice vokálního traktu (mezi hlasivkami a ústy) je pak základní rezonanční oblastí a úpravou jejího tvaru dochází k modifikaci hlasového signálu, což má de facto úlohu pásmového filtru[4]. Vznikají tak jednotlivé hlásky, které mají charakteristické formanty – lokální maxima ve frekvenčním spektru. To je dáno právě úspořádáním tvaru a objemu artikulačního ústrojí vokálního traktu – výše posazení jazyka a míra otevření ústní dutiny. Určujicími formanty jsou poté zejména ty základní – F1 a F2, případně F3. Na níže uvedeném obrázku je možné vidět jednotlivé samohlásky a jejich přidružená frekvenční spektra právě se základními formanty.



Obr. 2.2: Vyobrazení formantů jednotlivých samohlásek, užito z [1, str. 207]

[4] Lineární filtr propouštějící signál jen o omezeném rozsahu frekvencí

Souhlásky můžeme rozdělit na znělé a neznělé. Je u nich obecně složitější určit jejich formantovou strukturu, nejsou natolik tónového charakteru. U znělých konsonantů (z, ž, c, č, h, v, d, ď, m, j) se poloha základního formantu pohybuje okolo 300 Hz, u sykavek (s, z) bývá maximum kolem 6 kHz, zatímco u plozivních souhlásek (b, p) bývají přítomny i velmi nízké frekvence dosahující níže než 100 Hz.

Výsledný řečový signál je tedy vlastně plynulým tokem vokálů různě přerušovaných konsonantami. Jde o sled znělých tonálních částic s různými formanty, které jsou střídány částicemi netonálního, téměř až ruchového charakteru. Celkově jde o frekvenčně poměrně komplexní akustický signál, jehož všechny složky jsou velice důležité. Formanty určují charakteristickou barvu hlasu mluvčího, frekvenční průběh vypovídá o intonaci mluvčího, střední a vyšší frekvence jsou zásadní pro srozumitelnost projevu, hlubší kmitočty zas vypovídají o blízkosti mluvčího (jakožto zdroje zvuku) k posluchači a podobně. Navíc různé národní jazyky fungují na základě mírně odlišných zákonitostí. Kvůli této komplexnosti a frekvenční bohatosti lidské řeči je její pitch-shifting poměrně náročnou záležitostí, neboť všechny zmíněné prvky hlasu by měly ideálně zůstat nezměněné, pochopitelně kromě výšky výsledného zpracovaného zvuku. Pro dosažení uspokojivějších výsledků pak některé algoritmy rozdělují signál na vokální a konsonantní složky zvlášť, načež manipulují jen s potřebnými částmi signálu. Podrobněji jsou tyto metody rozebrány v kapitole 3.

2.3 Digitalizace zvuku

Pitch-shifting se dnes v naprosté většině případů provádí v digitální doméně. Z toho důvodu je důležité si uvědomit, co pro pitch-shifting vlastně digitalizace zvukového signálu znamená, jaká mu to poskytuje východiska, benefity a zároveň i jistá omezení. Ve stručnosti zde proto budou popsány základní zákonitosti související právě s digitálním zpracováním zvukového signálu.

Zvuk je spojité vlnění a z pohledu časového dělení má nekonečný počet stavů. Aby bylo možné s takovýmto signálem pracovat ve dnešních digitálních výpočetních zařízeních, je nutné jej z analogové domény převést do domény diskrétní – tedy že se vybere jen konečný počet bodů v čase. K tomu slouží tzv. převodníky signálu.

Průběh signálu se dá graficky vyjádřit na dvou osách, přičemž na ose X sledujeme čas a na ose Y okamžitou amplitudu signálu. Pro přenesení spojitého analogového signálu na diskrétní digitální signál je třeba tyto osy rozdělit na konečný počet bodů, ve kterých je možné parametry signálu číselně vyjádřit. Časová osa X je rozdělena podle tzv. vzorkovací frekvence, která určuje počet vzorků za vteřinu. Její jednotkou je Hz. V každém diskrétním čase se uchová jedna hodnota amplitudy. Tato informace však taktéž musí být škálována diskrétně – jedná se o proces nazvaný kvantizace. Ta rozdělí osu Y na konečné množství hodnot, přičemž jejich počet bývá N-tá mocnina čísla 2 a signál pak lze vyjádřit v N bitech[5]. 8bitový signál například představuje 2^8 kvantizačních úrovní, tedy 256. V dnešní době bývá však zvuk zpracováván zpravidla ve 24 či 32 bitech, což umožňuje 16 777 216 respektive 4 294 967 296 jednotlivých kvantizačních úrovní.

Aby byl diskrétní zvukový signál jednoznačně reprezentován, je nutné zaznamenat dostatek jeho jednotlivých vzorků, tedy použít dostatečně vysokou vzorkovací frekvenci. Ta musí být – podle Nyquistova-Shannonova vzorkovacího teorému – alespoň dvojnásobná oproti nejvyšší frekvenci spojitého vzorkovaného signálu [32, str. 40]. Nyquistovou frekvencí pak označujeme polovinu hodnoty

[5] Základní, nejnižší jednotka dat, nabývající pouze hodnot 1, nebo 0

vzorkovací frekvence. A jelikož je běžný frekvenční rozsah lidského sluchu 16 Hz až 20 kHz, musí být pro převod signálu použita vzorkovací frekvence alespoň 40 kHz. Zároveň je však pro eliminaci jevu zvaného aliasing[6] nutno odfiltrovat vyšší frekvence než právě frekvenci Nyquistovu. Neexistuje ale žádný dokonale strmý filtr, který by byl schopen od určité hranice všechny vyšší frekvence odstranit a zároveň neovlivnit amplitudy frekvencí pod touto hranicí. Je tedy třeba nechat frekvenční rezervu mezi požadovanou vzorkovanou frekvencí a Nyquistovou frekvencí. I z tohoto důvodu se běžně užívají hodnoty vzorkovací frekvence 44,1 kHz, 48 kHz či více – vyšší vzorkovací frekvence jako 96 kHz či 192 kHz pak teoreticky umožňují ještě lepší možnosti práce se signálem i v rámci posunování výšky zvuku, což bude prozkoumáno v kapitole 5.

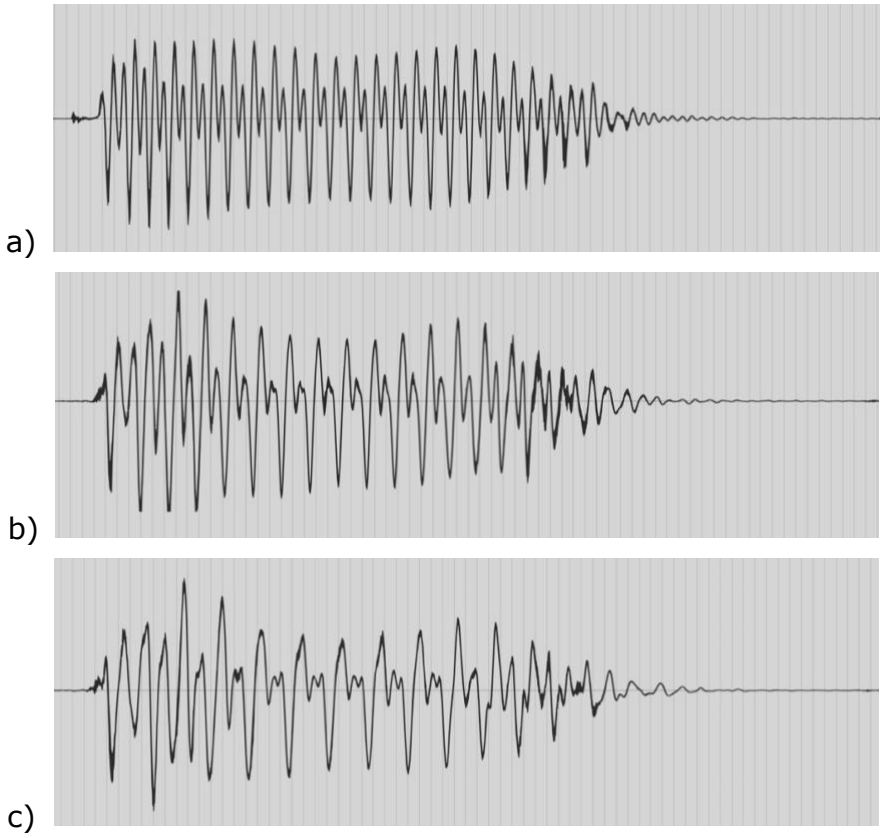
Výsledkem celého procesu digitalizace je tedy kvantovaný diskrétní signál o určité vzorkovací frekvenci, který je reprezentací původního spojitého analogového signálu a který je možné následně digitálně upravovat.

[6] Jev, který vzniká, pokud je vzorkovaný signál o vyšší frekvenci, než je použitá vzorkovací frekvence. Dochází k nepřesnému zaznamenání těchto vyšších frekvencí a tedy i k frekvenčním i fázovým nepřesnostem v konverzi signálu [32, str. 42]

3) Pitch-shifting a jeho technická stránka

V praxi je pitch-shifting vlastně sledem matematických operací, jejichž úkolem je změnit určité parametry signálu a jiné přitom zachovat. Záleží, čeho všeho při změnění výšky zvuku chceme dosáhnout. Existuje několik principiálně různých přístupů, kterak výšku zvuku měnit. Tyto přístupy pak zároveň mají odlišné možnosti použití pro rozdílné typy zpracovávaných signálů.

Změna výšky zvuku vždy znamená změnu periodického průběhu signálu. Pokud má být výsledný zvuk vyšší, musí mít vyšší fundamentální frekvenci, tudíž se perioda původního signálu zkrátí. A naopak, pokud má být nižší, periody musejí být delší a fundamentální frekvence tedy nižší než ta původní. Toto je patrné na obrázku 3.1, kde je možné zpozorovat proměnu v počtu opakujících se period signálu společně se změnou jeho výšky.



Obr. 3.1: hláska „í"
a) v původní verzi, b) posunutá o 6 půltónů níže, c) posunutá o 12 půtónů níže

3.1 Kritéria změny výšky zvuku

Kromě posunu fundamentální frekvence signálu je však třeba brát v potaz i další zvukové parametry dané jeho časovým průběhem – zejména vyšší harmonické frekvence komplexnějších signálů. Na obrázku 3.1 je možné zpozorovat, kterak se časový průběh signálu proměnil. Některé části period posunutých signálů mají jiný, složitější průběh. To značí o proměně charakteru zvuku, zejména jeho barvy. Dále je vidět celkem jasný výkyv v hlasitosti. Toto vše může být nežádoucí. Definuji zde tedy určité základní požadavky, jež by provedené posunutí výšky zvuku mělo naplnit.

1. *Kvalita výsledného zvuku*

   Jelikož je změna výšky zvuku velkým zásahem do původního signálu, dochází při ní mnohdy ke vzniku nežádoucích zvukových artefaktů. Pochopitelně čím větší změna, tím pravděpodobnější je jejich výskyt. Výsledný zvuk tak může působit nekvalitně, nepřirozeně či až roboticky „počítačově", což je ve většině případů nežádoucí a je tedy třeba se vzniku artefaktů či jiných forem zkreslení vyvarovat.

2. *Délka trvání výsledného zvuku*

   Pokud jde o změnu výšky mluveného slova, zpěvu či tónů hudebních nástrojů, obvykle chceme zachovat původní délku zvuku. Avšak například v oblasti sound-designu nám dává právě i zpomalování či zrychlování časového průběhu zvuku (tzv. time-stretching) společně s pitch-shiftingem mnohé kreativní možnosti a bývá naopak žádoucí. Pitch-shifting sám o sobě však spočívá pouze v posunutí výšky zvuku bez změny jeho trvání.

3. *Frekvenční složení výsledného zvuku*

   Pitch-shifting posouvá celé frekvenční spektrum původního signálu. Dalším parametrem při hovoření o výsledném změněném zvuku tedy bude způsob, jakým bude zacházeno s vyššími harmonickými frekvencemi, formanty

a celkově frekvenčním spektrem výsledného zvuku. Toto je klíčové právě pro mluvené slovo, avšak i hudební nástroje, které by mohly nevhodně provedeným posunem výšky zvuku ztratit svoji charakteristickou barvu a jiné zvukové vlastnosti.

Tyto 3 kritéria výsledného signálu je nutné vnímat současně. Dalším faktorem by mohla být výpočtová náročnost celého procesu a tím pádem i jeho výsledná rychlost a případně vzniklé zpoždění (latence), což hraje roli u zpracování v reálném čase. Tato práce se však zabývá především posunem výšky již zaznamenaného zvuku, kdy latence není takovým problémem. Výpočetní náročnost jednotlivých postupů nicméně alespoň naznačena bude.

3.2 Přístupy k dosažení změny výšky zvuku

Obecným principem, jenž charakterizuje vlastně všechny různé metody práce s výškou zvuku v digitální oblasti, je rozdělení původního (již diskrétního) signálu na množství jednotlivých dílčích částí, následná úprava těchto segmentů a posléze jejich spojení ve výsledný zvuk. Základním faktorem rozdělujícím jednotlivé přístupy k manipulaci s výškou zvuku je pak právě způsob rozdělení na tyto segmenty – a to sice jestli je se signálem pracováno v časové nebo frekvenční doméně. Práce se signálem v časové doméně je přímočařejší a ve své podstatě jednodušší, operuje přímo s navzorkovanými daty signálu. Nedosahuje však takových možností a kvalit, zvláště u složitějších signálů. Základním algoritmem zpracovávající zvukový signál v časové doméně je algoritmus SOLA (viz kapitolu 3.3.3), jeho modifikacemi pak vznikají precizněji fungující algoritmy PSOLA či WSOLA. Metody pitch-shiftingu pracující se signálem v doméně frekvenční pak fungují komplexněji, na bázi spektrální analýzy signálu. Ten je převeden do krátkých frekvenčně/amplitudových komponent, s nimiž je dále manipulováno. To celé umožňuje sofistikovanější práci s výškou zvuku. Výchozím algoritmem pitch-shiftingu ve frekvenční doméně je tzv. fázový vokodér (kapitola 3.4).

