

FILOZOFICKÁ FAKULTA UNIVERZITY KARLOVY
FONETICKÝ ÚSTAV

BAKALÁŘSKÁ PRÁCE

ROBIN HRUŠKA

Vlastnosti kontur základní frekvence
s ohledem na segmentální složení promluv

*Properties of fundamental frequency contours
in segmental contexts*

PRAHA 2016

VEDOUcí PRÁCE
ING. TOMÁŠ BOŘIL, PH.D.

PODĚKOVÁNÍ

Děkuji vedoucímu práce Tomáši Bořilovi za vstřícný přístup, konstruktivní zpětnou vazbu a podnětné rady k výzkumu.

Také děkuji celé své rodině za neutuchající podporu.

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 13. LEDNA 2016

PODPIS

ABSTRAKT

Tato práce se zabývá popisem kontur základní frekvence F_0 v českých promluvách ve vztahu k jejich segmentálnímu složení. Na základě seznámení s principy analýzy intonačních kontur navrhuje jednoduchý stylizační algoritmus, který používá k detekci vrcholových a sedlových bodů obratu v konturách slabičných nukleů. Poté provádí statistický popis časových posunů nalezených bodů vůči hranicím odpovídajících nukleů. Originální konturu pak stylizuje algoritmem implementujícím model tonální percepce a výslednou konturu popisuje co do zastoupení dynamických tónů a velikosti mezislabičných melodických skoků. Také navrhuje parametr kumulativní strmosti ke kvantifikaci celkové variability kontury.

KLÍČOVÁ SLOVA

intonace, kontura F_0 , stylizace, body obratu, tonální percepce, variabilita F_0

ABSTRACT

This paper attempts to describe the fundamental frequency (F₀) contours in Czech utterances in relation to their segmental context. After inspecting the principles of intonation contour analysis, a simple stylization algorithm is designed and used to detect peak and valley turning points in the contours of syllabic nuclei. A statistical description of the turning points' timing relative to the nucleus boundaries is provided. Stylization by a tonal perception model based algorithm is performed on the original contours and the result is then analysed for dynamic tones and intersyllabic pitch jumps. Also, cumulative steepness is proposed as a parameter to quantify total contour variation.

KEYWORDS

intonation, F₀ contour, stylization, turning points, tonal perception, F₀ variability

Obsah

Obsah	6
1. Úvod	8
2. Prozodie a intonace.....	10
2.1. Segmentace a prominence.....	10
2.2. Intonace.....	12
2.2.1. Funkce intonace	14
2.3. Kontura základní frekvence.....	14
2.3.1. Vlastnosti lidského hlasu	15
2.3.2. Vnímání a popis výšky tónu.....	16
2.3.3. Obecné vlastnosti kontury F_0	17
3. Intonace v češtině.....	19
3.1. Větná intonace	19
3.1.1. České melodémy.....	20
3.1.2. Přízvukový takt	22
4. Zpracování a interpretace intonace.....	23
4.1. Fonetické intonační modely a stylizace	25
4.1.1. Dolní propust.....	26
4.1.2. MoMel (Hirst & Espesser, 1993).....	26
4.1.3. Bagshawova stylizace (Bagshaw, 1993)	27
4.1.4. INTSINT (Hirst & Di Cristo, 1998)	27
4.1.5. Fujisakiho model (Fujisaki & Ohno, 1997)	28
4.1.6. Tilt model (Taylor, 1999)	30
4.1.7. Model tonální percepce (Mertens & d'Alessandro, 1995).....	31
4.2. Fonologické intonační modely	33
4.2.1. ToBI	33
4.3. Relevantní prvky kontury základní frekvence.....	35
5. Praktická část.....	38
5.1. Data.....	38
5.1.1. Klasifikace dat.....	40
5.1.2. Použité nástroje.....	40
5.2. Body obratu.....	41
5.2.1. Metoda	41
5.2.2. Posun vrcholů	43
5.2.3. Posun sedel	46
5.2.4. Stylizace kontury F_0 celé repliky.....	48
5.3. Strmost kontury.....	49
5.3.1. Metoda	49

5.3.2.	Dynamické tóny.....	50
5.3.3.	Melodické skoky	51
5.4.	Variabilita promluvy.....	53
6.	Závěr.....	56
6.1.1.	Poznatky z praktické části	56
7.	Zdroje	58
8.	Přílohy	61
8.1.	Obsah elektronické přílohy	61
8.1.1.	Naměřená data	61
8.1.2.	Skripty k programu Praat	61
8.1.3.	Skripty v jazyce R.....	62
8.2.	Dodatek: Vrstva sonority	62
8.3.	Dodatek: Diferenciální melodická kontura	63

1. Úvod

Výzkum intonace a suprasegmentální složky řeči vůbec je pro současný výzkum stále atraktivní téma, a to jak pro jazykovědce, tak pro řečové inženýry. Jako v každé jiné vědecké disciplíně i zde platí, že nové objevy s sebou přinášejí nové otázky pro další výzkum. Problémem je však nedostatek spolupráce mezi oběma skupinami vědců. Grabe et al. (2007) uvádí, že „*matematické modely intonace používané v řečových technologiích jsou lingvistům často nepřístupné*“. Situace na druhé straně je podobná, řečovní inženýři nepracují s fonologickými modely, protože „*nemohou být implementovány přímo*“. Lingvisté zřídka využívají obrovské řečové korpusy, s jakými pracují řečové technologie, dávají přednost malým databázím speciálně vytvořeným pro účely lingvistického výzkumu. Řečovní inženýři mají k dispozici obrovské množství akustických dat, ale nedokážou identifikovat, co je lingvisticky relevantní a přidělit tomu reálný význam. Mimoto aplikace pro analýzu řeči jen málokdy interpretují význam nesený intonací, přestože v mezilidské komunikaci hraje naprosto zásadní roli. Úkolem fonetického výzkumu intonace by mělo být vyplnění této mezery mezi technickým a jazykovědným přístupem a využití nejnovějších poznatků z obou disciplín.

Kontura základní frekvence (F_0) je považována za hlavní korelát intonace, proto se jí v lingvistických i technických výzkumech intonace věnuje nejvíce pozornosti. Zásadní otázkou je, které rysy této kontury nesou relevantní lingvistickou informaci a o jaký typ informace jde. Mnoho studií (Pierrehumbert & al., 1989; Rietveld & al., 1995) ukazuje na důležitost umístění bodů obratu, jiné (Taylor, 1999) poukazují na relevanci směru a tvaru kontury, případně výškových rozdílů mezi jednotlivými slabikami (Mertens, 2004).

Tato práce je popisná, klade si za cíl zjistit obecné tendence v intonačních konturách českých promluv ve vztahu k segmentální rovině – umístění bodů obratu, strmosti kontury, případně dalších ukazatelů, které mohou nést lingvistickou informaci.

Motivací pro hledání korelace mezi segmentální a suprasegmentální vrstvou je jednak možnost zpřesnění fonetického i fonologického popisu české intonace a jednak možnost zvýšit potenciál využití intonace v řečových technologiích. Fonetická podstata české intonace doposud nebyla popsána prostřednictvím žádného uceleného parametrického modelu, neboť k tomu je třeba detailnějšího popisu zákonitostí, kterými se v češtině řídí melodický průběh hlasu. Podchycení tendencí, které si tato práce klade za cíl, by mohlo být malým krokem k vytvoření takového modelu.

Nejprve se seznámíme s pojmem intonace a prozodie obecně, dále se zaměříme na konturu základní frekvence a její vlastnosti, na modely navržené k jejímu popisu a stylizaci a na lingvisticky relevantní rysy této kontury. Poté si přiblížíme systém intonace v českém jazyce. V praktické části provedeme výběr vhodného modelu pro stylizaci kontury a akustických dat k analýze, dále provedeme extrakci těchto dat a jejich statistickou analýzu s ohledem na segmentální rovinu promluv. V závěru se budeme věnovat vysvětlení a zobecnění případně nalezených korelací a vyvození východisek pro další výzkum.

2. Prozodie a intonace

Intonace je jedním z ústředních pojmů prozodie. Obecně, v nejširším smyslu slova, označuje melodii řeči, avšak abychom ji mohli definovat přesněji, nejprve si přiblížíme pojem prozodie. V souvislé řeči rozlišujeme dvě základní roviny: segmentální a suprasegmentální. Segmentální neboli hlásková rovina je tvořena lineární sekvencí hlásek. Suprasegmentální neboli prozodická rovina zahrnuje zvukové jevy, které pokrývají vyšší řečové jednotky než hlásky – někdy jen slabiku, jindy i celou promluvu či skupinu promluv. Ačkoliv v suprasegmentální funkci lidé využívají nejrůznější zvukové a artikulační prostředky, jako třeba druh fonace, labializaci či palatalizaci, existují právě tři prostředky, které se nejkonzistentněji využívají k vyjádření jazykových funkcí: *výška, délka a hlasitost*.

Cruttenden (1997, s. 2) výšku popisuje jako variaci frekvence hlasu v rámci jedné či více po sobě jdoucích slabik. Délka znamená vzájemně relativní trvání po sobě jdoucích slabik či relativní trvání slabiky v daném prostředí vůči trvání té samé slabiky v jiném prostředí. Hlasitostí se rozumí změny hlasitosti v rámci jedné slabiky či relativní hlasitost sousedících slabik v řeči. Všechny tři pojmy odkazují na percepční jevy, tak, jak je z akustického signálu jako posluchači interpretujeme v závislosti na jejich okolí, nikoliv na objektivní fyzikální veličiny. Bez kvantifikovatelných akustických veličin se však fonetický popis neobejde, proto definujeme *základní frekvenci* F_0 [Hz] jako akustický korelát výšky hlasu, *trvání* [s] jako korelát délky a *intenzitu*, resp. *hladinu intenzity* [dB] jako korelát hlasitosti. Ovládním těchto prostředků člověk vytváří základ řečové prozodie.

2.1. SEGMENTACE A PROMINENCE

Prozodické modulace nejsou nahodilé, ale tvoří hierarchický systém, který slouží k organizaci zvukové stránky jazyka. Díky tomu řeč vnímáme v rytmických či melodických celcích, což je nezbytné k jejímu porozumění. Podle Palkové (1994) toto

prozodické strukturování funguje na principu segmentace a prominence. Člověk má totiž tendenci řeč segmentovat na menší jednotky a těm připisovat různou relativní prominenci.

Za základní a nejmenší reálnou jednotku prozodického popisu je považována **slabika**. Po fonetické stránce je to hláskové seskupení se sonoritním vrcholem, sonorní hláskou (nejčastěji samohláskou), která tvoří její jádro.

Vyšší jednotkou, která funguje přibližně na úrovni jazykového slova, je **přízvukový takt**, nebo také *stopa*. Je to skupina slabik podřízená jednomu slovnímu přízvuku. **Slovní přízvuk**, ač pro naši intuici zcela zjevný, je jako pojem těžko popsateľný, neboť nemá stabilní zvukovou charakteristiku. Jedná se o vlastnost slabiky, kterou vnímáme jako zvukově prominentní vůči okolním slabikám. Anglická literatura tradičně rozlišuje související pojmy přízvuk (*stress*), akcent (*accent*) a prominence (*prominence*). Přízvuk je binární lingvistická vlastnost konkrétní slabiky ve slově, která ji označuje jako potenciálního nositele akcentu. Akcent je reálná vlastnost, kterou slabika získává při zvukové realizaci. Je to také vlastnost binární – slabika může být buďto akcentovaná, nebo neakcentovaná. Prominence je pak graduální, kvantitativní vlastnost pronesené slabiky, vnímaná relativně vůči jejímu okolí. Posluchač může například říci, zda mu konkrétní slabika připadá více nebo méně prominentní než slabika následující (Hermes, 2007: 29-30).

Na vyšších strukturních úrovních se principy členění řeči v různých jazycích liší. Popisy „zvukové věty“ vychází buďto ze zvukové realizace gramatické věty, nebo ji definují na základě průběhu prozodických kvalit v souvislé řeči. Oba přístupy však stojí na intonační analýze. Palková (1994: 160) na základě melodického průběhu definuje promluvový úsek a výpověď. **Promluvový úsek** je jednotka tvořená jedním či více přízvukovými takty, v jazykové doméně srovnatelná s rozvitým větným členem či krátkou syntaktickou větou. **Výpověď** je primárně jazyková jednotka, součástí jejíž charakteristiky je i zvuková podoba spojená s její komunikační platností. V řadě jazyků má tak na úrovni výpovědi intonace fonologickou funkci. Angličtina na těchto úrovních pracuje s pojmy střední fráze (*intermediate phrase*) a intonační fráze.

