Univerzita Karlova

Filozofická fakulta

Fonetický ústav

Bakalářská práce

Haštal Hapka

Harmonizace melodických kroků v mluvené češtině a její dopad na percepci

Harmonization of melodic intervals in spoken Czech and its perceptual impact

Praha 2019

Vedoucí práce: doc. PhDr. Jan Volín Ph.D.

Poděkování

Mé díky patří především doc. PhDr. Janu Volínovi Ph.D., který mi, kromě trpělivého zodpovídání mých četných dotazů a občasného upozornění na mé přílišné sklony k fejetonu, dal mnoho cenných rad ohledně zpracování a bez nějž by tato práce jen těžko vznikla. Dále bych rád samozřejmě poděkoval i všem 24 obětavým respondentům, kteří museli strávit půl hodiny v odzvučněné kabině a nebyli posléze nijak odměněni.

Prohlášení

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

V Praze, dne 12. 7. 2019 _____

Obsah

1. Úvod

Skladatel Leoš Janáček byl známý tím, že s sebou neustále nosil svůj notový zápisníček, a cokoliv zvukového ho zaujalo, se jal zapisovat. Zůstaly tak po něm mnohé zápisy zvuků přírody (od zpěvu ptáků po padání balvanů), ale hlavně melodie řeči lidí, co se kolem něj pohybovali na trhu, na přednáškách a podobně. Zapisoval je do notové osnovy a používal k tomu veškeré v té době známé hudebně notační prostředky jako udávání metra, předznamenání nebo důrazy. Omezoval se tedy na půltónovou škálu, vycházel z chromatiky, stejně jako téměř veškerá evropská hudba (Janáček, 1998).

Lidská řeč se chromatikou nicméně neřídí, spočte-li se základní hlasivková frekvence na slabičných jádrech, jen velkou náhodou mezi sousedními dvěma jádry vyjde čistý hudební interval. Bylo by proto zajímavé zjistit, jak moc by se řeč a její percepce změnila, pokud by vycházela opravdu jen z těchto intervalů. Tím se bude zabývat celá tato práce.

V kapitole 2 v oddílu 2.1 se zaměříme na obecné analogie mezi řečí a hudbou. Toto téma, kterým se zabývá velmi rozsáhle například Patel (2010) nebo Kivy (2007) nahlédneme z různých stran a zároveň se pokusíme vymezit hlavní rozdíly. V oddílu 2.2 se zaměříme již konkrétně na téma melodie řeči a v oddílu 2.3 na její percepci, kterou detailně popisují ‘t Hart, Collier & Cohen (1990). V posledním oddílu druhé kapitoly již stanovíme hypotézy našeho výzkumu.

Třetí kapitola je zaměřena na metodu výzkumu. V oddílech 3.1, 3.2 a 3.3 budou popsány výzkumný materiál, zadání testu a způsoby jeho následného vyhodnocování. Součástí popisu materiálu je i přehled statistických dat popisujících výškové kroky mezi sousedními slabikami v přirozené řečové produkci, který se může stát užitečným referenčním zdrojem. Percepční test bude zjišťovat rozdíly mezi hodnocením upravených a neupravených nahrávek rozličných mluvčích. Tyto úpravy budou spočívat právě v zarovnání základní hlasivkové frekvence na všech slabičných jádrech do půltónové škály.

Výsledky budou pak vyhodnoceny v kapitole čtvrté. Kromě potvrzení nebo nepotvrzení stanovených hypotéz bude zmíněno několik příkladných statistických úkazů a grafů je znázorňujících.

Na závěr, v diskuzi, bude shrnuto, jaké další výzkumy by tato práce mohla podnítit, případně jak jinak a za jakých podmínek by se výzkum mohl opakovat, aby se hypotéza lépe prokázala, a k čemu by se taková bádání mohla hodit.

2. Teoretické zázemí

2.1. Analogie mezi řečí a hudbou

Člověk má v jazyce přirozenou tendenci si navzájem metaforicky připodobňovat různé vjemy z vnějšího světa. Zmíněnou metaforou zde není myšleno její klasické pojetí, totiž že je to pouhý básnický tropus založený na použití výrazu mimo svůj obvyklý smysl za účelem vyjádření analogického konceptu. Podle teorie konceptuální metafory (Johnsen & Lakoff, 1980) totiž principy metafory nejsou jazykové, nýbrž kognitivní. Metaforou je zde tedy myšleno jakési „obecné mapování přes konceptuální domény".

Pro představu můžeme uvést jako příklad dvojici domén života a cesty (putování). To, že lidé vnímají život jako putování (což spadá pod ještě obšírnější metaforu mezi časem a prostorem), se právě posléze projevuje v jazyce, kde objevujeme taková pojmenování jako: *potloukat se životem; životní křižovatka; nevědět, kudy dál; směřovat k cílům; potkalo ho neštěstí* a podobně.

Díky těmto konceptuálním metaforám nám k pojmenovávání věcí v jazyce stačí daleko méně výrazů, než kdybychom vnímali každou doménu jako samostatný koncept.

Jednou z těchto dvojic domén jsou právě pak i hudba a řeč. A ačkoli uvádějí Lerdahl a Jackendoff ve své Generativní teorii tonální hudby (1996: 24), že „poukazovat na povrchní analogie mezi hudbou a jazykem, ať už s pomocí nebo bez pomoci generativní gramatiky, je stará a velmi marná hra", pokusíme se ve stručnosti na tyto různé souvislosti poukázat. Rozsáhlé hudební kompozice se dělí na věty. Věty mají jednotlivá témata. V tématech rozeznáváme uzavřené hudební myšlenky – fráze s tématy a motivy (Zenkl, 2014: 26). Podobných analogických pojmenování mezi hudbou a řečí je mnoho. Neodpovídají si stoprocentně a jejich hierarchie je v mnohých případech i obrácená. Nicméně jsou tu a jaksi nenápadně nám naznačují, že tyto dvě domény vnímáme velmi podobně.

Primárně bychom samozřejmě mohli říct, že tyto podobnosti ve vnímání plynou ze skutečnosti, že vůbec první lidmi vytvářená hudba byla vokální (Smolka, 2001: 31). Bohužel pro tuto domněnku nemůžeme dohledat přímé důkazy. „Nejstarší doklady o existenci tohoto druhu spadají do údobí mladšího paleolitu (poslední údobí starší doby kamenné asi 80 000-10 000 let př. n. l.)."[1] Těmito doklady jsou míněny různé nálezy předmětů připomínajících chřestítka, bubínky nebo píšťalky. Podle Smolky je velmi pravděpodobné, že hrou na tyto prastaré instrumenty doprovázeli lidé své vokální projevy při magických obřadech a obětních

[1] Fitch (2006: 1) uvádí s jistotou data kolem 40 000 př.n.l.

rituálech. Ale vzhledem k jakýmkoliv pochopitelně neexistujícím způsobům záznamu, notace, to nemůžeme s určitostí tvrdit.

Nejstarším a prvním vnímaným aspektem hudby zdá se být na druhou stranu opravdu rytmus. Trehub (2003: 669) uvádí jako nejzajímavější jevy v „písničkách“ vytvářených malými dětmi kromě aliterací, repetic a rýmu právě rytmické vzorce.

Stejně jako řeč, i hudba je samozřejmě jakýsi sled diskrétních „zvuků“ (o určitém průběhu), které se nám ve výsledku spojí a nesou kompaktní informaci. Tento sled probíhá v čase a řídí se jistými pravidly, která se týkají jednotlivých vrstev obou zmíněných domén, jejichž podobnosti si dovolím v následujících odstavcích velmi volně nastínit.

Pokud začneme u těch v hierarchii nejnižších, budeme tím u řeči mínit fonologii – vrstvu jednotlivých fonémů – a u hudby vrstvu samotných tónů. Fonémy i tóny (myšleno v širším slova smyslu, tedy míníme i tóny s danou pravidelnou frekvencí i ruchy s frekvencí nepravidelnou) jsou nejnižšími funkčními jednotkami obou domén a jejich výběr závisí na omezených prostředcích, které máme k dispozici. V případě řeči je to fonologický inventář příslušného jazyka a v případě hudby to mohou být rozsahy jednotlivých instrumentů. O akustické realizaci se zmíníme později.

Vrstva morfologie, kde se v řeči spojí jednotlivé fonémy v stále poměrně malé celky, které už ale nesou jisté významy, odpovídá menším shlukům tónů – kratším melodickým článkům. Tyto celky sice už informaci nesou, ale této informaci může být porozuměno v plné šíři až v kontextu další, hierarchicky vyšší, vrstvě.

Tou je v řeči vrstva lexikální a v hudbě proti ní stavím vrstvu harmonie. Jde opět o spojení jednotek předchozí vrstvy – pokud spojíme morfémy podle daných pravidel, vzniknou lexémy. Zároveň v tomto lexému nabudou významu funkce oněch morfémů (mluvnické kategorie a podobně). Stejně tak v hudbě spojíme několik článků dohromady (i seriálně i paralelně), čímž vznikne větší melodický celek (například fráze) podložený dalšími články – jejich souzvukem vznikají akordy, harmonie.

Tato vrstva už je poměrně soběstačná, nicméně plného pochopení se dočká až v další vrstvě, jež pracuje s časovým průběhem. Tou je syntax v řeči a forma v hudbě.

Nutno však podotknout, že hudba nenese znaky takzvaného *duality of patterning* (Fitch, 2006: 5). Hudba postrádá přímou sémanticitu (kterou řeč má), což znamená, že když se kupříkladu tón D připojí k sekvenci dalších tří tónů, stále to „nic neznamená.“

Všechny tyto vrstvy jsou pak u obou domén reprezentovány zvukovou realizací – tou je právě onen sled diskrétních „zvuků“. V hudbě a v prozodii řeči vnímáme čtyři hlavní akustické veličiny. Čas (metrum, tempo), amplitudu (dynamiku a hlasitost), frekvenci (výška tónů a základní frekvence) a spektrální vlastnosti (barvu). Zde se hudba a řeč odlišují, protože zatímco řeč jako hlavní veličinu pro percepci slova a jednotlivých segmentů používá spektrální vlastnosti – rozlišení jednotlivých souhlásek a samohlásek, hudba ve svém klasickém pojetí stále klade největší důraz na výšku tónů a rytmus.

Vraťme se ještě trochu nazpět a zaměřme se na zmíněná pravidla, jimiž se obě domény musí řídit. Zde se hudba a řeč již poměrně liší, převážně právě proto, že hudba postrádá onu přímou sémanticitu. U řeči je velmi striktně dáno, které morfémy s kterými se mohou pojit a jaké funkce budou plnit. U hudby toto možná platilo zhruba do začátku dvacátého století, nicméně v následném vývoji již určitě ne. Stále však lze u obou domén říct, že pokud tato pravidla porušíme, výsledná zvuková sekvence nebude pro komunikačního partnera příliš srozumitelná. Zde bychom mohli vedle sebe postavit jazyk a například hudební styl. Pokud by Mozart ve své době použil v klavírním koncertě dodekafonii[2], dalo by se to přirovnat k cizojazyčnému projevu.

Přestože je však třeba dbát pravidel, samotná produkce je velmi přirozená a nemusíme ovládat ani jednu z domén na odborné úrovni. Když mluvíme, nepřemýšlíme nad každým pohybem jazyka a jeho vzdáleností k patru stejně tak, jako nepřemýšlíme, kolikrát za vteřinu zavibrujeme hlasivkami, abychom zazpívali malé c. Něco jiného je například připravený projev nebo pečlivě nacvičená píseň, v takových případech lze při tréninku vycizelovat mnoho detailů.

V této oblasti podobností je největším rozdílem asi fakt, že hudba je daleko více uměleckým a emotivním projevem než řeč. Proto si může dovolit takové odchylky od „klasických pravidel“, které řeč jen těžko unese. Stejně jako je možno na různé nástroje hrát úplně jinak, než k čemu byly vytvořeny (což se v průběhu dvacátého století objevovalo častěji a častěji, setkáme se zde s rozličnými technikami jako klapání klapek flétny naprázdno, smýkání dřevěnou částí smyčce o houslové struny nebo klepání víkem od klavíru v rytmu skladby), hlasivkami a artikulačními orgány můžeme vytvářet nespočet tónů i ruchů. Jejich užití bude ale v řeči spíše nežádoucím elementem, kterému nepřikládáme žádný hlubší smysl. Naopak

[2] Technika takzvané dvanáctitónové sazby, která spočívá ve zrovnoprávnění všech dvanácti tónů temperované chromatiky.

v hudbě mají tyto odchylky v podstatě stejný význam, jako klasická hra (stále to „nic neznamená“).

Jinak se odchylky samozřejmě chovají v poezii. Zde se často (opět ve dvacátém století) setkáváme s citoslovci a různými hláskovými klastry – jedná se ale opět o umělecký projev, projev vyvolávající emoce. K čistě komunikační funkci řeči nenalézáme v hudbě dostatečnou obdobu.

Nezbytná a přirozená kategorizace jednotlivých prvků umožňuje u obojího grafický zápis, tedy písmo nebo notaci. Obojí má svou podobu pro běžné užití danou – opět se řídí ustálenými pravidly. Pro širší sdělení o akustickém průběhu pak existuje veliké množství symbolů, které běžný základ konkretizují. U řeči to mohou být rozšiřující diakritické a interpunkční znaky z mezinárodní fonetické abecedy (IPA) a u hudby pak všelijaká možná grafická znázornění kvality, místa vzniku nebo intenzity jednotlivých zvuků nebo jejich shluků. Ani písmo ani notace však nezaznamenává zvuk se stoprocentní přesností. Každý interpret přečte a reprodukuje zvuk jinak.

Další významnou a často zmiňovanou podobností je jedinečnost hudby i řeči jako kanálu komunikace člověka mezi ostatními živočichy. Na různých místech Země se vyvíjely kultury, které se v jistých obdobích nemohly nijak ovlivnit, přesto žádná z nich ani jednu z těchto domén nepostrádá. Dá se tedy říct, že to je záležitost týkající se opravdu celého lidského druhu. U řeči je to celkem pochopitelné, protože bez té by jen těžko nějaká fungující společnost mohla vzniknout. Nicméně platí to stejně tak i pro hudbu. Pro příklad můžeme uvést známý kmen Pirahã, který nejenže nemá ustálené výrazy pro barvy a v podstatě u něj nenajdeme výtvarné umění, ale dokonce postrádá jakákoli číselná vyjádření, výrazy má pouze pro číslovky jedna *(hói)* a dvě *(hoí)*, které se navíc liší jen tónem (Calapinto, 2007: 7). Hudba je však v tomto společenství hojně zastoupena a forma písní se nijak od jiných kultur neliší (Everett, 2005 in Patel, 2010: 3).

A jak je to tedy s onou jedinečností? Pojem řeč nebo hudba je u jiných živočichů těžké zavádět. Nacházíme u nich samozřejmě různé podobné formy komunikace, nicméně většinou na mnohokrát jednodušší (možná se tak však zdá pouze proto, že lidé jim nerozumí) nebo velmi odlišné bázi.

Co se týče řeči, vzhledem k tomu, že podle mnohé literatury (Yule, 1985; Fitch, 2006) zvířecí projevy nevykazují známky jakýchkoliv pokročilejších syntaktických celků, nejspíš toto označení nepřipadá v úvahu. Zároveň nejsou zvířata schopná komunikovat o samotné

komunikaci, což například Yule (1985) považuje za jeden ze základních rysů řeči. Jednou z dalších vlastností řeči je pak podle něj také to, že její mluvčí jsou schopni něčeho, co pojmenovává *displacement*. Tedy odkazování k něčemu, co není teď a tady. Pes jen těžko štěká o tom, co dělal předevčírem. Jedinou zajímavou výjimkou mohou být včely, které svým speciálním tancem umí navigovat ostatní jedince k nalezenému nektaru, jsou tedy schopny jakési prostorové abstraktní představivosti (Yule, 1985: 12). Tato schopnost však má svá úskalí, která se projevila v dost lítost vzbuzujícím pokusu, kdy se zjistilo, že včely umí sdělit informace o lokaci pouze v dvourozměrné podobě. Když měly pokusné včely úl na úpatí televizní věže a byly naváděny téměř až na vrchol, kde byl umístěn nektar, nedokázaly pak ostatním včelám sdělit, kam se ho mají vydat hledat. Ostatní včely trávily hodiny létáním několika kilometrů do stran. Yule později uvádí další důvody, proč se pojem řeč dá chápat pouze jako specificky lidskou záležitost (Yule, 1985: 13).

Co se týče hudby, z pohledu produkce, máme-li nějaký druh vokalizace nazvat výjimečným, určitě musíme zmínit ptačí nebo velrybí zpěv. U veškerých ostatních zvířat pozorujeme nesčetné způsoby zvukových projevů, abychom však něco mohli nazvat zpěvem (něčím, co má něco společného s hudbou), musí tento projev (podle Fitche) splňovat dvě kritéria. Zaprvé se musí vyznačovat čímsi, co do alespoň nějaké míry můžeme nazvat složitostí (komplexností). Tato složitost se zdá být poněkud vratkým pojmem, dá se však nějakými způsoby kvantifikovat, například jak minimálně dlouhá může být velrybí píseň, aby vůbec byla pro opačné pohlaví atraktivní (Fitch, 2006: 10). A zadruhé musí tento projev učit rodiče své potomky, musí být tedy přenášen z generace na generaci, mít podobu, která přetrvává život jednoho jedince. Toto například nesplňují madagaskarské žáby, které sice vyluzují až třicetišestislabičné sekvence, nicméně doklady o jakémkoli učení neexistují (Fitch, 2006: 10). Neopomeňme však nevokální hudbu. Mnozí primáti jsou schopni „bubnovat“ o své tělo a některé druhy kakadu dokonce mlátí větvičkami o strom (Fitch, 2006: 23).

Zajímavé je to u zvířat i z pohledu percepce. Právě primáti jsou schopni rozeznávat jednodušší diatonické melodie (například Old McDonald Had a Farm), ale na rozdíl od lidí mají daleko horší schopnost rozeznat melodie v transpozici (Trehub, 2003: 670). Pokud se posuneme o oktávu výše, reagují celkem obstojně. Posuny o kvintu nebo kvartu jsou pro ně již však velmi obtížné a nesrozumitelné. A toto platí obecně u většiny zvířat, což je jeden ze zásadních percepčních rozdílů lidí a ostatních živočichů. Lidé kladou o mnoho větší důraz na relativní výšku tónů (ibid.).