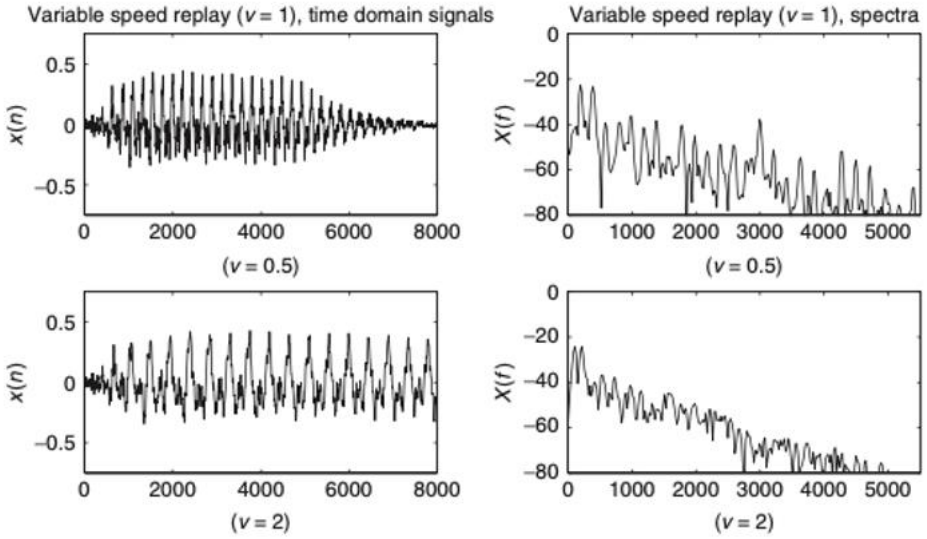
3.3 Metody manipulující se signálem v časové doméně

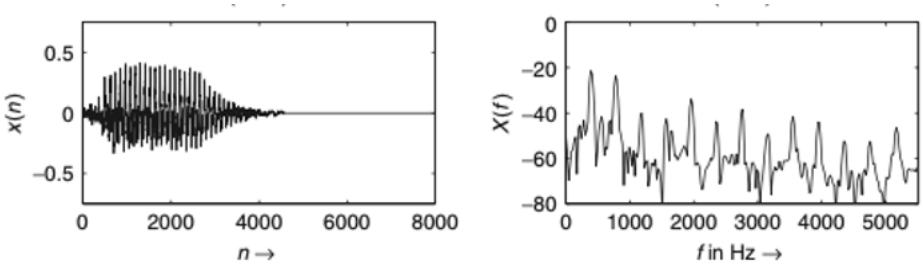
3.3.1 Proměnná rychlost přehrání

Již v době analogových přehrávačů zvukového záznamu (magnetofon, gramofon) bylo možné změnit výšku zvuku přehrávané nahrávky změnou rychlosti jejího přehrávání. Při rychlejším přehrání výška stoupne, při pomalejším zas klesne, což celkově pochopitelně promění i celkovou délku trvání nahrávky. Vysvětlení je fyzikálně prosté – pokud je nahrávka přehrávána rychleji, všechny kmitavé pohyby vyúsťující v reprodukovaný zvuk musejí taktéž proběhnout v kratším čase, tudíž bude jejich frekvence vyšší, a to stejnoměrně napříč spektrem. Tato zákonitost se dá popsat následujícími vztahy:

$$T_1 = T_0 / v$$
$$f_1 = f_0 \cdot v$$

kde T_0 a f_0 jsou původní doba přehrání a původní frekvence, T_1 a f_1 jsou výsledná doba přehrání a výsledné frekvence a v je relativní rychlost přehrání. Z toho vyplývá, že pokud bude $v > 1$, zvýší se i výsledná frekvence a naopak. Toto platí jak pro jednoduché, tak i komplexní signály, jelikož jsou úměrně zachovány poměry mezi jednotlivými složkami.

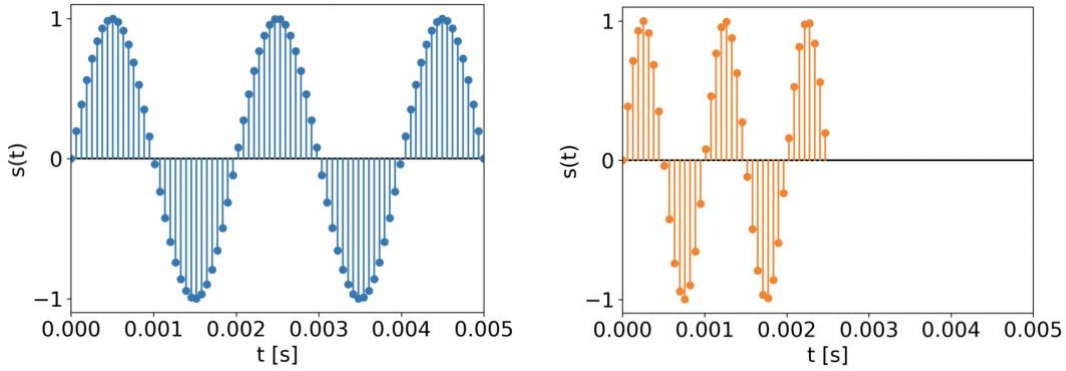




Obr. 3.2: Vyobrazení časového průběhu signálu společně s jeho spektrální obálkou při užití proměnné rychlosti jeho přehrání, užito z [2, str. 186]

3.3.2 Proměnná rychlost přehrání diskrétního signálu

Podobným principem jako u analogové změny rychlosti přehrání je možné se řídit i v oblasti digitalizovaného signálu – jeho převzorkováním. Pokud například chceme výsledný zvuk o oktávu zvýšit a tím pádem ho dvakrát zrychlit, musíme o polovinu snížit počet vzorků v signálu – každý druhý vzorek odstranit. Dojde tak ke „zhuštění" frekvence signálu (viz obr. 3.2) a tím, že přehrávací vzorkovací frekvence zůstane stejná jako původní, dojde v tomto případě ke zdvojnásobení výsledné frekvence signálu.



Obr. 3.3: Zhuštění frekvence signálu odstraněním poloviny vzorků

Tento proces (odstranění vzorků) se nazývá decimace a je nutné jej doplnit o dolnopropustní filtr, aby nemohlo dojít k aliasingu. Opačný proces, při kterém chceme zvuk snížit a zpomalit, vyžaduje naopak doplnění vzorků, což se vyřeší

tzv. interpolací[7] v poměru k variabilní rychlosti přehrání *v*. Klíčové je, že výsledný signál bude přehrán s původní vzorkovací frekvencí (např. 48 kHz), což vyústí v kýženou změnu frekvence zvuku, společně se změnou jeho délky trvání.

Tato metoda posunu výšky zvuku funguje jak pro jednoduché, tak i složité signály, nicméně její využití je poměrně omezené. Změna frekvence je totiž vždy podmíněna i změnou délky trvání zvuku. Taktéž tím, že některé samply jsou vynechány či duplikovány, dochází k nepřirozenému vyznění náhlých změn v signálu, například nějakých perkusních úderů s ostrým tranzientem. Navíc při posunu výšky zvuku nejsou zachovány formanty, spektrální obálka je posunována v celku, což má za důsledek značnou změnu charakteru výsledného zvuku. Při posunu výšky směrem níže pak zvuk působí dojmem, jako by ho pronesl někdo velmi veliký, zatímco při posunu směrem výše zvuk zní, jako by byl zdroj malinký. Při některých typech využití, například u sound-designu zpomalených zvuků, může být používání proměny rychlosti přehrávání poměrně užitečné (jak je prezentováno v kapitole 4.3.1), nicméně při práci s mluveným slovem tato metoda většinou není příliš praktická – zejména pro změnu délky trvání zvuku. Toto je řešeno pokročilejší skupinou algoritmů pracujících se signálem v časové (a frekvenční) doméně.

3.3.3 Algoritmus SOLA – Synchronnous Overlap and Add

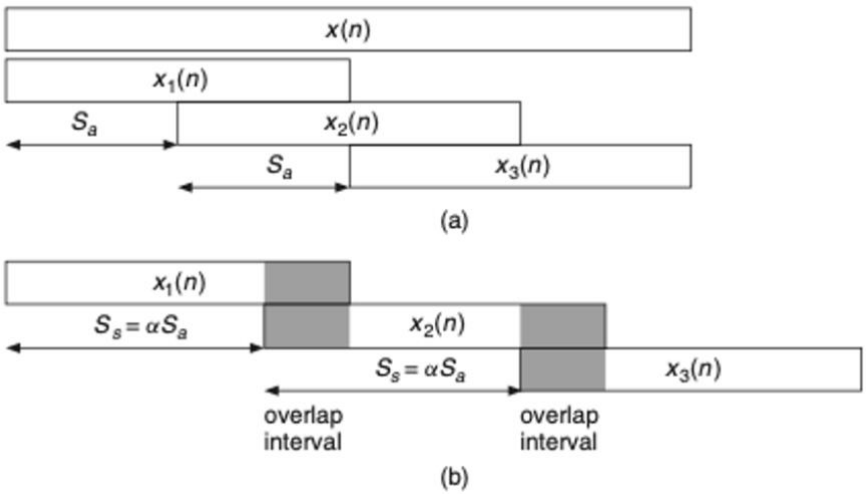
Níže popsané způsoby posunu výšky zvuku principiálně fungují na základě rozdělení diskrétního signálu na jednotlivé časově překrývající se úseky a jejich posunutí a následné spojení. Výhodou těchto algoritmů je, že umožňují pracovat jak s výškou zvuku, tak s jeho délkou trvání, a to poměrně nezávisle na sobě. Ve své podstatě totiž vycházejí ze základního algoritmu pro time-stretching [2, str. 191].

Algoritmus SOLA je základním výchozím bodem pro celou „rodinu" těchto algoritmů. Jeho princip je poměrně jednoduchý. Nejprve je signál v čase rozdělen na krátké překrývající se segmenty, jež jsou následně v čase posunuty

[7] Matematická operace spočívající v nalezení přibližné hodnoty funkce v rámci intervalu [33]

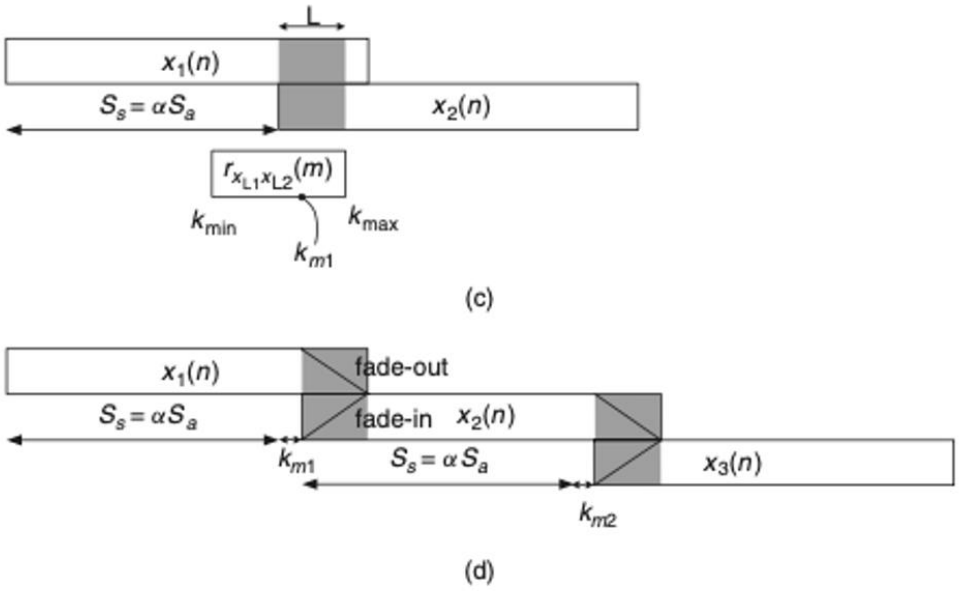
v požadovaném poměru (ať už směrem vpřed nebo zpět v čase). Poté proběhne vzájemná korelace[8] na sebe navazujících segmentů zohledňující časový posun – tím se úseky synchronizují, což zajišťuje nižší výskyt artefaktů. Nakonec se v těchto bodech maximální podobnosti jednotlivé překrývající segmenty prolnou křížovým prolnutím (viz obr. 3.4).

Výsledkem celého tohoto algoritmu je změněná délka trvání celkového zvuku bez posunutí jeho výšky. K právě aplikovanému time-stretchingu můžeme přidat pitch-shifting, pokud po aplikaci SOLA v poměru tedy N_2/N_1 signál převzorkujeme v poměru obráceném, tedy N_1/N_2 (obr. 3.5). Tím se časová proměna vzniklá vzorkováním vyrovná a výsledkem je pak zvuk o stejné délce trvání, avšak s odlišnou výškou. Ačkoliv je tento postup funkční i pro komplexní signály, stále přetrvává poměrně značný problém s artefakty a nezachováním formantů či tranzientů.



Obr. 3.4a): Schéma principu algoritmu SOLA
a) Rozdělení signálu na jednotlivé segmenty, b) posunutí bloků v čase
Užito z [2, str.192]

[8] Matematická operace umožňující porovnání dvou digitalizovaných signálů, zjišťuje jejich podobnost a periodicitu [4].



Obr. 3.4b): Schéma principu algoritmu SOLA

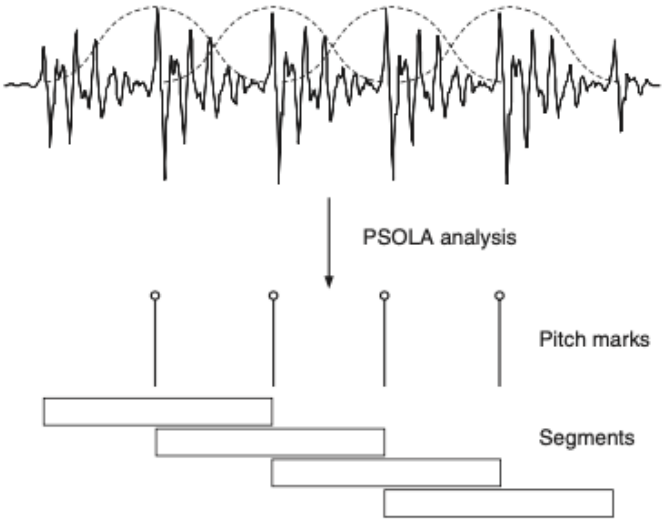
b) c) vzájemná korelace a synchronizace bloků, d) prolnutí segmentů (crossfade).

Užito z [2, str.192]

Obr. 3.5: Schéma provedení pitch-shiftingu společně s algoritmem SOLA spočívající v převzorkování algoritmem zpracovaného signálu [2, str. 203]

3.3.4 Algoritmus PSOLA – Pitch-synchronous overlap and add

PSOLA je ve svém principu algoritmus vycházející ze SOLA. Byl však navržen speciálně pro zpracování hlasového signálu, tudíž má napomoci zachování formantů a tím pádem i hlasového charakteru mluvčího. Oproti SOLA se liší hlavně provedením analýzy původního signálu – ta probíhá již v závislosti na jeho frekvenci (výšce) – algoritmus je schopný rozpoznat periodický průběh signálu a na jeho základě jej rozdělit do segmentů, čímž je umožněno segmenty následně přesněji napojit a zároveň jsou tak frekvence zpracovávány přesněji.