Univerzální a jasně definovanou popisnou jednotkou je **nádechový úsek**, tedy úsek řeči pronesený mezi dvěma nádechy. Jeho nevýhodou při popisu je ovšem jeho nepravidelnost a spíše fyziologická než jazyková motivovanost.

2.2. INTONACE

Ve srovnání se segmentální vrstvou, která je v mnoha jazycích velmi podrobně zmapovaná jak po fonetické, tak i po fonologické stránce, má výzkum suprasegmentálií stále velké rezervy, a to především ve fonologickém popisu. Vztah zvukových prostředků a významu v prozodické rovině není tak silný jako v rovině hláskové – ve vyjádření konkrétního významu se zvukové prostředky zastupují navzájem, ale také s prostředky jazykovými, což hledání fonologických zákonitostí značně komplikuje. Doposud není zcela jasné, které prozodické rysy jsou pro lidský mozek distinktivní, proto k popisu existuje nespočet přístupů, které vycházejí z odlišných předpokladů. Díky tomu také nalezneme více přijímaných pojetí intonace.

Tradičně je intonace spojována s průběhem větné melodie (viz pohled Palkové níže), ale běžně se setkáme s pojetími, které pod intonaci zahrnují i trvání a intenzitu, a blíží se tak tomu, co jsme si definovali jako prozodie (viz Laddovo pojetí). Takový přístup je opodstatněný, neboť F_0 , trvání a intenzita jsou jednak do jisté míry fyziologicky závislé jedno na druhém – zvýšení intenzity hlasu i zrychlení mluvního tempa se projevují také zvýšením F_0 – a jednak jsou spojeny i funkčně – v rámci jazyka se navzájem doplňují ve vyjadřování lingvistických významů.

Palková (1994) intonaci definuje jako „*průběh výšky tónu ve větě nebo její části, tj. melodické změny hlasu, které uživatel jazyka hodnotí jako součást výpovědi (nebo její části)*“. Patří k univerzálním charakteristikám struktury jazyka. Ve všech známých jazycích plní různé gramatické funkce – indikuje přítomnost přízvuku, napomáhá syntaktickému a diskurzivnímu členění řeči, a často také rozlišuje komunikační platnost výpovědi nebo její části, takže určuje, zda jde o otázku, či sdělení, a zda je větný celek ukončený, či neukončený. Poměrně univerzální, avšak komplexnější, je pragmalingvistický aspekt intonace, tedy její využití k vyjádření postoje mluvčího.

Některé postoje korespondují s určitými intonačními schémata relativně stabilně, v češtině je to například tzv. varovná intonace. Naproti tomu různá emoční rozpoložení mluvčího, ač se do intonace slyšitelně promítají, nemají v suprasegmentální rovině stabilní a jednoznačný korelát.

Ladd (2008: 4) intonaci definuje z fonologického pohledu, tedy s důrazem na její funkci, jako „užívání suprasegmentálních fonetických jevů na úrovni věty k vyjádření pragmatického významu lingvisticky strukturovaným způsobem“. Suprasegmentálními jevy rozumí modulaci základní frekvence, intenzity a trvání, přičemž není podstatné, zda mluvíme o akustických, objektivně měřitelných veličinách, nebo o jejich percepčních, psychoakustických korelátech. Termín „lingvisticky strukturovaný“ znamená přibližně totéž, co „fonologický“. Týká se tedy těch rysů, které lze systematicky uspořádat do funkčních kategorií (melodický přízvuk × hraniční tón) a odlišit na základě jejich vzájemných vztahů (vysoký × nízký tón). Vylučuje tím jevy paralingvistické – ty, v nichž jsou spojitě proměnné akustické veličiny (intenzita, mluvní tempo) obrazem spojitě proměnného stavu mluvčího (zaujetí, vzrušení). Často je však obtížné paralingvistické jevy od jevů intonačních odlišit. Ladd také zdůrazňuje, že jde o jevy na úrovni věty neboli jevy postlexikální, aby vyloučil případy, kdy suprasegmentální jevy rozlišují věcný význam slov. V tónových jazycích, jakým je například vietnamština, význam slova závisí kromě segmentálního složení také na tónovém průběhu.

Hirst a Di Cristo (1998: 3–7) rozlišují intonaci v širším a v užším slova smyslu. Širší pojetí zahrnuje faktory jako slovní přízvuk, tón či délku, které mohou zasahovat do lexikální roviny jazyka, a blíží se tak významově k prozodii. Užší pojetí označuje pouze intonaci postlexikální. Dále rozlišují intonaci na fyzické a na formální úrovni. Na fyzické, akustické úrovni je intonace měřitelná a objektivně kvantifikovatelná, zatímco na formální, lingvistické úrovni označuje abstraktní jev, který je v kompetenci mluvčích daného jazyka. Intonaci pak přiznávají dva významy: fonologický, který označuje abstraktní postlexikální složku prozodie, a fonetický, který označuje fonetické realizace této složky.

Vzhledem k tomu, že cílem této práce je fonetická analýza české intonace, přikloníme se ke konkrétnějšímu, fonetickému pojetí intonace jako percepčně relevantní melodie řeči bez ohledu na jejich jazykovou motivaci.

2.2.1. FUNKCE INTONACE

Intonace jako melodie řeči plní v komunikaci nejrůznější funkce – od těch úzce jazykových po ty obecnější mimojazykové. Volín a Bořil (2014) intonaci připisují těchto šest funkcí:

- 1) LEXIKÁLNÍ. V tónových jazycích a v jazycích s tónovým přízvukem se melodický průběh používá k rozlišení lexikálního významu.
- 2) GRAMATICKÁ. Intonace může rozlišovat komunikační platnost výpovědi, tj. odlišit otázku od prostého tvrzení. Zároveň utváří syntaktické členění mluveného projevu.
- 3) AFEKTIVNÍ. Intonace je také nositelem pragmatické informace. Jejím prostřednictvím dává mluvčí najevo své nálady, emoce, postoje či vztahy k ostatním účastníkům komunikace.
- 4) DISKURZNÍ. V kontextu dialogu lze intonačně signalizovat například předání slova komunikačnímu partnerovi.
- 5) AKCENTUÁLNÍ. Intonace umožňuje zdůrazňovat konkrétní slova či jiné jednotky.
- 6) INDEXNÍ. Intonace také posluchači poskytuje mimolingvistické informace o mluvčím. Lze z ní například odhadnout jeho pohlaví, věk či společenský status.

Omezíme-li se pouze na základní jazykové funkce, pak je pro intonaci zásadní funkce prominenční a segmentační.

2.3. KONTURA ZÁKLADNÍ FREKVENCE

Využití intonace v řečových technologiích předpokládá automatizaci její analýzy a objektivní přístup. Intonaci jako percepční jev je však obtížné kvantifikovat, neboť je ze

své podstaty vždy značně subjektivní. Proto se již od svých počátků experimentální fonetika zabývá jejím hlavním akustickým, instrumentálně měřitelným, a tedy objektivním korelátem – průběhem základní frekvence hlasivkového tónu, konturou F_0 .

2.3.1. VLASTNOSTI LIDSKÉHO HLASU

Hlas tvoří tónovou složku řeči. Vzniká periodickým kmitáním hlasivek, které tak modulují výdechový proud vzduchu. Vzniklé zvukové vlny pak sluchem vnímáme jako hlasový tón. Ačkoli je přítomnost hlasu základní složkou řeči, při mluvení „*dochází dosti často k jeho přitlumení nebo úplnému přerušení; také jeho kvalita není vždy zcela stejná. V řadě případů je hlas suplován zněním třetího šumu*“ (Hála, 1962: 61). Dále uvádí následující pravidla činnosti hlasu v závislosti na hláskovém složení řeči:

- 1) *Plnozvučně zní hlas jenom při samohláskách.*
- 2) *Při souhláskách znělých (hlasných) zní jenom polozvučně.*
- 3) *Úplně je přerušen při absolutní pauze a dále při neznělých souhláskách závěrových.*
- 4) *Je přitlumen, tj. redukován na poloznění ve formě neurčitého vokálu při tzv. pauze relativní.*
- 5) *Je suplován třecím šumem při neznělých souhláskách třených f, s, š, ch.*

(Hála, 1962: 61)

Z akustického hlediska je hlasivkový tón kvaziperiodický signál složený z mnoha harmonických složek. Nejnižší z nich, nultou harmonickou složku, nazýváme základní frekvencí. Ta udává výšku hlasu a je zásadní pro popis intonace. Ostatní, vyšší harmonické složky jsou v signálu zastoupeny s menšími amplitudami a určují barvu hlasu.

Chceme-li zkoumat základní frekvenci hlasového signálu, musíme ji nejprve oddělit od zbytku signálu, tedy od vyšších harmonických složek tónu a od neperiodických šumových složek konsonantů. Pro tyto účely bylo navrženo mnoho algoritmů; některé pracují se signálem v časové, jiné ve frekvenční doméně, existují i

pokročilé metody, které oba přístupy kombinují, případně je zpřesňují statistickou analýzou (Gerhard, 2003).

Extrahovanou konturu je vhodné dodatečně poslechem porovnat s originální nahrávkou a případně ji manuálně doladit, neboť extrakční algoritmy nejsou neomylné a v řeči se objevují jevy, které úspěšnou extrakci komplikují. Problémy působí například třepená fonace – nízkofrekvenční nepravidelná činnost hlasivek, která se v řeči může objevit jako realizace glotálního rázu nebo často slouží k rozšíření intonačního rozsahu směrem dolů. Vzhledem k její nepravidelnosti bývají třepené řečové úseky při automatické extrakci F_0 vyhodnoceny jako neznělé nebo částečně znělé, případně jejich spektrální složení může způsobovat oktákové skoky.

Při detekci znělosti jsou často problematické frikativy (v češtině pak zejména frikativní vibranta /ř/), protože jejich formantový šum bývá mylně interpretován jako průběh F_0 . V extrahované kontuře pak vznikají výškové skoky vyšší než oktáva, které se v řeči standardně nevyskytují.

2.3.2. VNÍMÁNÍ A POPIS VÝŠKY TÓNU

Objektivně výšku tónu – periodického signálu – popisujeme jeho frekvencí, tedy jako počet period za sekundu. Základní jednotkou frekvence je 1 *hertz* (Hz), který odpovídá 1 periodě za sekundu. Hertzová stupnice je lineární, ale člověk výškové rozdíly tónu nevnímá v rozdílech, nýbrž v poměrech, proto se při popisu hlasové frekvence častěji používá logaritmická, půltónová stupnice převzatá z hudební teorie. Interval o velikosti 1 *půltónu* (ST) odpovídá $\sqrt[12]{2}$ -násobnému zvýšení frekvence, 12-půltónový interval neboli oktáva tedy představuje dvojnásobné zvýšení frekvence. Intervaly nižší než půltón se vyjadřují v centech, tj. intervalech stokrát menších než půltón. Půltónová stupnice je ze své podstaty relativní, takže k určení absolutní výšky tónu je třeba stanovit tón referenční, což obvykle bývá 100 Hz, nebo případně 440 Hz (komorní *a*).

De Looze & Hirst (2014) pro popis intonace navrhli tzv. stupnici OMe (oktáva–medián). Je to rovněž stupnice logaritmická, její základní jednotkou je 1 *oktáva*, jelikož představuje pro lidské vnímání přirozený interval (na rozdíl od půltónu, který je

z oktávy poměrně složitě odvozen). Drobnější intervaly se poměřují v *milioktávách*, intervalech 1000× menších než oktáva. Referenčním tónem stupnice je *medián* F_0 analyzované nahrávky, který má odpovídat střední hlasové poloze mluvčího.

Dále se ve fonetických výzkumech využívají také psychoakustické stupnice (barková, melová a erbová), které modelují nelinearity lidského vnímání výšky ve frekvencích vyšších než cca 800 Hz. Těmi se zde zabývat nebudeme, jednak pro složitost jejich implementace a zpracování a jednak proto, že v těchto frekvencích se F_0 lidského hlasu nepohybuje.