Je ale hudba stejně nebo alespoň podobně evolučně důležitá jako řeč? Vědce toto téma trápí již od dob Darwina, který označil lidské produkční i percepční schopnosti v hudbě za „jednu z nejzáhadnějších věcí, kterou bylo lidstvo obdařeno“. (Patel, 2010: 367) A ačkoli dnes převládá názor, že hudba je v lidské existenci velmi zajímavým a stále ne dost pochopeným prvkem, který však pro přežití a samotné existování není nezbytný, v historii jsme se setkali s mnoha pracemi i myšlenkovými směry, které této domněnce odporují.

Nejspíš prvním, kdo vůbec začal publikovat na téma, jestli se hudba například nějak nepodílí na přirozeném výběru, byl právě Darwin, na kterého navázalo bezpočet vědců včetně Millera (2000 in Patel, 2010: 368), který tvrdí, že hudba je analogií k ptačímu zpěvu a že se pomocí jí snaží samci zapůsobit na samice a soutěžit o ně s ostatními. Jako argumenty k tomuto tvrzení uvádí, že vrchol zájmu o hudbu a její produkci zažívají lidé ve věku dospívání a že muži hudebníci produkují daleko více hudby než ženy (na základě počtů vydaných nahrávek). Tato teorie má samozřejmě velké množství nedostatků, Huron (2003 in Patel. 2010: 369) například poukazuje na to, že zatímco u zvířat je produkce „hudby“ dána mnohdy tím, že pro ni samci mají speciální fyziologické dispozice, u lidí se vokální ústrojí jednotlivých pohlaví nijak zvlášť anatomicky neliší. S dalším tvrzením přišel Cross (2003 in Patel, 2010: 369), který uvádí, že hudba hraje velkou roli v mentálním rozvoji a ve zdokonalování sociálních schopností. Nicméně nikdy nebylo prokázáno, že by třeba lidé s amúzií (porucha vnímání výšek tónů, rytmu nebo metra) měli problémy se socializací.

Tím se však dostáváme k další rovině hudby i řeči, kterou je bezesporu jakási jejich schopnost sjednocovat lidi. Jak uvádí Kivy v *Music, language and cognition*, proslýchá se, že když se Franz Josef Haydn[3] v padesáti osmi letech chystal na své cesty do Anglie, před odjezdem se setkal s Wolfgangem Amadeem Mozartem, který mu řekl: „Ale papá, vždyť toho tolik nevíte o celém širém světě a mluvíte tak málo jazyky!“ Načež mu Haydn odpověděl: „Ale mému jazyku rozumí lidé po celém světě.“ (Kivy, 2007: 215)

Těžko domýšlet, co si představovali „papá“ Haydn a Mozart pod pojmem celý svět, patrně by v oné době Japonci, Číňané či Austrálci[4] jen těžko rozuměli Haydnově hudbě, stejně tak jako by Haydn rozuměl hudbě jejich. Nicméně Haydn tím správně zachytil myšlenku, že pokud lidé dané hudbě, žánru hudby rozumí, pak je spojuje. Ačkoli si třeba nerozumí v řeči. A naopak, pokud si dva lidé nerozumí v hudbě, ale jsou schopni ovládat společný jazyk, jsou si

[3] Klasicistní hudební skladatel žijící mezi roky 1732 a 1809.
[4] Původní obyvatelé Austrálie, někdy také Aboridžinci.

pochopitelně schopni výborně porozumět. Výhodu má hudba v tom, jak už jsme zmínili, že má jistou (pro někoho) srozumitelnou syntax, ale postrádá přímou sémanticitu.

Dá se jen těžko soudit, jak moc je možno docílit této schopnosti u dalších uměleckých odvětví, v nichž postrádáme verbální složku. Nejspíš v tomto případě hraje velkou roli ona „časovost", kterou má právě řeč i hudba společnou. Předpokládáme, že obraz, socha nebo budova nevyvolává v lidech takový pocit sounáležitosti jako společný prožitek z hudebního nebo tanečního vystoupení. Koneckonců na této jednotě přímo staví i různé druhy pohybového umění, jako například eurytmie. Eurytmie má v člověku rozvíjet fantazii, vůli a empatii, přičemž Rudolf Steiner tento druh umění postavil právě na základech podstaty řeči a hudby.[5]

Ještě než se zaměříme konkrétněji na složku melodie v řeči a hudbě, je nutno poukázat na to, že smyslem této části práce zaměřené na analogie není tvrzení, že hudba a řeč je to samé nebo že hudba je vlastně druh řeči či jazyka. Úmyslem bylo pouze nastínit všechny možné úhly pohledu na souvislosti těchto dvou domén. Sám Kivy tvrdí, že ono Haydnovo „rozumění" nemá se skutečným rozuměním řeči co dělat. Toto chápání hudby je spíš schopnost užít si a ocenit dílo v těch složkách, ve kterých to pravděpodobně bylo zamýšleno. Pokud si člověk libuje v kombinování dvou témat v Bachově fuze nebo ocení v Beethovenově Eroice předčasný nástup lesního rohu, pak se dá říct, že poslouchané hudbě rozumí. Tento způsob percepce je tak velmi odlišný od percepce u lidské řeči (Kivy, 2007: 218).

Kivy je obecně ke všem těmto analogiím velmi skeptický. Obzvlášť pak k samotným pojmenováním jedné domény na základě druhé. Kde se vlastně vzaly takové definice, jakože hudba je „mezinárodní jazyk/řeč" nebo „jazyk emocí"? Historie těchto pojmenování sahá až hluboko do sedmnáctého století k vzniku žánru zvaného „opera", nicméně abychom pochopili celý problém, musíme se vrátit ještě o kus dál do dob Tridentského koncilu. Jedno z mnoha témat tohoto koncilu byla vzrůstající nesrozumitelnost vokální hudby. Polyfonie se stala tak složitou a samotná hudební složka tak dominantní, že skoro nebylo možné rozumět, co je vlastně obsahem zpívaného. Dlouhé slabiky přes několik taktů, více než šest hlasů na sobě nezávislých, to vše srozumitelnosti jistě nepomáhalo. K vyřešení tohoto problému, tedy jak skládat hudbu tak, že je rozumět všemu, co se v ní zpívá, se přišlo s jednoduchým návrhem. Je prostě potřeba z řeči vycházet a napodobovat její přirozenou intonaci i rytmický průběh. Nyní se již můžeme posunout ke vzniku opery – tehdy vznikající žánr se těmito pravidly řídí víc, než kterýkoli jiný. Jestliže má někde hudba co nejvíce napodobovat

[5] https://cs.wikipedia.org/wiki/Eurytmie

přirozenou mluvu, pak je to právě v opeře. Vzhledem k tomu, že se v sedmnáctém století také už hudba oprostila od modálního myšlení (které se vrátilo zpět až o mnoho let později) a zakotvila pořádně v dur-mollovém systému, bylo zde celkem omezené množství výrazových prostředků, jak v hudbě doprovázet onu náladu, kterou zpěvák vyjadřoval a zpíval. Jakýmsi přirozeným vývojem (nemůžeme říct, že arbitrárním, nicméně nemůžeme tvrdit ani opak) se ze dvou možností moll přisoudil smutné náladě a dur veselé. Od té doby se úzus tak zažil, že člověk, vyrostší v naší kultuře, neměl na vybranou a samozřejmě tento jev vnímal. A vnímáme to v evropské, a nyní už nejen v evropské, kultuře dodnes. I člověk v hudbě nikterak vzdělaný má vžité, že dur je veselý, moll smutný. Zajímavé by bylo zkoumat tuto percepci zpětně. Pokud bychom pustili třeba Griegovo *In the Hall of Mountain King*, kolik procent lidí by poznalo, že to je v mollové tónině (i přes veselost skladby). Pokud tedy Ind (v době Haydna) neumí poznat, která část symfonie je veselá a která smutná, není to tím, že by zažíval smutek nebo štěstí jinak než Evropané, ani tím, že by je jinak vyjadřoval. Je to tím, že prostě nevnímá onu hudbu samotnou. Je mu tedy cizí evropský *hudební* smutek a evropské *hudební* veselí (Kivy, 2007: 221).

Hudba proto není ani „jazykem emocí“ ani „mezinárodní řečí“. Pouze se v ní musíme učit číst emoce, stejně jako se učíme číst francouzsky nebo německy. Emotivní charakter dělá tedy pak podle Kivyho hudbu „*language-like in that respect.*“ (Kivy, 2007: 222)

2.2. Melodie řeči a její komunikační funkce

Některé naprosto základní termíny ale hudba s řečí sdílí. Vzhledem k tomu, že prozodie se občas uvádí jako „věda o těch jevech, které řeč sdílí s hudbou“ (Volín, 2018 – přednáška v rámci Základů prozodie), názvy některých prozodických složek (melodie a tempo) jsou i hudebními termíny. V naší práci je jistě nejdůležitější prozodickou složkou melodie, které budeme věnovat pozornost v celé praktické části. Dá se však vytvořit taková definice, která by popisovala melodii hudební i melodii řeči?

Ringer (2001: 363 in Patel, 2010: 182) definuje melodii jako „zvuky o určité výšce umístěné v hudebním čase v souladu s danými kulturními konvencemi a omezeními“. Patel upomíná na to, že to vypadá, že by se z této definice dalo pouze vynechat ono „v hudebním čase“ a dostaneme obecnou definici melodie. Nicméně dodává, že by pak tento popis splňovala například i evropská ambulantní houkačka, kterou můžeme těžko nazvat melodií (Patel, 2010: 182).

Dalo by se diskutovat, zda by byl opravdu takový problém nazývat houkačku melodií, ale to přenechme jiným. Patel dále uvádí jako lepší vymezení „uspořádaný sled tónových výšek, který sděluje širokou škálu informací“. Kromě zdůraznění důležitosti oné bohatosti sdělení zde také vyzdvihuje fakt, že posluchač vnímá onen sled jako celek – ze kterého pak právě čte komplexní informace (Patel, 2010: 182).

Ať tedy však existuje nějaká obšírná definice, která by důsledně „obě“ melodie popsala, či nikoliv, je jisté, že melodie v hudbě je něco jiného než melodie v řeči. Než se proto dostaneme ke krátkému shrnutí melodie řeči a jejích funkcí, ještě se na chvíli vrátíme k porovnání obou domén, tentokrát s větším důrazem na hledání rozdílů, nikoliv analogií.

Především melodie v hudbě na rozdíl od melodie řeči vychází z jistého setu daných použitelných intervalů. Ačkoli jsou tyto intervaly v různých kulturách velmi odlišné, nemáme ve světě žádný doklad o hudbě, která by měla melodii a neměla určené, které intervaly smí použít. Tato vlastnost samozřejmě pak hudbě umožňuje chovat se tonálně a směřovat k určitému centru. Jako pravděpodobný důvod, proč se řeč nedrží pouze povoleného inventáře intonačních kroků, se jeví to, že řeč míchá lingvistickou intonaci s intonací afektivní do jediného kanálu (Patel, 2010: 205). (Na otázku, zda by se percepce řeči změnila, pokud by bylo množství použitelných intonačních kroků omezené, možná částečně odpoví výsledky našeho výzkumu.)

Z toho částečně plyne další důležitý rozdíl, totiž že jsou tóny v hudbě (kromě glissanda) odděleny pomocí intervalů skokově, kdežto melodie řeči je plynulá, mezi jednotlivými slabikami klouže podobně jako třeba hudební nástroj těremin. (Nicméně pouze fyzikálně, percepčně ne, jak bude dále vysvětleno.)

Mezi různými výškami tónů existuje pak v hudbě hierarchie (nebereme-li v potaz techniky dvacátého století, jako již zmíněnou dodekafonii), kdežto v řeči se nic obdobného nikdy neprokázalo.

Také díky omezenému množství těchto intervalových, ale i metrických a temporálních prostředků si spíš zapamatujeme hudební melodii než melodii řeči. U hudby totiž, jakožto u estetického objektu, končí sled tónů (podle Patela) sám u sebe, kdežto melodie řeči je teprve prostředkem k docílení příslušného efektu. Jak píše Patel možná trochu příliš mysticky, pokud je hudební melodie „skupinou tónů, které se navzájem milují“ (Shaheen in Hast, Cowdery & Scott, 1999 in Patel, 2010: 184), pak je melodie v řeči „skupinou tónů, které spolupracují, aby zdařile dokončily svůj úkol“ (Patel, 2010: 184).

Vraťme se k melodii řeči neboli intonaci v užším slova smyslu. Termín intonace je sice kratší, ale lehce zavádějící, neboť ve fonetické literatuře odkazuje k různým skutečnostem. Takzvanou intonací v širším slova smyslu je myšlena prozodie tj. agregát melodie, hlasitostních průběhů, proměnlivých charakteristik barvy hlasu a vlastností časového uspořádání (Skarnitzl, Šturm & Volín, 2016: 124), zatímco intonace v užším slova smyslu odkazuje právě k melodii řeči, k průběhu základní frekvence (F0) v čase.

> *Většina zvuků, z nichž se běžná řeč skládá, je znělá. V každé slabice musí být sonorní jádro, což je nejčastěji samohláska. Dále může být ve slabice souhláska sonorní, znělý obstruent nebo obstruent neznělý. Během znělých segmentů (samohlásek, sonorních souhlásek a znělých obstruentů) kmitají hlasivky a vytvářejí zvuk tónového charakteru, který je bohatý na frekvence, ovšem jeho nejsilnější složkou je frekvence základní, tj. ta, která odpovídá rychlosti kmitání hlasivek. Tato frekvence, označovaná jako F0, je také nejvíce odpovědná za vnímanou výšku, tedy např. to, zda slyšený hlas nebo pronesenou promluvu posluchač vnímá jako vysokou, nízkou nebo ve střední poloze* (Skarnitzl, Šturm & Volín, 2016: 125).

Pokud má tedy většina segmentů schopnost nést informaci o základní frekvenci, nutně tato vnímaná výška musí mít nějaký průběh – to je právě melodie řeči. Tento průběh bude jen těžko náhodný, odráží náš psychický stav i pragmatické aspekty promluv. Proto je také pro děti snazší nejprve si osvojit intonaci a pak až se učit jednotlivá slova a jejich významy (Bolinger 1978 v Duběda 2005: 175). Bohužel vzhledem k tomu, že je intonace v užším slova smyslu suprasegmentálním jevem, je pro ni velmi obtížné najít tak přesné a kategorizovatelné funkce a pravidla, jako pro jiná lingvistická odvětví typu morfologie, syntax a podobně. Skarnitzl, Šturm a Volín (2016) však uvádějí jakési shrnutí do pěti základních obecnějších funkcí, které by bylo na místě pro přehlednost zmínit.

Autoři zmiňují jako první funkci lexikální, nicméně ta se jako jediná netýká českého jazyka, na němž jsou v této práci prováděny experimenty, proto ji zmíníme a rozvedeme až jako poslední. Přejděme tedy k funkcím gramatickým.

Gramatické funkce zahrnují funkce větně významové a funkce členící. Funkce větně významová je schopná rozlišit u některých jazyků (jako je čeština) například některé typy otázek od oznamovací věty. V češtině mohou zjišťovací otázky a oznámení mít stejnou syntaktickou i lexikální podobu, v textu je odlišujeme pouze otazníkem, v řeči však

interpunkci zavedeme jen těžko. Proto je potřeba tyto dvě jednotky odlišit intonací. Věta *Jmenuji se Bernard* může být otázkou i oznámením, pokud se však tážeme, na konci promluvy nutně zvýšíme F0. Funkce členící se pak týká vhodného uspořádávání slov do melodických celků, nazývaných intonační jednotka, intonační fráze, promluvový úsek (Duběda, 2005) nebo prozodická fráze (Skarnitzl, Šturm & Volín, 2016: 131). To může mít vliv na zapamatovatelnost sdělení, ale v některých případech to může i rozlišit význam. *„Ale ošetřovatel," poznamenal vrátný, „je sadista."* Oproti *Ale ošetřovatel poznamenal: „Vrátný je sadista."* (Skarnitzl, Šturm & Volín, 2016: 131)

Další funkcí je funkce diskurzní. Pomocí té můžeme řídit v komunikaci střídání mluvčích nebo vyjadřovat, zda jsme už myšlenku dokončili, či ji ještě chceme rozvést. Účastníci komunikace jsou tak schopni odhadnout, kdy už mohou začít mluvit a kdy mají ještě vyčkávat na dokončení repliky druhého. Tímto druhem intonace se podrobněji zabývá konverzační analýza.

Nezbytnou součástí promluv ve všech jazycích světa je pak funkce afektivní. Podle Skarnitzla, Šturma a Volína reálná každodenní komunikace vlastně nezná neutrálnost. Jakákoli promluva je afektivně zabarvena, což se pochopitelně neprojevuje pouze v intonaci, ale ve všech prozodických složkách. Odrážejí tak emoce, nálady, postoje, aktuální ladění a osobnostní vlastnosti mluvčího. Prostředky pro vyjádření afektivních funkcí však nejsou po celém světě stejné, každý jedinec jedná tak, jak je to v jeho společenství považováno za náležité.

Což lehce souvisí i s čtvrtou funkcí, tedy funkcí indexovou. Ta vyjadřuje příslušnost mluvčího k sociální skupině, ze které pochází. Například jinak intonačně vyjadřuje doplňovací otázky člověk z Plzeňska než člověk z Prahy a specificky tvoří otázky zjišťovací lidé z Olomouce (Skarnitzl, Šturm & Volín, 2016: 132).

Na závěr se vraťme k funkci lexikální, která se týká takzvaných tónových jazyků. V takových jazycích se za tón považuje fonologický intonační příznak vázaný na slovní jednotku (Duběda, 2005: 159). Ačkoli je tento koncept Evropanům poměrně cizí, ve skutečnosti je takových jazyků více než netónových (Gussenhoven, 2004). Tóny jsou samozřejmě relativními změnami ve výšce F0. Absolutní tóny žádný známý jazyk nevyužívá z mnoha důvodů. Jako jeden z nich zmiňme, že není mnoho tónů, které jsou v hlasovém rejstříku všech lidí ve společenství. Oblast okolo tónu c1 mohou zazpívat i basy i soprány, nicméně pro muže je to daleko namáhavější a mluvit permanentně v takové oblasti by bylo prakticky nemožné.

Oblíbeným příkladem pro představu, jak tónové jazyky fungují, je mandarínská čínština, kde může například slovo transkribované artikulačně pouze jako [ma] znamenat čtyři významy (matka, kůň, konopí a nadávat) a liší se pouze tonálním průběhem (Duběda, 2005: 159).