Obr. 3.6: Rozdělení signálu na segmenty podle jeho periodického průběhu, užito z [2, str. 195]

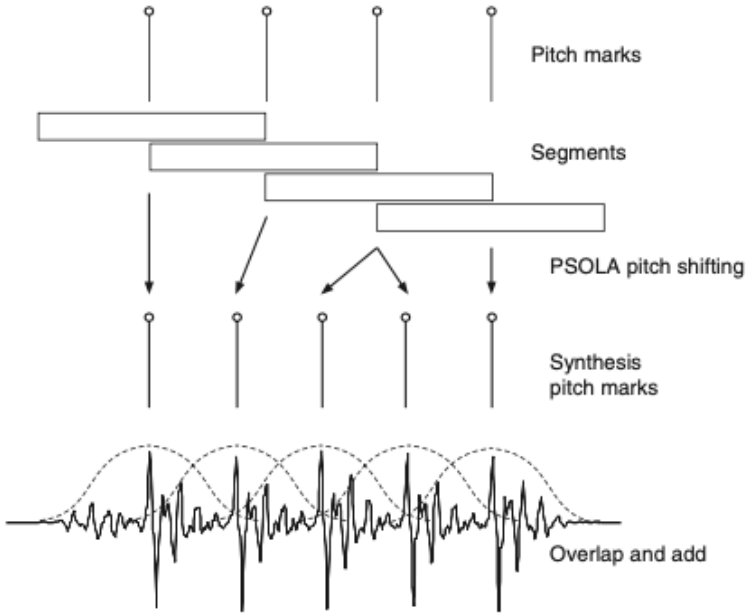
Jednotlivé segmenty jsou o délce 2T, zahrnují tedy 2 periody signálu, přičemž úsek je vycentrovaný na maximální amplitudu (tzv. pitch marks) v rámci jedné periody. Délky těchto period se v průběhu signálu pochopitelně mohou měnit, čemuž se přizpůsobí i délka segmentů. Úseky jsou tvarovány Hanningovým oknem[9] pro jejich následné plynulé navázání. Hlavní myšlenkou využití algoritmu PSOLA pro pitch-shifting je zachování charakteristických formantů zvuku. Toho je pak dosaženo posunutím pozice center segmentů (pitch marks) o požadovaný

[9] Druh okénkové funkce – amplitudově upravuje signál na malých úsecích podle tvaru okna

poměr frekvencí *β*, přičemž tvary vln (waveforms) v rámci jednotlivých segmentů jsou zachovány [2, str. 206].

„The basic idea consists of time stretching the position of pitch marks, while the segment waveform is not changed. The underlining signal model of speech production is a pulse train filtered by a time-varying filter corresponding to the vocal tract. The input segment corresponds to the filter impulse response and determines the formant position. Thus, it should not be modified. Conversely, the pitch mark distance determines the speech period, and thus should be modified accordingly." [2, str. 206][10]

Tím, že sled pitch marks determinuje výslednou frekvenci zvuku, tak jejich zhuštěním či zředěním dojde ke změně této výsledné frekvence, zachováním waveforem v rámci segmentů pak dojde k zachování charakteristických formantů. Následně probíhá syntéza segmetů.

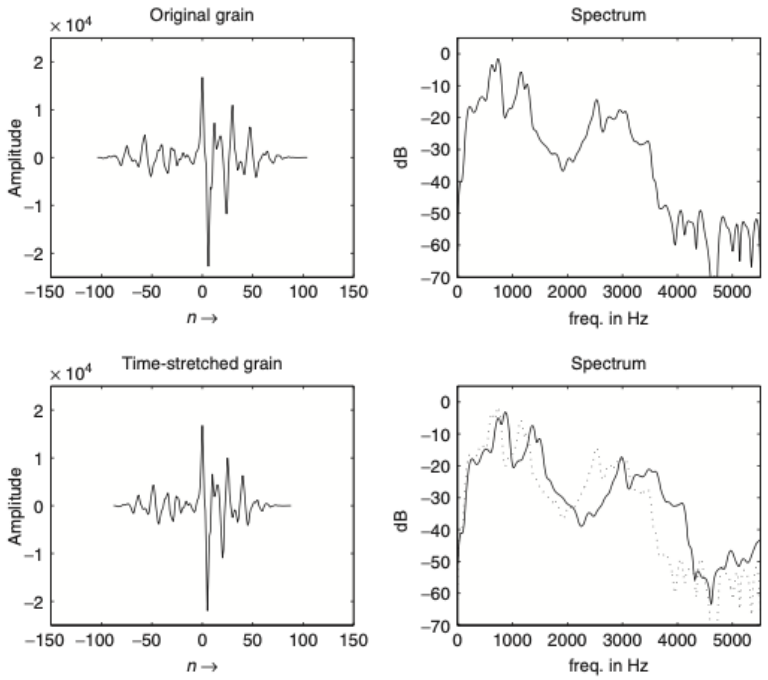


Obr. 3.7: Syntéza segmentů algoritmu PSOLA pro pitch-shifting

[10] *Překlad: Základní myšlenka se sestává z time-stretchingu pozic výškových značek (pitch marks), přičemž tvary vln v rámci jednotlivých úseků zůstávají neměnné. Základním prvkem produkce lidské řeči je proud dechu pozměněný různými časově-proměnnými filtry v podobě vokálního traktu. Jednotlivý vstupní úsek koresponduje s charakterem takovýchto filtrů a určuje tak pozici formantů. Následkem toho by neměl být pozměněn. Zároveň, vzdálenost výškových značek určuje periodu (a tedy i frekvenci) řeči, proto by měla být požadovaně změněna.*

Během syntézy dochází pro zachování celkové délky trvání zvuku k opakování (pokud $\beta > 1$, tudíž jde o zvýšení výšky) či vypuštění ($\beta < 1$) některých segmentů. Ty jsou následně přesně napojeny tak, aby byly zachovány lokální frekvence v oblastech prolnutí. A jelikož jsou segmenty tvarované podle Hanningova okna, je jejich napojení amplitudově plynulé, obdobně jako by bylo aplikované napojení pomocí křížového prolnutí.

Výsledkem tohoto procesu je zvukový signál o stejné délce trvání, se změněnou základní frekvencí, avšak s poměrně dobře zachovanými pozicemi vyšších složek harmonického spektra – formantů. S těmi je však v rámci užívání algoritmu PSOLA i dále pracovat – časovým přeškálováním jednotlivých segmentů před jejich syntézou. Tím může být dosaženo lineárního posunutí formantů, což má za efekt změnu charakteru mluvčího či interpreta (tzv. formant shifting) [2, str. 208].



Obr. 3.8: Posunutí formantů ve frekvenčním spektru užitím algoritmu PSOLA

Algoritmus PSOLA je užitečný především pro posun výšky lidské řeči či jednohlasého instrumentu. V této oblasti vede k uspokojivým a užitečným výsledkům. Ke zpracování polyfonních signálů však příliš vhodný není, a to pro jeho nutnost analýzy signálu podle jedné konkrétní základní frekvence. Zároveň pokud jde o signál s rychlými výraznými amplitudovými či frekvenčními změnami, dochází ke vzniku artefaktů či pocitu krátké ozvěny. Stejně tak dochází k jistým problémům během analýzy úseků signálu, které jsou šumového či jiného stochastického charakteru – nelze u něj jednoduše definovat jeho základní periodu. Toto se dá řešit například pokynem, aby takovéto úseky byly rozděleny prostě na stejně dlouhé segmenty nehledě na jejich vnitřní průběh, nicméně pro algoritmus toto může být složité přesně rozpoznat.

Dalším faktorem je výpočetní náročnost algoritmu PSOLA, která je sice relativně stále ještě nízká, nicméně vyžaduje předběžnou znalost průběhu signálu, což vyúsťuje v komplikovanější využití tohoto algoritmu pro posun výšky zvuku v reálném čase. V [5] je uvedena průměrná latence okolo 100 ms, nicméně s rozvojem výpočetní technologie se tato prodleva již snížila.

3.3.5 Další variace algoritmů OLA

Klíčovým úskalím pro výslednou efektivitu algoritmů pracujících s překrývajícími se segmenty v časové doméně je právě rozdělení signálu na tyto segmenty tak, aby respektovaly časový průběh zvuku a bylo možné je následně co nejlépe na sebe napojit. Zatímco SOLA ve své základní verzi segmenty dělí prostě jen rovnoměrně, nehledě na průběh signálu, PSOLA se snaží respektovat jeho periodický průběh a segmenty rozdělí podle jeho frekvence. Další metodu, kterak signál rozdělit na segmenty, přináší algoritmus WSOLA.

WSOLA je založený na tzv. „waveform similarity" neboli na podobnosti časového průběhu signálu. Než aby se snažil determinovat základní frekvenci signálu a té pak přizpůsobil i délku segmentů, WSOLA hledá dopředu pomocí posouvání fixní délky segmentu a vypočtení křížové korelace nejbližší možná místa

napojení. Detailnější fungování tohoto algoritmu je popsáno v [6]. Výsledkem je možnost pracovat s robustnějším signálem, bez nutnosti přesného odhadnutí základní frekvence signálu – tudíž je WSOLA užitečný pro polyfonní zvukové signály. Nicméně se však jedná o time-stretchingový algoritmus a při posunutí výšky zvuku za současného užití převzorkování a WSOLA dochází ke ztrátě formantů. Oproti základní SOLA je tento algoritmus však efektivnější. Dále existují i četné modifikace všech těchto algoritmů pracujících na bázi překrývajících se segmentů, které se snaží různými způsoby napravit jejich nedostatky či trochu jinak přistoupit k některým částem analýzy či syntézy.

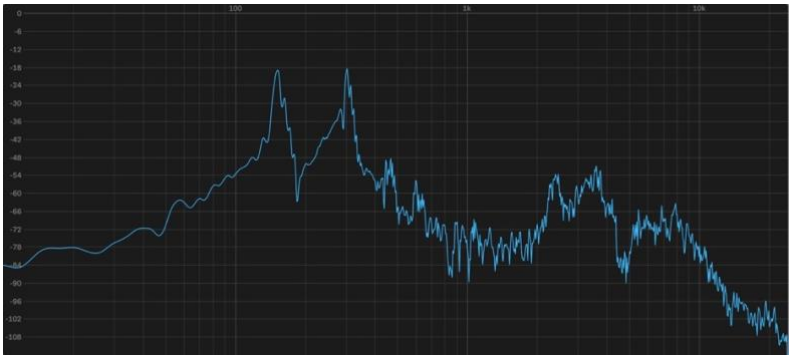
Obecně řečeno jsou metody pitch-shiftingu pracující se signálem v časové doméně poměrně efektivní, pokud jde o menší změny ve výšce zvuku, cca do 20 %, což odpovídá přibližně 3 půltónům. Při rozsáhlejších posunech již dochází ke vzniku artefaktů, nejčastěji v podobě krátkých ozvěn či dozvuků, zejména v oblastech, kdy v signálu dochází k náhlým změnám ve frekvenci či amplitudě, což je dáno složitostí napojení jednotlivých segmentů v těchto kritických momentech. Pokud je však kvalitně provedena analýza signálu v rámci PSOLA, mohou být tyto nežádoucí artefakty do značné míry potlačeny a výsledkem pak je kvalitní výsledný zvukový signál, jež může mít zároveň zachovanou svoji formantovou charakteristiku. Zároveň je možné i tyto formanty posouvat, pokud je to žádoucí.

3.4 Metody manipulující se signálem ve frekvenční doméně

„Every signal has a spectrum and is determined by its spectrum. You can analyze the signal either in the time (or spatial) domain or in the frequency domain." [7][11]

Na zvukový signál je tedy možno v zásadě nahlížet těmito dvěma různými pohledy. V předešlé kapitole je zvuk vnímán jakožto časový průběh měnící se amplitudy – je vidět jeho periodicita, z níž je možné vyčíst jeho základní frekvenci a do jisté míry se takto dá rozeznat i harmonické složení signálu – posoudit jeho tonalitu či atonalitu, odhadnout barvu zvuku (zejména ostrost/měkkost), v případě řeči zkušenější oko rozezná i jednotlivé samohlásky či souhlásky. Pokud však chceme o charakteru zvuku mít přesnější představu, je možné na něj nahlížet z pohledu frekvenčního spektra. To nám odhalí, jakou má daný zvuk v konkrétním čase frekvenční strukturu – základní frekvenci a rozložení vyšších harmonických složek, či případně šumové či jiné složky.

a)



b)

Obr. 3.9: frekvenční spektra a) hlásky *„í"*, b) sinusoidy o frekvenci 333 Hz

[11] *Překlad: Každý signál má své spektrum a je určený svým spektrem. Signál může být analyzován buď v časové anebo frekvenční doméně.*

Pokud by na zvukový signál takto dokázal nahlížet i nějaký algoritmus, nebylo by pak možné s jednotlivými frekvencemi manipulovat? Zhruba takto fungují procesy zpracovávající zvukový signál ve frekvenční doméně. Pomocí analýzy převedou zvukový signál postupně po částech na jeho frekvenční reprezentaci – kmitočtové složení. Následně jsou jeho jednotlivé složky modifikovány, načež je spektrum transformováno zpět do časového vyjádření průběhu, jednotlivé části se spojí a vznikne požadovaný výsledný zvuk. Tomuto celému procesu se v základě říká metoda fázového vokodéru. Jeho klíčovou součástí je právě frekvenční analýza signálu. Ta bývá provedena využitím matematické operace zvané Fourierova transformace.

3.4.1 Frekvenční analýza pomocí Fourierovy transformace (DFT, FFT)

Podle tzv. Fourierova teorému je možné jakoukoliv[12] funkci vyjádřit pomocí množství jednotlivých trigoniometrických funkcí sinus a cosinus [8, str. 2]. Toto množství je u diskrétních signálů taktéž konečné – každý vzorkovaný zvukový signál je tedy možno vyjádřit jakožto kombinaci jednotlivých jednoduchých sinusoid o různých frekvencích [9]. Pochopitelně čím komplexnější signál, tím větší množství sinusových signálů bude potřeba. Na základě tohoto teorému funguje tzv. Fourierova transformace, což je matematická operace převádějící signál mezi časově a frekvenčně závislým vyjádřením pomocí harmonických signálů, tj. funkcí sinus a cosinus.

Principem Fourierovy transformace je zjednodušeně řečeno porovnávání dílčích referenčních vzorových sinových signálů (o konkrétních frekvencích a amplitudě) s původním analyzovaným signálem. Pokud se daná frekvence v analyzovaném signálu nachází, je zaznamenána, společně se svojí amplitudou. Toto funguje na principu excitace referenční frekvence vstupním analyzovaným signálem. Jednotlivé frekvenční oblasti referenčních signálů jsou pak v běžné terminologii nazývány bins neboli „koše", do kterých jsou zahrnuty amplitudy

[12] V případě, že se jedná o neperiodickou funkci, je pro účel DFT považována za 1 periodu jakožto celek od počátku do konce.

a fáze daných frekvencí. (Tyto koše by se daly zjednodušeně přirovnat ke sledu velmi úzkých frekvenčních filtrů napříč spektrem, přičemž v každém je měřena amplituda.) U diskrétní Fourierovy transformace je pro každý koš užita jak sinová, tak cosinová referenční vlna, zároveň je pracováno i s jejich fázemi, aby bylo možné přesně analyzovat signály složené ze sinusoid o libovolných fázích. Výsledkem je reprezentace původního vstupního signálu jakožto množství sinusoid. Ty jsou definovány svojí frekvencí, magnitudou (amplituda) a fází. Jedná se tedy o magnitudové spektrum měřeného signálu.