2.3.3. OBECNÉ VLASTNOSTI KONTURY F_0

Většina univerzálních vlastností průběhu F_0 je dána fyziologicky, způsobem produkce hlasu. **Průměrná hlasová výška** neboli poloha „*je podmíněna anatomickým utvářením hrtanu i fonačního ústrojí*“. „*Hrtany prostorné, velké, s dlouhými hlasivkami vydávají hlas spíše hlubší, kdežto hrtany malé, s krátkými hlasivkami vydávají hlas spíše vyšší.*“ (Hála et al., 1955: 51). Pro mužský hlas je průměrná hlasová výška přibližně 120 Hz, pro ženský 220 Hz a pro dětský 265 Hz (Cruttenden, 1997: 3). Anatomicky je dáno i **hlasové rozpětí**. V kompetenci mluvčího je s rozpětím manipulovat, proto je při zpěvu výrazně širší než při řeči. Hála (1962: 59) uvádí, že intonační rozpětí hlasu při klidném mluvení je u mužů zhruba 100 až 150 Hz a u žen a dětí 200 až 300 Hz.¹ Cruttenden (1997: 3) mužům připisuje celkový rozsah 60 až 240 Hz, ženám 180 až 400 Hz. Rozpětí a poloha však fungují i jako spojitě variabilní dynamické parametry s různou funkcí. Zvýšení rozpětí může znamenat afektovanost mluvčího, náhlé snížení polohy může ve struktuře projevu signalizovat například vsuvku. Oba parametry navíc fungují jako prostředky utvářející mluvní styl (spolu s mluvním tempem, barvou hlasu atd.). Jinou polohu hlasu člověk používá při veřejném projevu, jinou při telefonním hovoru s blízkou osobou.

¹ Hála rozsahy popisuje pomocí hudebního názvosloví. U mužů uvádí rozsah $G(A) - d(e)$, u žen a dětí $g(a) - d^1(e^1)$.

Kromě vnímané melodické, tzv. makrointonační složky obsahuje kontura F_0 také složku **mikrointonační**. Ta je tvořena drobnými i většími výkyvy F_0 , způsobenými změnami tlaku ve vokálním traktu a přerušováním činnosti hlasivek, k nimž dochází při produkci hlásek. Mikrointonace je v procesu vnímání odfiltrována, proto se vyhlazuje i při instrumentálním zpracování řeči (viz *stylizace*, str. 23).

Za obecnou vlastnost kontury F_0 se také často považuje *deklinace*, tedy postupný melodický pokles v rámci nádechového úseku.² Podobně jako mikrointonace je motivována fyziologicky, způsobuje ji pokles tlaku výdechového proudu.

² Je ovšem třeba rozlišovat pokles fonetický (popsaný v textu) od poklesu fonologického, který nese jazykovou funkci (signál ukončenosti výpovědi, segmentace řeči). Ač takové tendence nalezneme v mnoha známých jazycích, nelze je považovat za univerzálie. (Ladd, 2008: 81)

3. Intonace v češtině

Čeština patří mezi jazyky, v nichž modulace hlasové výšky nerozlišuje lexikální význam slov, ale z jazykového hlediska plní především prominenci a segmentační funkci. Na úrovni věty má pak i funkci gramatickou, přeneseně zastává funkci, jakou má v psaném textu interpunkce.

V popisu členění souvislé řeči v češtině navážeme na členění uvedené ve 2.1. a uvedeme si rysy pro češtinu charakteristické. Jádro české slabiky může tvořit vokál, nebo slabičný sonorní konsonant ([r], [l], případně [m] či [n]). V préture i v kodě mohou stát samostatné souhlásky nebo souhláskové shluky. V případě shluků se souhlásky většinou řadí tak, že s jádrem sousedí ty sonornější a na okrajích slabiky jsou ty méně sonorní. Pokud tomu tak není, vzniká tzv. pobočná slabika (př. *m-dlo*, *l-sti-vý*).

Čeština se klasifikuje jako jazyk s pevným slovním přízvukem (např. na rozdíl od němčiny, která má přízvuk pohyblivý). Zjednodušeně lze říci, že přízvuk je umístěn pevně na první slabice slova, což platí obecně pro dvou- a víceslabičná slova. U jednoslabičných slov je situace složitější. Zatímco plnovýznamová slova nesou na své jediné slabice přízvuk, synsémantická slova (především zájmena, spojky a tvary slovesa „být“) přízvuk zpravidla nemají a chovají se jako klitika (předklonky a příklonky). Vážou se pak k předchozímu nebo následujícímu přízvukovému taktu, případně pokud se jich vyskytne více při sobě, semknou se v jeden společný takt. Výjimku tvoří jednoslabičné předložky vlastní, které při spojení s gramatickým jménem přebírají jeho přízvuk (Palková, 1994: 280-282). V plynulé řeči se však umístění slovního přízvuku často řídí rytmem řeči spíše než jazykovou klasifikací. Běžně se tak stávají příklonkami i plnovýznamová jednoslabičná slova či víceslabičná synsémantická slova.

3.1. VĚTNÁ INTONACE

V češtině má věta stabilní zvukovou charakteristiku, tvoří tedy nejenom syntaktický, ale i zvukový celek. Z fonetického hlediska můžeme u intonačních průběhů zvukových

realizací některých vět pozorovat podobné rysy. Tím máme na mysli především melodické pohyby základní frekvence, jejich umístění vzhledem k segmentálnímu složení promluvy a jejich míry a trvání. Abstraktní schéma melodického průběhu vymezené určitými rysy nazýváme zvukově popisnou kadencí či zkráceně **intonační kadencí**. Obecně je definována jako nejmenší jednotka, jakou v intonačním popisu rozlišujeme. Pro jednoduchost systému každá kadence zahrnuje jen jednu změnu melodického pohybu. Podle toho rozlišujeme základní typy kadencí: stoupavá, rovná, klesavá, klesavo-stoupavá a stoupavo-klesavá. (Palková, 1994: 162)

Z hlediska fonologického za nejmenší jednotku považujeme funkční kadenci neboli **melodém**, soubor všech kadencí, které v daném jazykovém systému plní stejnou funkci. V češtině rozlišujeme tři základní melodémy: melodém ukončující klesavý, melodém ukončující stoupavý a melodém neukončující. Každý z nich má několik kadenčních variant lišících se mírou příznakovosti a stylovým zařazením (Palková, 1994: 162, 307).

Intonační kadence ve funkci melodému obvykle zvukově pokrývá jen část promluvového úseku. Standardně se váže na jeho poslední větný přízvuk a pokračuje přes libovolný počet taktů až do konce výpovědi nebo výpovědního úseku. Druh melodému nezávisí na délce ani na věcném obsahu výpovědi, řídí se pouze jejím typem.

Melodická kadence věty je také využívána k signalizaci větné prominence. Pokud chce mluvčí zvukově zdůraznit význam některého slova a učinit ho rématem výpovědi, může přesunout melodém z výchozí pozice na konci úseku na příslušný přízvukový takt, čímž vytváří tzv. kontextový přízvuk.

3.1.1. ČESKÉ MELODÉMY

Ukončující klesavý melodém (také značený M1) je typický pro realizaci oznamovacích a rozkazovacích vět a doplňovacích otázek. Charakterizuje ho pokles melodie od slabiky nesoucí větný přízvuk až po konec výpovědi. Tento pokles může být postupný i skokový a minimální rozdíl F_0 je 8 ST. Často se však dotýká dolní hranice rozsahu mluvčího. Melodém může být realizován kadencí příznakovou nebo bezpříznakovou.

Bezpríznamová kadence značí neutrální věcnou výpověď a charakterizuje ji, že první slabika kadence není melodicky výš než poslední slabika předchozího taktu. První slabika příznakové kadence je naopak zřetelně výš než slabika předcházející. (Palková, 1994: 309)

Ukončující stoupavý melodém (M2) odlišuje zjišťovací otázku od oznamovací věty. Charakterizuje ho náhlé stoupnutí melodie (7-9 ST). Ve spisovném standardu bývá realizován dvěma kadencemi: antikadencí stoupavou nebo antikadencí stoupavo-klesavou. Pro antikadenci stoupavou je typické stoupnutí melodie na poslední slabice výpovědi. V případě antikadence stoupavo-klesavé se zvyšuje první slabika po přízvuku nesoucím melodém a dále melodie klesá. Preference v používání jedné či druhé antikadence jsou dané především oblastně, přičemž někteří mluvčí užívají obou variant. Za nespisovnou antikadenci je považována kadence tzv. „pražské otázky“, charakteristická náhlým stoupnutím hned na první, přízvučné slabice. (Palková, 1994: 311-312)

Melodém neukončující (M3) se vyskytuje na konci větných celků, které předjímají pokračování. Kadence neukončujícího melodému jsou výrazně méně stabilní, mají mnoho variant a někdy jsou těžko identifikovatelné. Obecně je charakterizuje pokles na slabice, která předchází přízvučné slabice nesoucí melodém. V praxi je identifikace melodému složitější, obzvláště záleží na velikosti tónových intervalů. (Palková, 1994: 313-314)

Průběh intonace v souvětí závisí převážně na jeho délce. Souvětí složená z krátkých vět mohou být realizována jedním promluvovým úsekem a jsou zakončena melodémem odpovídajícím typu hlavní věty. Delší souvětí se dělí do více promluvových úseků spojených neukončujícími melodémy. Poslední úsek pak opět nese ukončující melodém dle typu hlavní věty.

Ve složitějších kontextech se také často komplikují vztahy mezi typem věty a melodémem. Například doplňovací otázka může být realizována stoupavým melodémem. V takovém případě se ovšem mění komunikační platnost výpovědi

a vnímáme ji jako řečnickou otázku, projev údivu nebo překvapení, případně může jít o zopakování položené otázky. (Palková, 1994: 315)

3.1.2. PŘÍZVUKOVÝ TAKT

V rámci promluvového úseku melodém určuje melodický průběh taktu, který jej nese, a pokud to není poslední takt úseku, tak i taktů, které po něm následují, a to až k hranici úseku. V taktech, které melodému předcházejí, ovšem také dochází k melodickým pohybům, které se řídí pozicemi a prominencemi přízvučných slabik a hranicemi taktů.

Palková a Volín (2003) ve studii zaměřené na percepci hranic přízvukových taktů v závislosti na kontuře F_0 ukázali, že stoupavě klesavé, lineárně stoupavé a lineárně klesavé kontury podporují kohezi jednotky. Naopak klesavě stoupavé podporují dělení taktů. Slabika klesavě stoupavé kontury s nejnižší F_0 bývá zpravidla první slabikou taktu.

4. Zpracování a interpretace intonace

Aby bylo možné z kontury F_0 získat lingvistickou informaci, je třeba ji vhodně upravit, a odhalit tak rysy, které jsou z hlediska popisované informace distinktivní. Základem pro takovéto zpracování je předpoklad, že kontura obsahuje redundantní prvky, které lze odstranit, aniž by došlo ke změně nebo ke ztrátě významu. K jejich odfiltrování a k předání lingvistické informace k abstraktnějšímu zpracování slouží intonační modely.

Podobně jako intonace samotná, ani pojem **intonační model** není zcela jednoznačně vymezen. Můžeme ho chápat obecně jako systém pro popis intonace, nebo můžeme vymezení specifikovat. Pfitzinger et al. (2009) pojem intonačního modelu chápou jako systém, který intonaci popisuje na fonologické úrovni, nejvyšší úrovni zpracování intonace jako zvukového jevu. Mezi surovými hodnotami F_0 extrahovanými ze signálu a intonačním modelem stojí další tři mezifáze: oprava extrakčních chyb, stylizace/vyhlazení kontury a modelování F_0 . Oprava chyb vzniklých při extrakci je záležitost spíše technická, otázka zpracování zvukového a digitálního signálu. Procesy stylizace a modelování kontury jsou již úkolem fonetiky.

Stylizace je zjednodušení a vyhlazení kontury F_0 . Má simulovat proces percepce intonace, takže jejím úkolem je z akustické kontury vytvořit konturu tak, jak ji pravděpodobně vnímáme. Snaží se odfiltrovat element mikrointonace a interpolovat konturu v místech, kde je přerušena neznělými hláskami. Model IPO, jeden z prvních stylizačních modelů, stanovil základní pravidlo stylizace, tj. že „*promluva resyntetizovaná se stylizovanou konturou by měla být percepčně stejná jako promluva resyntetizovaná s původní konturou*“ (Hermes, 2006). Ačkoli je možné provádět stylizaci manuální, dnes existuje mnoho algoritmů, které ji provádí automaticky. Asi nejjednodušším příkladem takového algoritmu je filtr typu dolní propust s mezní frekvencí cca 10 Hz, jaký najdeme i v programu Praat (Boersma & Weenink, 2015).