Než se dostaneme k samotnému vnímání základní frekvence, zmiňme ještě, že melodie řeči má kromě svých funkcí i několik inherentních vlastností, které neovládáme vědomě. Jednou takovou vlastností jsou takzvané spádové jevy. V důsledku toho, že s mluvením ubývá dechu i energického potenciálu mluvních orgánů, je zde tendence k tomu, aby hodnota F0 v průběhu nějak vymezené intonační jednotky klesala.[6] Lidský mozek však tento pokles vnímá jako nepříznakový, naopak v některých jazycích má lingvistickou funkci jeho zabránění (Di Cristo & Hirst, 1998: 25).

2.3. Percepce základní hlasivkové frekvence

Vzhledem k tomu, že jakýkoliv tón vytvořený fonačním ústrojím člověka je zvuk s neharmonickým průběhem (není možné vytvořit hlasivkami například zvukovou vlnu v podobě sinusoidy), vždy je ve zvukovém spektru tvořeného tónu přítomno i mnoho vln o frekvencích s hodnotami celočíselných násobků základní hlasivkové frekvence. Pokud tak hlasivky vibrují frekvencí 120 Hz (120 kmitů hlasivek za sekundu), s vlnou o této frekvenci spolu zároveň vytvářejí i vlnu o frekvenci 240 Hz, 360 Hz, 480 Hz a podobně. Vlny o těchto frekvencích pak vytvářejí takzvanou alikvótní řadu. V jednotkách hertzů jsou rozdíly mezi těmito hodnotami stále stejné, v půltónech, které jsou subjektivní jednotkou, samozřejmě ne. Pokud je skok o oktávu posun o 12 půltónů a zároveň zdvojnásobení frekvence v Hz, pak nám vychází, že pro posun o jeden půltón je potřeba frekvenci v Hertzech vždy vynásobit dvanáctou odmocninou ze dvou. Hodnoty v oné alikvotní řadě (120 Hz, 240 Hz, 360 Hz, 480 Hz a dále) jsou si pak vzdáleny odpovídající počet půltónů. V základní alikvótní řadě jsou tyto počty 12, 7, 5, 4, 3, a dále. Slyšíme tedy fundamentální tón spolu s oktávou nad ním, kvintou nad touto oktávou, další kvartou nad touto kvintou a podobně.

Mozek však pro vnímání základní frekvence nepotřebuje vnímat všechny tyto vlny dohromady. Jejich kombinace, a hlavně jejich jednotlivé amplitudy, ovlivňují percepci barvy tónu, nikoliv výšky. Proto stačí rozpoznat rozdíl mezi pouhými dvěma sousedními vlnami. V naší již zmíněné řadě jsou všechny vlny celočíselnými násobky té základní, takže se všechny rozdíly mezi sousedními vlnami rovnají hodnotě základní hlasivkové frekvence.

[6] https://www.czechency.org/slovnik/VĚTNÁ%20INTONACE#antikadence

Podle Goldsteina (1973) a Terhardta (1979) je v mozku jakýsi centrální procesor, který tento největší společný násobek rozpoznává. Nejefektivnější je pak podle nich rozpětí od třetího do šestého alikvótu. V naší příkladové řadě by to tak byly vlny o hodnotách 480 až 840 Hz.

Tento způsob je velmi ekonomický, neboť můžeme vnímat základní frekvenci, aniž by se ve spektru vlna o takové frekvenci vůbec nalézala. Proto také například v pevné telefonní lince, ačkoli je schopna tato síť přenášet pouze frekvence o hodnotách 300 – 3400 Hz, můžeme rozeznat nižší mužské hlasy. Tón o základní frekvenci 80 Hz by náš mozek tak rozpoznal z alikvótů 320, 400 a 480 Hz, které jsou nejnižšími jejími celočíselnými násobky v tomto rozmezí.

Pokud tyto vlny nejsou příliš blízko u sebe (rozdíl mezi jejich frekvencemi v půltónech není příliš malý), dokonce mohou postačit, jak už bylo naznačeno, pouhé dvě pro rozpoznání základní frekvence i za podmínky, že jedna je o 20 až 25 dB slabší než druhá (Cardozo, 1972; Houtsma, 1980 in ‘t Hart, Collier & Cohen, 1990: 26).

Nejnižší vnímatelná a rozpoznatelná frekvence je pro člověka někde kolem hodnoty 20 Hz, zatímco nejvyšší hodnoty se pohybují kolem 20 kHz. Nejnižší frekvence vnímaná jako tón byla v laboratorních podmínkách naměřená jako 12 Hz, nicméně celé rozmezí 4 – 16 Hz vnímáme převážně senzomotorickým systémem (Olson, 1967).

Co se týče trvání, základní frekvence může být vnímána už od úseku o délce 30 ms (Cardozo & Ritsma, 1965 in ‘t Hart, Collier & Cohen, 1990: 26). Tento údaj se však nedá vztahovat na percepci intonace v mluvené řeči, neboť při dotyčném výzkumu měli respondenti k dispozici izolované nahrávky ohraničené tichem. Řeč je oproti tomu kontinuální záležitostí, kde je každý vnímaný intonační bod ohraničený dalšími (‘t Hart, Collier & Cohen, 1990: 26).

V Bachemově (1937) experimentu se ukázalo, že přesného určení výšky samostatného tónu jsou schopni pouze takzvaní „absolutní sluchaři“, u ostatních respondentů se průměrná chyba pohybovala mezi 5 až 9 půltóny. Pro percepci základní hlasivkové frekvence u řeči je však daleko zásadnější vnímání rozdílu mezi dvěma sousedními intonačními jednotkami. J. ‘t Hart, R. Collier a A. Cohen uvádějí v *A perceptual study of intonation*, že jsou respondenti schopní snadno rozlišit mezi dvěma tóny rozdíl 1 až 1,005 Hz, nicméně neuvádí, v jakých frekvencích. 1 Hz mezi 50 a 51 Hz je úplně jiný rozdíl než mezi 720 a 721 Hz. Dále percepci rozdílu mezi dvěma tóny opět ovlivňuje trvání, čím kratší dobu jsou zvuky respondentům pouštěny, tím hůře poznají, zda byl mezi nimi rozdíl (Cardozo & Ristma, 1965). J. ‘t Hart všechny tyto experimenty shrnuje se slovy, že nebyly nalezeny žádné velké rozdíly mezi

vnímáním frekvence „psychoakusticky“ a „psychofoneticky“. Tedy žádný pokus nevykazuje podstatnou diskrepanci mezi percepcí syntetizovaných zvuků a percepcí syntetizované řeči.

U vnímání přirozené řeči se nicméně ukázalo, že člověk vůbec nepotřebuje vnímat F0 každých oněch 30 ms, které je schopen zaznamenat. Skarnitzl, Šturm a Volín (2016) například uvádějí, že u znělých souhlásek jako [b], [z] nebo [ɦ] je už tak obtížné udržet hlasivky v chodu, že ještě soustředit se, jakou výšku tónu člověk zvolí, by bylo extrémně náročné.

Hermes (2006) pomocí percepčních experimentů ukázal, že jsou z hlediska percepce melodie řeči nejdůležitější hodnoty F0 uprostřed samohlásek, tedy nejvíc v prostřední třetině jejich trvání. Do 100 ms vnímá mozek pouze jednu hodnotu, nad 100 ms je schopen vnímat i případný pohyb F0 v rámci vokálu (Skarnitzl, Šturm & Volín, 2016: 128). Tento Hermesův výzkum bude hrát poměrně klíčovou roli i v naší praktické části.

Nejenže tedy mozek nemusí vnímat F0 v každé hlásce, ale naopak dokonce v těch segmentech, kde hlasivky nevibrují (a není tak možno určit základní frekvenci), si melodický průběh sám doplňuje, vytváří tak souvislou intonační konturu pro celou prozodickou jednotku.

2.4. Hypotézy pro předkládanou studii

V předchozích oddílech jsme si nastínili rozdíly a souvislosti dvou domén, hudby a řeči. Následně jsme se zaměřili na jednu prozodickou složku, totiž melodii řeči. U ní jsme zmínili, že na rozdíl od hudby nevychází z daného počtu intonačních kroků, distribuce intervalů je tedy spojitá.

Jak by se ale vnímání změnilo, pokud bychom inventář použitelných intervalů omezili? Jak by se změnilo, pokud bychom v řeči rozpoznali pouze hudební intervaly jako malá a velká sekunda, malá a velká tercie a podobně?

Tuto otázku nám zjednodušuje již zmíněný výzkum, že percepce F0 probíhá intenzivně v prostředních třetinách slabičných jader. Pokud tedy v těchto úsecích bude zprůměrovaná F0 zarovnána do půltónové škály, melodie řeči bude v této škále vnímána – mohli bychom ji bez používání čtvrtkřížků a čtvrtbéček zapsat do notové osnovy.

Lidé v evropských kulturách jsou na půltónovou (chromatickou) stupnici již zvyklí a rozumí jí. Bylo by tedy možné uvažovat, že v případě, že by někdo mluvil pouze v hudebních

intervalech, našemu sluchu by to bylo příjemnější, neboť mozek má přirozené sklony ke kategorizaci.

Předpokládejme tedy nulovou hypotézu H0, že po harmonizaci intonačních kroků se percepce nijak nezmění. Alternativní hypotéza Ha1 tvrdí, že percepce se změní kladně. Lidé tedy budou vnímat upravené promluvy příjemněji. Případná druhá alternativní hypotéza Ha2 pak tvrdí, že mozek takové úpravy bude vnímat negativně.

3. Metoda

3.1. Materiál

Nahrávky byly vyňaty z dvanácti zvukových stop mluveného slova Českého rozhlasu. Tyto stopy byly konvertovány z youtube kanálu Český Rozhlas Dvojka ze sekcí Mluvené slovo[7] a Historie českého zločinu[8]. Z nich byly pečlivě vybrány takové promluvy, které odpovídají zhruba délce 2,5 až 5 sekund a které, pokud možno, nejsou ničím příliš nápadné, tedy že se v nich nevyskytují žádná příliš expresivní, archaická ani onomatopoická slova, vlastní jména, složená čísla a podobně.

Počet promluv na každého mluvčího se lišil, nejméně byla jedna – tito mluvčí byli posléze použiti na zácvik. Maximum bylo pět, nicméně pro výzkum byly použity nejvýše dvě. Všechny úryvky od každého mluvčího byly seřazeny za sebou do jedné nahrávky, která se stříhala až později.

Promluvy byly následně vybírány tak, aby se v nich nevyskytoval žádný ruch v pozadí – hudba nebo zvuky prostředí. Dalším prvkem, který by mohl nepříznivě ovlivnit výzkum, by mohly být příliš známé hlasy mluvčích. Byly proto vybrány nahrávky s méně známými herci.

Vytříděné nahrávky formátu **.wav**[9] byly v Adobe Audition upraveny ze dvou kanálů (falešné stereo[10]) do jednoho (mono). Dále bylo manipulováno se vzorkovací frekvencí – z původních 44 100 Hz na 32 000. Bitová hloubka byla snížena z 32 bitů na 16.

Vzhledem k nevyrovnanostem ve vnímané hlasitosti byly všechny nahrávky amplifikovány tak, aby se vlny o největší amplitudě pohybovaly kolem absolutní hodnoty 6 dBFS. Pokud se i v rámci jedné nahrávky příliš lišila hladina zvuku jednotlivých promluv, byly amplifikovány zvlášť.

Tímto způsobem upravené celé nahrávky byly posléze otevřeny v programu Praat, kde jim byly přiděleny textgridy[11]. V textgridech byly ohraničeny jednotlivé promluvy z obou stran, aby bylo možno posléze použít skript, kterým se tyto ohraničené úryvky ořezaly a samostatně uložily jako zvukové stopy.

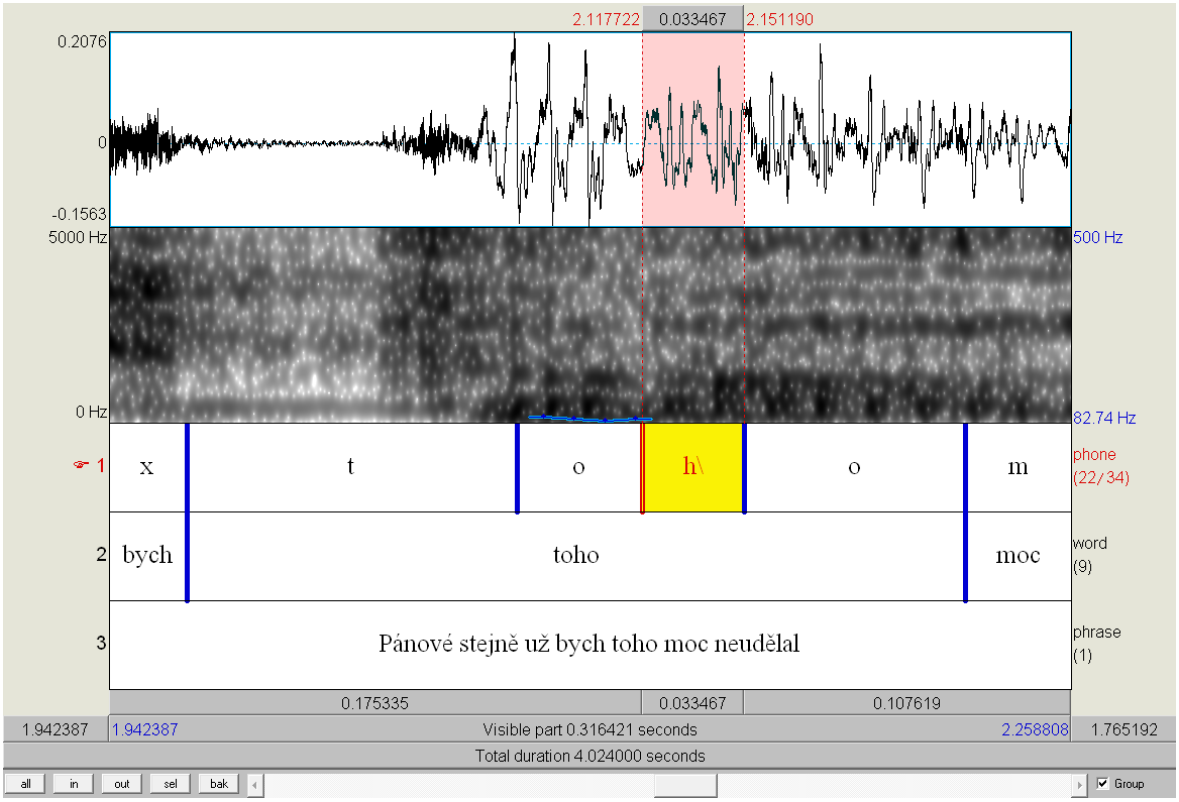
[7] https://www.youtube.com/playlist?list=PL8Vad8nI5Kr2xINiM_2C3jX0cwsP3le6M
[8] https://www.youtube.com/playlist?list=PL8Vad8nI5Kr2NnQrAV1GaSur4JkzcrTgP
[9] Zkratka výrazu *Waveform audio file format.*
[10] Zvuk vychází ze dvou kanálů, které jsou nicméně totožné.
[11] Typ souborů, který program Praat umí číst, a pomocí nějž je možno anotovat a segmentovat nahrávky.

Pomocí programu Prague Labeller (Pollák, Volín & Skarnitzl, 2007) byly funkcí ALIGN všechny tři vrstvy textgridu rozsegmentovány podle příslušných jednotek, tedy vrchní vrstva *phone* na hlásky, *word* na slova a *phrase* na celé promluvy. Takto byly textgridy uloženy.

Dalším krokem bylo manuálně doupravit veškeré potřebné hranice hlásek. Skript je umí rozpoznat velmi přesně, nicméně ne dokonale. Pro naše účely stačilo zarovnat co nejpřesněji hranice všech samohlásek, popřípadě slabikotvorných souhlásek. Jak již bylo mnohokrát prokázáno, základní intonační jednotkou je slabika. Je to tedy nejmenší jednotka řeči, na které je lidské ucho schopno vnímat změny ve výšce tónu, intonaci (‘t Hart & Collier & Cohen, 1990). Nositeli základní frekvence jsou pak v jednotlivých slabikách jejich jádra – právě ony samohlásky. V rámci samohlásky je nejinformativnější prostřední třetina celého jejího trvání. První třetinu člověk ještě nestihne percepčně pojmout a poslední už bývá zase ovlivňována následující hláskou a jejími specifiky (Volín, 2009). Právě z tohoto důvodu bylo nutné samohlásky opravdu přesně ohraničit, neboť budou vypovídat o stěžejních vnímaných bodech intonačního průběhu u každé promluvy.

Samohlásky na počátcích slov či jejich kořenů byly ohraničeny s nástupem základní frekvence, tedy po případném glotalickém rázu (manuálně přidaném do textgridu, neboť náš segmentovací skript ho nebyl schopný vytvářet). Tímtéž pravidlem se řídilo i ohraničování samohlásek na konci celé promluvy. Pravá hranice vokálu byla vytvořena tam, kde už autokorelační metoda praatu neidentifikovala základní frekvenci. Problémem v těchto finálních samohláskách mohla být třepená fonace celého vokálu. Ta je přirozená a často se týká nejen posledních hlásek, ale i několika posledních slov promluvy. V takových případech jsme se řídili plností formantové struktury.

Žádné větší potíže se neobjevovaly při určování hranic frikativ či exploziv a vokálů. Frikativy jsou obecně jedny z těch nejsnáze identifikovatelnějších hlásek, jejich trvání přesně odpovídá trvání frikativního šumu, který můžeme v oscilogramu i spektru velmi zřetelně nalézt. Jediná z frikativ, u které se občas dalo váhat, bylo /h/. Pokud se navíc /h/ třeba vyskytovalo intervokalicky v úseku s třepenou fonací, frikativní šum (už tak u /h/ méně zřetelný) nebylo možné vůbec pozorovat. V takových příkladech jsme se řídili právě formantovou strukturou a hranicemi úseků, kde program aspoň částečně spočetl F0 (viz obr. 3.1).