Pro přesnost tohoto spektra je zásadní mít co nejjemnější rozlišení košů, ideálně aby pokrývaly celou šíři spektra. Na to by však byl potřeba nekonečný počet košů, což samozřejmě není možné. Navíc, tento počet košů je při diskrétní fourierově transformaci limitován další zákonitostí – délkou trvání vstupního signálu. Aby mohla být určena perioda (a tudíž i frekvence) referenční vlny konkrétního koše, musí se vměstnat do délky okna DFT. Pokud budeme provádět transformaci o hodnotě 1000 samplů, nejdelší hodnota periody koše může být taktéž o délce 1000 samplů. Toto při vzorkovací frekvenci 48 kHz odpovídá 20,83 ms, tudíž nejnižší frekvence koše je v tomto případě 1/0.02083 = 48 Hz. A jelikož další frekvence košů musí začínat a končit v bodě začátku a konce transformace, musejí tyto frekvence lineárně odpovídat násobkům této nejnižší frekvence. Hodnota frekvenčního rozlišení košů se tedy dá vyjádřit tímto vztahem:

$$f_b = f_s / N,$$

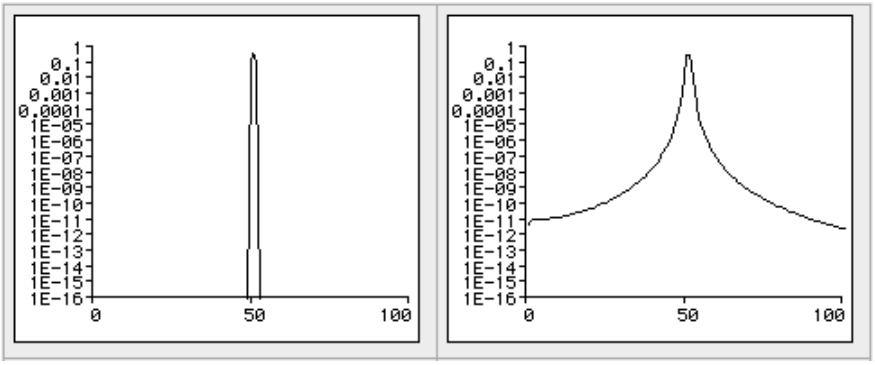
kde f_s je vzorkovací frekvence vstupního signálu a N je tzv. velikost Fourierovy transformace (kolik vzorků transformujeme). Zároveň z tohoto všeho vyplývá, že Fourierova transformace má za důsledek rozdělení spektra do fixní tabulky frekvenčních košů, jež nepokrývají celé spektrum – mezi koši jsou mezery. Rozlišení se sice dá zvýšit, ale na úkor prodloužení signálu (navýšení počtu samplů, jež transformujeme), tudíž i na úkor delšího časového záběru transformace, což může být nežádoucí. Existují i další, pokročilejší způsoby navýšení frekvenčního rozlišení DFT, navržené například v [10].

U Fourierovy transformace je podstatné, že se jedná o reverzibilní operaci. pokud byla její Fourierova transformace vypočítána z funkce, lze provedením inverzní transformace tuto původní funkci jednoznačně obnovit. Namísto této klasické DFT se spíše používá její rychlejší varianta – FFT neboli Fast Fourier Transformation. Jedná se o efektivnější algoritmus schopnější pracovat v podstatně kratším čase, obzvláště při větší hodnotě vzorků transformace. Výsledek FFT je však v zásadě identický jako u DFT. Tyto druhy Fourierovy transformace zpracovávají signál jako celek od začátku do konce a není tedy možné zjistit určité jeho spektrální složení v daném čase či sledovat jeho postupný vývoj. Pro účely pitch-shiftingu využitím metody fázového vokodéru se proto užívá ještě jeden speciální druh Fourierovy transformace.

3.4.2 Krátkodobá fourierova transformace – STFT

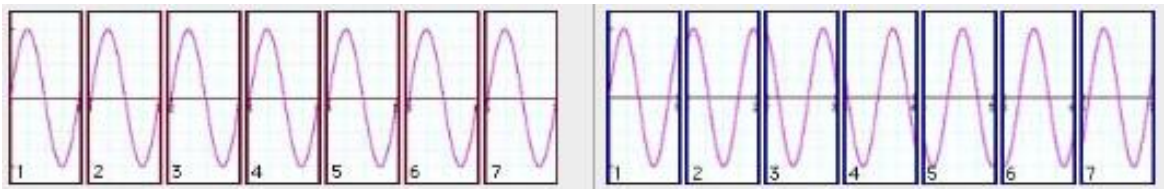
Jelikož je velká většina zvukových signálů takzvaně kvaziperiodická (sice se v čase mění, nicméně na krátkých úsecích bývá téměř dokonale periodická), je pro účely zpracování zvuku výhodné provádět DFT právě na takto krátkých úsecích neboli oknech. Tento způsob transformace se nazývá STFT – Short time Fourier transform.

Nastává zde však problém s frekvenčním rozlišením transformace. Pokud jde o krátké úseky o počtu vzorků N, z nichž je prováděna, vyplývá nám z výše zmíněné zákonitosti $f_b = f_s / N$, že výsledné rozlišení nebude příliš jemné. Při délce transformačního okna 2000 vzorků a vzorkovací frekvenci 48 kHz se jedná o rozlišení 24 Hz – jednotlivé koše jsou tedy v tomto případě 24 Hz od sebe. Je tak vytvořena umělá fixní tabulka frekvencí. I přes různá vylepšení frekvenčního rozlišení nebude nikdy zastoupena pomocí košů každá frekvence zvlášť. Důsledkem toho je fakt, že pokud se v analyzovaném signálu nachází frekvenční složka nezapadající přesně do některého z košů, její magnituda se rozmělní mezi sousedních několik košů – nejvíce přímo do těch přímo sousedících, avšak i do těch za nimi, jak vidno v obr. 3.8. A v praxi jen velmi málo frekvencí přesně zapadá do košů. Tento jev je nazýván spectral leakage neboli spektrální únik.



Obr. 3.10: Magnitudová spektra a) pokud frekvence spadá přesně do koše, b) pokud nespadá a její magnituda je rozmělněna do okolních košů [9]

Takovéto rozmělnění magnitud mezi okolní koše je zásadním úskalím implementace pitch-shiftingu pomocí krátkodobé Fourierovy analýzy. Dalším problémem je fakt, že frekvence nespadající přímo do košů budou mít v různých jednotlivých oknech STFT různé fáze, viz obr. 3.9.



Obr. 3.11: Schéma signálu rozděleného do jednotlivých oken STFT,
a) pokud frekvence přesně odpovídá frekvenci koše, b) pokud frekvence nespadá přesně do frekvence koše, je možné všimnout si posunutí fáze v rámci oken [9]

Řešením obou těchto úskalí krátkodobé Fourierovy transformace je aplikace okénkové funkce na jednotlivá okna signálu před vstupem do STFT a následný překryv těchto oken. Díky tomu je možné určit rozdíl ve fázi odlišných frekvencí a pomocí tohoto rozdílu pak určit pravé hodnoty těchto frekvencí. Při vyšším překryvu oken dochází k přesnější determinaci jednotlivých pravých frekvencí, z nichž je signál složen, ideální je pak překryv o délce 75 % okna. Podrobněji je problematika překrývání oken a vypočtení fázového rozdílu popsána v [11]. Výsledkem je získání relativně přesných frekvenčních komponent analyzovaného signálu.

3.4.3 Posunutí výšky zvuku ve frekvenční doméně

Když je signál pomocí překrývajících se oken STFT převeden na jeho jednotlivé frekvenční činitele, je pak posunutí výsledné výšky tohoto signálu poměrně jednoduchou záležitostí. Pokud není účelem zachovat charakteristické formanty zvukového signálu, stačí jednoduše magnitudy spektrálních komponent ve frekvenčním spektru posunout – při změně výšky v poměru 0.5 bude například původní sinus o hlasitosti -2 dB a frekvenci 1000 Hz posunut na frekvenci 500 Hz o stejné magnitudě -2 dB. A nápodobně pak se všemi jednotlivými komponenty. Je nutné vyhnout se aliasingu – při posunu směrem nahoru ve výsledku nezahrnovat frekvence vyšší, než je ta Nyquistova. V případě, že je žádoucí zachovat formanty původního zvukového signálu, je nutné jednotlivé frekvenční komponenty upravit podle jeho původní spektrální obálky, jíž je nejprve nutno co nejlépe odhadnout [12]. S poměrem zachování formantů je možno libovolně manipulovat.

Závěrečnou fází celého procesu pich-shiftingu pomocí metody fázového vokodéru je pak syntéza signálu – jeho přeměna z frekvenční reprezentace zpět do časové. Její postup je pak vlastně inverzní ku všem krokům, jež předcházely posunutí frekvenčních komponentů. Je zpětně vypočítán správný fázový rozdíl jednotlivých frekvencí, je provedena inverzní krátkodobá Fourierova transformace za použití okénkové funkce, následně jsou jednotlivá okna překryta a jejich součtem vznikne požadovaný zvukový signál s posunutou výškou. Během syntézy je možné i manipulovat s celkovou délkou trvání signálu – time stretching, jednoduše pomocí proměny napojení překrývajících se oken. [5]

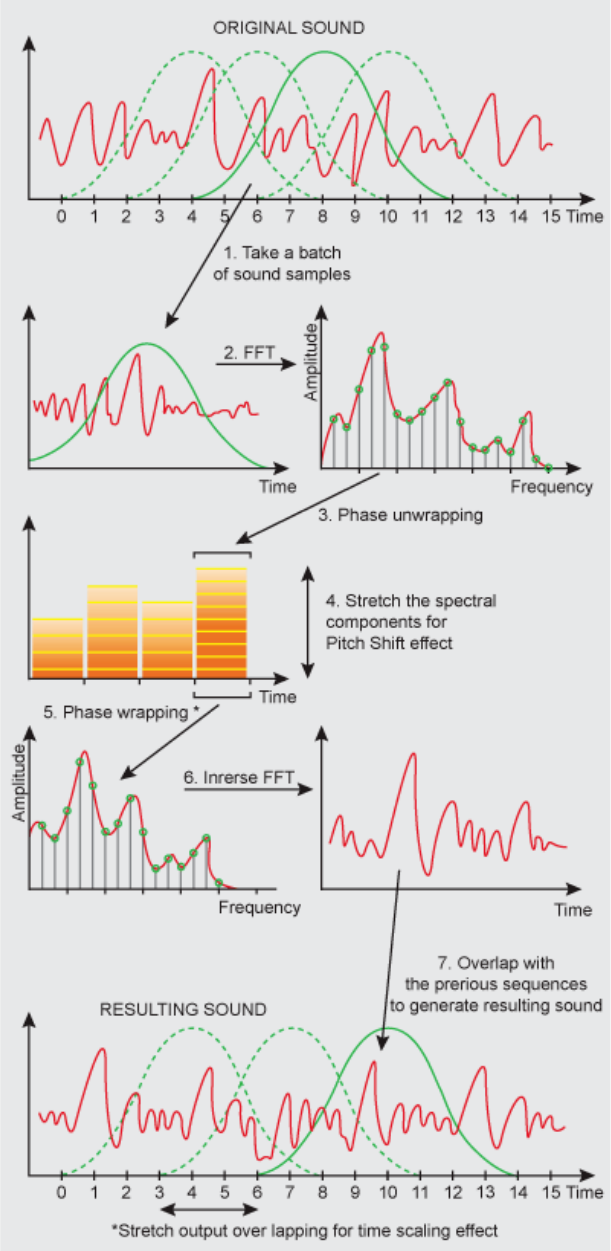
3.4.4 Shrnutí metody fázového vokodéru

Posunutí výšky zvuku pomocí metody fázového vokodéru je již poměrně náročnou výpočetní operací, jež však dokáže dosáhnout relativně kvalitních výsledků. Oproti OLA metodám zpracovávajícím signál v časové doméně nedochází ke vzniku artefaktů ozvěnového charakteru, je možné kvalitně zachovat formanty původního signálu a obecně je umožněna preciznější manipulace s jeho frekvenčními složkami. Další výhodou je i možnost pracování s polyfonními harmonickými signály, jelikož fázový vokodér dokáže právě tuto polyfonii zohlednit.

Nicméně ani tato metoda posunu výšky zvuku není dokonalá. Hlavním nedostatkem je vznik fázových zkreslení způsobujících „tupost" či „matnost, neostrost" výsledného zvuku, obzvláště u vyšších frekvencí. Příčinou těchto zkreslení je napojování jednotlivých transformovaných oken, což způsobuje změnu vztahů mezi fázemi odlišných frekvenčních složek [5,12]. Dále může docházet k celkové nerovnoměrné frekvenční odezvě celého procesu, kdy jsou kraje spektra značně utlumeny, což může být taktéž nežádoucí. Ostře znějící hudební nástroje či různé perkusní údery pak znějí méně kvalitně a nepřirozeně. Další nevýhodou této metody by pak mohla být i samotná její výpočetní náročnost, jež má za důsledek obtížnější využití v reálném čase. Nicméně s postupujícím rozvojem výpočetních technologií bych toto nepovažoval za příliš závažné.

Mnohé modifikace a vylepšení fázového vokodéru se snaží všechny tyto nedostatky eliminovat. Ty pokročilejší a v zásadě i nejefektivnější modifikace jsou již v dnešní době obestřeny obchodními tajemstvími. Na veřejnost konkrétněji často vyplynou až s jistým časovým odstupem, během něhož zas budou vyvinuty pokročilejší a efektivnější varianty algoritmů.

Celkové schéma posunutí výšky zvuku za využití metody fázového vokodéru tedy vypadá tak, jak je vyobrazeno na obr. 3.10. Všechny jednotlivé kroky jsou detailněji popsány výše.