Stylizovaný průběh F_0 je stále pouze souvislý řetězec frekvenčních hodnot. K interpretaci je nutné jej rozložit na parametrizovatelné jednotky, které se

v promluvě daného jazyka, případně více jazyků, objevují pravidelně a tvoří jejich fonetický inventář. Tento rozklad na dílčí prvky nazýváme modelování F_0 a to je úkolem **fonetických intonačních modelů**. Fonetické modely vytváří rozhraní mezi konkrétním akustickým průběhem a abstraktní fonologickou reprezentací, se kterou pracují fonologické modely. Diskrétním kategoriím jednoznačně přiřazují zvukové reprezentace, definují je pomocí spojitých, objektivně kvantifikovatelných parametrů a stanovují pravidla pro jejich realizaci. Těmito parametry mohou být přímo měřené fyzikální veličiny (F_0 , intenzita, čas apod.) nebo častěji parametry sekundárně z těchto veličin odvozené při aproximaci kontury. Tilt model (Taylor, 2000) například rozkládá intonační konturu na intonační události, jejichž tvar charakterizuje tzv. *tilt* parametrem. Ten určuje podíl stoupavé a klesavé složky na základě jejich frekvenčních a časových rozměrů. Parametrizované události pak přiřazuje funkčním kategoriím definovaným například fonologickým systémem ToBI (Hirschberg & Beckman, 1994). Fonetické modely postupují od konkrétních vrstev k abstraktnějším, takže jazykovou funkci akustickým jevům přiřazují až sekundárně.

Fonologické intonační modely vytvářejí abstraktní jazykový systém a následně pro něj hledají konkrétní fonetické koreláty, případně interpretují foneticky stylizované průběhy a hledají pro ně místo v systému. K popisu intonace využívají diskrétní kategorie vymezené svou distinktivní funkcí. Systém ToBI, model vycházející z autosegmentálně-metrické fonologie, intonační konturu popisuje pomocí omezeného inventáře tónů a jejich konfigurací. Např. v americké angličtině, pro kterou byl systém původně navržen, jsou to nízký tón (L) a vysoký tón (H), jiné jazyky mají v inventáři také střední tón (M) – najdeme jej například i v systému ToBI pro češtinu (Duběda, 2014). Fonologické modely často pracují i s jevy, které v kontuře F_0 není možné pozorovat (Taylor, 1999), či s kategoriemi, které mají v kontuře stejný korelát, takže je nelze přímo implementovat v řečových technologiích – je třeba vytvořit pro ně fonetické rozhraní a diskrétní kategorie parametrizovat.

4.1. FONETICKÉ INTONAČNÍ MODELY A STYLIZACE

Každý fonetický model vychází z trochu jiných předpokladů o vnímání intonace, a využívá proto jiný způsob stylizace. Uvedeme si několik rysů, podle nichž lze modely klasifikovat, a přiblížíme si některé konkrétní z nich.

Modely můžeme rozlišit podle jejich uspořádání na **lineární** a **vrstvené**. V lineárních modelech (sekvenčních, jednovrstvých) je kontura tvořena jedinou sekvencí tónu či intonačních událostí. Tilt model například intonační průběh promluvy rozkládá na dílčí kontury intonačních událostí ukotvené na přízvučných slabikách promluvy a spojené rovnou čarou. Vrstvené modely (vícevrstvé, superpoziční) konturu modelují superpozicí dvou či více tónových vrstev. Příkladem je Fujisakiho model (Fujisaki & Ohno, 1997), který konturu tvoří součtem frázové kontury, přízvukové kontury a základové hlasové frekvence mluvčího.

Dále rozlišujeme modely **konturové** a **bodové**. Konturové modely předpokládají, že pro lidské vnímání je relevantní tvar kontury F_0 , tedy zda a jak moc je stoupavá, rovná či klesavá, případně zda je konvexní či konkávní. Bodové modely považují za relevantní rys klíčové body, kterými kontura prochází, a které tak určují její tvar. Tyto body mají reprezentovat abstraktní tónové cíle, ke kterým mluvčí postupně směřuje. Ve stylizacích jsou většinou identifikovány s *body obratu* (*turning points*), tedy body, ve kterých se konzistentně mění směr kontury. S tím souvisí i dělení modelů dle toho, zda konturu stylizují **hladkou křivkou** nebo **lomenou čarou**. Bodové modely většinou používají stylizaci lomenou čarou, aby byly body obratu ve stylizaci zřetelné, zatímco konturové modely často využívají stylizaci křivkami, např. polynomy (Grabe et al., 2007). Percepční rozdíl mezi hladkou a lomenou konturou je zanedbatelný (Hermes, 2006), proto záleží především na tom, zda přikládáme percepční důležitost přesnému umístění bodů obratu nebo celkovému tvaru kontury.

Modely mohou být buď **spojité**, nebo **tonální**. Spojité modely předpokládají, že člověk vnímá intonaci v intonační frázi jako spojitý průběh F_0 a veškerá přerušování způsobená neznělými segmenty řeči jsou při percepci doplněna. Do této kategorie

spadají všechny doposud zmíněné modely. Podle tonálních modelů vnímáme intonaci jako sled diskrétních tónů. S tímto předpokladem pracuje například Prosogram (Mertens, 2004).

Dále se podíváme podrobněji na některé používané stylizace a fonetické modely a na principy, na kterých zpracovávají a popisují intonaci.

4.1.1. DOLNÍ PROPUST

Filtr typu dolní propust je primitivní způsob stylizace, který v kontuře vyhlazuje všechny změny rychlejší než daná mezní frekvence³. Scheffers (1988, cit. v Pfitzinger & al., 2009) však uvádí, že prostý filtr není dostatečný pro odstranění mikrointonace a naopak ovlivňuje strmosti kontury a další důležité prvky kontury. Na druhou stranu modulace F_0 je málokdy rychlejší než 3,5 Hz (Pfitzinger & al., 2009). Výhodou dolní propusti je její jednoduchost, univerzálnost a snadná implementovatelnost. Nevýhodou je, že jde o čistě matematické vyhlazení, které nereflektuje zákonitosti percepce tónů.

4.1.2. MoMEL (HIRST & ESPESSER, 1993)

MoMel (zkratka z „*modélisation de mélodie*“, tj. modelování melodie) je algoritmus pro spojitou hladkou stylizaci kontury pomocí cílových bodů. Jeho smyslem je odfiltrování mikrointonační složky, která je při percepci irelevantní. Pracuje tak, že ze surové kontury nejprve odstraní zjevně chybné hodnoty F_0 , které se od okolních hodnot liší více, než je stanovený práh. Poté pomocí kvadratické regrese stanoví kandidáty na cílové body a jejich počet postupně zredukuje srovnáváním s lokálními trendy v kontuře. Výsledné cílové body pak víceméně odpovídají lokálním maximům, minimům a bodům inflexe vyhlazené kontury (Hermes, 2006). Stylizovaná kontura vznikne proložením cílových bodů kvadratickými splajny. Podobně jako v předchozím případě je i zde nevýhodou, že algoritmus je čistě statistická procedura, která nereflektuje lidské vnímání tónů.

³ Podotýkáme, že se zde nejedná o základní frekvenci F_0 , nýbrž o frekvenci, která základní frekvenci moduluje, a udává tak rychlost jejích změn.

4.1.3. BAGSHAWOVA STYLIZACE (BAGSHAW, 1993)

Stylizace dle Bagshawa (1993) navazuje na stylizaci Scheffersovu (1988) a je příkladem stylizace lomenou čarou. Surové hodnoty F_0 jsou nejprve filtrovány mediánovým filtrem a Hannovým vyhlazovačem, které mají odstranit většinu oktávových skoků a snížit mikrintonační výkyvy. V takto vyhlazené kontuře algoritmus hledá regresní metodou kandidáty na body obratu. Chronologicky prokládá hodnoty F_0 regresní přímkou a kontroluje, jak moc se další vzorky F_0 od této přímky odchylojí. Pokud absolutní odchylka překročí povolený práh (1 ST), a toto překročení trvá déle, než je povolená doba (100 ms), algoritmus se vrací k prvnímu vzorku F_0 , který překročil frekvenční práh, a označí jej jako bod obratu. Celý proces se opakuje znovu od tohoto bodu, dokud není dosažen konec kontury. Body obratu jsou poté lineárně propojeny.

Součástí Bagshawovy stylizace je také automatická detekce slabičných jader a hranic na základě sonority signálu. Při fonetické analýze jsou brány v potaz jen ty úseky kontury, které se kryjí s detekovanými slabičnými jádry.

4.1.4. INTSINT (HIRST & DI CRISTO, 1998)

INTSINT je zkratkou pro „*international transcription system for intonation*“, tedy „mezinárodní transkripční systém pro intonaci“. K popisu INTSINT používá inventář 8 tónů, které lze generovat automaticky ze stylizované kontury (systém byl navržen především pro popis stylizace algoritmem MoMel), jejichž výška se odvozuje v půltónové stupnici z výškového rozsahu mluvčího.

ABSOLUTNÍ TÓNY (definované vůči intonačnímu rozsahu):

- *Top (T)*. Horní hranice rozsahu.
- *Middle (M)*. Intonační střed, tj. logaritmický střed rozsahu.
- *Bottom (B)*. Spodní hranice rozsahu.

RELATIVNÍ TÓNY (definované relativně vůči rozsahu a předchozímu tónu):

- *Higher (H)*. Průměr frekvence předchozího tónu a horní hranice rozsahu, tj.:

$$f_H = f_n = \frac{f_{n-1} + f_T}{2}.$$

- *Upstep (U)*. Průměr frekvence předchozího tónu a tónu *H*, tj.:

$$f_U = f_n = \frac{f_{n-1} + f_H}{2} = \frac{3f_{n-1} + f_T}{4}.$$

- *Same (S)*. Tón se stejnou výškou jako bod předchozí.
- *Downstep (D)*. Průměr frekvence předchozího tónu a tónu *L*.
- *Lower (L)*. Průměr frekvence předchozího tónu a spodní hranice rozsahu.

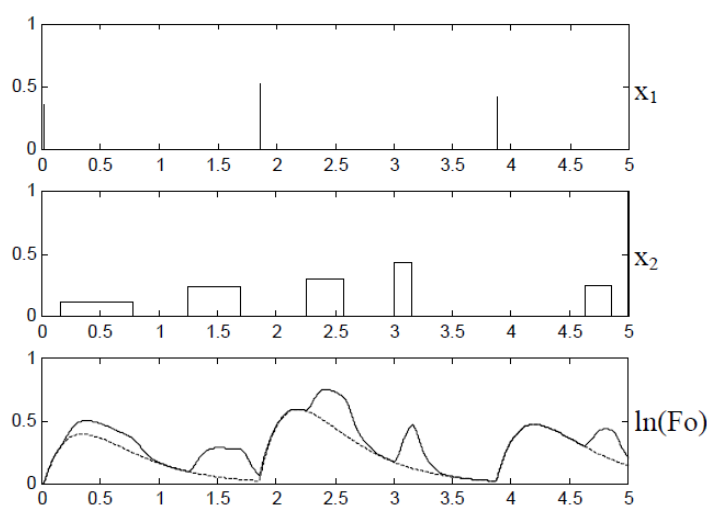
Algoritmus INTSINT je navržen pro transkripci již stylizované kontury, ideálně vytvořené algoritmem MoMel. Hledá v kontuře body obratu a kóduje je tóny z uvedeného inventáře, čímž dochází k zaokrouhlení jejich frekvencí a tedy k další stylizaci. Výsledné tóny pak mohou být v transkripci zarovnaný buďto s konturou ve spojitém čase (třeba prostřednictvím TextGridu v programu Praat), nebo diskrétně s ortografickým přepisem textu.

INTSINT navíc využívá doplňující symboly signalizující posunuté načasování tónu vůči levé hranici slabiky (např. <*H* značí brzký (*early*) vyšší tón, >*U* značí zpožděný (*late*) upstep). Systém dále používá hranatých závorek k reprezentaci hranic vyšších prozodických jednotek – intonačních frází a nádechových úseků. Extrémní hraniční tóny mohou být vyjádřeny kombinací absolutního tónu s hranatými závorkami, např. *T]* nebo *B]]*.

4.1.5. FUJISAKIHO MODEL (FUJISAKI & OHNO, 1997)

Fujisakiho model je vrstvený spojitý hladký konturový model. Vychází z předpokladu, že kontura F_0 je tvořena dvěma komponenty: „frázovým“ („*phrase component*“) a „přízvukovým“, resp. „tónovým“ („*accent component*“ nebo „*tone component*“), podle toho, zda se jedná o jazyk s tónovým přízvukem, nebo o jazyk tónový. Frázový komponent je pomalu proměnlivý průběh pokrývající celou větu nebo intonační úsek (frázi). Přízvukový (či tónový) komponent je mnohem více proměnlivý a je úzce spojen se slovním přízvukem. Pokud od kontury F_0 odečteme tzv. základovou hlasovou frekvenci, která je pro každého mluvčího individuální a konstantní, zjišťujeme, že frázové a přízvukové komponenty jsou na mluvčím nezávislé.