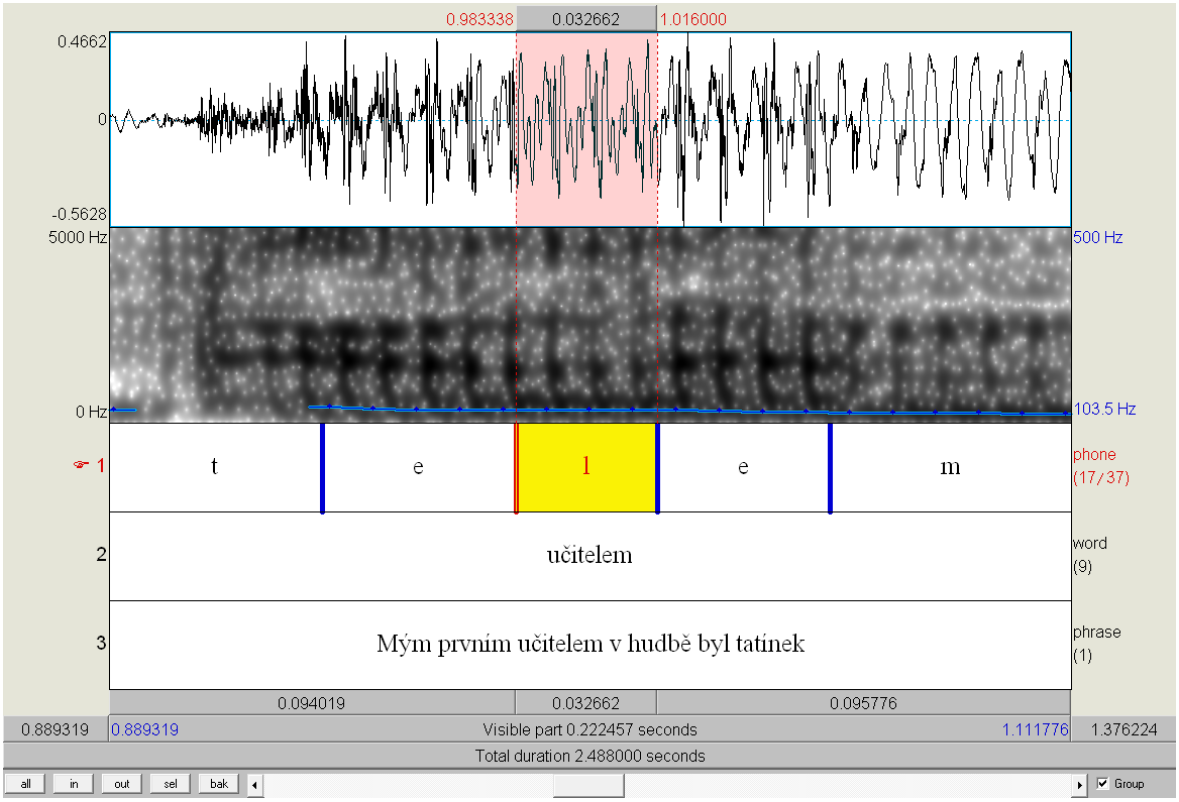


Obr. 3.1. Příklad určování hranice laryngální frikativy v intervokalické pozici.

Méně jasné je ohraničování samohlásek v okolí sonor. Nejmenší problémy se objevovaly u nazál, /m/ má velmi specifickou spektrální strukturu a /n/ pouze silně ovlivňuje sousedící vokály, nicméně i poslechem lze hlásky od sebe celkem přesně rozdělit. Nejplynulejší, a tím pádem nejproblematičtější přechody vykazuje /ň/ a následná přední samohláska /i/ nebo /e/. Zde jsme se řídili pravidlem rozdělovat tyto hlásky přesně v polovině úseku o nejasné příslušnosti k té či oné hlásce. I po oddělení je totiž ve vokálu silně slyšet palatalizace, a pokud chceme docílit „čistého“ vokálu, pak už je v předchozím /ň/ zase slyšet zřetelný nástup samohlásky.

Co se týče likvid, hláska /r/ většinou nedělala žádné potíže, zato /l/ ano. I u ní ale šlo pomocí postupné kontroly sluchem (zároveň s hledáním slabších formantů ve spektru) hranice nakonec poměrně přesně určit (viz příklad na obr. 3.2).

Nejproblematičtější skupinou sonor jsou pro ohraničování obecně aproximanty. V češtině tím míníme především /j/. V okolí samohlásek /a/, /o/ a /u/ se dle formantové struktury daly předěly poměrně dobře umístit. Nicméně v okolí /e/ a /i/ (předních vokálů) jsme museli v naprosté většině případů umístit hranici do již zmíněné poloviny délky úseku nejasné příslušnosti.



Obr. 3.2. Příklad určování hranice laterální aproximanty v intervokalické pozici.

Po manuálním dorovnání samohlásek ve všech osmdesáti úryvcích jsme rozdělili mluvčí do pěti kategorií za účelem co nejmenšího chybování při extrakci základní hlasivkové frekvence od tzv. pitchtierů.[12] Dvacet jedna mužů bylo rozděleno do tří kategorií a sedm žen do dvou.

Kategorie byly určeny vždy místem s nejvyšší základní frekvencí. Pokud byla například v nejvyšším bodě nahrávky frekvence 167 Hz, zařadili jsme mluvčího do kategorie s nejnižšími hlasy. Ve výsledku jsme vytvořili tedy horní hranice mužských kategorií 180 Hz, 230 Hz a 270 Hz. Do nejnižší kategorie jsme přiřadili 5, do střední 11 a do nejvyšší opět 5 mluvčích. U žen byly hranice 280 Hz a 310 Hz. Do první připadly 4 mluvčí do druhé 3.

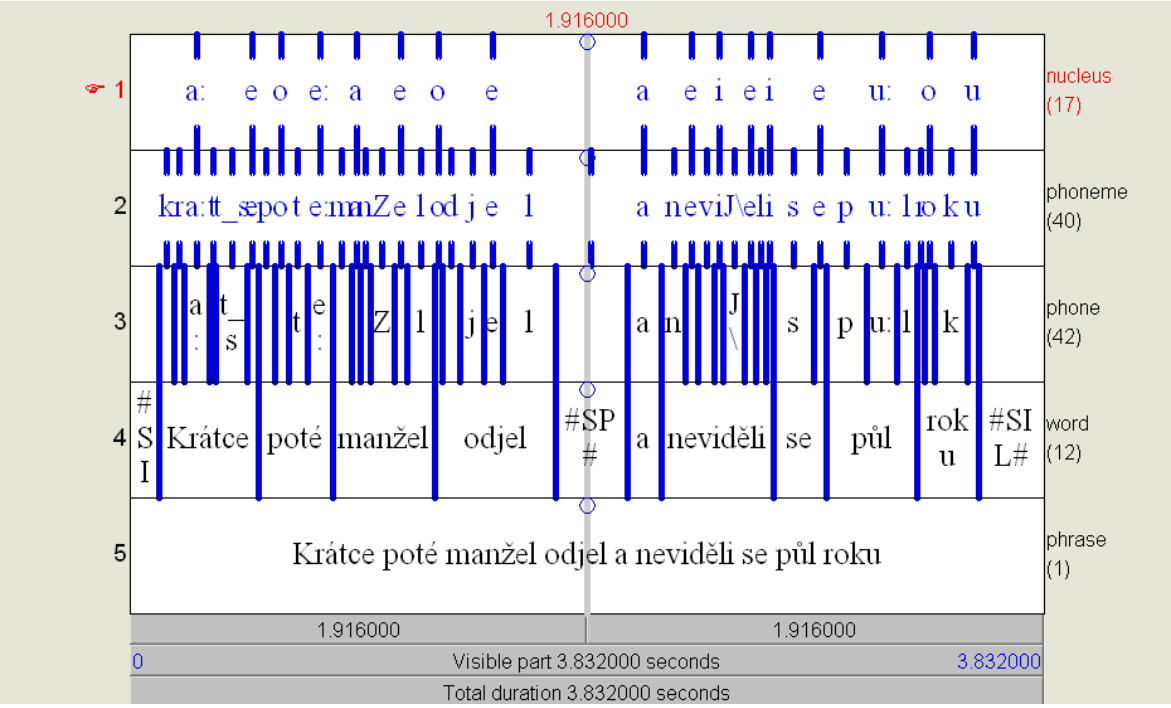
Po těchto skupinách jsme vygenerovali pitchtiery a zkontrolovali jsme, jestli se v nich neobjevily chybné hodnoty. Autokorelační metoda, kterou program Praat používá k určení F0 občas mylně vybere jednoho z 15 kandidátů a v křivce F0 pak vznikají nepatřičné skoky do odlehlých (často o mnoho vyšších) hodnot. Těmto chybám jsme se snažili vyhnout právě v onom ořezání nejvyšší reálné F0.

Dalším krokem pro samotné vynesení frekvenčních hodnot každé slabiky bylo vytvoření bodové vrstvy v textgridech, která vynesla z každého hláskového intervalu přesnou polovinu

[12] Pitchtier je křivka F0 dané promluvy vygenerovaná programem Praat.

trvání. Jiným skriptem jsme poté z těchto bodů do páté vrstvy vynesli pouhá jádra slabik – samohlásky a diftongy. Bohužel tento skript neuměl rozpoznat slabikotvorné souhlásky /r/ a /l/. Z celkových 80 úryvků naštěstí tyto hlásky obsahovalo pouze 8 nahrávek, přičemž v každé se nacházela jedna. Takovéto množství tedy nebylo obtížné manuálně dooznačit.

Výsledné textgridy pak obsahovaly 5 vrstev: základní vrstvu obsahující celou větu podle ortografie, vrstvu slov, vrstvu hlásek, vrstvu bodů v polovině trvání hlásek a vrstvu bodů – jader slabik (viz obr. 3.3).



Obr. 3.3. Znázornění pěti vrstev v textgridu.

K osmdesáti textgridům jsme načetli všech osmdesát pitchtierů a aplikovali jsme na ně skript, který vypočte průměrnou F0 z prostřední třetiny trvání každého slabičného jádra. Vynesené hodnoty se zobrazily v okně praat info, ze kterého se dají jednoduše zkopírovat do tabulkového procesoru Microsoft Excel 2007, a následně se s nimi dá pracovat. Než jsme pokračovali v manipulacích s nahrávkami, v tomto bodě jsme se pozastavili a vynesli ze zjištěných statisticky poměrně zajímavých dat několik poznatků a grafů.

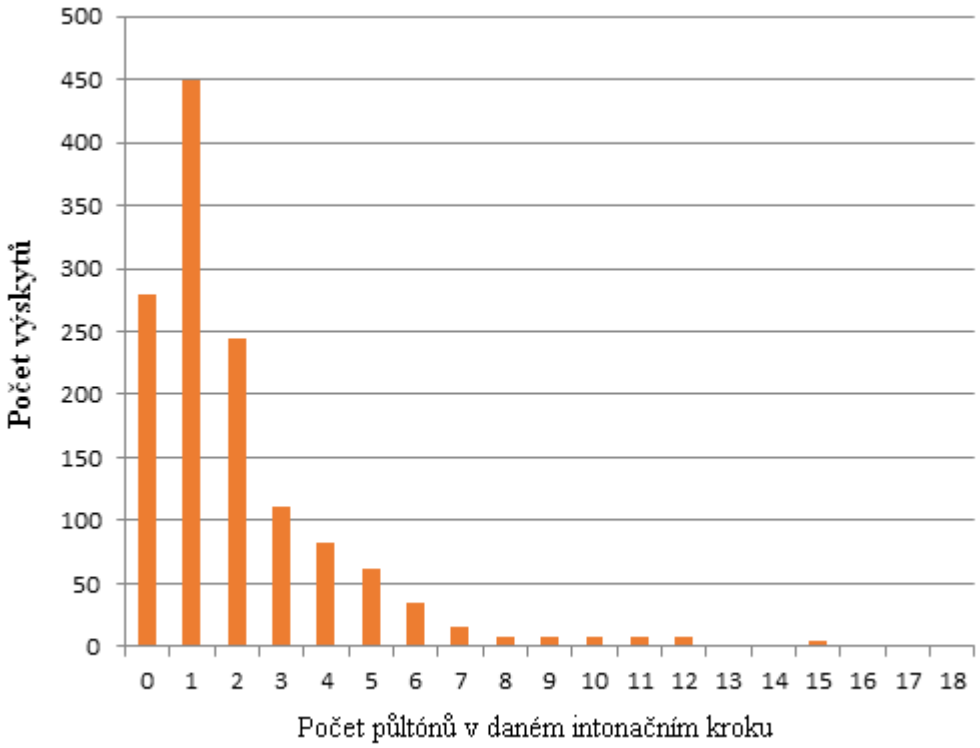
Vůbec nejvyšší frekvence byla naměřena u jedné z ženských mluvčích. Tato frekvence dosahovala hodnoty 286 Hz, což je něco mezi cis1 a d1 (podle ladění na 442 Hz). To je celkem překvapivé, neboť v běžné mluvě ženy hovoří někdy daleko výše, dokonce i ve dvoučárkované oktávě. Naše úryvky byly nicméně vyňaty z nahrávek mluveného slova, kde panují úplně jiné podmínky než v normálním rozhovoru. Klade se zde veliký důraz na

srozumitelnost a emotivní složka, která většinou hodně ovlivňuje intonaci, je zde notně potlačována.

Naopak nejnižší naměřenou frekvenci jsme pochopitelně zaznamenali u mužského mluvčího, konkrétně to byla hodnota 52,4 Hz, což odpovídá zhruba kontra as. Tyto velmi hluboké hodnoty najdeme u mnoha mluvčích na konci frází, nicméně ne vždy se praatu v těchto úsecích podařilo identifikovat základní frekvenci.

Průměr z hodnot všech frekvencí ve všech jádrech slabik vykazuje 128,7 Hz, což odpovídá malému c. Tento údaj se samozřejmě nedá nijak generalizovat na populaci, neboť v našich datech převládají muži a nahrávky byly pořízeny za okolností odlišných od běžné spontánní mluvy.

Zjištěné frekvenční hodnoty všech slabičných jader jsme následně pomocí vzorce přenesli na vzdálenost od frekvence 100 Hz v půltónech. Tato hodnota (100 Hz) je čistě náhodná, je to hodnota referenční. Jinými slovy jsme vytvořili půltónovou stupnici, zahrnující v sobě hodnotu 100 Hz. Každý frekvenční údaj tak má buď kladnou hodnotu (počet půltónů nad 100 Hz), nebo zápornou (počet půltónů pod 100 Hz).



Obr. 3.4. Frekvence výskytů intervalů mezi sousedními slabičnými jádry (zaokrouhleno na celé půltóny).

Toto referenční rozdělení nám skvěle poslouží pro určení vztahů mezi jednotlivými slabikami. Nejprve jsme se zaměřili na samotné půltónové rozdíly mezi sousedními vokály. Vynesli jsme si proto tabulku s veškerými rozdíly ve všech 80 úryvcích.

Pro demonstrování názorných statistik jsme nejprve zaokrouhlili všechny tyto rozdíly k nejbližšímu celému číslu – tedy do naší umělé referenční půltónové stupnice (obr. 3.4).

Není překvapivé, že skoky o malý interval jsou daleko častější než o velký. 73 % všech třinácti set devatenácti intervalů po zaokrouhlení jsou menší nebo rovny velké sekundě. Možná je zajímavější, že skoky o malou sekundu o tolik převyšují primy, tedy setrvání na stejném půltónu (tab. 3.1).

Velkých sekund je podobně jako prim, zatímco malých tercií je už přibližně dvakrát méně. Zajímavostí je, že intervalů o 8, 9, 10, 11 a 12 půltónech jsme všech našli sedm. Z toho však nelze usuzovat nic statisticky významného.

půltónů	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Zastoupení (v %)	21,2	34,0	18,5	8,4	6,3	4,7	2,7	1,1	0,5	0,5	0,5	0,5	0,5	0,2	0,0	0,3

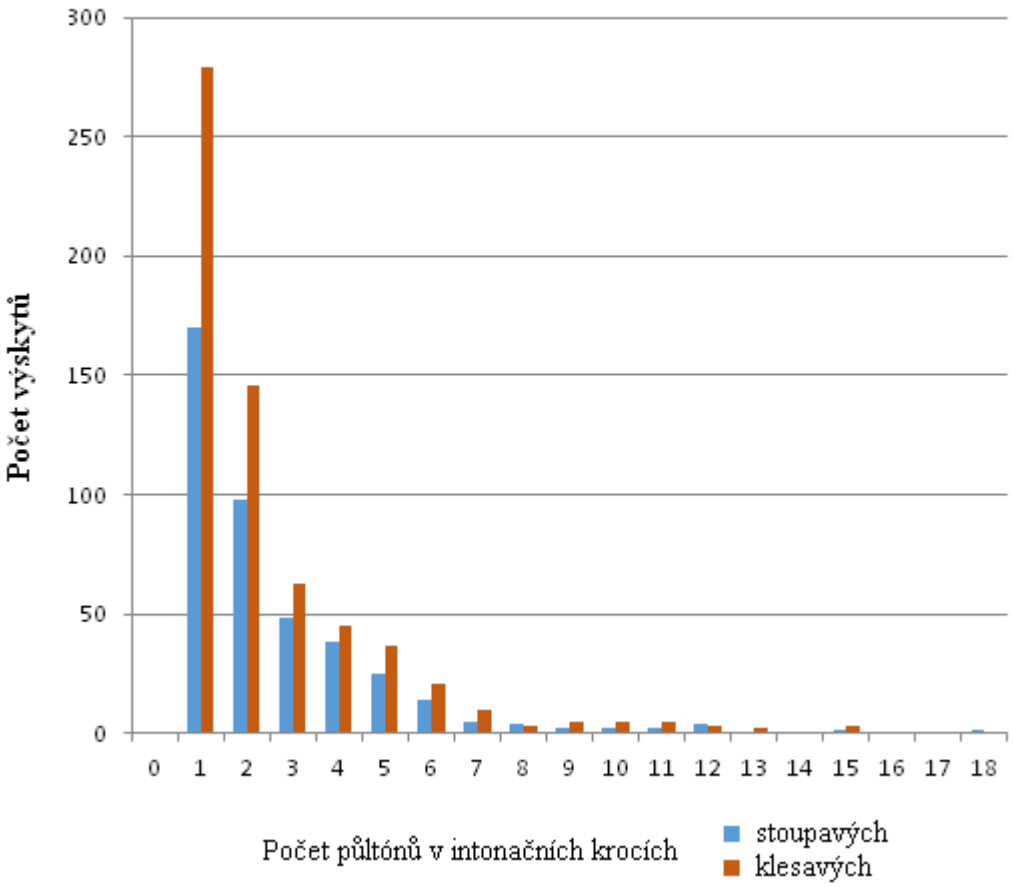
Tab. 3.1. Procentuální zastoupení jednotlivých intervalů (v půltónech).

Je vcelku překvapivé, že se v datech vyskytly i skoky o 15 půltónů – tedy malou decimu. Tyto údaje opravdu nejsou chybnými, našli jsme je u dvou mužských mluvčích.

Zatím nebylo zmíněno, jestli jsou intervaly klesavé nebo stoupavé. To nám osvětlí následující graf (viz obr. 3.5).

Vzhledem k zaokrouhlování na nejbližší číslo jsme nemohli u nulových kroků rozlišovat klesavost či stoupavost, uvádíme ji tedy až od půltónového kroku.