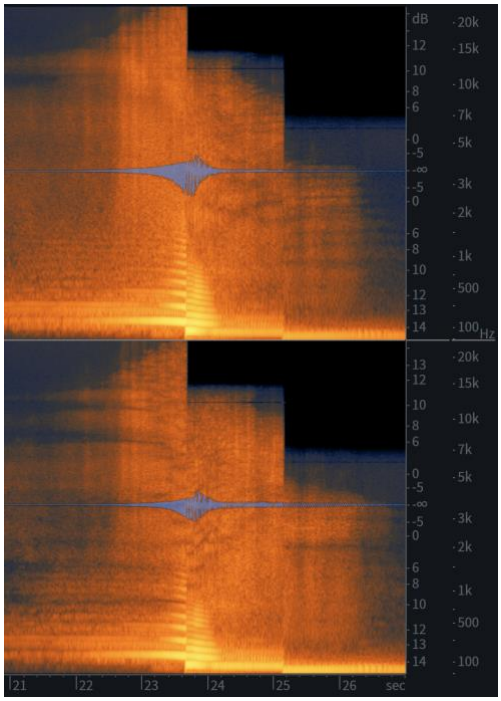


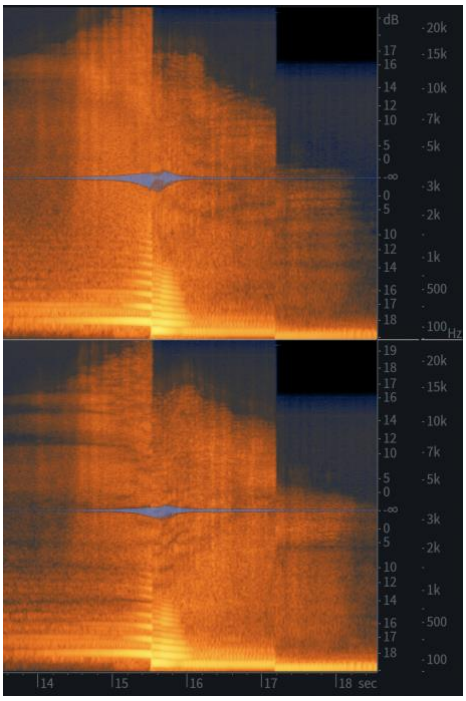
Obr. 3.12: Celkové schéma metody fázového vokodéru [5]

3.5 Rozlišení zvukového záznamu a jeho vliv na pitch-shifting

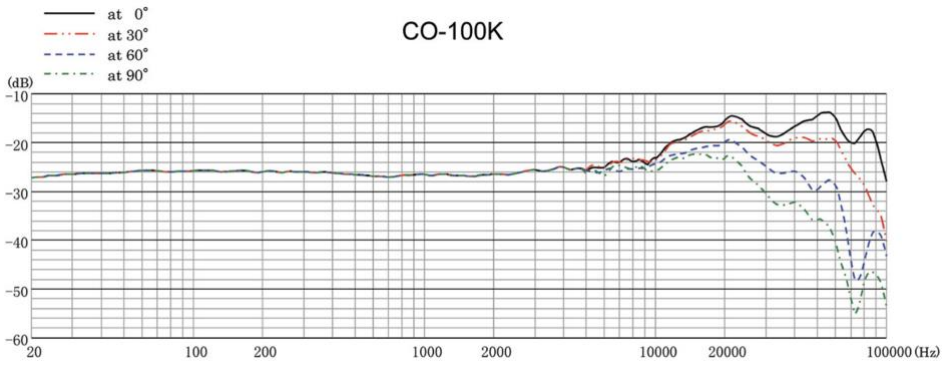
V důsledku posunu výšky zvuku směrem dolů dochází ke snížení horní hranice frekvenčního rozsahu. Zvukový záznam o vzorkovací frekvenci 48 kHz může podle Nyquistova teorému nabývat nejvyšší hodnoty frekvence 24 kHz (reálně je vlivem anti-alias filtru o trochu nižší). Pokud tedy dojde ke snížení zvuku o jednu oktávu, horní hranice nejvyšší frekvence je snížena na 12 kHz. Při snížení o dvě oktávy je to 6 kHz. Pokud by však byl transformován záznam o vzorkovací frekvenci dvojnásobné (96 kHz), i tyto horní frekvenční hranice posunutého zvuku budou dvojnásobné. Do výsledného zvuku tedy budou zahrnuty ultrasonické frekvence (pokud jsou v záznamu přítomny). Tato skutečnost je patrná na obrázku 3.11, kde si lze mezi jednotlivými grafy povšimnout odlišné horní frekvenční hranice signálu. Zatímco zvuk vyobrazený na grafu *a)* posunutý o dvě oktávy níže dosahuje nejvyšší hodnoty frekvence okolo 6 kHz, zvuk vyobrazený na grafu *b)* dosahuje při stejném posunu výšky frekvencí až ke 12 kHz. Rozdíl takto posunutých zvuků je možné slyšet ve zvukové ukázce *Z1*. Ačkoliv tento rozdíl není zcela markantní, transformovaná nahrávka s vyšší vzorkovací frekvenci dosahuje frekvenčně bohatšího výsledného zvuku. V praxi pak další vyšší vzorkovací frekvence (192 kHz) vyúsťují v ještě kvalitnější možnosti posunu výšky zvuku.

Důležitou součástí této problematiky je frekvenční charakteristika mikrofonu použitého pro záznam zvuku. Pokud mikrofon není schopen dobře zaznamenat kmitočty přesahující 20 kHz, nemůže být pro pitch-shifting plně využito potenciálu vyšších vzorkovacích frekvencí. V záznamu totiž žádné ultrasonické frekvence, které by mohly být posunuty do slyšitelného spektra, nebudou. V dnešní době již existují vysokofrekvenční mikrofony schopné zaznamenat frekvence až 100 kHz. Takovýmto zařízením je například mikrofon Sanken CO-100K.

a) 

b) 

Obr. 3.13: Spektrogram zobrazující posunutí horní frekvenční hranice při pitch-shiftingu
a) při užití vzorkovací frekvence 48 kHz
b) při užití vzorkovací frekvence 96 kHz



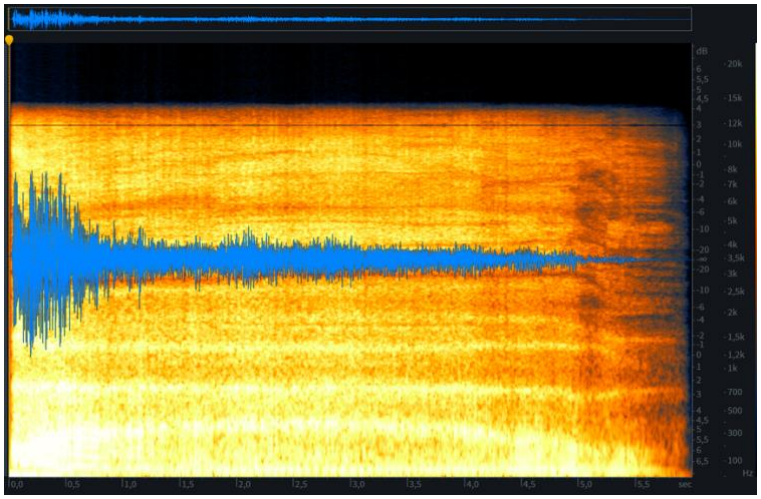
Obr. 3.14: Frekvenční charakteristika mikrofonu Sanken CO-100K [30]

Pokud je zvuk mikrofonem tohoto typu zaznamenán o vzorkovací frekvenci 192 kHz, horní frekvenční hranice záznamu dosahuje 96 kHz. Zvuk je tedy možné o dvě oktávy snížit a zachovat přitom plně frekvenčně zaplněné slyšitelné spektrum. Toto může být u posouvání výšky některých ruchů obzvláště užitečné. Obecně řečeno jsou jakékoliv mikrofony schopné dobře zaznamenávat frekvence

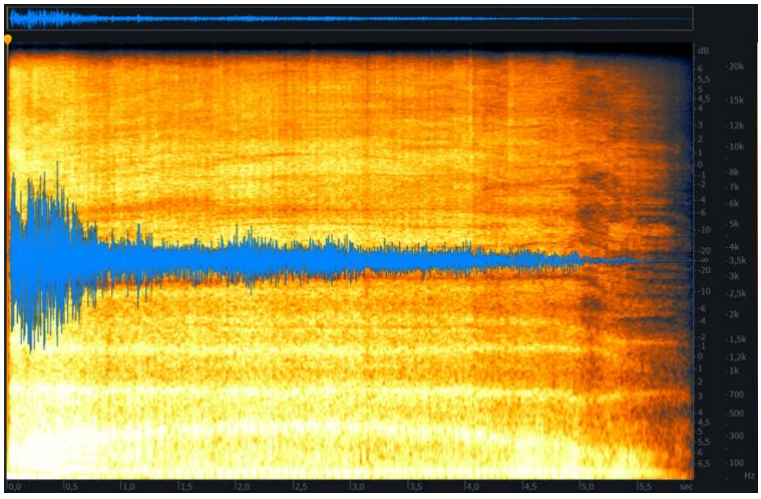
vyšší než 20 kHz v kombinaci s vyššími vzorkovacími frekvencemi velmi užitečné pro účely posouvání výšky zvuku. Mezi takovéto mikrofony patří například Sennheiser MKH8040 či MKH8050.

Ukázka *Z2* demonstruje rozdíl mezi použitím vzorkovací frekvence 48 kHz a 96 kHz na stejném zvuku při jeho snížení o jednu oktávu.

a)



b)



Obr. 3.15: Zvuky z ukázky Z2, oba snížené o 1 oktávu
a) při původní vzorkovací frekvenci 48 kHz, b) při pův. vzorkovací frekvenci 96 kHz

4) Využití pitch-shiftingu

Zatímco v předchozích kapitolách bylo popsáno, kterak je možné dosáhnout posunu výšky zvuku po teoretické stránce za užití několika různých algoritmů, v této kapitole bude ukázáno, k čemu a jak bývá pitch-shifting aplikován prakticky. V dnešní době se jedná o hojně využívanou metodu zpracování zvukového signálu napříč všemi různými odvětvími zvukové produkce. Jelikož je však tato práce zaměřena především na oblast audiovize, bude posléze uvažováno o využití posouvání výšky zvuku zejména v tomto odvětví.

4.1 K čemu je posunutí výšky zvuku?

Základní podstatou využití jakékoliv formy úpravy zvukového signálu je následující otázka: K čemu vlastně tato daná úprava slouží? Čeho je díky ní možné dosáhnout? Jaké možnosti jejím využitím získáme? Pokud se budeme bavit o podstatě prostého vícepásmového filtru, je možné tvrdit, že slouží například k odstranění nežádoucích frekvencí obsažených ve zvukovém signálu či třeba ke změnění určitých frekvenčních poměrů signálu za účelem jeho frekvenčního zkreslení a dosažení tak např. efektu zvuku vycházejícího z mobilního telefonu. Nápodobně dozvukový procesor slouží k vytvoření efektu prostoru, v němž se zdroj zvuku náchází a dynamický kompresor ve svých podobách zas ke kontrole poměrů hlasitosti výsledného signálu. Jak je tomu tedy u pitch-shiftingu?

Posunování výšky zvuku zahrnuje takovýchto možností využití hned několik, tyto možnosti se navíc mezi sebou značně liší. Pitch-shifting je nástroj komplexní, jedná se o celkovou transformaci zvukového signálu v mnoha různých proměnných aspektech. Jeho podstata využití může poté být jednak čistě tvůrčího charakteru (kdy je možno se zvukovým signálem manipulovat za účelem dosažení pozměněného nového signálu) a jednak charakteru opravného, kdy je provedena menší změna v původním signálu za účelem jeho korektury.

Do první kategorie (té kreativní, tvůrčí), můžeme pak řadit například využití pitch-shiftingu v rámci sound designu – pro tvorbu nových zvuků či modifikaci těch

již zaznamenaných, upravení lidského hlasu za účelem dosažení proměny charakteru mluvčího či obecně pozměnění vlastností zdroje zvuku, který ve výsledku zní výše nebo hlouběji.

Co se týče opravných možností posunu výšky zvuku, zde bývá nejtypičtějším příkladem jeho využití pro korekce nepřesně intonujícího hlasového projevu zpěvačky či zpěváka nebo pro dodatečné opravení nedokonale naladěného hudebního nástroje. Nemusí se však jednat pouze o využití v oblasti hudby, pomocí pitch-shiftingu je možné i pozměnit intonaci mluvčího ať už při pouhém přednesu nějakého textu, tak i třeba během jeho dialogových replik v rámci audiovizuálního díla. Stejně tak je možno i podobně pracovat s ruchy či atmosférami, tudíž posouvání výšky zvuku nabízí opravdu široké možnosti v rámci filmové zvukové postprodukce – což má tato práce za účel zpracovat. Nejprve však představím využití pitch-shiftingu v rámci samotné hudební produkce – hudební složka bývá často důležitou součástí audiovizuálních děl a je možné na ní demonstrovat základní možnosti využití posunutí výšky zvuku.

4.2 Využití v rámci hudební produkce

4.2.1 Analogové metody posunu výšky zvuku

Historicky sahá užívání pitch-shiftingu již do dob raných 50. let 20. století, kdy bylo možné měnit výšku zvuku společně s délkou jeho trvání pomocí proměnných rychlostí čtecích či přehrávacích hlav magnetofonů jak při nahrávání, tak zpětném přehrávání pásky.

Zmíněný efekt může být slyšen například ve skladbě Lese Paula „*Whispering*" již z roku 1951 [2, str. 188]. Postupem času byla tato technika zdokonalována a nadále hojně využívána. Mezi nejznámější příklady jejího využití patří nahrávací proces mnohých písní skupiny Beatles – konkrétně třeba skladby *Strawberry Fields Forever*. S touto nahrávkou kapela dosti bojovala a mnohokráte měnila její instrumentaci. Nakonec vzniklo 26 verzí, přičemž Johnu Lennonovi se sice nejvíce líbila ta finální šestadvacátá, nicméně začátek preferoval ze starší a jemnější sedmé verze. Požadoval pak po zvukových inženýrech Georgi Martinovi a Geoffu Emerickovi, aby tyto verze napojili, avšak problém byl, že měly jak jiné tempo, tak tóninu – rozdíl byl o jeden půltón. Technikům se to nakonec podařilo – pomocí proměnné rychlosti přehrávání pásek dokázali tempo obou částí přiblížit k sobě a kritické místo napojení po harmonické stránce vyřešit postupným snížením výšky zvuku první části. Tím sice dochází i ke zpomalení v dané oblasti a následnému zrychlení v napojené části, nicméně harmonické kontinuity bylo poměrně dobře dosaženo a místo napojení je vhodně zvoleno tak, že naprostá většina posluchačů jej nerozpozná. Pokud je však posluchač dostatečně pozorný, může napojení zaznamenat v 60. vteřině nahrávky [15] (*ukázka Z3 – v čase 20s*).

Dalším známým příkladem využití práce s proměnnou rychlostí pásky – a tentokráte nikoliv pro opravné, nýbrž tvůrčí účely – je píseň *Fame* od Davida Bowieho, vydaná roku 1975 v rámci alba Young Americans. V její závěrečné části je možné slyšet, kterak tato metoda úpravy výšky hlasu může znít při správném vypočítání poměrů mezi rychlostí záznamu a požadovaného výsledného tempa i na dnešní poměry vskutku kvalitně, čistě a bez artefaktů (*ukázka Z4*).

Mimo to v 50. letech navrhl Jacques Poulin ve společnosti s Pierrem Schaefferem zařízení zvané Phonogène Universal, jež pomocí rotující soustavy hlav dokázalo měnit výšku zvuku nezávisle na jeho délce trvání [13]. Efekt využívající principu rotačních hlav pak může být slyšen například v nahrávce *She's Going Bald* z roku 1967 od skupiny Beach Boys v čase od 0:50 dále (*ukázka Z5*).

Obr. 4.1: Zařízení Phonogène [13]

4.2.2 Vývoj digitálních technologií pro pitch-shifting

V první polovině 70. letech, společně s postupným rozvojem digitálních technologií, bylo společností Eventide představilo zařízení zvané *Harmonizer H910*. Jednalo se o první komerčně dostupnou digitální efektovou jednotku. Kombinovala několik dílčích efektových možností spočívajících v kombinaci zpožďovacích linek (delay, reverb, echo) a posunutí výšky zvuku. Bylo tak možné dosáhnout široké škály dosud neslýchaných zvukových transformací. Co se samotného pitch-shiftingu týče, zařízení umožnilo posun výšky v rozsahu dvou oktáv, avšak společně s utvořením charakteristického ozvěnového či fázového efektu. Jakožto

obohacení zvukové nahrávky o další harmonické prvky však tato efektová jednotka sloužila na danou dobu dobře, zněla neotřele, a tak se stala v druhé polovině 70. let a později v raných letech 80. značně populární. Utvořila tak charakteristický zvuk mnohých nahrávek. Obzvláště patrné je pak použití funkce pitch-shiftingu u bicí soupravy, konkrétně malého bubínku[13], v albu Low od Davida Bowieho, jež je první komerční nahrávkou silně využívající zařízení Harmonizer [14]. Jak je možné slyšet ve skladbě *Sound and Vision*, zvuk tohoto bubnu je hlubší, a dokonce se v době trvání úderu jeho výška snižuje (*ukázka Z6*). Z dnešního pohledu je tato původní verze Harmonizeru již značně překonanou technologií, jelikož artefakty při jejím použití byly velmi výrazné. Společnost Eventide však technologii dále postupně modifikovala a vylepšovala. Nová verze představená roku 1979 – *Harmonizer H949* – částečně snižovala množství vzniklých artefaktů, umožňovala mírnější změny ve výšce zvuku. V tomto smyslu byla vylepšena i následující *Eventide H3000 Factory* z roku 1986. Podobně se snažily možnosti digitálních efektů včetně pitch-shiftingu rozvíjet mnohé další společnosti, jejichž zařízení se pak nazývaly přímo Pitch-Shifter anebo Pitch Transposer. Obecně se po dlouhou dobu jednalo o zpracování zvuku v časové doméně, tato zařízení byla i koncipována pro jejich využití v reálném čase. Při změnách výšky v rámci půltónu či tónu byl výsledkem poměrně kvalitní zvuk, nicméně s větší změnou rostla přítomnost artefaktů (viz kapitolu 3).