V mnoha známých jazycích pozorujeme v rámci intonační fráze tzv. *deklinaci*, postupný pokles F_0 . Fujisaki proto frázový komponent modeluje jako odezvu stabilní lineární soustavy druhého řádu na impuls (který představuje tzv. frázový příkaz). Tento systém je poměrně flexibilní v tom smyslu, že snížením amplitudy vstupních impulsů lze modelovat mluvčí, u nichž se deklinace příliš neprojevuje. Přízvukové komponenty se modelují jako odezva stabilní lineární soustavy druhého řádu na obdélníkovou funkci (která představuje tzv. přízvukový příkaz). Autoři uvádí, že i jinými aproximacemi kontury F_0 (pomocí spline křivky nebo po částech definované parabolické funkce) můžeme dojít k podobně přesným výsledkům, avšak pouze popsané modelování pomocí odezev „*má základ ve fyzikálním mechanismu*“ (Fujisaki & Ohno, 1997). Výsledná stylizovaná kontura je pak určena amplitudami vstupních impulsů a skokových funkcí a jejich časovým ukotvením. Proces syntézy je graficky znázorněn na Obr. 1.



Obrázek 1: Grafické znázornění Fujisakiho syntézy stylizované křivky ze dvou příkazových vrstev.

Horní graf představuje vrstvu frázových příkazů, prostřední graf vrstvu akcentových příkazů, spodní graf výslednou konturu vzniklou filtrace a superpozicí obou vrstev. Převzato z Rossi et al. (2002).

Mixdorff (2000) uvádí metodu pro plně automatickou extrakci parametrů Fujisakiho modelu. K extrakci a vyhlazení kontury od mikrintonační složky používá výše zmíněný MoMel. Jeho výstup pak rozfiltruje na nízkofrekvenční a vysokofrekvenční složku. V nízkofrekvenční vrstvě, která představuje frázový komponent, detekuje lokální maxima a z jejich amplitudy a časového ukotvení určí amplitudu a časové ukotvení

frázových příkazů. Podobně z vysokofrekvenční složky určí parametry přízvukových příkazů. Algoritmus poté metodou analýzy syntézou zkouší upravovat parametry za účelem získání kontury, která se od té původní liší s menším rozptylem.

4.1.6. TILT MODEL (TAYLOR, 1999)

Tilt model je fonetický model pro spojitou konturovou stylizaci. Jedním z jeho hlavních účelů bylo usnadnit robustní komputační analýzu a syntézu řeči. Základní jednotkou modelu je intonační událost. Tyto události se v čase dějí jako diskrétní okamžiky, nejde tedy o spojitý řetězec jednotek jako na segmentální rovině. Model vychází z autosegmentálně-metrické fonologie (dále viz sekce ToBI, 4.2.1). Rozlišuje dva základní typy událostí: melodické přízvuky (*pitch accents*) a předělové tóny (*boundary tones*). Melodický přízvuk je melodický pohyb svázaný s konkrétní slabikou a slouží k vytváření důrazu. Předělový tón je stoupavý melodický pohyb vyskytující se na okrajích intonační fráze a kromě toho, že signalizuje konec fráze, také může určovat její komunikační platnost.

Popis intonačních událostí navazuje na tzv. RFC model (*Rise/Fall/Connection*, tj. „stoupání/pokles/spojení“), (Taylor, 1994). Podle něj se intonační událost obecně skládá ze stoupavé a klesavé části. Obě části jsou definovány svým trváním a svou amplitudou, takže celou událost definují čtyři parametry. Pokud se událost skládá např. pouze z poklesu, potom uvažujeme amplitudu a trvání jejího stoupnutí rovné nule.

Tilt model se od RFC modelu liší v tom, že používá jen tři parametry zvané Tilt parametry: *trvání*, *amplituda* a *náklon* (anglicky „*tilt*“). Trvání je dáno součtem trvání stoupavé a klesavé části. Amplituda je rovna rozdílu velikostí amplitudy stoupání a amplitudy klesání. Náklon je bezrozměrný parametr udávající tvar intonační události a je definován jako průměr relativního trvání a relativních amplitud stoupání a poklesu:

$$tilt = \frac{tilt_{amp}}{2} + \frac{tilt_{dur}}{2} = \frac{|A_{rise}| - |A_{fall}|}{2(|A_{rise}| + |A_{fall}|)} + \frac{D_{rise} - D_{fall}}{2(D_{rise} + D_{fall})}$$

(Taylor, 1999)

A_{rise}, A_{fall} ... amplituda stoupnutí, amplituda poklesu

D_{rise}, D_{fall} ... trvání stoupnutí, trvání poklesu

Hodnota parametru náklonu se pohybuje v intervalu $\langle -1; +1 \rangle$. Hraniční hodnota $tilt = -1$ odpovídá čistě klesavé kontuře, $tilt = 1$ stoupavé kontuře. Hodnota $tilt = 0$ reprezentuje stoupavě klesavou konturu. Klesavě stoupavá kontura musí být vyjádřena dvěma samostatnými intonačními událostmi.

Taylor ve stejné studii také navrhuje detektor a klasifikátor intonačních událostí. Detektor slouží k identifikaci výškových přízvuků a hraničních tónů v kontuře a funguje na bázi skrytých Markovových modelů trénovaných na ručně označovaném řečovém korpusu. Detekované události poté vstupují do klasifikátoru, který z vlastností jejich kontury vypočítá jejich Tilt parametry.

4.1.7. MODEL TONÁLNÍ PERCEPCE (MERTENS & D'ALESSANDRO, 1995)

Model tonální percepce vychází z následujících zákonitostí vnímání výšky tónu:

- 1) Variace amplitudy a spektrálního složení řečového signálu způsobuje tzv. segmentační efekt, díky kterému konturu F_0 nevnímáme jako spojitou, ale jako sekvenci slabičných tónů (Mertens & d'Alessandro, 1995). K největším změnám v kontuře dochází právě na hranicích slabik, tedy v místech, kde je intenzita a sonorita signálu relativně nízká (Mertens, 2004).
- 2) Abychom byli schopni vnímat stoupavý či klesavý pohyb F_0 , musí strmost kontury překročit mezní hodnotu strmosti⁴ (*glissando threshold*) g_{thr} , která je přibližně $g = \frac{0,16}{T^2}$, kde g je strmost kontury v půltónech za sekundu a T je trvání frekvenční změny v sekundách. Pokud je sklon kontury tónu nižší než tato hodnota, vnímáme jej jako statický tón, pokud hodnotu překročí, je klasifikován jako dynamický. Mertens (2004) dále předpokládá, že aby mohl mít sklon dynamického tónu kategorickou funkci, musí být alespoň dvojnásobný oproti prahu vnímatelnosti, tj. minimálně $g > \frac{0,32}{T^2}$. V implementaci navíc stanovuje práh pro trvání tónu d_{min} , při němž je možné vnímat v jeho průběhu změny výšky (35 nebo 50 ms).

⁴ Přesnější překlad pojmu „*glissando threshold*“ by zněl „práh klouzavosti“, avšak protože v této práci dále operujeme s pojmem *strmost*, rozhodli jsme se kvůli ozřejmění souvislosti obou pojmů pro „práh strmosti“.

- 3) Abychom byli schopni vnímat změnu strmosti kontury, musí se strmost konturových segmentů lišit alespoň o $DG = |g_2 - g_1| = 20$ ST/s, kde DG je *differential glissando threshold*, práh difference strmosti, a g_1 a g_2 jsou strmosti segmentů v jednotkách ST/s. Hodnota DG doposud nebyla ověřena experimentálně, d'Alessandro a Mertens (1995) ji stanovili arbitrárně z rozsahu 12 až 40 ST/s odvozeného na základě relevantní literatury. Pokud je rozdíl strmostí menší než DG , tón je identifikován jako jednoduchý a je modelován jedinou úsečkou, pokud je rozdíl větší, jde o tón složený a je modelován dvěma úsečkami.
- 4) Krátkodobé frekvenční variace, které nepřesahují práh klouzavosti, jsou vnímány jako statické tóny. Pravděpodobně dochází k percepční integraci kontury F_0 , kterou lze simulovat časovým průměrováním (WTA, *windowed time average function*).

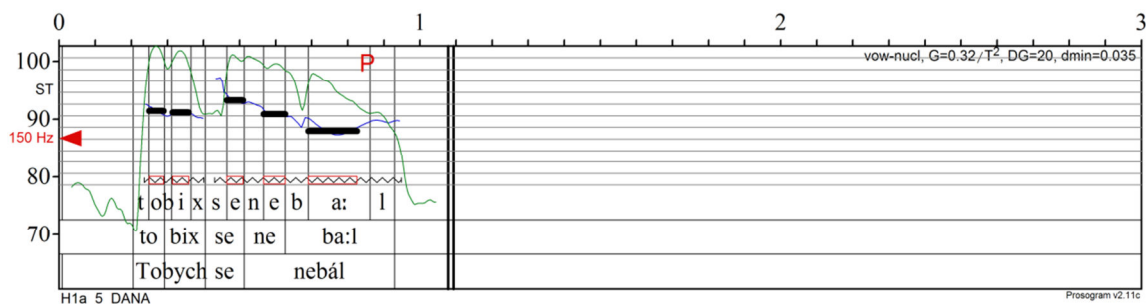
Stylizační algoritmus tonálního modelu postupně provádí

- 1) automatickou akustickou **segmentaci signálu** na slabiky a detekci nukleů na základě intenzity signálu a detekované znělosti;
- 2) krátkodobou percepční **integraci kontury F_0** pomocí WTA, která má za následek vyhlazení kontury;
- 3) **segmentaci slabičných kontur**, při níž je každá slabičná kontura identifikována jako jednoduchá, nebo složená, a každý segment kontury jako statický, nebo dynamický;
- 4) **stylizaci**, při níž jsou dopočítány tonální cíle kontur a pomocí lineární interpolace vzniká výsledná kontura.

Mertens (2004) popsaný algoritmus implementoval formou sbírky skriptů pro program Praat, kterou pojmenoval *Prosogram*.⁵ Ten provádí extrakci, segmentaci a stylizaci průběhu F_0 a umožňuje úpravu některých parametrů (periody vzorkování F_0 ,

⁵ Prosogram je volně dostupný ke stažení na webu <http://bach.arts.kuleuven.be/pmertens/prosogram/download.html>

prahu klouzavosti, diferenciálního prahu klouzavosti a dalších). Kromě toho poskytuje grafickou reprezentaci stylizované kontury včetně popisu (viz Obr. 2) a další užitečné informace získané z nahrávky, jako například intonační rozpětí, intonační střed a mluvní tempo mluvčího. Navíc umožňuje využít externí segmentaci signálu, což je užitečné, pokud máme k dispozici manuálně segmentovaný řečový korpus.



Obrázek 2: Prosogram realizace věty „To bych se nebál.“

Tenká modrá čára znázorňuje průběh F_0 , zelená čára průběh intenzity a tučné černé linky reprezentují stylizované tóny. Vytvořeno v programu Praat s pomocí sady skriptů Prosogram.

4.2. FONOLOGICKÉ INTONAČNÍ MODELY

Po percepčním zpracování intonace a identifikaci relevantních rysů dochází k jejich klasifikaci, diskretizaci a vyhodnocení jejich lingvistické funkce, což simulují fonologické modely. Protože se však jedná o velmi abstraktní procesy, které nelze zkoumat přímo, existuje o jejich povaze množství teorií, každá s vlastními fonologickými modely. Navíc každý jazyk – ač v mnoha případech nalezneme společné rysy – má svá vlastní fonologická pravidla, proto není možné modely používat univerzálně, ale je nutné adaptovat je potřebám konkrétního jazyka.

Ačkoli se tato práce věnuje zkoumání fonetickému, pro lingvisticky relevantní popis řeči je třeba brát v potaz abstraktnější vrstvy popisu. Proto si zde zběžně přiblížíme i fungování fonologického modelu.

4.2.1. TOBI

ToBI, zkratka pro „*tones and break indices*“ (tóny a předělové indexy), je systém původně vyvinutý k transkripci prozodie americké angličtiny, dnes již adaptovaný pro popis více než desítky dalších jazyků. Je vytvořen na principech autosegmentálně-metrické

fonologie. Pro anglickou intonaci je dnes ToBI již zažitý způsob fonologického zápisu, o čemž svědčí i to, že bývá úspěšně propojen s fonetickými modely (Grabe et al., 2007; Taylor, 1999).