Můžeme sledovat jasnou převahu klesavých intervalů nad stoupavými. Lidská řeč prokazuje v promluvách různé druhy deklinace (Patel, 2010: 184). Jedna z nich je ta intonační, melodie každého promluvového úseku klesá a ke konci se občas přerodí až v třepenou fonaci. To bude velmi pravděpodobně také hlavní důvod, proč klesavé intervaly tolik převažují. Nicméně je zajímavé sledovat podíl klesavosti a stoupavosti u jednotlivých intervalů (viz tab. 3.2).



Obr. 3.5. Frekvence výskytů stoupavých a klesavých intervalů mezi sousedními slabičnými jádry (zaokrouhleno na celé půltóny).

půltónů	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
podíl k/s	1,64	1,49	1,31	1,18	1,48	1,5	2	0,75	2,5	2,5	2,5	0,75	--	--	3

Tab. 3.2. Podíl klesavých a stoupavých zastoupení jednotlivých intervalů v půltónech. (Podíl *2* k/s u hodnoty *7* půltónů tak znamená, že klesavých čistých kvint bylo dvakrát více než stoupavých.)

V prvních čtyřech, nejpočetnějších, intervalech (malá a velká sekunda a tercie) můžeme pozorovat značný pokles v hodnotách tohoto podílu (klesavých a stoupavých výskytů). Dat u větších intervalů není nejspíš dostatečné množství, aby to mohlo být považováno za statisticky relevantní.

Poměr klesavých sekund je daleko větší než poměr klesavých tercií. Můžeme usuzovat, že to je tím, že pokud se mezi dvěma slabikami posuneme o tercii, většinou jde už o nějaké příznakové jednání. Expresivní, direktivní či jiné. Tyto větší intervaly nepostihuje celková deklinace promluv tolik.

Naopak deklinace ovlivňuje daleko více stagnující intonaci v prostředních (nejklidnějších) částech intonačních frází – melodie tedy mírně klesá. Nejmírnějším harmonickým intervalem je pak právě malá sekunda. Možná proto v sekundových intervalech tedy převažují ty klesavé.

Z jiného úhlu pohledu, pokud je v promluvách přítomna jakási přirozená deklinace, stoupání je příznakovějším jevem (Vassiere, 2005). Je méně přirozené, spontánní, více vědomě ovládané. Pokud už tedy vědomě stoupáme, abychom docílili žádaného efektu, dalo by se soudit, že je malá sekunda příliš malá na to, aby byl tento efekt zřetelný. Proto není stoupavá malá či případně velká sekunda v porovnání s klesavým protějškem tak častá.

Pokud nebudeme intervaly zaokrouhlovat na celá čísla, můžeme vysledovat, kolik procent jejich celkového počtu leží v jakých půltónových rozmezích (viz tab. 3.3).

půltónové rozmezí	**0-1**	**1-2**	**2-3**	**3-4**	**4-5**	**5-6**	**6-7**	**7-8**	**8-9**	**9-10**	**10-11**	**11-12**	**12-13**	**13-14**	**14-15**	**15-16**
Podíl všech intervalů (%)	40,0	25,2	12,7	7,3	5,8	3,0	1,7	0,8	0,5	0,5	0,5	0,8	0,2	0,2	0,2	0,2

Tab. 3.3. Podíl všech intervalů pohybujících se v daném půltónovém rozmezí.

Pozorujeme podobné statistiky jako se zaokrouhlenými hodnotami. Tentokrát necelých 78 % intervalů je v rozmezí primy až malé tercie. Zároveň se do této statistiky promítnou i malé intervaly, které v zaokrouhlených hodnotách nefigurovaly – celých 40 % všech intervalů je menší než půltón.

Po tomto statistickém pozastavení jsme se dále ubírali k hlavnímu bodu naší práce, tedy k manipulaci s nahrávkami a jejich percepčnímu testu.

Vypočtené základní frekvence pro každé slabičné jádro vyjádřené v kladné či záporné půltónové vzdálenosti od frekvence 100 Hz jsme následně zaokrouhlili na celá čísla. Tato celá čísla nám nyní určovala hodnoty, kterých budeme chtít pomocí manipulací následně dosáhnout. Tedy pokud bylo původně jisté slabičné jádro 9,6 půltónu od frekvence 100 Hz, nyní bude vzdálené půltónů 10.

V tabulkovém procesoru jsme tedy měli dvě tabulky. Jednu s reálnými hodnotami, jednu se zaokrouhlenými. Nyní jsme vytvořili třetí tabulku, ve které byly zaznamenány rozdíly odpovídajících si hodnot tabulky 2 a tabulky 1. Pokud jsme tedy zaokrouhlili nahoru, výsledný rozdíl byl kladný, pokud dolů, rozdíl byl záporný.

O tyto rozdíly jsme později u každé příslušné slabiky posunuli základní frekvenci. V podstatě jsme tedy vetkli veškeré naše frekvenční hodnoty slabičných jader do oné naší umělé půltónové stupnice zahrnující hodnotu 100 Hz.

Percepci těchto pozměněných nahrávek samozřejmě může ovlivnit mnoho faktorů. Jedním z nejzásadnějších by mohla být jakási míra pozměněnosti nahrávek. Problémem však bylo, jak tuto míru hodnotit. Jako zásadní změnu jsme shledali pozměnění intonačních kroků, nikoliv výšku jednotlivých slabičných jader. Pokud by kupříkladu nějaký mluvčí mluvil v naprosto čistých půltónových krocích, ale všechny by byly o čtvrttón výš, než jsou půltóny v naší umělé stupnici, zaznamenali bychom u něj největší míru pozměněnosti, přitom bychom vůbec neupravili melodii. Jako další možnost se nabízelo tedy nebrat jako referenční hodnotu oněch 100 Hz, ale frekvenci prvního vokálu. Zde by ovšem opět mohl nastat problém, a to tehdy, kdyby například první vokál byl o čtvrttón vyšinutý a všechny zbylé by byly zarovnány v naší stupnici. Pak by nám vyšla míra pozměněnosti opět vysoká, ačkoli bychom měli vyjma prvního vokálu naprosto čistý intonační průběh.

Rozhodli jsme se proto, že budeme zkoumat míru změny u samotných intervalů mezi sousedními vokály. Kromě již zmíněné a popsané tabulky s původními vypočtenými intervaly jsme tedy vytvořili i další, v níž jsou vypočteny intervaly mezi již zaokrouhlenými hodnotami slabičných jader. Z těchto dvou tabulek jsme pak opět vytvořili třetí, která uvádí rozdíly prvních dvou. Nyní jsme tedy mohli vidět, o kolik se změní každý intonační krok mezi dvěma vokály. Další tabulka, kterou jsme zhotovili, pak uvádí absolutní hodnoty všech těchto změn. U každé nahrávky jsme pak spočetli součet všech těchto změn a vydělili ho počtem slabik, abychom dostali průměrnou hodnotu pozměnění u každého intervalu.

Podle těchto hodnot jsme následně vybírali nahrávky, které použijeme do percepčního testu. Hodnoty se pohybovaly od 0,155 po 0,447 půltónu na interval. Nutno poznamenat, že zatímco u jednotlivých frekvenčních hodnot slabičných jader jsme nedostali nikdy rozdíl větší než 0,5 (protože jsme zaokrouhlovali desetinná na celá čísla), rozdíly u jednotlivých intervalů mohly teoreticky dosáhnout až hodnot blížících se 1. Pokud bychom například měli dva po sobě jdoucí vokály s hodnotami 4,51 a 2,49 (půltónů od 100 Hz), interval mezi nimi by byl 2,02

půltónu. Nicméně po zaokrouhlení jednotlivých slabik na celá čísla, tedy na hodnoty 5 a 2, je interval již třípůltónový. Hodnota pozměnění intervalu je tedy 0,98 půltónu.

Pro snazší kategorizaci jsme vybírali pouze nahrávky ze zhruba první, třetí a páté pětiny pořadí těchto průměrných hodnot pozměnění. Pro přehlednost jsme určili první skupinu, kde je tato míra menší než 0,25 půltónu na slabiku (dále jen p/s), druhou skupinu o míře větší než 0,29, ale menší než 0,32 p/s a třetí skupinu o míře větší než 0,35 p/s. Z těchto poměrně vyrovnaných skupin (co se týče počtu položek v nich obsažených) jsme následně vybrali samotné krajní hodnoty (z každé skupiny tedy dvě) a několik dalších, přičemž jsme se snažili, aby každý vybraný mluvčí měl ve výběru alespoň dvě nahrávky.

Dohromady jsme tedy chtěli docílit počtu třiceti nahrávek (+ tři zaučovací). Každá nahrávka bude mít ve výběru dvě formy – jednu upravenou, jednu neupravenou. Nutno dodat u neupravených, že to nejsou nahrávky původní, ale resyntetizované. Při manipulacích s nahrávkami v programu Praat se vždy trochu zhorší kvalita a vlastnosti zvukového souboru. Upravené nahrávky by pak měly vždy horší kvalitu, což by mohlo být značně rušivým elementem při vyhodnocování závěrů. I neupravené jsme proto resyntézou LPC „zhoršili“ a uložili pro percepční test.

Pokud bylo tedy třeba docílit třiceti nahrávek o dvou formách, dostali jsme celkových 60 úryvků, přičemž jsme alespoň pět z nich chtěli zopakovat, neboť se to pro ověření konzistence účastníků percepčního testu se u podobných studií žádá.

Pokoušeli jsme se o vyrovnanost i v jednotlivých hlasových rozpětích mluvčích. Tedy o podobný počet mužských nízkých, mužských středních, mužských vysokých, ženských nízkých i ženských vysokých hlasů (obr. 3.6).

x>0.35	x<0.25	0.29<x<0.3
K3	S3	Ď2
T1	I2	R4
S2	Ď1	W1
O3	U2	D2
P4	X1	M2
P3	X2	Q1
W3	D1	N4
N1	L1	K1
E1	L2	B2
E2	R2	A2

Obr. 3.6. Rozdělení nahrávek do skupin. Jednotlivé barvy odpovídají hlasovým rejstříkům mluvčích, velká písmena označují mluvčí a číslice označují konkrétní nahrávku.

Jelikož náš celkový materiál nebyl úplně vyrovnaný v pohlaví a hlasovém rejstříku, nemůže být pochopitelně vyrovnaný ani náš výběr, tyto aspekty ale nejsou součástí výzkumu. Je třeba na ně později tedy brát ohled, ale neřídit se jimi.

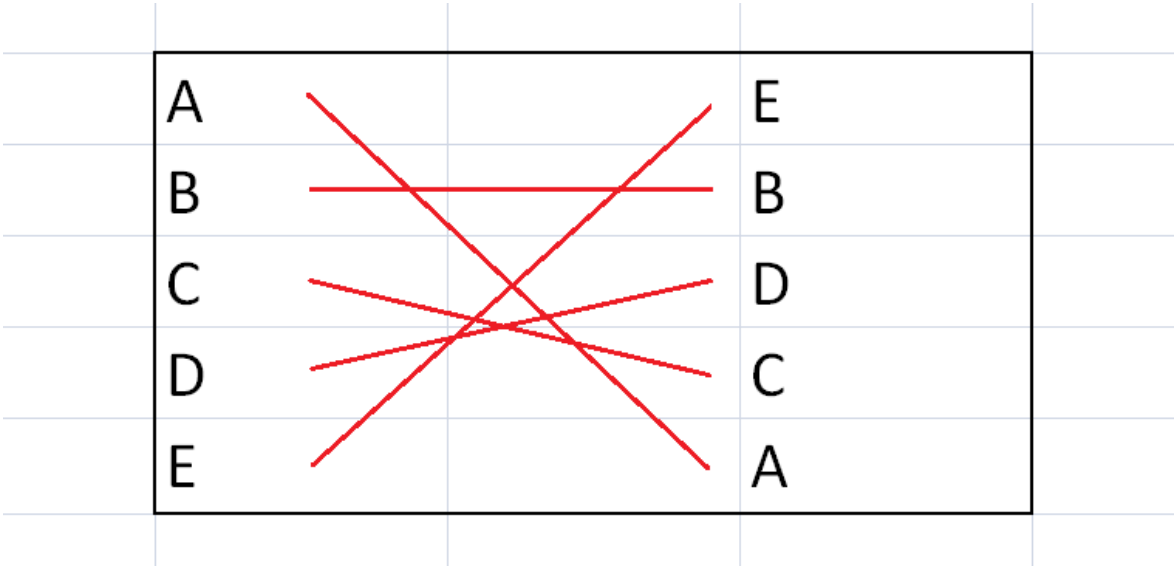
Nyní jsme přistoupili k určení pořadí nahrávek. Pro čistě náhodné pořadí jsme použili karty na hru prší. Vybrali jsme třicet z nich, zamíchali je, rozložili náhodně rubem navrch, ty jsme náhodně opět seskupili do balíčku a ještě jednou zamíchali. Každé kartě jsme přisoudili číslo od 1 do 30. V abecedním seznamu nahrávek určených do testu jsme tak posléze postupně od A do X přiřazovali pořadí určené kartami. Když byla první vytažená karta kulová desítka, nahrávce A2 jsme přisoudili pořadí 20. (Karty jsme ohodnotili tak, že nejmenší hodnotu mají žaludy, pak listy, kule a největší srdce.)

Toto náhodné pořadí jsme si následně zobrazili v tabulkovém procesoru. Vzhledem k příliš kompaktním shlukům nahrávek ve stejné kategorii míry pozměněnosti jsme dvě dvojice nahrávek navzájem vyměnili – naše pořadí bylo ve výsledku tedy pseudonáhodné.

Těchto třicet položek tedy tvořilo zhruba polovinu percepčního testu. Pomocí opětovného náhodného rozdělení (rozhodli jsme se, že v první polovině bude náhodných patnáct nahrávek již upravených a patnáct resyntetizovaných, v druhé polovině pak naopak) jsme určili u všech úryvků hodnotu upravenosti. Tedy U (upravená) / N (neupravená). Kulové a srdcové karty znamenali upravenost, listové a žaludové neupravenost. Toto pořadí zůstalo náhodné, nebyly zde příliš nápadné shluky nutící nás k rozhodnutí se k pseudonáhodnosti.

Počet třiceti nahrávek je poměrně velký, nicméně pokud bychom ve stejném pořadí zopakovali i dalších třicet, nejspíš by respondenti zaregistrovali, že jdou po sobě nahrávky v úplně stejném pořadí. Abychom tomu zabránili, pořadí jsme v druhé půli promíchali. Nebylo však možno míchat pořadím moc, protože by se mohly některé úryvky moc přiblížit svému protějšku v první půli. Proto jsme se rozhodli promíchat nahrávky po pěticích. Nemohlo se tedy stát, že by se navzájem si odpovídající nahrávky příliš přiblížily a zároveň nebylo mnoho nahrávek, sousedících spolu v první půli, sousedy i v druhé.

V pěticích jsme pořadí tedy rozházeli takto:



Obr. 3.7. Jednotlivé pětice v pořadí jsme v druhé půli rozhodili vždy následujícím způsobem. Pouze několik nahrávek tak spolu stále sousedilo, nicméně v opačném pořadí.

Nyní již bylo potřeba pouze někam vměstnat ony nahrávky „přespočet“ (pro pozorování konzistence respondentů). U nich jsme se opět snažili, aby si nebyly příliš blízko. Proto jsme vždy vybrali nahrávku, určili polohu jejího protějšku a zhruba do poloviny jejich vzdálenosti umístili onu třetí.

Tyto nahrávky „přespočet“ byly také celkem rovnoměrně rozděleny. Tři byly upraveny, dvě neupraveny. Tři z nich jsou mužští mluvčí střední výšky hlasu, zbylé dvě nahrávky pak jeden mužský nízký a jeden ženský nízký hlas.

Po výběru potřebných nahrávek do testu jsme začali s manipulacemi. V tabulkovém procesoru jsme již měli zaneseny veškeré potřebné změny, šlo tedy pouze o manuální nastavování základní frekvence u každého vokálu.

U každé nahrávky jsme postupovali stejným způsobem. Nejprve jsme si v praatu otevřeli příslušný zvukový soubor, textgrid a pitchtier. Následně jsme v kolonce *Manipulate* vytvořili soubor *Manipulation*. Při vytváření souboru *Manipulation* jsme volili hraniční frekvenční hodnoty rozpoznávání velmi podobně jako při generování pitchtierů, vycházelo to pochopitelně z hlasového rejstříku daného mluvčího. *Time step* jsme měli nastavený na 0.01 sekundy.

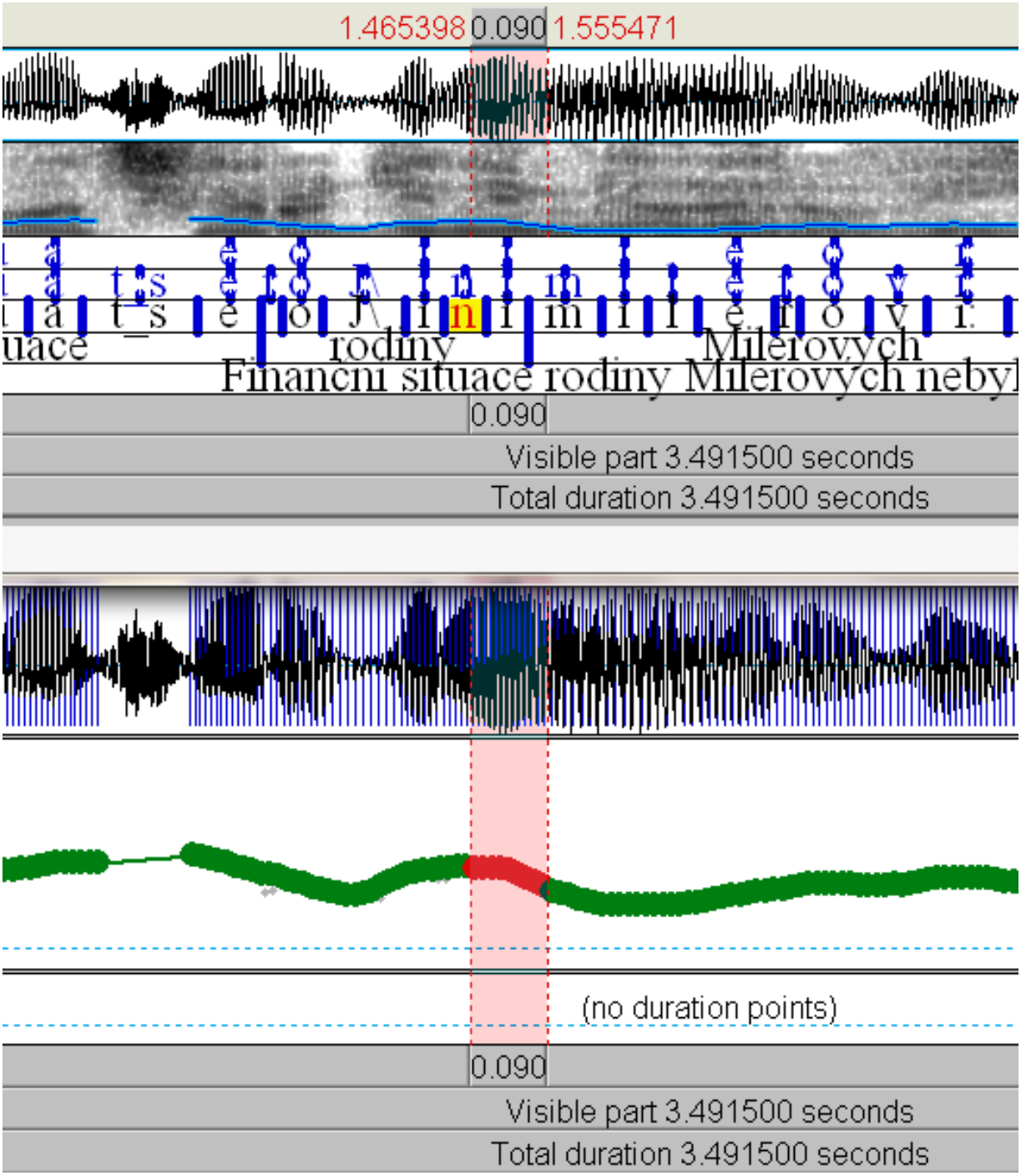
Poté jsme označili soubor *Manipulation* spolu s pitchtierem a zvolili funkci *Replace pitchtier*. Nyní se nám v souboru *Manipulation* zobrazovaly body onoho pitchtieru, z něhož jsme původně počítali frekvenční hodnoty slabičných jader.