Společně s Harmonizery a Pitch-shiftery se nadále vyvíjela i technologie využívající práci s tzv. samply, tedy nahranými zvuky či zvukovými elementy, jež lze následně pomocí klaviatury přehrávat v ruzných výškách. Průkopníkem této technologie byl *Fairlight CMI* – zkratka pro *Computer Musical Instrument* [16]. To byl vlastně takový předchůdce pozdějších DAW[14]. Umožňoval digitálně zaznamenávat signál, na displeji vyobrazit jeho spektrum a následně s ním dále manipulovat a přehrávat jej. Jedním z prvních interpretů využívajících tuto nákladnou technologii byl Peter Gabriel, posléze se zakomponování samplů stalo běžnou součástí hudební produkce. Podobným zařízením umožňujícím práci se

[13] Jinými slovy virbl, snare, šroťák nebo pochoďák
[14] Digital Audio Workstation – software umožňující práci se zvukovým záznamem

zvukovými samply byl Synclavier, jenž později našel četná využití i v oblasti filmového sound designu.

Další oblastí pitch-shiftingu, jež se postupně s rozmachem digitálních technologií rozvíjela, byla intonační korekce, tedy pojem v běžné řeči známější pod termínem auto-tune. V roce 1997, kdy již byl studiovým standardem DAW Pro Tools, byl uveden na trh společností Antares plug-in[15] zvaný Antares Auto-Tune, jehož původním účelem byla drobná korektura nepřesností v intonaci interpreta či hudebního nástroje.

„Systém analyzuje zvukový signál a hledá základní frekvenci monofonického – tedy jednohlasého – signálu samostatné melodie. Následně se ji snaží doladit na nejbližší tón odpovídající zvolenému ladění." [17]

Tento efekt byl následně zpopularizován popovou skladbou *Believe* od Cher z roku 1998. Ta byla první komerční popovou písní, jež využila efekt posunutí výšky hlasu nepřirozeně očividným, tvůrčím způsobem, a díky své nesmírné popularitě rozvinula trend, který přetrvává v mnohých hudebních žánrech až dodnes (*ukázka Z7*).

[15] Doplněk rozšiřující jakýmkoliv způsobem možnosti základního softwaru [34].

4.2.3 Současná využití pitch-shiftingu v hudební produkci

V současné době jsou možnosti využití posunu výšky zvuku v rámci hudební produkce velmi široké a bývají používány ve velké většině dnes produkované hudby. Nejzákladnější je využití pitch-shiftingu jakožto korekce intonačních nepřesností, ať už hlasu anebo hudebního instrumentu. K tomu slouží mnohé pokročilé nástroje, mezi nejznámější patří již výše zmíněný Antares Auto-Tune či Celemony Melodyne. Většina dnešních DAW (např. Cubase, Logic, FruityLoops Studio) má zároveň již zabudované možnosti upravování výšky zvuku. V dnešní době bývá tzv. auto-tune i často užívaný v reálném čase při živé hudební produkci.

Kromě opravy intonace může však pitch-shifting v rámci hudby být využit i tvůrčím způsobem. Jednoho ze základních efektů je možné dosáhnout pomocí proměnného mírného posunutí výšky zpožděné linky signálu a její následné kombinace s původním signálem. Tento efekt je nazván *Chorus* a jeho výsledkem je obohacený signál typického charakteru. Dále může být změna výšky zvuku samozřejmě využita k proměně hlasového charakteru zpěvačky nebo zpěváka. V souvislosti s pitch-shiftingem bývá často zmiňován i tzv. formant shifting, jenž bývá taktéž hojně využíván k proměně hlasu interpreta, nejčastěji v prostředí rapu, avšak i v jiných žánrech. Zde však nedochází k posunu fundamentální výšky hlasu, nýbrž k posunutí pozic jeho formantů, tudíž je výsledkem efekt změny pocitu velikosti zpěváka při zachování jeho intonace.

Krom toho je možné díky posunu výšky zvuku dosáhnout nových, zajímavých zvuků hudebních nástrojů, využívat zvuky z reálného světa jakožto tzv. samply a dále při kompozici pracovat s jejich výškou, změnit hlas zpěváka nebo zpěvačky zcela k nepoznání, obohatit ho o mnohé další harmonicky posunuté složky a dosáhnout tak efektu sboru a tak dále. V oblasti hudební produkce jsou tedy možnosti využití pitch-shiftingu prakticky neomezené, záleží na fantazii a záměru daného autora.

4.3 Využití pitch-shiftingu v audiovizi

Na rozdíl od čistě hudební tvorby bývá zvuková složka audiovizuálních děl úzce spjatá se složkou obrazovou – nestojí tedy sama o sobě, je o něco více svázána a nucena spolupracovat v rámci širšího celku. Poté tedy pochopitelně záleží na typu či žánru daného audiovizuálního díla. Je samozřejmé, že například experimentální psychedelická audiovizuální instalace nabízí bohatší využití různých zvukových prostředků nežli v realitě ukotvené sociální drama. Obecně řečeno – každé odvětví audiovize a každý žánr kinematografie má svá jistá specifika v oblasti zvukové dramaturgie. Možnosti využití pitch-shiftingu v rámci audiovize se tedy budou taktéž odvíjet od těchto specifik. Dále bude pro účely práce pojem audiovize zúžen zejména na oblast kinematografie a její různá odvětví.

Zvukovou složku filmu je možné rozdělit na její jednotlivé dílčí složky – nejčastěji mluvené slovo, ruchy a atmosféry a hudbu. Možnosti využití pitch-shiftingu v rámci hudby byly již prozkoumány v předchozí kapitole, nyní bude tedy pozornost věnována zejména různým formám mluveného slova a ruchům/atmosférám. Obdobně jako je tomu u hudební produkce, i zde se dá pitch-shifting využít jak pro tvůrčí účely, tak pro účely opravné.

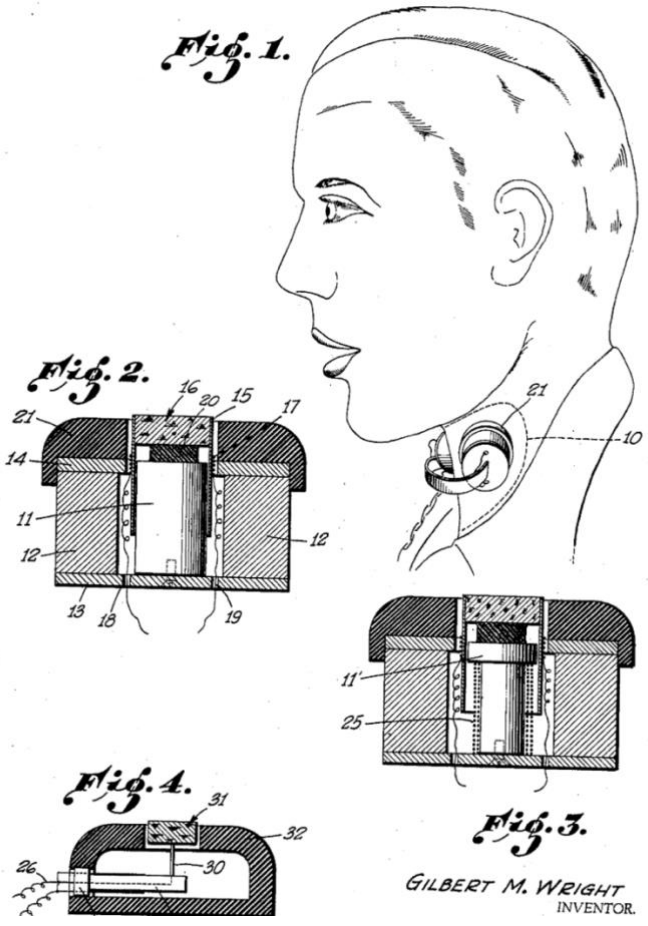
4.3.1 Historický vývoj pitch-shiftingu v audiovizi

Stejně jako tomu bylo v oblasti hudební produkce, i v kinematografii docházelo v průběhu jejího vývoje k implementacím různých způsobů zacházení s výškou zvuku. Technologické metody byly obdobné jako u hudby – nejprve bylo užíváno proměny rychlosti přehrávání záznamu, později s nástupem digitálních technologií umožněno manipulace se samply a posléze i různé formy zpracování zvuku v DAW.

Jednou z prvních zmínek o využití posunu výšky zvuku v kinematografii je utvoření charakteristického řevu King Konga ze stejnojmenného filmu z roku 1933. Zvukový technik Murray Spivak nahrál v ZOO zvuky lvů a tygrů, následně je přenesl do studia a následně využil proměnné rychlosti přehrání zvuku. Výsledkem

byly snížené, mohutnější zvířecí zvuky, jež lépe odpovídaly charakteru obřího lidoopa [23] (*ukázka V1*).

Ještě před rozšířením magnetofonové technologie bylo od konce 40. let k utvoření zvukového efektu změny hlasu mluvčího užíváno zařízení zvané Sonovox[16]. Příkladem budiž metalický zvuk vydávaný lokomotivou Casey Jr v animovaném filmu Dumbo z roku 1941 (*ukázka V2*). Nejedná se však o pitch-shifting ve své pravé podstatě, jelikož samotný mluvčí není zdrojem zvuku – tím je hudební nástroj či jiný oscilátor a samotný mluvčí jen moduluje tento zvukový signál prostřednictvím svého hlasového ústrojí. Výška zvuku se tedy nemění.



Obr. 4.2: Schéma zařízení Sonovox [18]

[16] Sonovox – zařízení představené v roce 1939 Gilbertem Wrightem umožňující modulaci zvuku hudebního nástroje skrze lidské hlasové ústrojí [18]

Další možnosti využití pitch-shiftingu v kinematografii přišla s rozvojem technologií proměnné rychlosti přehrávání pásky a spočívala převážně v proměně zvukového charakteru herce či herečky nebo utvoření nového zvukového efektu. Těchto metod bylo zprvu užíváno nejčastěji v animovaném filmu – například díky nim bylo ve snímku *Popelka* od studií Walta Disneyho z roku 1950 dosaženo abnormálně vyššího hlasu myší (*ukázka V3*). Později bylo díky zařízení zvanému ELTRO Information Rate Changer[17] umožněno zacházet s výškou zvuku a dobou jeho trvání nezávisle na sobě. Toto zařízení, původně zamýšlené jakožto time-stretchingový nástroj k edukačním účelům [20, 22], našlo svá využití i v oblasti hudební produkce, reklamách a ve filmové zvukové tvorbě.

Obr. 4.3: Zařízení ELTRO Information Rate Changer [21]

Nejznámějším příkladem použití zařízení ELTRO Information Rate Changer v kinematografii je kultovní film *2001: Vesmírná Odysea* od Stanleyho Kubricka z roku 1968. Posun výšky zvuku je zde použit ve scéně, kdy je počítač Hal 9000 postupně deaktivován a v průběhu tohoto procesu se pozvlona snižuje výška jeho hlasu (*ukázka V4*). Společně s ní se i prodlužuje jím vydávaný zvuk, avšak v jiném poměru. Nakonec je posun zvuku již vskutku extrémní, až přejde v ticho – počítač

[17] Zařízení vyvíjené v 50. letech 20 století Libereckým rodákem Antonem Marianem Springerem, fungující na principu rotující soustavy magnetofonních hlav [19]

je vypnut. Postupně se snižující zvuk zůstává po většinu času poměrně kvalitní, s rostoucí změnou sice roste množství slyšitelných artefaktů, ty však v rámci této scény nepůsobí nikterak rušivě, naopak umocňují pocit deaktivujícího se počítačového systému. Režisér filmu Stanley Kubrick doku 1971 v konverzaci se skladatelkou Wendy Carlos posléze prozradil, že zařízení ELTRO IRC bylo použito ve všech scénách s hlasem počítačového systému HAL 9000, avšak bez změny výšky, pouze jako prodloužení doby přehrávání o 10-20 %. Samotná hlasová sekvence s deaktivací pak byla zařízením zpracována nadvakrát – nejprve pro graduální snižování výšky hlasu a podruhé pro jeho postupné zpomalení [21]. Zařízení ELTRO IRC bylo poté častěji využívané, jedním z dalších příkladů je změna výšky hlasu postavy Strážce trůnu v pilotním dílu seriálu Star Trek z roku 1965.

S postupným rozvojem digitálních technologií byly uvedeny v provoz efektové jednotky a zařízení obsahující pitch-shifting popsané v kapitole 4.2.2. Značným vývojem procházely zejména softwarové technologie umožňující práci se samply. U filmového sound designu se oblíbeným zařízením mnohých zvukových inženýrů stal Synclavier. Ten se, ačkoliv byl původně spíše syntezátorem, postupem času vyvinul v poměrně schopný sampler. V druhé polovině 80. let jej bylo možné propojit s počítačem Macintosh skrze rozhraní MIDI[18] a vzorkovat samply o vzorkovací frekvenci až 100 kHz [24]. Ty mohly být editovány ať už v samotném Synclavieru, tak v dalších programech na systému Macintosh (např. Digidesign Sound Designer[19]) a následně do Synclavieru přeneseny. Tohoto zařízení poté bylo využíváno ve značném množství velkých studiových blockbusterů jako třeba Jurský Park z roku 1993 či Titanic z roku 1997, kde zvukový inženýr Gary Rydstrom pomocí přehrávání samplů v různých výškách vytvořil rytmický hluk vydávaný lodním motorem [25] (*ukázka V5*). Svoji práci se samply popsal následovně:

„The idea of using a sampler for sound effects work had astonishing potential. With sampled sounds in RAM, you can instantly pitch-bend it and layer

[18] Musical Instrument Digital Interface – softwarový protokol umožňující digitální komunikaci mezi hudebními nástroji a počítači

[19] Digidesign Sound Designer – Jeden z prvních programů umožňujících editaci zvukových samplů na počítači, postupně se vyvinul v dnešní Pro Tools

it and play it and shape it, without using any processing time. You can layer on the same key and very finely manipulate the pitch and delay and merge them together in ways that were harder to do in the tape-to-tape days. It allowed me to create the dinosaurs in Jurassic Park, in which I took several layers and blended different animal sounds into what sounds like one animal. The Titanic engine sounds took advantage of how the Synclavier could speed up and slow down a sound pattern." [25][20]

Obr. 4.4: Zařízení Synclavier v roce 1989

Synclavier samozřejmě nebyl jediným zařízením umožňujícím využití samplů. Rozvoj softwarových nástrojů operujících na různých platformách (Macintosh, Atari, Amiga) umožňoval díky protokolu MIDI pracovat se samplery a dalšími elektronickými nástroji.