V autosegmentálně-metrickém popisu je fonologická podoba řeči rozložena do synchronních hierarchických vrstev. Jednotky každé vrstvy jsou seřazeny sekvenčně. Mezi každými dvěma sousedícími vrstvami leží vrstva spojů, které podřizují nižší jednotky vyšším jednotkám. Tak jsou např. segmenty sdružovány ve slabiky a slabiky v takty.

Základem transkripce ToBI jsou 4 vrstvy: tónová, ortografická, předělová a „různé“.⁶ Ortografická vrstva poskytuje ortografický přepis oddělených slov. Předělová vrstva popisuje předěly mezi slovy číselnou stupnicí od 0 do 4 indikující sílu předělu. Vrstva „různé“ slouží ke značení paralingvistických jevů (smích, změna hlasu atd.) a ke vkládání upřesňujících komentářů.

K samotnému popisu melodie řeči slouží tónová vrstva. Americký ToBI rozlišuje 2 základní úrovně tónů: vysokou (*H*) a nízkou (*L*), a dva typy tónů: výškový přízvuk (*pitch accent*), který se váže k přízvučné slabice, a frázové tóny (*phrasal tones*), spojené s hranicí prozodických jednotek. Frázové tóny se dále dělí na frázový přízvuk (*phrase accent*) a hraniční tón (*boundary tone*). **Frázový přízvuk** je intonační pohyb spjatý s koncovou hranicí tzv. střední fráze (*intermediate phrase*), prozodické jednotky vyšší než přízvukový takt a nižší než intonační fráze. Podle úrovně tónu se značí *H-* nebo *L-*. **Hraniční tón** je melodický pohyb na konci intonační fráze a značí se znakem procenta, tj. *H%* nebo *L%*. Protože hranice intonační fráze je zároveň hranicí střední fráze, vyskytuje se hraniční tón vždy ve spojení s frázovým přízvukem, například *L-H%*. **Výškový přízvuk** je spjat s každou přízvučnou slabikou promluvy a značí se hvězdičkou. Například „vrcholový přízvuk“ *H** značí přízvukování slabiky pomocí stoupnutí směrem k horní hranici rozsahu a následného poklesu. Inventář systému

⁶ Transkripce se obvykle realizuje formou TextGridu v programu Praat, případně v jiném softwaru, který umožňuje časově zarovnaný popis zvuku.

dále obsahuje symboly pro upřesnění načasování tónu (< a >), pro downstep (*!H*) a další jevy. (Hirschberg & Beckman, 1994)

Pro českou intonaci systém ToBI adaptoval Duběda (2014). Vychází z tradičního Danešova popisu intonace a formálně ho přizpůsobuje konvencím ToBI.

4.3. RELEVANTNÍ PRVKY KONTURY ZÁKLADNÍ FREKVENCE

Jak jsme se dozvěděli ve 3.1, konturu F_0 lze rozkládat na různé dílčí jednotky. Pokud se omezíme na stylizaci lomenou čarou, pak za základní, nejjednodušší prvky popisu lze považovat body obratu (*turning points*), melodické pohyby (*pitch movements*) a melodické skoky (*pitch jumps*), (Hermes, 2006: 43-45). Všechny tyto prvky jsou určeny svými „souřadnicemi“ ve frekvenční doméně a v časové doméně a svou asociací s konkrétní slabikou či jinou prozodickou jednotkou. Podstatné jsou vždy jejich relativní hodnoty vůči okolí, např. v případě bodu obratu je to hodnota jeho F_0 ve srovnání se sousedními body a načasování vůči nukleu slabiky, k níž se bod váže.

Body obratu jsou takové body, v nichž kontura F_0 konzistentně mění svůj směr. Často to bývají její lokální maxima (*vrcholy*) a minima (*sedla* či *údolí*). Zároveň jsou body obratu logicky krajními body melodických pohybů a určují tak jejich začátek a konec.

Načasování bodů obratu může rozhodovat o načasování celého melodického pohybu, a tím i určit, která slabika získá akcent a o jaký typ akcentu půjde. V nizozemštině, angličtině a švédštině je rozhodující zarovnání začátku melodického pohybu vůči levé hranici nuklea slabiky. Pokud začíná pohyb nanejvýš několik desítek milisekund před hranicí a nejpozději několik desítek milisekund po ní, slabika je vnímána jako přízvučná. Pokud jde o stoupavý pohyb a začíná před levou hranicí nuklea, přízvuk je vnímán jako vysoký, pokud až po hranici, přízvuk je nízký. Pro klesavý pohyb to platí naopak (Hermes, 2006: 44).

Podobně umístění bodu obratu tvořícího vrchol kontury může mít kategorickou funkci (Pierrehumbert & Steele, 1989). Některé studie (Dilley & Brown, 2007; Xu, 1999) však ukazují, že načasování vrcholu vůči hranici slabičného jádra souvisí se strmostí

kontury. Zpožděný vrchol tak může být jen důsledkem většího výškového skoku nebo rychlejšího mluvního tempa, tedy záležitostí fyziologie produkce intonace.

Melodický pohyb je plynulá změna výšky F_0 v čase. Jednoduchý melodický pohyb je v kontuře vymezen úsečkou spojující dva body obratu, má tedy počáteční frekvenci a čas a konečnou frekvenci a čas. Složený pohyb je tvořený více bezprostředně navazujícími úsečkami. Podle směru rozlišujeme pohyby stoupavé, klesavé, klesavě-stoupavé a stoupavě-klesavé. Aby však bylo možné plynulou změnu F_0 vnímat jako melodický pohyb, musí probíhat ve spektrálně stabilních segmentech řeči, tj. slabičných jádrech. Pokud se děje mimo jádro, vnímáme jen výškový rozdíl mezi sousedními jádry, který pohyb způsobuje – melodický skok. Melodické pohyby jsou charakterizované parametrem (průměrné) strmosti g [ST/s], tedy výslednou poměrnou změnou F_0 za dobu trvání pohybu.⁷

Melodický pohyb můžeme chápat i na úrovni taktů či výpovědi jako tendenci určující postupnou změnu výšky v po sobě jdoucích slabikách. Takto bývají základním a intuitivním prvkem popisu intonace pro mnohé jazyky, viz např. české kadence (Palková, 1994) nebo tradiční britský popis anglické intonace (Cruttenden, 1997). Tento delší pohyb můžeme také vyjádřit několika výškovými skoky, jak to dělá například Prosogram.

Strmost kontury může mít podobný percepční účinek jako načasování intonační události. Barnes et al. (2008) navrhli pro zohlednění strmosti nástupu melodického přízvuku H^* v angličtině určovat pozici vrcholového bodu obratu pomocí tonálního těžiště (*tonal center of gravity*). To počítá jako časové těžiště bodů kontury vážených jejich frekvencí. Například vrcholový přízvuk, jehož nástup je strmější než sestup, bude mít tonální těžiště oproti vrcholovému bodu zpožděné. Umístění těžiště odráží také konvexnost či konkávnost kontury.

⁷ Prahové hodnoty vnímání strmosti jsme uvedli v sekci Model tonální percepce (4.1.7).

Kategorizaci přízvuků na základě načasování a tvarování vrcholů a sedel využívá i kielský intonační model pro popis němčiny. Niebuhr (2014) uvedl, že jak načasování, tak tvarování lze použít se stejným účinkem bez percepčně znatelných rozdílů. Lze vystopovat jisté individuální tendence, někteří lidé využívají častěji načasování vrcholu, jiní zase spíše tvarují tím, že mění strmost nástupu a sestupu.

Melodickou konturu můžeme popisovat i z pohledu vyšších jednotek, nejen jako sled dílčích intonačních událostí. Celkové trendy v kontuře pak mohou souviset s paralingvistickými funkcemi intonace – zda na nás projev působí expresivně, zdali ho vnímáme jako živý nebo monotónní apod.

Jedním z častých aspektů popisu je intonační rozpětí a hlasová poloha. Palková (1997) uvádí, že zvýšení intonačního rozpětí a zvětšení melodických skoků je v češtině charakteristické pro expresivní promluvy. Emocionální řeč se projevuje nadprůměrně častými stoupavými kadencemi, zvyšováním celého melodému a zmírněním klesavých tendencí.

5. Praktická část

V této části se zaměříme na instrumentální fonetickou analýzu průběhu F_0 v českých promluvách ve vztahu k jejich segmentálnímu složení. Na základě poznatků z teoretické části práce se pokusíme analyzovat umístění bodů obratu vůči hranicím slabičného jádra, analyzujeme využití dynamických tónů (dle tonálně percepčního modelu) a dále se podíváme na melodické skoky a strmost kontury. Naměřené hodnoty statisticky zpracujeme a pokusíme se je dát do souvislosti s prozodickými jednotkami a případně tak odhalit tendence v české intonaci.

5.1. DATA

Pro účely analýzy jsme využili nahrávky z korpusu *Minidialogy* vytvořeného a rozvíjeného na Fonetickém ústavu na FF UK v Praze. Jde o krátké dialogy o délce čtyř až pěti replik, předem nacvičené a realizované dvoučlennými skupinami studentů filologických oborů na FF UK v Praze ve věku 20–25 let. Při nahrávání bylo dbáno na to, aby byly dialogy zahrány co nejpřirozeněji.

Dialogy v korpusu jsou separovány na jednotlivé repliky a seříděny dle jejich obsahu a dle mluvčích. Ke každé je k dispozici soubor typu *TextGrid* s ortografickým přepisem repliky, rozdělením na promluvvé úseky, slova, slabiky, hlásky a fonémy a rozlišením konsonantů a vokálů. K segmentaci na hlásky byl použit program *Prague Labeller* (Pollák et al., 2007), hranice segmentů pak byly manuálně doopraveny dle Machače a Skarnitzla (2009). V rámci této bakalářské práce jsme se sami na segmentaci některých dialogů podíleli.

Z části korpusu, která nám byla přidělena, jsme vybrali 10 replik, každou realizovanou 20 stejnými mluvčími. Pro zjednodušení zpracování jsme vybírali z replik složených nejvýše ze dvou syntaktických vět (klauzí) a jejichž realizace by tedy neměla být tvořena více než dvěma promluvvými úseky. Delší repliky znamenají větší variabilitu v prozodickém členění a umístění melodémů a větných přízvuků, čemuž

bychom se rádi vyhnuli. Zároveň jsme výběr volili tak, aby se v realizacích objevily všechny tři základní druhy melodémů. Vybrané repliky uvádíme v Tabulce 1. Mluvčí byli zvoleni náhodně, pouze s ohledem na zastoupení obou pohlaví, takže vzorek tvoří 5 mužů a 15 žen.

Kód	Věty	Typ	Znění
H1a_5	1	sdělení	To bych se nebál.
H2d_1	2	otázka	Máte nějakou představu, jak to bude probíhat?
H3a_1	2	sdělení	Až začnou o tom transportu, nastražíš uši.
H3b_5	1	otázka	Na internetu to není?
H3c_1	2	sdělení	Trochu se bojím, že je tím rozčillíte.
H4b_1	1	sdělení	Hodilo by se trochu víc informací.
H5c_1	1	sdělení	Takže všechno to teď závisí na vás.
H5d_5	1	otázka	A ze vzduchu nic vidět není?
H6c_5	1	sdělení	A právě s tím nikdo nepočítá.
H6d_1	2	sdělení	Jen abyste měli dost peněz, až to přijde.

Tabulka 1: Repliky vybrané pro analýzu

Z celkem 200 realizací jsme jich 7 vyřadili kvůli vynechanému nebo přidanému slovu, případně kvůli nápadným řečovým redukciím.

Z výsledných 193 replik jsme v programu Praat extrahovali konturu základní frekvence, konkrétně autokorelačním algoritmem *To Pitch (ac)*...⁸ se vzorkovací periodou 5 ms, detekčním rozsahem 60 až 450 Hz a ostatními parametry v jejich výchozích hodnotách. Získané kontury jsme opravili algoritmem pro odstranění oktávových skoků a použili jsme vyhlazovací funkci *Smooth*... s parametrem 10 Hz k primitivnímu odfiltrování mikointonačních výkyvů. Kontury jsme převedli na objekty *PitchTier*, které jsou pro analýzu vhodnější, neboť intervaly mezi jednotlivými vzorky F_0 interpoluje a prokládá tak neznělé segmenty.⁹ Jejich další zpracování jsme řídili podle konkrétních potřeb analýzy (např. ke zkoumání bodů obratu jsme konturu segmentovali na slabičné nukleý, o tom však dále).

⁸ K extrakci kontury jsme vyzkoušeli i další dvě metody, které Praat nabízí (vzájemná korelace a subharmonický součet), ale autokorelační algoritmus se ukázal jako zdaleka nejspolehlivější jak z pohledu detekce znělosti, tak i v odolnosti vůči mikointonaci.