Z tohoto souboru jsme nejprve pomocí funkce *Get resynthesis (overlap-add)* vygenerovali zvukový soubor, který má mírně horší kvalitu než ten původní, nicméně tuto kvalitu sdílí s těmi později upravenými. Označili jsme ho písmenem R (resyntetizovaný).

Po opětovném otevření souboru *Manipulation* jsme začali s upravováním. Vzhledem k tomu, že, jak už bylo několikrát řečeno, člověk vnímá intonaci pouze na slabičných jádrech, upravovali jsme vždy celé trvání vokálu a případný překryv do sousedních souhlásek. Hlavní pointou bylo, aby se jádro vokálu přesunulo na naši umělou půltónovou stupnici. Že se s tím posunula základní frekvence i ve zbytku vokálu a v okolních souhláskách nebylo na závadu, naopak jsme alespoň nevytvořili zbytečně neúměrné skoky.

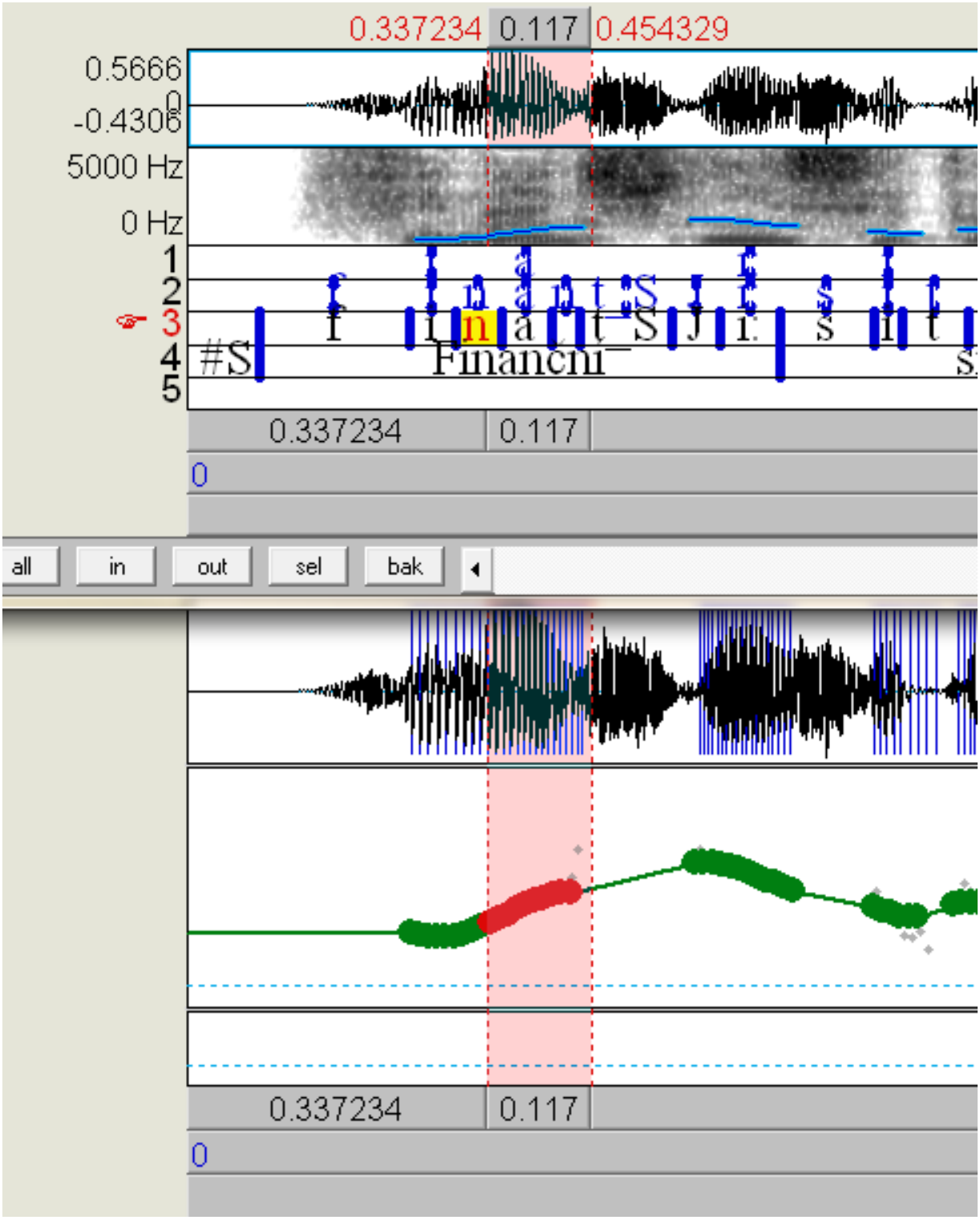
Spolu se souborem *Manipulation* jsme otevřeli ve stejně širokém okně textgrid se zvukovým souborem a funkcí *Group* tato okna spárovali. Pokud jsme tedy označili nějaký úsek v jednom okně, označil se ten samý v okně druhém.

Samohlásky uvnitř delšího úseku, v němž praat rozpoznal základní frekvenci, jsme upravovali tak, jak bylo zmíněno výše. Tedy menší kus předcházející a menší kus následující souhlásky plus samozřejmě celé trvání samohlásky (viz obr. 3.8).



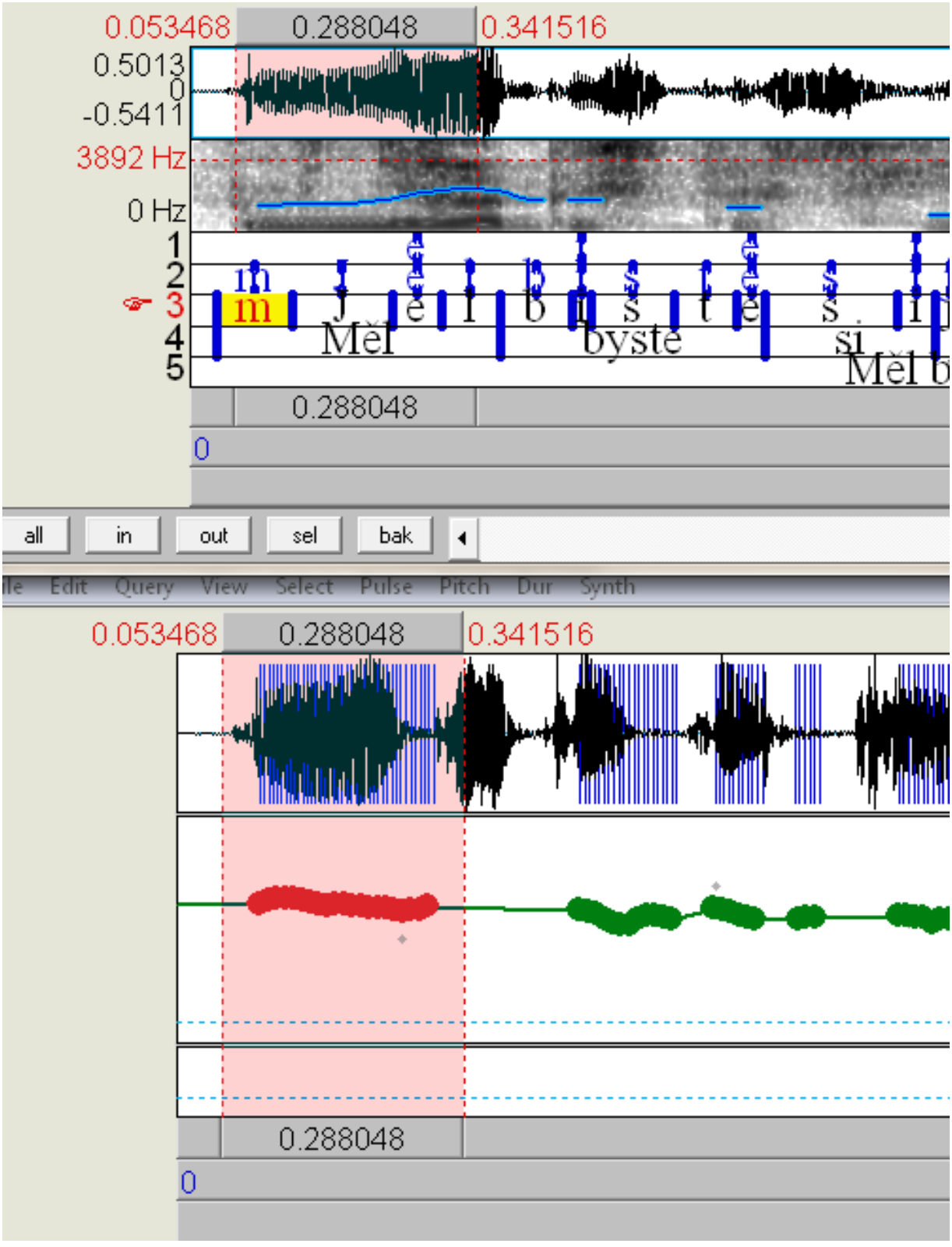
Obr. 3.8. Označení vokálu v delším úseku o rozpoznané základní frekvenci.

Pokud byla na jedné straně od vokálu část, kde praat F0 nerozpoznal (a do níž se nevyskytoval ani jeden další vokál), například pokud se tam vyskytovala neznělá souhláska nebo pauza, označili jsme toto místo až k hranici rozpoznání (viz obr. 3.9).



Obr. 3.9. Označení vokálu slabiky, na jejímž kraji není rozpoznaná F0. Na obrázku vidíme, že označení pokrývá ještě skoro celé druhé /n/, které sousedí s vokálem.

A pokud se slabika nacházela na úseku rozpoznání ohraničeném zleva i zprava, označili jsme tento úsek celý (viz obr. 3.10).



Obr. 3.10. Označení vokálu ve slabice, která nese samostatný úsek rozpoznané F0.

U několika málo nahrávek se stalo, že se po funkci *Replace pitchtier* se u některého vokálu špatně interpretovala frekvence a skákala do třepené fonace například o oktávu výš. Tyto

vzácné případy jsme ručně dorovnali, tedy vycházeli jsme z frekvence, kterou naměřila funkce *Manipulation*.

Výše uvedená pravidla jsme si tedy stanovili pro označování úseků, u kterých se měla upravovat základní frekvence. V okně *Manipulation* je možno zvolit funkci *Pitch -> Shift pitch frequencies*. Tato funkce nabízí změnu frekvence v hertzech, ERBech, melech, ale i v půltónech. Tuto poslední možnost jsme pochopitelně využili, neboť veškeré naše výpočty provádíme v půltónové stupnici. Údaje, o kolik máme zvýšit či snížit F0 u jednotlivých slabičných jader, jsme již vypočítali v tabulkovém procesoru, stačilo tedy ručně postupně všechna tato jádra pomocí funkce *Shift pitch frequencies* upravit.

Upravené soubory jsme poté opět pomocí funkce *Get resynthesis (overlap-add)* převedli na soubory zvukové a ty jsme uložili s poznámkou U (upravené). Pro případná řešení různých nedopatření jsme si uložili i veškeré soubory *Manipulation*.

Zbývalo tedy pouze zhotovit jeden veliký zvukový soubor pro percepční test, do nějž by se za sebe s příhodnými pauzami mezi sebou vložilo všech 65 úryvků. Soubor bylo dále nutné rozdělit na čtyři rovnoměrné části, neboť při rozdělení testu na pouhé třetiny nebo poloviny by mohla strmě klesat pozornost respondentů.

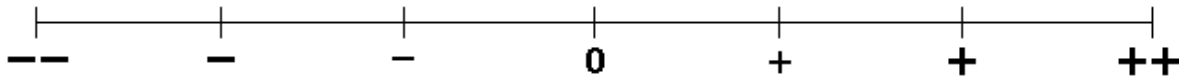
Jako desenzitizační pasáže jsme použili vždy dvousekundové úryvky klasické klavírní skladby, které pak měly vzhledem k nahrávkám zhruba třetinovou amplitudu. 1,5 sekundy před každou větou uváděl v pozornost upozorňovací zvuk podobný zvuku tlačítek na starších mobilních telefonech o přibližně poloviční amplitudě než následná nahrávka. Mezi nahrávkou a následující desenzitizační pasáží pak bylo 2,5 sekundy ticha. Celkově tak vypadalo oddělení dvou sousedních nahrávek takto: nahrávka, 2,5 sekundy ticha, 2 sekundy desenzitizační pasáže, upozorňovací zvuk, 1,5 sekundy ticha. Klavírní nahrávky pro desenzitizaci jsme použili čtyři, aby plnily účel, při stálém opakování jedné zvukové stopy by mozek začal desenzitizaci ignorovat.

65 není číslo dělitelné čtyřmi, nahrávky jsme tedy do čtyř bloků rozdělili nerovnoměrně. V jednom bylo 17 nahrávek, ve zbylých vždy po šestnácti. Bloky jsme nazvali A, B, C a D. Pro zachování potřebných vzdáleností mezi odpovídajícími si nahrávkami vedle sebe nemohly nikdy ležet blok B s blokem D a blok A s blokem C. Zároveň nemohli mít všichni respondenti stejné pořadí bloků, neboť i ono samo ovlivňuje percepci. Možná pořadí (tedy ABCD, DCBA, BADC, CDAB, ADCB, DABC, BCDA, CBAD) jsme proto rozdělili mezi subjekty rovnoměrně.

3.2. Zadání testu

Respondentům byl předložen text s instrukcemi (viz příloha I). Po pilotním testování na dvou subjektech jsme shledali, že času pro ohodnocování nahrávek je dost, pauzy mezi jednotlivými větami byly vyhovující.

Úkol respondentů byl zakroužkovat na stupnici o sedmi stupních, jak moc je jim příjemný celkový projev mluvčího. Tato stupnice byla graficky znázorněna jako osa se sedmi zvýrazněnými body. Nejmenší hodnota byla znázorněna dvěma tučnými minusy, největší dvěma tučnými plusy (viz obr. 3.11).



Obr. 3.11. Škála hodnocení respondentů o sedmi stupních.

Respondentům bylo vysvětleno, že pokud na ně mluvčí působí velmi dobře, mají zakroužkovat nejvyšší hodnotu, pokud na ně naopak působí velmi špatně, mají kroužkovat hodnotu nejnižší. Pro představu bylo uvedeno, že si mohou respondenti představit, jak by se jim líbilo, kdyby daný mluvčí byl vypravěčem v audioknihách pro děti.

Vzhledem k tomu, že pilotní respondenti si všimli, že se nahrávky opakují, a značně je to rozptýlilo, bylo v instrukcích zmíněno, že pokud uslyší nějakou nahrávku, u které mají dojem, že už ji jednou slyšeli, mají ji hodnotit nezávisle stejně jako ostatní a nesnažit se vzpomenout si, jak ji hodnotili poprvé.

Respondenti byli povolání do odzvučené kabiny, kde byl umístěn pouze stůl s laptopem, mikrofix a sluchátka Sennheiser HD 206. Zácvikový test absolvovali za přítomnosti experimentátora, aby mohli mít případné upřesňovací otázky. Posléze jim byl puštěn již samotný percepční test, kdy experimentátor odešel za dvojité dveře a po každém bloku přicházel a pouštěl jim blok další.

Test byl zadán 24 respondentům, mužům i ženám v rozmezí 20 – 34 let.

3.3. Vyhodnocování testu

Zakroužkované hodnoty všech respondentů byly převedeny na celá čísla od -3 do 3 do tabulkového kalkulátoru. Z této obsáhlé tabulky bylo vypočítáno průměrné hodnocení a směrodatnou odchylku každé z 65 položek testu.

Zajímavým jevem byly dvojice nahrávek určené k pozorování konzistence respondentů (viz tab. 3.4).

A2	Ď2	P3	R4	U2
0,542	0,917	0,833	0,167	1,375
1,250	1,042	0,292	0,417	1,583

Tab. 3.4. Průměrná hodnocení u identických nahrávek zkoumajících konzistenci respondentů.

U nahrávek označovaných jako A2 a P3 (tab. 3.4) pozorujeme opravdu veliký rozdíl, který přesahuje 0,5 jednotky na hodnotící škále. Nicméně na tabulce 3.5, která ukazuje průměrná hodnocení nahrávek s opačnou hodnotou upravenosti (tedy pokud v tabulce 3.4 U2 ukazuje dvě upravené nahrávky, U2 v tabulce 3.5 ukazuje hodnocení neupravené), lze vyčíst, že se hodnoty blíží daleko více právě těmto „opačným“ nahrávkám než těm identickým. Kupříkladu P3 a R4 mají dokonce naprosto stejné průměrné hodnocení, u A2 se blíží jedné z hodnot o mnoho víc, než odpovídající nahrávka identická.

Lze tedy již předvídat, že byl percepční test velmi subjektivní a z výsledků nebudou vycházet žádné převratné závěry.