[20] *Překlad: Myšlenka užívání sampleru pro účely zvukových efektů měla úžasný potenciál. Nasamplované zvuky v RAM bylo možné okamžitě posouvat ve výšce, vrstvit, přehrávat je a tvarovat bez jakékoliv prodlevy. Můžete je navrstvit ve stejné harmonii, jemně manipulovat s jejich výškou a zpožděním a spojovat je dohromady způsoby, kterými to za dob užívání magnetofonních pásek šlo hůře. Umožnilo mi to vytvořit dinosaury v Jurském parku, kde jsem spojil několik vrstev různých zvířecích zvuků v jedno nové. Zvuk motoru Titaniku využil schopností Synclavieru zrychlovat či zpomalovat sledy zvuků.*

4.3.2 Využití Pitch-shiftingu v audiovizi v současnosti

V dnešní době je díky pokročilým softwarovým nástrojům již možné poměrně značně manipulovat s výškou nahraného zvuku, aniž by vznikaly artefakty v nežádoucí míře. To vyúsťuje v širokou škálu možností zvukových efektů a zvukově-dramaturgických prostředků, kterých můžeme dosáhnout.

Tím základním je nadále změna výšky hlasu herce či herečky. Účelem je povětšinou v kombinaci s dalšími zvukovými efekty vytvořit nový hlas odpovídající požadovanému charakteru filmové postavy. Tento nový hlas může být taktéž nakombinován s hlasem původním, který je tak pouze obohacen o další vrstvy. Ukázkovým příkladem tohoto postupu je scéna z filmu *Pán prstenů: Společenstvo Prstenu* (2001), ve které elfská čarodějka Galadriel znenadání promlouvá hlubším děsivým hlasem (*ukázka V6)*. Tohoto efektu bylo dosaženo kombinací původního hlasu herečky a poté jejího hlasu pitch-shiftovaného níže, přičemž tento druhý hlas byl namluven dvakrát rychleji pro minimalizaci vzniku artefaktů. Dalšími příklady obohacení hercova původního hlasu posunem jeho výšky pak je výsledný hlas draka Šmaka ve filmu *Hobit: Šmakova dračí poušť* (2013) či třeba hlas postavy Kylo Rena z filmu *Star Wars: Síla se probouzí* (2015). Obzvláště je to patrné ve scéně, kdy tato postava vyslýchá postavu Rey, přičemž si sundá svoji masku, která změnu hlasu způsobuje (*ukázka V7*). Zde je slyšet zřejmý rozdíl mezi hercovým přirozeným hlasem a jeho hlasem posunutým níže v kombinaci s mnohými dalšími efekty. Zajímavostí je, že v tomto případě byl kromě pitch-shiftingu záměrně použit i tzv. formant shifting, což dodává výslednému zvuku charakteristickou „potemnělost". Trochu atypickým příkladem posunutí výšky hlasu je pak scéna z filmu *Matrix* z roku 1999, při které je hlavní postava Neo odpojena od systému Matrixu, přičemž se kamera dostává do hercovy ústní dutiny (*ukázka V8*). Jeho výkřik je pak modifikován v digitalizovanou, robotizovanou podobu, jež se stupňovitě snižuje ve své výšce. Cílem změny výšky zde tedy nebyla přirozenost či kvalita nahrávky, naopak šlo o její záměrnou degradaci.

„My approach was to start with the apparent high resolution of his Matrix sensory inputs, and then degenerate into more and more quantized, granular bits of his own sound to simulate that transition of virtual sense breaking down." [28][21]

Další běžnou praxí kromě posouvání výšky lidského hlasu bývají taktéž často změny výšky zvukových projevů zvířat – pro získání nového neexistujícího zvuku. V tomto případě může být příkladem již zmíněný film King Kong z roku 1933 či třeba zvuk mimozemské příšery Rancora ze snímku Star Wars: Návrat Jediho z roku 1983, který je utvořen značným zpomalením zvuků vydávaných drobným jezevčíkem [25]. Podobné principy jsou pro vytváření nových zvuků příšer či dalších stvoření využívány dodnes. Ukázkovým příkladem jsou všemožné dračí zvukové projevy v sérii animovaných filmů *Jak vycvičit draka*, jejíž první díl byl uveden do kin roku 2010. Jelikož se ve filmu vyskytuje veliké množství různých draků, kteří nemluví a nejsou tedy namluveni herci, bylo nezbytné vytvořit širokou škálu nových zvuků vyjadřujících jejich všemožné emoční zvukové projevy. Sound-designerský tým v čele s Randym Thomem proto využil nahrávky mnohých druhů zvířat (konkrétně například slonů, koček, psů, velryb, koz, mrožů a velbloudů) a lidských emočních hlasových projevů. Tyto nahrávky pak následně různě kombinoval a manipuloval s jejich výškami, aby bylo dosaženo charakteristické palety zvuků jak pro hlavní dračí postavu, tak pro desítky dalších rozmanitých draků. V rozhovoru z roku 2019 pro online magazín *A Sound Effect* pojednávajícím o třetím dílu této serie prozradil Randy Thom následující:

„It's very difficult for human beings to make themselves believably sound like animals. There's always something about it that tells the listener that it's a person trying to sound like an animal and not a real animal. So we try to start with actual animal vocalizations, like elephants and whales and camels — every conceivable type of animal you can think of. The next trick is to integrate all of those things, so that it doesn't sound like you are going from an elephant to a goat to a human in the span of one phrase that the dragon might utter. There is some

[21] *Překlad: Můj přístup vycházel ze zjevného vysokého rozlišení jeho Matrixových vstupních vjemů, jež jsou následně degenerovány ve více a více kvantizované, granulované kousky jeho vlastního hlasu pro simulaci proměny virtuálního vnímání, které se bortí.*

technical trickery that we use in an effort to change the source sounds quite a bit. For instance, we change the pitch as much as two or three octaves to make a dog sound like a creature that weighs 2,000 pounds." [26] [22]

V rozhovoru dále uvádí, kterak je při pitch-shiftingu stále zásadním faktorem uvěřitelná přirozenost výsledného zvuku, obzvláště pokud má být napodobením zvířecího hlasového projevu. Dobrou ukázkou proměny různých zvířecích zvuků v dračí „řeč" je scéna z filmu *Jak vycvičit draka*, ve které hlavní postava večeří společně s hlavní dračí postavou Bezzubkou – při pozorném poslechu je zde možné i rozeznat některé původní zdroje zvuku v podobě různých živočichů (*ukázka V9*).

Obdobně jako se zvuky zvířat je pak možné pracovat i obecně s ruchovou složkou filmu. Zde může být účelem pitch-shiftingu konkrétních zvuků buď jejich proměna ve zcela jiný požadovaný zvukový objekt anebo jistá forma obohacení původního zvuku pro dosažení chtěného efektu. Dobrým příkladem vytvoření nového ruchu zcela odlišného od jeho původního zdroje může být zvuk letící stíhačky Tie Fighter z filmu *Star Wars: Nová naděje* z roku 1977, který byl vytvořen z pitch-shiftovaného sloního křiku v kombinaci se zvukem projíždějícího auta na mokrém asfaltu [25]. Podobným principem je možné vytvářet i neotřelé zvukové atmosféry. Takto byly utvořeny hluboké kovové rezonance uvnitř vesmírné lodi Nebuchadnezzar ve filmu *Matrix* z roku 1999. Původním zdrojem tohoto zvuku je podle sound-designéra Danea Davise skřípající kovová brána, jejíž nahrávka byla následně o mnoho oktáv snížena.

„There was a giant metal gate that I recorded in Texas a few years ago that made this singing resonance, and so I pitch-shifted that down many, many, many octaves using a program called SoundHack, one of the few stand-alone applications that I used, can do pitch-shifting a long way without hearing any

[22] *Překlad: Pro lidské bytosti je velmi obtížné znít uvěřitelně jako zvířata. Něco na tom vždy zní jako člověk předstírající, že je nějaký živočich, a není to pak reálné. Proto se snažíme začít se skutečnými projevy zvířat, například slonů, velbloudů, velryb – každé myslitelné zvíře, jež vás může napadnout. Dalším trikem je všechny tyhle věci namíchat, tudíž to pak nezní jako že prostříláte slona, kozu a člověka během jedné dračí fráze. Je v tom i jistá technologická magie, kterou užíváme ve snaze pozměnit zdroj zvuku. Například měníme výšku zvuku až o dvě nebo tři oktávy pro proměnění psa ve stvoření, které váží 900 kilogramů.*

nonsense, and so I used that to create all of these extremely low metal resonances that you hear inside the ship." [27] [23]

Tento způsob sound-designu, kdy se požadovaný zvuk sestává z kombinací jiných zvuků se změněnou výškou, je v dnešní praxi využíván často a je díky němu možné vyhotovit velkou škálu užitečných zvuků.

4.3.3 Slow-motion scény

Druhou možností, kterak využít pitch-shifting v rámci práce s ruchovou složkou, je manipulace s výškou zvuku konkrétního zvukového objektu – zdroj zvuku je tedy shodný s tím ve filmové diegézi[24], jen má jiný frekvenční charakter ze zvukově-dramaturgických důvodů, například v takzvaných slow-motion[25] záběrech. Zde bývá výškou zvuku obvykle manipulováno společně s dobou jeho trvání, aby bylo dosaženo efektu zpomalení zvuku souběžně s obrazovou složkou. Jedním z ukázkových příkladů je využití zpomalování zvuku v sérii filmů *Matrix*. Jde zejména o zvuky výstřelů, nárazy a švihy při soubojích či různé kovové údery, které byly zaznamenány ve vyšších vzorkovacích kmitočtech, načež mohly být kvalitněji pitch-shiftovány či prostě jen zpomalovány, jak uvádí zvukoví inženýři Dane Davis a Eric Lindenmann:

"*I'm very interested in extreme high frequency, extreme level and the complexity of the acoustic waveform. We do a lot of pitching up and down; generally, when you do that, you lose a lot of naturalness. The 96 kHz resolution just happens to play into one of our fascinations: capturing the harmonics of sounds that we don't normally hear. We also did extensive recording at 192 kHz using mics with extended upper range, like the Sennheiser MKH800 and some calibration microphones.*"

[23] *Překlad: V Texasu jsem nahrával takovou obří kovovou bránu, která vydávala melodický rezonující zvuk, a tak jsem ho posunul o mnoho, mnoho oktáv níže za využití programu jménem SoundHack, který umožňuje takovéto posuny, aniž by vytvářel nějaké nesmysly. No a takto jsem vytvořil hluboce znějící rezonance, jež můžete slyšet uvnitř vesmírné lodi.*

[24] Diegéze – fikční filmové prostředí, v němž se film odehrává

[25] Slow-motion – obrazový efekt zpomaleného záběru

"For the large metal sounds of the machines that are in the real world, we recorded big metal bangs and hits at very high sample rates to capture ultrasonic frequencies. That way, we could pitch them down while still maintaining the whole harmonic structure." [29][26]

Sekvencí, která dobře ukazuje výše popsané postupy, je úvodní scéna filmu *Matrix Reloaded* z roku 2003 (*ukázka V10*). Nejprve je možné slyšet zpomalení a snížení zvuku motocyklu, načež následuje mohutná exploze doprovázená hlubokým šlehnutím plamenů. Poté dochází ke krátkému souboji mezi strážníky a protagonistkou Trinity, kde je použito několik zpomalených záběrů podpořených pitch-shiftovanými zvuky pohybů. Po chvíli následuje záběr na skleněnou stěnu budovy, jež je proražena vyskočivší Trinity, načež je záběr zpomalen a souběžně s tím i snížena výška zvuku tříštícího se skla. Během toho dochází ke zpomalené přestřelce mezi Trinity a nepřátelským agentem. Jednotlivé výstřely jsou zpomaleny a posunuty níže, zároveň jsou zachovány vyšší frekvenční složky výstřelů. Projektily mají taktéž svůj charakteristický zpomalený turbulentní zvuk utvořený vrstvením mnoha různých zvuků, u nichž bylo s výškou taktéž manipulováno. Výsledkem je komplexní vjem zpomaleného zvukového prostředí, jehož bylo dosaženo užitím pitch-shiftingu reálných zvuků v kombinaci s dalšími zvukovými efekty.

Scény využívající slow-motion záběry jsou běžné v mnohých jiných akčních filmech, přičemž přístup k jejich zvukovému ztvárnění může vždy být odlišný. Obvykle je pro efekt zpomalení použito snížení výšky daného zvuku, jeho prodloužení v čase a určitá forma dozvukového efektu. U série Matrix je toto vše užito ve velice intenzivní, někdy až přehnané míře. Naproti tomu ve filmu *Počátek* z roku 2010 je práce s pitch-shiftingem ve zpomalených scénách o poznání

[26] *Překlad: Velmi mne zajímají extrémně vysoké zvukové frekvence, hlasitosti a kompexita tvarů zvukových vln. Provádíme hodně posunu výšky zvuku nahoru i dolů a obvykle tím ztrácíte hodně přirozenosti. Rozlišení 96 kHz je právě proto fascinující, zachycuje harmonické složky zvuků, jež normálně nelze slyšet. Také jsme provedli četná nahrávání o vzorkovací frekvenci 192 kHz za využití mikrofonů s vyšší hranicí frekvenční charakteristiky, jako třeba Sennheiser MKH800 a některé kalibrační mikrofony.*

Pro mohutné kovové zvuky stojů v reálném světě jsme nahráli velké kovové rány a údery ve velmi vysokých vzorkovacích frekencích, abychom zachytili ultrazvukové frekvence. Touto cestou jsme poté mohli snížit jejich výšku a stále přitom zachovat plnou harmonickou strukturu.

decentnější. Zde je využit zejména při přechodu mezi jednotlivými vrstvami snů – třeba ve formě zpomalení/zrychlení tikání hodin, což doprovází i výšková změna hukotu zvukové atmosféry (*ukázka V11*) U tohoto filmu je však velmi zajímavé i zacházení s výškou diegetické hudby[27]. Jelikož v různých vrstvách snu ubíhá čas jinak rychle, postavy vnímají znějící skladbu *Non, je ne regrette rien* zpomaleně. Zvukově je tento efekt nejvíce patrný v sekvenci, během níž je slyšet hned po sobě jednotlivé vrstvy snu (*ukázka V12*). Provedení tohoto efektu je poměrně chytré – tóny dechových nástrojů jsou zpomaleny a sníženy až o několik oktáv, avšak zpěvaččin hlas má výšku nezměněnou, je pouze prodloužený. Jinak by totiž mohlo být pro diváka nepochopitelné, že se jedná stále o tu samou skladbu.