⁹ V některých bodech této práce (analýza dynamických tónů a melodických skoků) využíváme konturu extrahovanou a stylizovanou skriptem Prosogram, takže se jich popsané zpracování netýká.

5.1.1. KLASIFIKACE DAT

Abychom mohli zkoumat závislosti naměřených hodnot na jazykových faktorech, provedli jsme klasifikaci zkoumaných dat. Jednotlivé segmenty byly rozděleny do tříd konsonant/vokál, znělý/neznělý, je/není jádrem slabiky.¹⁰ Slabiky jsme označili z hlediska znělosti prétury, z hlediska přítomnosti slovního přízvuku a z hlediska pozice slabiky vůči přízvuku (přízvučná slabika byla označena číslem 1, slabika následující číslem 2 a slabika následující po ní číslem 3; pozdější slabiky a předklonky nebyly s ohledem na rozsah této práce rozlišeny a označili jsme je NA). Na základě identifikace přízvuků jsme popisný TextGrid replik rozšířili o dělení replik na přízvukové takty a poslechem jsme určili, které z taktů nesou větný melodém a o jaký základní typ melodému jde.¹¹ Jednotlivé nahrávky pak byly seříděny podle repliky, mluvčího a u každé jsme pomocí skriptu TXT-artTempo.praat¹² spočítali artikulační tempo.

5.1.2. POUŽITÉ NÁSTROJE

K analýze zvukových souborů a k práci s jejich konturami a příslušnými TextGridy jsme využili již zmíněný program *Praat* (Boersma & Weenink, 2015) a doplňky *Prosogram* (Mertens, 2004) a *INTSINT* (Hirst, 1998). Ke složitější analýze byly vytvořeny skripty v jazyce R pomocí programu *RStudio* (R Core Team, 2015).

Základní matematická zpracování naměřených hodnot a orientační statistické výpočty jsme prováděli v programu Microsoft Excel. Finální statistickou analýzu a vykreslení grafů jsme provedli opět v jazyce R.

¹⁰ Zjistili jsme, že ve výběru se nevyskytují žádné slabičné sonorní konsonanty, což znamená, že každé slabičné jádro je vokál a každý vokál tvoří slabičné jádro. Proto si dále vystačíme jen s rozlišením konsonant/vokál.

¹¹ Vzhledem k tomu, že detekci a klasifikaci přízvuků a melodémů prováděla pouze jedna osoba, bude třeba brát ji pouze jako orientační. V případě, že bychom objevili nějaké tendence související s těmito kategoriemi, bylo by třeba experiment zopakovat s daty klasifikovanými vícečlenným týmem foneticky trénovaných osob.

¹² Bližší informace o skriptech, které jsme vytvořili pro účely této práce, uvádíme v přílohách, sekce Skripty k programu Praat (8.1.2) a Skripty v jazyce R (8.1.3).

5.2. BODY OBRATU

Ve zkoumání bodů obratu jsme se rozhodli vycházet z předpokladu, že nejvíce intonační informace při poslechu získáváme ze spektrálně stabilních částí řeči, tj. slabičného jádra. Abychom mohli vnímat přítomnost a pozici bodu obratu, musí se nacházet ve slabičném jádře nebo blízko jeho hranic. Zároveň, jak uvádí např. Hermes (2007), relevantní bývá jeho posun vůči levé hranici jádra.

Naším cílem je zjistit, zda v češtině existuje tendence v umístění bodů obratu vůči hranici přetury a jádra a případně zda tento parametr nevytváří kategorie, které by mohly mít distinktivní platnost, jako např. v němčině.

5.2.1. METODA

K detekci bodů obratu a stylizaci kontury F_0 jsme se nejprve pokusili vybrat vhodný fonetický model. Prvním kritériem výběru byl princip modelu, totiž aby detekci bodů vůbec umožňoval. Tím jsme vyloučili modely konturové (př. Tilt, Fujisaki) a modely tonální (Prosogram). Dalším kritériem byla jednoduchost použití, ideálně model ve formě již hotového a volně dostupného programu nebo skriptu. Tím odpadl Bagshawův algoritmus, jehož funkčnost je sice dobře popsána (Bagshaw, 1993), ale který by byl poměrně náročný na implementaci. Sítím požadavků prošel algoritmus MoMel¹³ a vestavěný stylizační algoritmus Praatu.

Po otestování na několika replikách z korpusu se MoMel ukázal jako nevhodný a to hned z několika důvodů. Algoritmus nebere v potaz manuální segmentaci a klasifikaci (na rozdíl např. od Prosogramu) a při stylizaci nerozlišuje podstatnou konturu slabičného nuklea od méně důležité kontury zbytku signálu. S tím souvisí i jeho další nedostatek, totiž že body obratu neumísťuje s ohledem na segmentální složení promluvy a stylizace tak ztrácí své jazykové opodstatnění. Při poslechovém porovnání repliky resyntetizované s původní konturou a se stylizovanou konturou byly

¹³MoMel ve formě pluginu pro Praat byl získán z webu <http://uk.groups.yahoo.com/group/praat-users/files/Daniel_Hirst/>

slyšet značné rozdíly, nelze zde tedy hovořit o percepční rovnosti stylizace. MoMel navíc nenabízí žádné nastavitelné parametry, které by řídily míru stylizace.

Stylizace v programu Praat zpracovává konturu z objektu PitchTier. Testuje jednotlivé body kontury tím, že body sousedící s testovaným bodem proloží přímkou a zjistí, o kolik se od ní testovaný bod odchyluje. Bod, pro nějž je absolutní odchylka nejmenší a zároveň nepřekročí stanovený práh rozlišení stylizace (výchozí hodnota je 4 ST), je odebrán a celá procedura se opakuje tolikrát, dokud odchylka bodu s nejnižší odchylkou není vyšší než zmíněný práh.

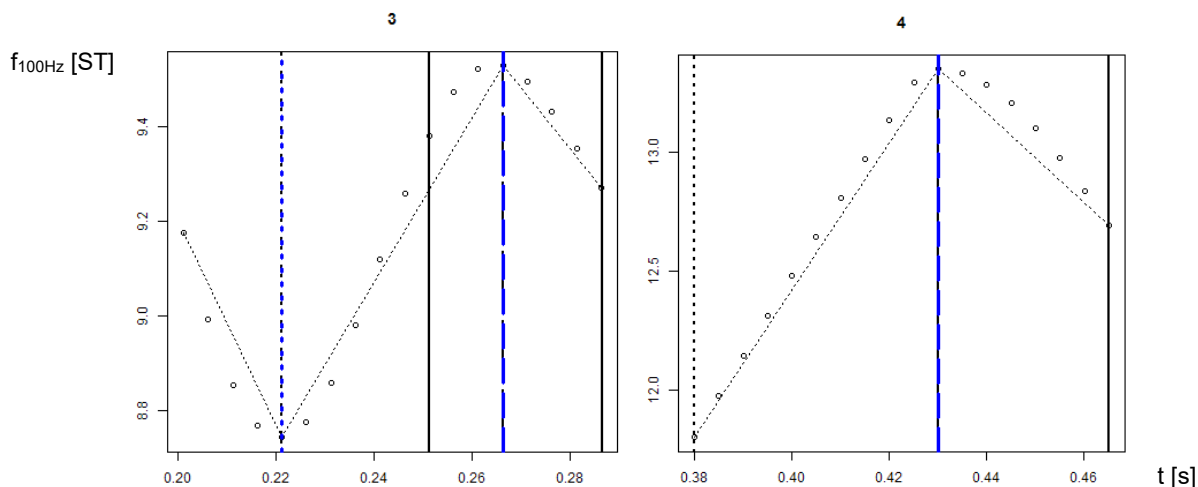
Testování algoritmu pomocí resyntézy dopadlo lépe než v případě MoMelu. Při snížení prahu na 1,5 ST byly rozdíly mezi stylizovanou a původní konturou takřka zanedbatelné. Bohužel i v tomto případě je stylizace pouze matematická, nezohledňuje segmentální složení a umístění bodů obratu je pravděpodobně lingvisticky irelevantní.

Rozhodli jsme se proto pro odlišný přístup: extrahovat z materiálu kontury slabičných jader a jejich bezprostředního okolí a detekovat body obratu v každé z nich zvlášť. K samotnému nukleu jsme přidali levý okraj o délce 50 ms, abychom tak mohli detekovat i ty body, které se nacházejí i těsně před začátkem nukleu. K automatizaci extrakce jsme vytvořili skript PTR+TXT-vokalKontury.praat. Z našeho vzorku jsme získali celkem 2079 dílčích kontur nukleů.

S ohledem na rozsah práce jsme se rozhodli identifikovat pouze body obratu, které určují extrémy kontury, tedy její vrcholy a sedla. V získaných dílčích konturách jsme pokaždé detekovali nejvýše jeden vrcholový bod a jeden sedlový bod, který se nachází nejbližší k levé hranici nukleu. V případech, kdy byl nalezený extrém zároveň prvním nebo posledním bodem kontury, nebyl brán v potaz a nukleus byl označen jako bez vrcholu, resp. bez sedla. V opačném případě algoritmus zaznamenal frekvenční a časovou souřadnici extrému a zahájil analýzu dalšího nukleu. Algoritmus byl implementován formou skriptu v jazyce R. Výsledek zpracování dvou nukleů i s naznačenou stylizovanou konturou je zobrazen na Obr. 3.

Dále jsme pracovali jen se souřadnicí časovou. Pro každý nukleus byl odečtením časové souřadnice jeho levé hranice od časové souřadnice extrému stanoven **posun**

extrému [s]. Podělením tohoto posunu trváním nukleu jsme získali **normalizovaný posun** extrému [-].

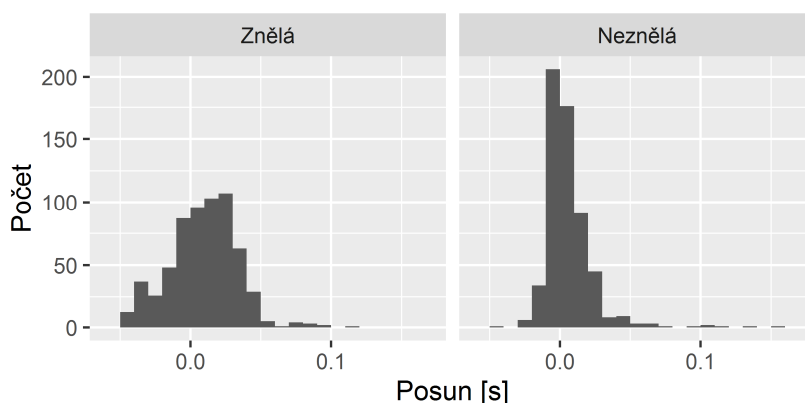


Obrázek 3: Detekce bodů obratu a stylizace průběhu F_0 v nukleu dvou různých slabik.

Prázdné tečky značí vzorky F_0 původní kontury, tečkovaná čára lomenou stylizovanou konturu. Časové umístění bodů obratu indikují modré čáry. Svislé plně černé čáry vymezují hranice nukleu.

5.2.2. POSUN VRCHOLŮ

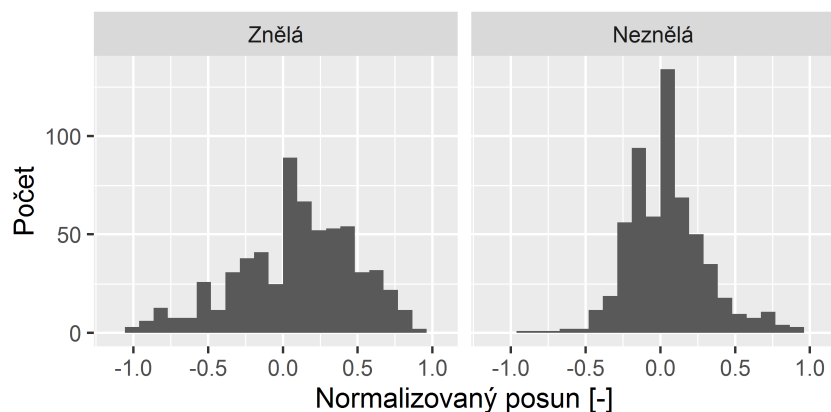
Z 2079 nukleů byly vrcholy nalezeny u 1229 z nich. Případy jsme rozdělili podle znělosti přetury a našli jsme celkem 638 položek se znělou přeturou a 591 s přeturou neznělou. V případě neznělé přetury lze totiž očekávat, že umístění vrcholu kontury bude určeno dobou nástupu hlasivkového tónu (*voice onset time*). Tento předpoklad podporuje i Graf 1.