A2	Ď2	P3	R4	U2
1,125	0,75	0,292	0,417	1,792

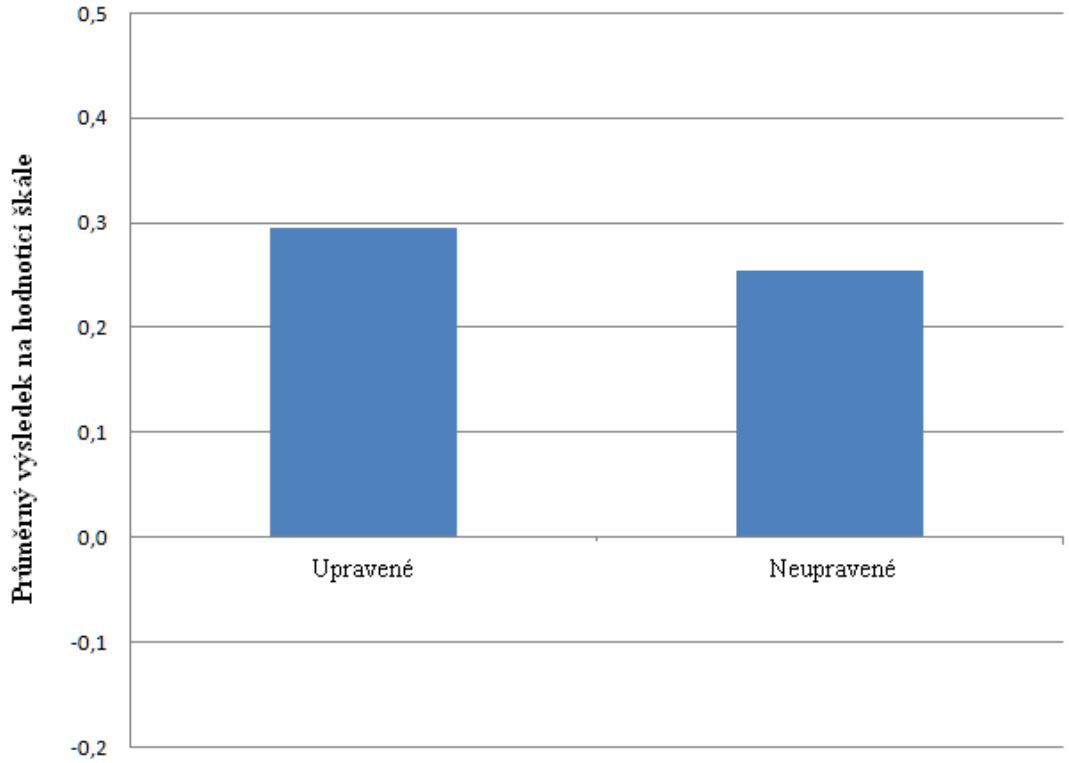
Tab. 3.5. Průměrná hodnocení nahrávek s opačnou hodnotou upravenosti (U/N) vůči nahrávkám zkoumajícím konzistenci respondentů.

Pro vyhodnocování výsledků hypotézy jsme pak zprůměrovali hodnoty průměru a směrodatné odchylky u identických nahrávek, abychom mohli zavést pro všech 30 nahrávek mluvčích pro nás nejdůležitější proměnnou, a to je rozdíl hodnocení upravené a odpovídající neupravené položky.

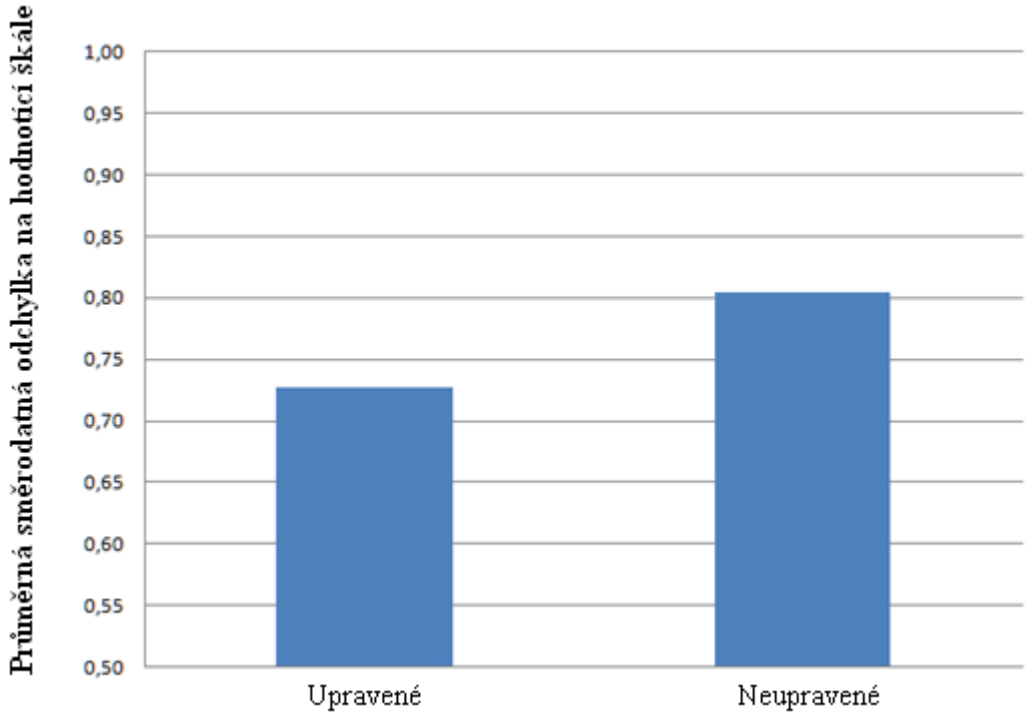
4. Výsledky

Jako alternativní hypotézu Ha1 jsme stanovili, že mluvčí s upravenou F0 na slabičných jádrech do půltónové škály by měli být obecně lépe hodnoceni než ti samí mluvčí bez úprav. Podle pouhého výpočtu průměrů veškerých upravených a neupravených nahrávek lze pozorovat, že upravené byly opravdu o něco lépe hodnoceny než neupravené (obr. 4.1).

Zároveň se u upravených nahrávek objevila menší směrodatná odchylka, respondenti se na nich tedy lépe shodli (obr. 4.2).



Obr. 4.1. Průměrné hodnocení všech upravených a všech neupravených položek.



Obr. 4.2. Směrodatná odchylka od průměrného hodnocení u upravených a neupravených položek.

Po výpočtu t-testu pro korelovaná měření, kdy byly tyto průměry a směrodatné odchylky porovnávány, však vyšly tyto hodnoty:

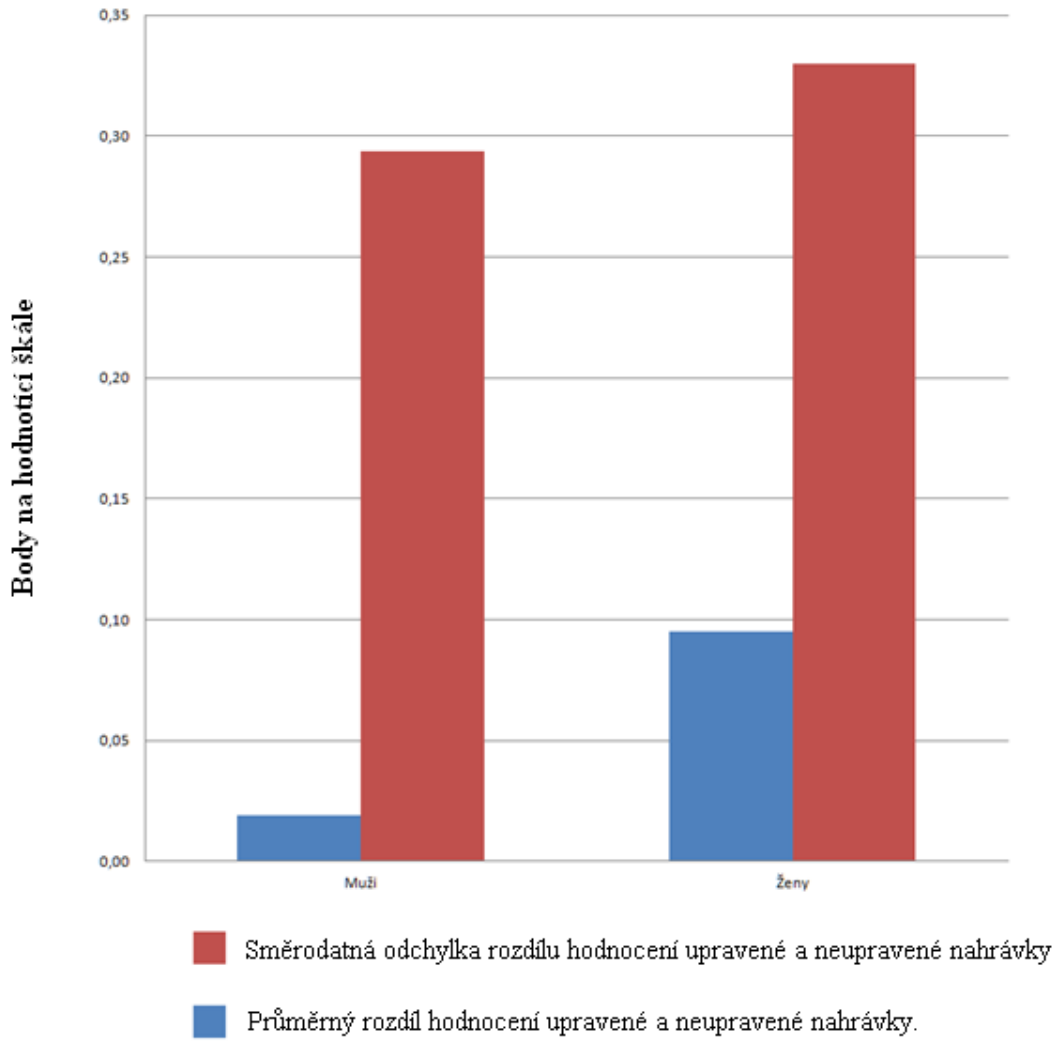
Počet	Rozdíl	Směrodatná odchylka	t	Sv	p	Int. sp. - 95%	Int. sp. 95%
30	0,042	0,301	0,757	29	0,455	0,154	0,071

Tab. 4.1. Výsledné hodnoty t-testu v programu Statistica.

Pro výzkum nejdůležitější jsou výsledné hodnoty *t* a *p*. Vzhledem k tomu, že *p*-hodnota je větší než 0,05, mezi upravenými a neupravenými odpovídajícími si nahrávkami neexistuje (v našem výzkumu) statisticky významný vztah korelace.

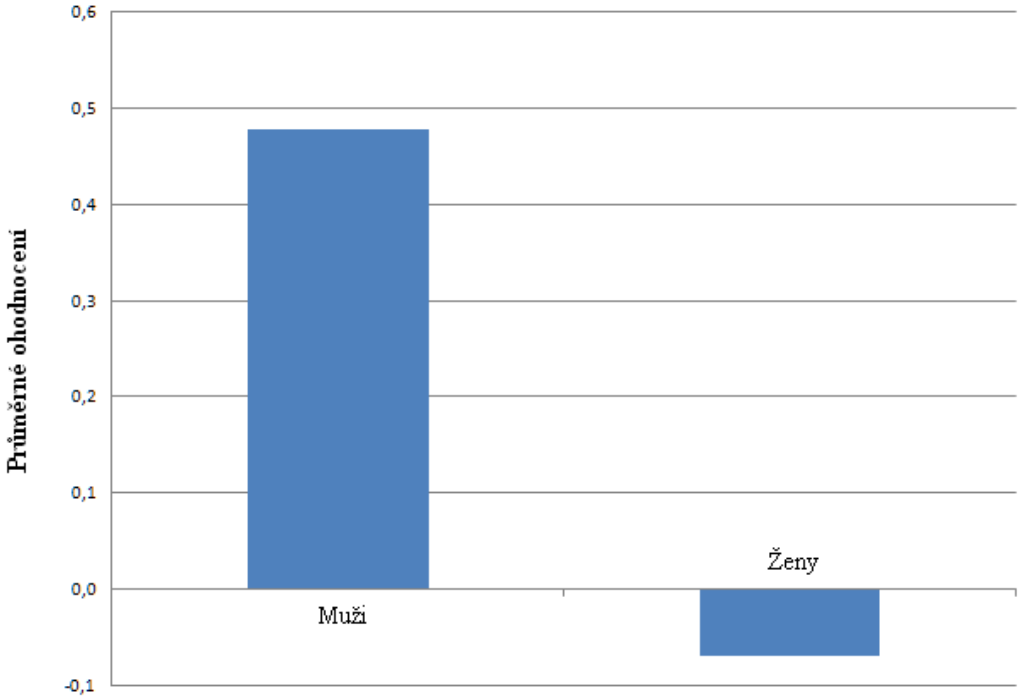
Provedli jsme nicméně další statistické analýzy a objevilo se množství zajímavých výsledků a porovnání. Například se projevilo, že se u mužů objevil daleko menší rozdíl mezi průměrným hodnocením upravených a neupravených nahrávek. Ženské hlasy tedy splňovaly stanovisko naší hypotézy lépe než mužské. Nutno ale podotknout, že ženských hlasů bylo méně. Mužských mluvčích bylo 21 a ženských pouze 9. Z toho důvodu nalézáme pravděpodobně u žen o něco větší směrodatnou odchylku (obr. 4.3). Zároveň je vhodné poznamenat, že rozdíly

byly velmi často i záporné a dosahovaly až hodnoty pěti jednotek. Proto na grafu v obr. 4.3 sloupec směrodatné odchylky tolik převyšuje průměrný rozdíl.

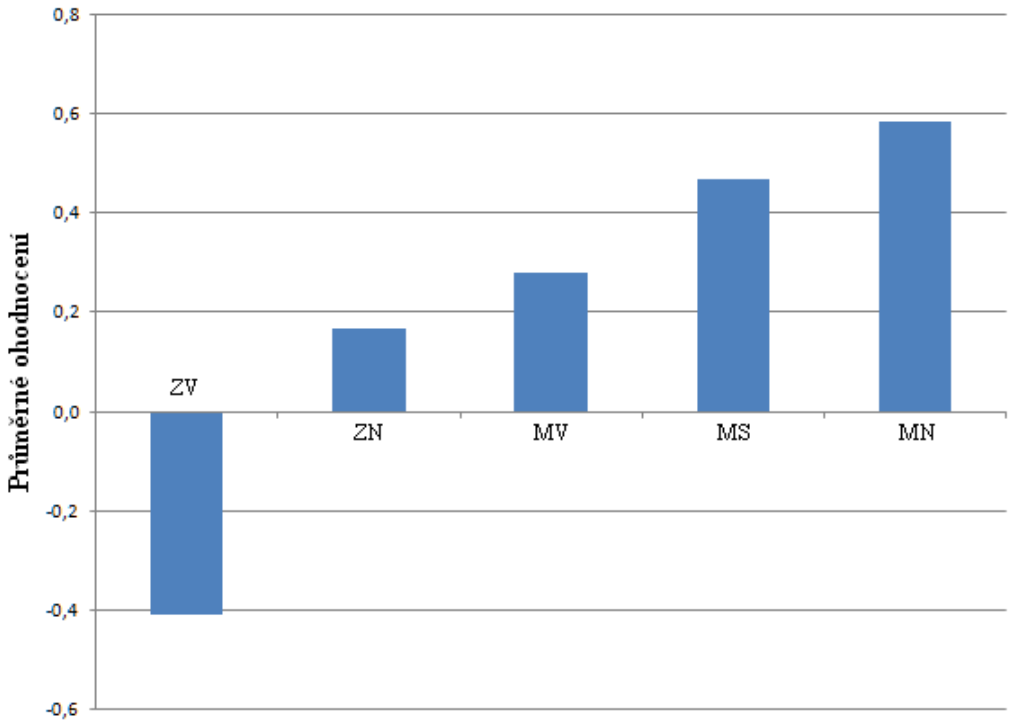


Obr. 4.3. Průměrná hodnota a směrodatná odchylka rozdílu hodnocení upravené a neupravené nahrávky zvlášť u mužů a žen.

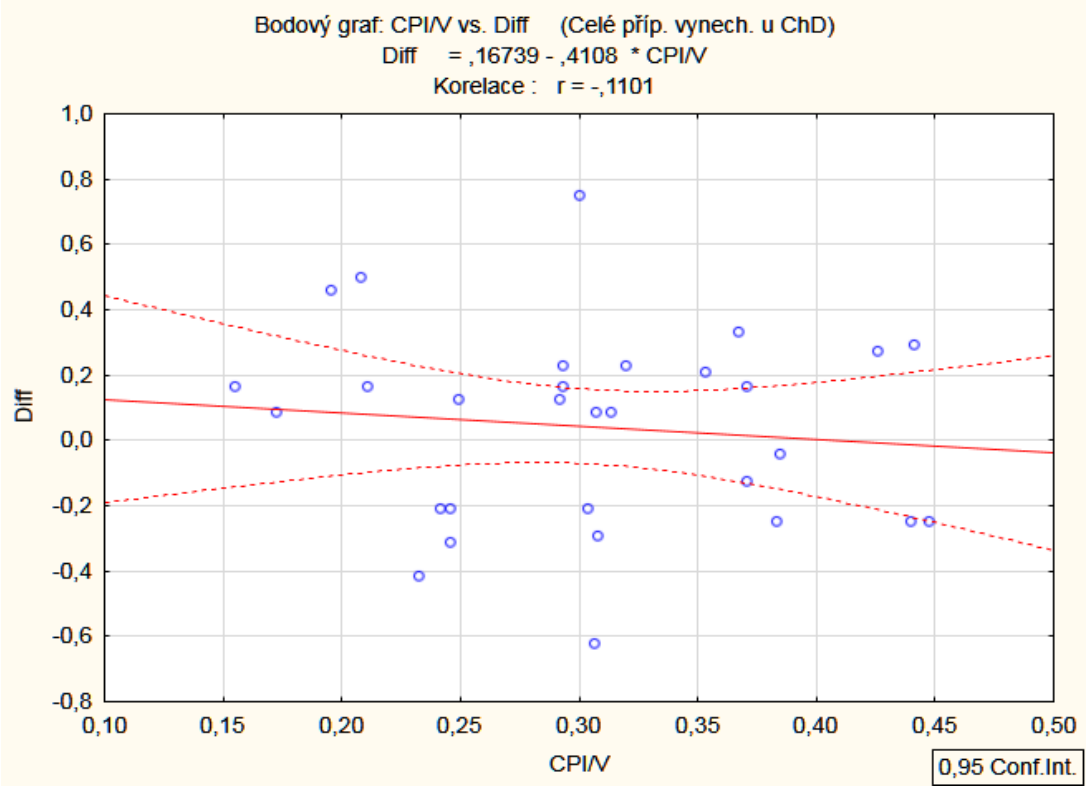
Co se týče celkového průměru hodnocení, respondenti hodnotili lehce častěji na kladné straně hodnotící škály. Průměr hodnocení všech respondentů byl 0,318. Z toho u mužů 0,478 a u žen – 0,070 (viz obr. 4.4). Z tohoto grafu nelze ani v nejmenším vyvozovat to, že by ženy měly obecně pro percepci nepříjemnější hlasy, protože v našem výzkumu byl jen velmi malý vzorek. Je však zajímavé, že v pěti kategoriích, které byly pro manipulaci s nahrávkami vytyčeny, s hlubším hlasem opravdu pravidelně stoupá průměrné hodnocení (viz obr. 4.5).