[27] Diegetická hudba – hudba, jejíž zdroj se nachází přímo v prostředí, jež film vyobrazuje [32, str. 25]

4.3.4 Další využití pitch-shiftingu v kinematografii i mimo ni

Posouvání výšky zvuku je v kinematografii velmi důležitým postupem pro umocnění audiovizuálního zážitku diváka. Prakticky všechny výše popsané způsoby využití posunu výšky zvuku ve filmu mají společné dvě věci. Zaprvé se jedná o využití ve specifických žánrových odvětvích – sci-fi, fantasy anebo akční film – což umožňuje poměrně pestré a svobodné zacházení s různými zvukovými postupy. Je zde často nutné vytvořit nové zajímavě znějící ruchy, proměnit hlas mluvčího anebo dosáhnout takové zvukové atmosféry, jež není ve skutečném světě běžná. Proto je tedy pitch-shifting často užívaným prostředkem k dosažení těchto požadovaných zvuků. Samozřejmě může být posun výšky zvuku využit i v nějakém specifickém případě v rámci jiných filmových žánrů, nebývá zde ale zpravidla tolik výrazným prvkem.

Druhou skutečností vyplývající z dosud předvedených ukázek je, že pitch-shifting v kinematografii bývá využíván především pro tvůrčí účely – obohacení hlasu, zvukový efekt a podobně. Nicméně posun výšky zvuku je možné využít i pro účely opravné. Podobně jako u intonačních korektur v rámci hudební produkce, i v oblasti filmové zvukové postprodukce mohou být tyto korektury provedeny. Příkladem budiž hypotetická situace, kdy by bylo užitečné při editaci dialogů použít zvukové stopy z různých verzí natočených záběrů, avšak jejich prosté navázání na sebe by nebylo funkční kvůli drobným intonačním rozdílům v hereckém projevu. Když se však jemně a citlivě pohne s výškou navazujících částí tak, aby byl intonační projev srovnán do jedné roviny, bude možné tyto zdánlivě nenavazující části spojit. Anebo jiný příklad – situace, kdy je potřeba vhodně posílit či naopak potlačit emoční projev herce a herečky – i zde může přijít vhod jemné proměnné posunutí výšky hlasu. Samozřejmě musí být provedeno tak, aby výsledný hlasový tok nezněl nepřirozeně (pokud to není účelem). Tímto způsobem je možné i intonačně zakončit větu v místě, kde původně ještě nekončila a byla dále rozvedena. Ačkoliv dochází k jisté degradaci kvality záznamu hlasu, mnohdy může být být důležitější prvotní přenesený emoční vjem na diváka, který je daný právě intonací hercova projevu. Užívání takovýchto posunů výšky zvuku je obzvláště užitečné například v rámci dokumentární tvorby či při tvorbě podcastů anebo

jiných formátů využívajících delší výpovědi mluvčích/respondentů, kde často bývá nutné věty stříhat a zakončovat nezávisle na intonaci mluvčího.

Dalším, čistě praktickým případem využití posunu výšky zvuku, je konverze AV díla z 24 snímků za vteřinu do 25 či obráceně – vlivem rozdílu mezi standardem promítání v kině a v televizi. Touto konverzí se zvuk pochopitelně taktéž zrychlí anebo zpomalí, rozdíl činí 4 %, čemuž následně odpovídá i změna ve výšce zvuku o necelý půltón. Tu je možné následně opravit, nicméně to přináší i jistá úskalí. U multikanálového zvuku může docházet k fázovému zkreslení, jelikož není zachována prefektní fázová shoda původního signálu. Vlivem toho mohou vznikat prostorové artefakty, výkyvy v hlasitosti či rozmazaná lokalizace.

Pitch-shifting je pochopitelně možné využívat i v jiných oblastech než pouze v hudbě a v kinematografii. Například v rámci herního designu – ať už v tzv. cutscénách[28] (zde jsou jeho implementace vlastně velmi obdobné jako v kinematografii), tak během samotné hry. Zde může být posouvání výšky různých zvuků vhodné buď pro rozšíření palety užívaných zvuků (aby se neopakovaly dokola ty samé) anebo pro jejich proměnlivost v rámci vývoje děje hry – například pokud postava posbírala více věcí, její hmotnost se zvýší, může pak vydávat hlubší zvuky kroků, ta se může postupně proměňovat společně s tím, jakou hmotnost věcí nese – a tak podobně. Právě v této programovatelnosti změny výšky zvuků spočívá specifický potenciál využití pitch-shiftingu v herním designu.

Odlišným typem využití posunu výšky zvuku pak může být anonymizace hlasu – použitelná například z bezpečnostních důvodů v dokumentární tvorbě či televizním nebo jiném zpravodajství, pokud nechce být mluvčí rozeznán. V závislosti na míře potřeby skrýt původní hlas mluvčího lze hlas anonymizovat většinou posunutím jeho výšky o oktávu či níže, zároveň s posunutím formantů. Výsledný hlas by neměl mít ani přílišné známky intonace.

[28] Cutscéna – zpravidla animovaná sekvence vyprávějící příběh hry, nebývá interaktivní

5) Závěr

Pitch-shifting je v dnešní době velmi využívaným nástrojem, jenž umožňuje manipulovat se zvukovým signálem. V průběhu práce byl představen jeho historický vývoj, rozebrána technická stránka jednotlivých pitch-shiftingových algoritmů a následně ukázány možnosti jeho využití jak v oblasti hudební produkce, tak zejména v kinematografii. Z uvedených ukázek je možné se přesvědčit, že v oblasti filmové zvukové postprodukce je pitch-shifting klíčovým nástrojem pro tvorbu nových zvuků – ať už hlasů lidí nebo projevů jiných stvoření, či ruchů a atmosfér. Toto stvrzují i citované rozhovory s předními světovými sound-designery. Jeho využití v rámci takzvaných slow-motion scén je pak velmi výrazným prvkem zvukové dramaturgie filmu. Při použití vyšších vzorkovacích frekvencí signálu v kombinaci s dostatečně citlivou frekvenční charakteristikou mikrofonu je v dnešní době možné posouvat výšku zvuku i o několik oktáv níže při zachování bohatého frekvenčního spektra. Krom těchto kreativních účelů je však možné posun výšky zvuku i v audiovizi užívat pro účely opravné.

Bakalářská práce tedy představuje pitch-shifting v několika rovinách a podává poměrně komplexní přehled o principech fungování a možnostech využití tohoto nástroje.

Seznam užitých zdrojů

1. SYROVÝ, Václav. *Hudební akustika*. 2., dopl. vyd. V Praze: Akademie múzických umění, 2008. Akustická knihovna Zvukového studia Hudební fakulty AMU. ISBN 978-80-7331-127-8.

2. ZÖLZER, Udo. *DAFX: digital audio effects*. 2nd ed. Chichester: Wiley, 2011. ISBN 0470665998.

3. CHEN, Julian Chengjun. *Elements of human voice*. World Scientific, 2017. ISBN 9789814733915

4. VOJÁČEK, Antonín. Algoritmus korelace v digitálním zpracování signálů. *Automatizace.hw.cz* [online]. 2006 [cit. 2022-08-14]. Dostupné z: https://automatizace.hw.cz/clanek/2006031701

5. PARVIAINEN, Oli. *Time and pitch scaling in audio processing*. *Software Developer's Journal* [online]. 4/2006 [cit. 2022-08-13]. Dostupné z: https://www.surina.net/article/time-and-pitch-scaling.html

6. VERHELST, Werner, ROELANDS, Marce*. An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech. 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing* 2 (1993): 554-557 vol.2.

7. OSGOOD, Brad. The Fourier Transform and its Applications. *Lecture Notes for EE 261* [online]. [cit. 2022-08-13]. Dostupné z: https://see.stanford.edu/materials/lsoftaee261/book-fall-07.pdf

8. MORIN, David. *Fourier Analysis* [online]. 28.11.2009 [cit. 2022-08-13]. Dostupné z: https://scholar.harvard.edu/files/david-morin/files/waves_fourier.pdf

9. BERNSEE, Stephan. The DFT "à Pied": Mastering The Fourier Transform in One Day. *Stephan Bernsee's Blog* [online]. 21.9.1999 [cit. 2022-08-13]. Dostupné z: http://blogs.zynaptiq.com/bernsee/dft-a-pied/

10. GASIOR, Marek & GONZALEZ, Jose. *Improving FFT Frequency Measurement Resolution by Parabolic and Gaussian Spectrum Interpolation*, 2004. DOI: 10.1063/1.1831158. Dostupné z: https://doi.org/10.1063/1.1831158

11. BERNSEE, Stehphen. Pitch Shifting Using The Fourier Transform. *Stephan Bernsee's Blog*[online]. 21.9.1999 [cit. 2022-08-13]. Dostupné z: http://blogs.zynaptiq.com/bernsee/pitch-shifting-using-the-ft/

12. ROEBEL, Axel, RODET, Xavier. *Efficientt Spectral Envelope Estimation and its application to pitch shifting and envelope preservation*, International Conference on Digital Audio Effects, 2005. Madrid, Španělsko. pp.30-35, hal-01161334

13. CRAB, Simon. The 'Groupe de Recherches Musicales' Pierre Schaeffer, Pierre Henry & Jacques Poullin, France 1951. *120 Years of Electronic Music* [online]. Hastings, UK, 2021 [cit. 2022-08-13]. Dostupné z: https://120years.net/wordpress/the-grm-group-and-rtf-electronic-music-studio-pierre-schaeffer-jacques-poullin-france-1951/

14. CASE, Alex U. Sound and vision. *Recordingology* [online]. c2010-2022 [cit. 2022-08-13]. Dostupné z: https://120years.net/wordpress/the-grm-group-and-rtf-electronic-music-studio-pierre-schaeffer-jacques-poullin-france-1951/

15. LEWISOHN, Mark. *The Beatles recording sessions*. New York: Harmony Books, c1988. ISBN 978-0517570661.

16. GRANT, Jim. The Fairlight Explained. *Electronics & Music Maker: Computer Musician*. UK: Future Publishing, 1984. Dostupné z: https://www.muzines.co.uk/articles/the-fairlight-explained/7992#

17. KRAVAŘÍK, Jindřich. Antares Auto-Tune 8. *Audiozone.cz* [online]. 2015 [cit. 2022-08-13]. Dostupné z: https://www.audiozone.cz/recenze/antares-auto-tune-8-t22394.html

18. WRIGHT, Gilbert. *Means and Methods for Producing Sound Effects*. USA. US2273078A. Uděleno 17.2.1942. Zapsáno 27.3.1939.

19. VOIGTSCHILD, Fabian, Jonathan STERNE a Mara MILLS. Anton Springer and the Time and Pitch Regulator. *Sound and Science* [online]. 2018 [cit. 2022-08-13]. Dostupné z: https://soundandscience.de/contributor-essays/anton-springer-and-time-and-pitch-regulator

20. TEMMER, Steve. *Eltro Descriptive Brochure*. Dostupné z: https://www.wendycarlos.com/other/Eltro-1967/Eltro-1967.pdf

21. CARLOS, Wendy. *The Eltro and the Voice of HAL* [online]. c1996-2020 [cit. 2022-08-14]. Dostupné z: https://www.wendycarlos.com/other/Eltro-1967/

22. SCOTT, Jason. Demonstration of Compressed and Expanded Speech for Education using the Eltro Information Rate Changer Mark II. *Archive.org* [online]. [cit. 2022-08-14]. Dostupné z: https://archive.org/details/TNM_infotronic_eltro_information_rate_changer/infotronic_eltro_information_rate_changer_demo-side_b.wav

23. HANSON, Helen. *Hollywood Soundscapes: Film Sound Style, Craft and Production in the Classical Era*. UK: British FIlm Institute, 2017. ISBN 978-1844575046.

24. MEYER, Chris. The Synclavier Story. *Music Technology*. 1989. Dostupné z: http://www.muzines.co.uk/articles/the-synclavier-story/97

25. RINZLER, J. W. *The sounds of Star Wars*. San Francisco: Chronicle Books, c2010. ISBN 978-0811875462.

26. WALDEN, Jennifer. How Randy Thom & Al Nelson Crafted 'How to Train Your Dragon: The Hidden World's Impressively Evocative Sound. *A sound effect* [online]. [cit. 2022-08-14]. Dostupné z: https://www.asoundeffect.com/how-to-train-your-dragon-3-sound/

27. BUSKIN, Richard. The Matrix Young guns, new tricks. *Studio Sound, filmsound.org* [online]. 1998 [cit. 2022-08-14]. Dostupné z: http://filmsound.org/studiosound/post_matrix.html

28. KAUFMAN, Debra. Dane A. Davis on 'The Matrix'. *Cinemontage.org* [online]. 2014 [cit. 2022-08-14]. Dostupné z: https://cinemontage.org/dane-davis-matrix/

29. DRONEY, Maureen. The Matrix Reloaded. *Mixonline.com* [online]. 2003 [cit. 2022-08-14]. Dostupné z: https://www.mixonline.com/sfp/matrix-reloaded-369031

30. CO-100K Supersonic Wide-Range Omni-Directional Condenser Mic. *Sanken Chromatic* [online]. [cit. 2022-08-14]. Dostupné z: https://www.sankenchromatic.com/products/co-100k/

31. SMITH, Steven W. *The scientist and engineer's guide to digital signal processing*. San Diego, Calif.: California Technical Pub., 1997. ISBN 978-0966017632

32. BLÁHA, Ivo. *Zvuková dramaturgie audiovizuálního díla*. 3., upr. vyd. V Praze: Nakladatelství Akademie múzických umění, 2014. ISBN 978-80-7331-303-6

33. WAHAB, Muhammad. Interpolation and Extrapolation. 2017. University of Paderborn. Dostupné z: https://www.researchgate.net/publication/313359516_Interpolation_and_Extrapolation

34. ŘÍHA, Petr. *Slovník počítačové informatiky: výkladový slovník pro práci s informacemi: hardware a software včetně počítačových sítí, internetu a mobilních technologií*. Ostrava: Montanex, 2002. Informační technologie. ISBN 8072250833

Seznam ukázek

A) Zvukové ukázky

- **Z1** – Porovnání originální nahrávky a pitch-shiftingu při vzorkovací frekvenci 48 kHz a 96 kHz (zdroj nahrávky: Soundsnap)
- **Z2** – Porovnání originální nahrávky a pitch-shiftingu při vzorkovací frekvenci 48 kHz a 96 kHz (zdroj nahrávky: Soundsnap)
- **Z3** – Strawberry Fields Forever, Beatles (1967)
- **Z4** – Fame, David Bowie (1975)
- **Z5** – She's Goin' Bald, Beach Boys (1967)
- **Z6** – Sound and Vision, David Bowie (1977)
- **Z7** – Believe, Cher (1998)

B) Ukázky z filmů

- **V1** – King Kong (1933)
- **V2** – Dumbo (1940)
- **V3** – Popelka (1950)
- **V4** - 2001: Vesmírná odysea (1968)
- **V5** – Titanic (1997)
- **V6** – Pán prstenů: Společenstvo prstenu (2001)
- **V7** – Star Wars: Síla se probouzí (2015)
- **V8** – Matrix (1999)
- **V9** – Jak vycvičit draka (2010)
- **V10** – Matrix Reloaded (2003)
- **V11** – Počátek (2010)
- **V12** – Počátek (2010)

Ukázky jsou dostupné i pod odkazem:
https://drive.google.com/drive/folders/1VU2XTbYLvH9sb4S0kEBNoBiyj7NUSWky?usp=sharing