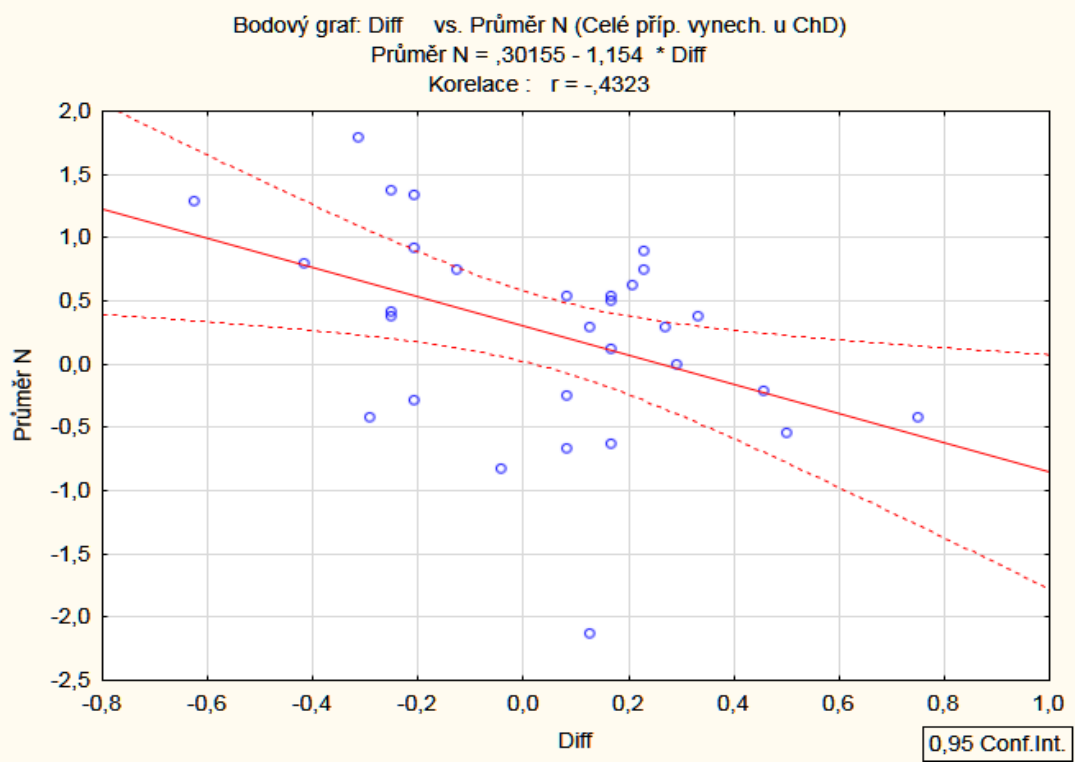
Obr. 4.4. Průměrné ohodnocení mužských a ženských hlasů.



Obr. 4.5. Průměrné ohodnocení jednotlivých kategorií. (ZV = ženský vysoký hlas, MS = mužský střední hlas, MN = mužský nízký hlas a podobně).



Obr. 4.6. Bodový graf znázorňující korelaci mezi velikostí rozdílu hodnocení a mírou upravenosti položky.



Obr. 4.7. Bodový graf znázorňující korelaci mezi průměrným hodnocením neupravené nahrávky a mezi rozdílem hodnocení.

Při hledání dalších korelací mezi rozdílem hodnocení a jinou veličinou se toho mnoho neukázalo. Žádný efekt na rozdíl neměl ani počet vokálů ani například míra upravenosti nahrávky (viz obr. 4.6), u níž byl nějaký vztah očekáván.

Objevila se nicméně zajímavá korelace mezi průměrným hodnocením neupravené položky a rozdílem upravené a neupravené položky (viz obr. 4.7). Z obrázku je patrné, že čím hůř byl daný mluvčí hodnocený, tím větší je onen rozdíl. Není tím však myšlena absolutní hodnota rozdílu. Nejméně oblíbení mluvčí si tak po úpravě nahrávky nejvíce polepšili.

Nyní přistoupíme k položkové analýze. Zaměříme se konkrétně na extrémní hodnoty v průměrném hodnocení, směrodatné odchylce a rozdílu hodnocení.

Suverénně nejhorší hodnocení získal od respondentů mluvčí označený jako I2. Zde není pochyby, čím to je, v nahrávce nejsou žádné ruchy, ale mluvčí má velmi zvláštní barvu hlasu, která se opravdu pro namlouvání audioknih nehodí. Konkrétně jde o výstřižek z audiopořadu Historie českého zločinu, kde nemluví pouze školení herci, ale i různí správci muzeí a podobně. Tento člověk byl jedním z nich.

Druhé nejhorší hodnocení získala ženská mluvčí označená jako S2. Tuto mluvčí dokonce respondenti často zmiňovali po testu, že jim byla velmi nepříjemná. Dalo by se soudit, že by to mohlo být tím, že věta v promluvě mluvčí S2 je jako jedna z mála tázací a zároveň ji mluvčí pronese s melodickým průběhem, který otázce moc neodpovídá. Velmi úzké intonační rozpětí navozuje dojem, že mluvčí sedí na nějakém velmi ostrém hřebíku.

Naopak nejlepší hodnocení získal mluvčí U2, který má velmi hluboký měkký hlas. Vzhledem k tomu, že velké množství dokumentů, reklam a podobných domén, kde se setkáváme s namlouvaným doprovodným hlasem, komentují právě muži s hlubokým hlasem, je možné, že to ovlivnilo respondenty k myšlence, že je takový člověk vhodný i k namlouvání audioknih pro děti. Význam jeho věty *Nikdo, vůbec nikdo to nesmí vědět, ani vaše rodina* percepci zřejmě nijak pozitivním způsobem příliš neovlivňuje.

Největší směrodatnou odchylku vykazuje položka T1, což je zároveň nejdelší položka (33 slabik). Je tu tedy rozpor možná až příliš odlišné délky, kdy se může každý respondent zaměřit na jinou část nahrávky, a zároveň hlubokého mužského hlasu, který je respondentům podle sesbíraných dat příjemnější. Naopak nejmenší směrodatnou odchylku vykazuje mluvčí B2, ovšem zde se dá opravdu jen těžko soudit, proč se na něm subjekty tolik shodly.

Zajímavějším zkoumáním by mohly být extrémní hodnoty v rozdílu hodnocení upravené a neupravené podoby nahrávky. Nejmenší rozdíl nacházíme u již zmíněného mluvčího B2. Zde

je rozdíl -0,625, což znamená, že o 0,625 hodnotili respondenti lépe neupravenou nahrávku. Při bližším srovnání bylo zjištěno, že se při manipulacích základní frekvence na zlomek sekundy vychýlila do vyšších hodnot a subjektivně to působí, jako by mluvčí na konci věty lehce mutoval.

Podobná chyba se vyskytla i u položky s druhým nejmenším rozdílem, v tomto případě u mluvčího Ď2, nicméně zde se chyba vyskytovala i v neupravené podobě, zůstává tedy otázkou, co přimělo respondenty hodnotit upravenou verzi hůř.

Mluvčí s naopak nejvyšším rozdílem (0,750) byla žena označená jako W1. Žádné chyby se na nahrávce neobjevují, možné vysvětlení je to, že průměrné hodnocení nahrávky předcházející v testu upravené položce bylo velmi nízké a hodnocení nahrávky předcházející neupravené položce bylo naopak velmi vysoké. Mohlo by se tak jednat o tzv. efekt pořadí. Tytéž okolnosti potkaly mluvčího R2, nicméně zde je veliký rozdíl možná i důsledek toho, že intonace v neupravené podobě byla velmi plochá a po zaokrouhlení se intonační kroky zvětšily. Mluvčí tak mohl respondentům znít příjemněji.

Harmonizace průběhů s tak malými rozdíly, ke kterým vedla naše metodologie, má zjevně tedy malý vliv na hodnocení mluvčího ve srovnání s vlivy ostatními. Alespoň to však dává popud k dalšímu výzkumu s markantnějšími manipulacemi.

5. Diskuse

Předkládaná studie byla zaměřena na vnímání atraktivity mluvčího a možnost jejího ovlivnění manipulací průběhu základní frekvence. Bylo vybráno 30 nahrávek od 19 mluvčích, každý mluvčí byl ve výběru zastoupen maximálně dvakrát. Pomocí různých manipulací byly vytvořeny kontrastní páry. V každém takovém páru byla nahrávka samotná (pouze resyntetizovaná) a nahrávka upravená. Úprava takové nahrávky spočívala v zarovnání základní frekvence (F0) veškerých slabičných jader do půltónové škály. Byla stanovena hypotéza, že takové úpravy budou mít pozitivní dopad na percepci.

Jádro výzkumu bylo realizováno v podobě percepčního testu na 24 respondentech. Těm bylo postupně ve čtyřech blocích puštěno všech 60 položek (+5 pro kontrolu vnitřní konzistence subjektů). Respondenti měli za úkol zhodnotit na sedmistupňové škále, jak moc příjemně jim daný mluvčí zní. Pro přiblížení této poněkud vágní instrukce bylo řečeno, aby si představili, jak vhodný by jim přišel daný hlas pro namlouvání audioknih pro děti.

Vzhledem k tomu, že byly porovnávány pouze nahrávky v jednom kontrastním páru navzájem, vedlejší vlivy jako obsah věty, obecné přiklánění se k hlubším hlasům a podobně, neměly na výzkum vliv. Všechny tyto vlivy byly totiž přítomny v obou položkách kontrastního páru. Jediné, nač si později respondenti stěžovali, bylo to, že byli často notně ovlivňováni předchozí položkou. Je tedy možné, že důkladnější desenzitizace (delší a složitější) a delší pauzy mezi položkami by mohly vést k jiným výsledkům. Celý percepční test se odehrával v odzvučněné kabině s pomocí sluchátek a bez vlivu okolních ruchů. Přesto zjevně zachycení pocitu z mluvčího a převedení tohoto pocitu na škálu není jednoduchým úkolem, což je zřejmé z dat zkoumajících vnitřní konzistenci.

Z výsledků bohužel nevyplynulo potvrzení platnosti stanovené hypotézy, tedy že se po harmonizaci intonačních kroků nahrávky různých mluvčích stanou percepčně příjemnějšími. Je však nutno uznat, že míra pozměnění byla opravdu velmi malá. Výzkum stavěl na podvědomém vnímání, neboť ani při přehrání dvou kontrastních položek vedle sebe nebylo jisté, která je která.

Zajímavým výsledkem však bylo, že u méně přijatelných mluvčích hypotéza vycházela lépe. Bylo by možná záhodno udělat podobný výzkum na neprofesionální mluvčí nebo mluvčí nepříliš dokonalého řečnického umu. Je možné, že by se hypotéza prokázala, nebo by korelace alespoň mírně zesílila.

Případnou slabinou výzkumu by mohl být fakt, že se operovalo s půltónovou škálou běžně používanou ve fonetickém výzkumu (tedy zahrnující hodnotu 100 Hz), nicméně desenzitizační pasáže byly laděny na hudební temperované ladění (a1 je 440 Hz). Je teoreticky možné, že bez tohoto rozdílu by výzkum dopadl o něco málo lépe.

V této oblasti fonetiky češtiny bylo provedeno podobných výzkumů nemnoho, bylo by tedy nasnadě se o to pokusit, ať na základě této práce, nebo případně z nějakého jiného úhlu pohledu. Naše práce se zaměřovala na naprosto nejmenší možné manipulace nahrávek, možná, že při markantnějších změnách by se projevil efekt silněji.

Pokud by se v budoucnu nějaký vztah mezi harmonizací intonačních kroků v mluvené češtině a jejím dopadem na percepci ukázal a případně pokud by tento vztah byl takový, že jsou nahrávky po harmonizaci percepčně příjemnější, mohl by se vynalézt nesložitý algoritmus, který by jakoukoli nahrávku automaticky harmonizoval a tím ji zkrášlil. Dal by se využít jak v řečové syntéze, tak v audioknihách nebo veřejných projevech. Zároveň by se mohla vytvořit domněnka, že evropské půltónově orientované stupnice nejsou pouhým konstruktem, ale jakýmsi přirozeným matematickým prvkem přírody. Bylo by pak záhodno podobný výzkum provést například na respondentech z kultur, kde hudba vychází ze stupnic o dvaadvaceti nebo více stupních v rámci jedné oktávy.

Literatura

Bachem, A. (1937). Various types of absolute pitch. *Journal of the Acoustical Society of America 9:* 146 – 151.

Bolinger, D. (1978). Intonation Across Languages. In: Greenberg, J. H. (ed.). *Universals of Human Language 2 – Phonology*: 471-524. Stanford: Stanford University Press.

Calapinto, J. (2007). The Interpreter: Has a remote Amazonian tribe upended our understanding of language? *The New Yorker*. Dostupné z:

https://www.newyorker.com/magazine/2007/04/16/the-interpreter-2

Cardozo, B. L., & Ritsma, R. J. (1965). Short time characteristics of periodicity pitch. In: Commins, D. E. (Eds.) *Proceedings of the 5th International Congress on Acoustics, Liège.* (1964): B37.

Cardozo, B. L. (1972). Topics in audition. *IPO Annual Progress Report* 7: 1-4.

Cross, I. (2003). Music, cognition, culture, and evolution. In: Wallin N. L., Merker B., & Brown, S. (Eds.), *The Origins of Music*: 42-56. Cambridge: MIT Press.

Di Cristo, A., & Hirst, D. (1998). *Intonation Systems: A Survey of Twenty Languages.* Cambridge: Cambridge University Press.

Duběda, T. (2005). *Jazyky a jejich zvuky*. Praha: Karolinum.

Fitch, W. T. (2006). *The biology and evolution of music: A comparative perspective*. Dostupné z:

https://homepage.univie.ac.at/tecumseh.fitch/media/files/Fitch2006BiomusicCognition.pdf

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America 54*: 1496-1516.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hast, D. E., Cowdery, J. R., & Scott, S. (Eds.). (1999). *Exploring the World of Music*. Dubuque: Kendall Hunt. (Citace z interview se Simonem Shaheenem ve video programu 6: Melodie).

Hermes, D. J. (2006). Stylization of pitch contours. In: Sudhoff, S. et al. (Eds.). *Methods in Empirical Prosody Research*: 26-61. Berlin: Walter de Gruyter.

Henning, G. B. (1966). Frequency discrimination of random-amplitude tones. *Journal of the Acoustical Society of America 39*: 336-339.

Houtsma, A. J. M. (1980). Pitch of harmonic two-tone complexes of unequal amplitudes. *Journal of the Acoustical Society of America 68*: abstract.

Huron, D. (2003). Is music an evolutionary adaptation? In: Wallin N. L., Merker B., & Brown, S. (Eds.), *The Origins of Music*: 57-75. Cambridge: MIT Press.

Janáček, L. (1998). *Fejetony*. Praha: Ars Bohemica.

Kivy, P. (2007). *Music, Language, and Cognition*. New York: Oxford University Press.

Lakoff, G., & Johnsen, M. (2003). *Metaphors we live by*. London: The University of Chicago press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: MIT Press.

Miller, G. (2000). Evolution of human music through sexual selection. In: Wallin N. L., Merker B., & Brown, S. (Eds.), *The Origins of Music*: 329-360. Cambridge: MIT Press.

Olson, H. F. (1967). *Music, Physics and Engineering*. New York: Dover Publications.

Patel, A. D. (2010). *Music, Language, and the Brain*. New York: Oxford University Press.

Pollák, P., Volín, J. & Skarnitzl, R. (2007). HMM-based phonetic segmentation in Praat environment. In: *Proceedings of XIIth Speech and Computer – SPECOM 2007*: 537 – 541.

Ringer, A. L. (2001). Melody. In: Sadie S. (Eds.). *The New Grove Dictionary of Music and Musicians 16*: 363-373. New York: Grove.

Skarnitzl, R., & Šturm, P., & Volín, J. (2016). *Zvuková báze řečové komunikace*. Praha: Karolinum.

Smolka, J. (2001). *Dějiny hudby*. Brno: Togga.

't Hart, J. T., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research 1*: 155-182.

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience 6*: 669-673.

Vassiere, J. (2005). Perception of Intonation. In: Pisoni, D., & Remez, R. (2005). *Handbook of Speech Perception*: 236-263. Oxford: Blackwell Publishing.

Yule, G. (2010). *The Study of Language*. Cambridge: Cambridge University Press.

Zenkl, L. (2014). *ABC hudebních forem*. Praha: Editio Bärenreiter.

Příloha I

Instrukce k vyplňování percepčního testu

V následujících čtyřech blocích uslyšíte pokaždé 16 – 17 vět namluvených různými mluvčími. Vaším úkolem je zhodnotit na škále o sedmi stupních (**--** je nejméně, **++** je nejvíce), jak příjemně na vás daný mluvčí působí z hlediska svého celkového projevu.

Pro lepší představu můžete například uvažovat nad tím, jak moc si myslíte, že by bylo vhodné, aby daný mluvčí namlouval knížky pro děti.

Pokud je vám tedy opravdu velmi příjemný, zakroužkujte dvě tučná plus (**++**), pokud je vám naopak velmi nepříjemný, zakroužkujte dvě tučná minus (**--**).

Mějte na paměti, že každá odpověď je správná. Pokud vám přijde, že už jste nějakou větu jednou zaslechli, nenechte se vyvést z míry, ohodnoťte ji stejným způsobem jako ostatní. Nesnažte se při hodnocení vzpomínat, co jste zakroužkovali, když jste ji slyšeli poprvé.

V čem přesně test spočívá, vám rád povím po jeho absolvování. V opačném pořadí by to mohlo ovlivnit výsledky výzkumu.

Pokud se naše hypotéza prokáže, budeme zase o krok dál v poznání percepce lidské řeči!