Graf 1: Histogramy posunu vrcholu F_0 dle znělosti slabičné přetury.

Na pravém grafu je prominentní maximum kolem 0, takže maximum se v konturách nukleů se znělou přeturou často kryje s hranicí konsonant-vokál. Znělá přetura oproti

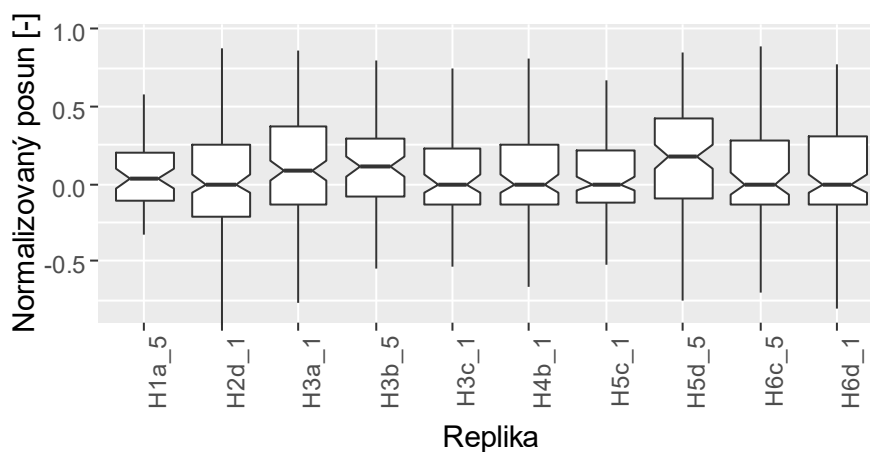
tomu vykazuje větší tendenci ke zpoždění. Abychom vyloučili vliv trvání nukleu, vytvořili jsme stejné histogramy i pro normalizovaný posun vrcholů (Graf 2).



Graf 2: Histogramy normalizovaného posunu vrcholu F_0 dle znělosti slabičné préturey.

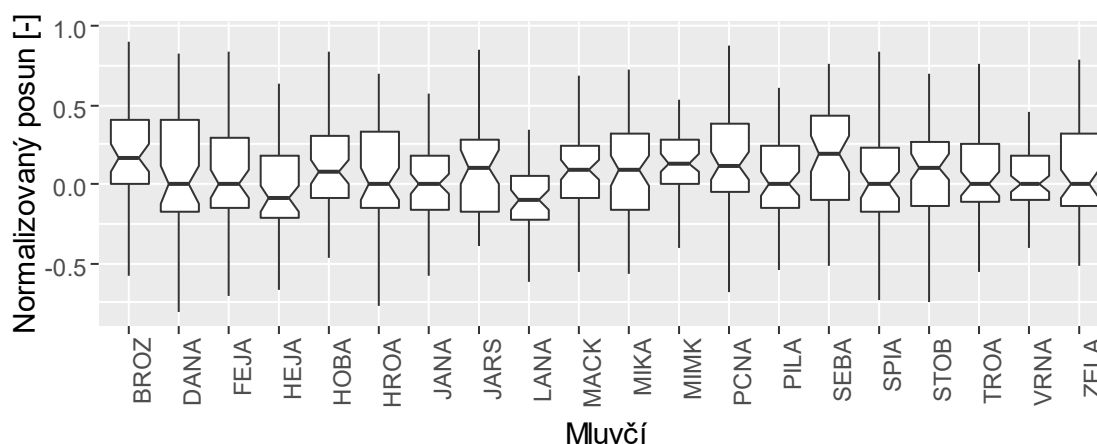
Na Grafu 2 se oproti předchozímu grafu zobrazila zřetelnější maxima rozdělení a to v obou kategoriích znělosti. Vzhledem k přidanému 50ms okraji nalevo od nukleu se u některých vokálů s velmi krátkým trváním objevily i hodnoty normalizovaného posunu nižší než -1. Ty jsme v Grafu 2 odfiltrovali. Můžeme si povšimnout tendence, že slabiky se znělou préтурой vykazují větší rozptyl hodnot, zatímco ty s neznělou se těsněji koncentrují podél hranice.

V následujících krabicových grafech se věnujeme již jen normalizovanému posunu vrcholů se znělou préтурой, protože odhadujeme, že neznělé případy jsou ovlivněny faktorem *voice onset time*. Všechny krabicové grafy v této práci jsou grafy Tukeyho typu: „krabice“ znázorňují mezikvartilové rozpětí (IQR, rozdíl mezi 25% a 75% percentilem). Konec horního vousu ukazuje hodnotu horního kvartilu přičtenou k 1,5násobku IQR, nebo maximum, pokud je menší než tato hodnota. Konec spodního vousu analogicky ukazuje minimum nebo hodnotu 1,5 IQR odečtenou od dolního kvartilu. Hodnota vyznačená uvnitř krabice reprezentuje medián vzorku. Hranice zkosené části uprostřed krabice indikují interval spolehlivosti (CI) na 5% hladině významnosti. Odlehlé hodnoty v krabicových (případně ani v houslových) grafech nezobrazujeme.



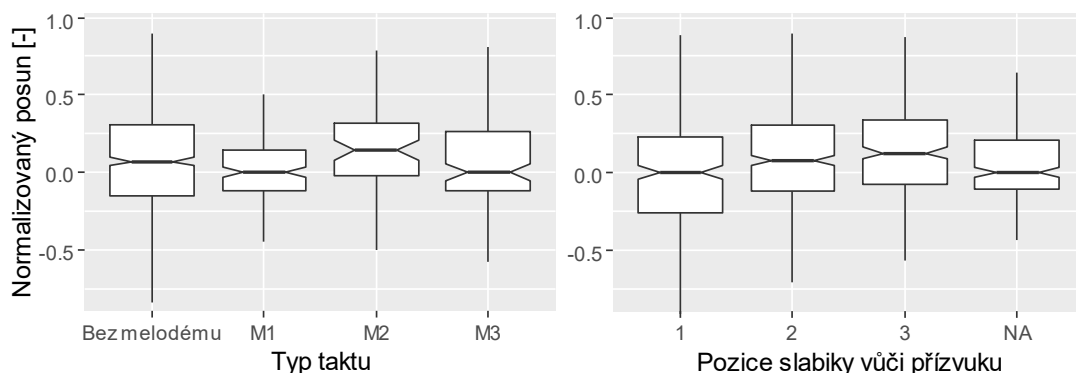
Graf 3: Norm. posun vrcholu F_0 dle repliky.

Graf 3 žádné viditelné tendence nezobrazuje, pouze si můžeme povšimnout, že některé repliky mají menší rozpětí hodnot než jiné. Replika H1a_5 s nejmenším rozpětím hodnot je zároveň replikou nejkratší (5 slabik), zatímco repliky H2d_1 a H3a_1 s velkým rozpětím jsou mnohoslabičná souvětí, tak můžeme předpokládat, že je výsledek ovlivněn délkou repliky.



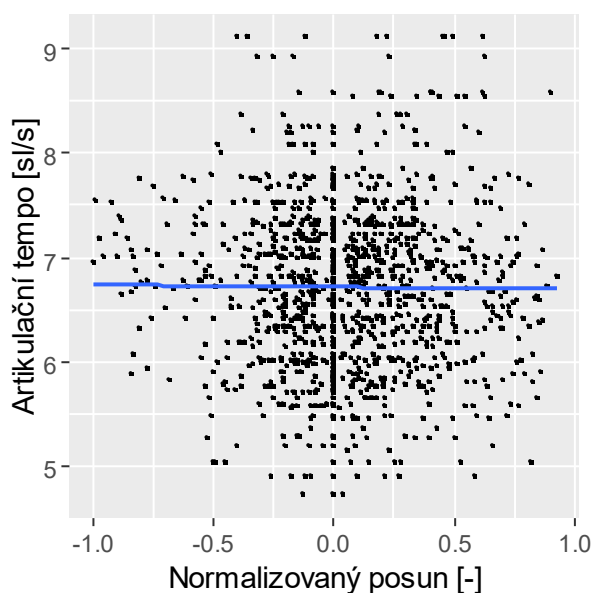
Graf 4: Norm. posun vrcholu F_0 dle mluvčích.

Graf 4 je oproti předchozímu podstatně variabilnější. Mluvčí se liší jak rozpětím hodnot, tak jejich střední hodnotou. Zatímco LANA vykazuje brzké umístění vrcholů, MIMK se s podobně úzkým IQR pohybuje spíše v kladných hodnotách a má tedy vrcholy převážně zpožděné.



Grafy 5 a 6: Norm. posun vrcholu F_0 dle typu taktu a dle pozice slabiky vůči přízvuku.

Na Grafu 5 vidíme, že slabiky taktu nesoucího melodém 1 mají vrcholy celkem konzistentně v blízkosti hranice konsonant-vokál. Melodém 2 nenápadně tíhne ke zpožděným vrcholům. Velké IQR taktů bez melodému bude pravděpodobně projevem jejich různorodosti (některé z nich mohou nést větný přízvuk, což jsme v klasifikaci nezohlednili). Graf 6 žádné nápadné tendence nezobrazuje.



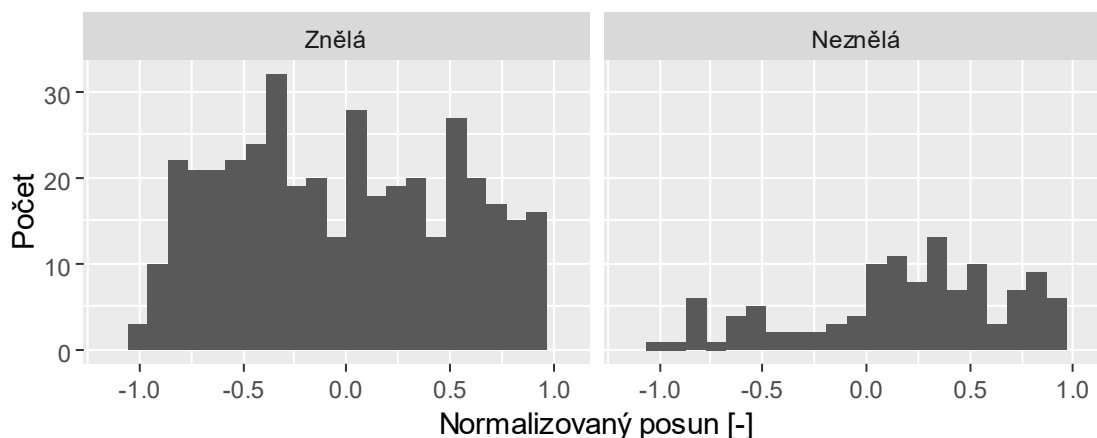
Graf7: Korelace norm. posunu vrcholu F_0 s artikulačním tempem promluvy.

Grafem 7 jsme zkusili ověřit vztah mezi normalizovaným posunem a artikulačním tempem, žádná korelace zde ovšem vidět není.

5.2.3. POSUN SEDEL

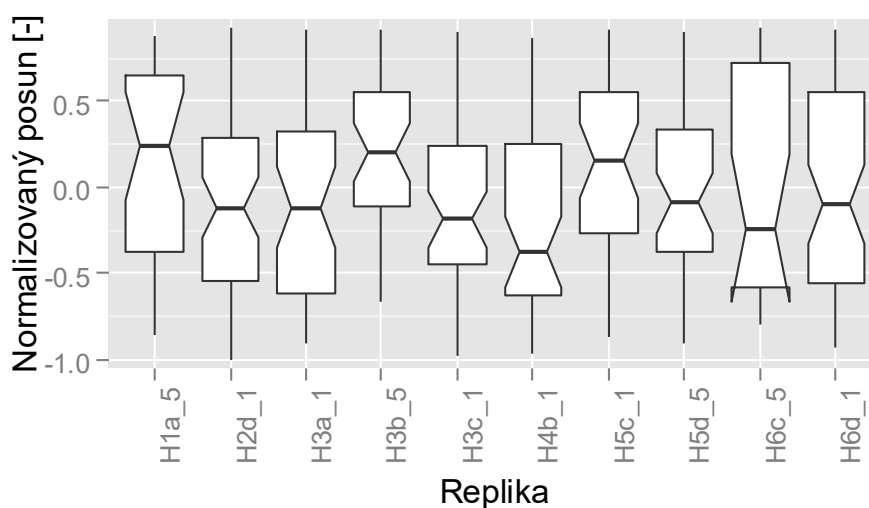
Podobnou analýzu jako pro vrcholy jsme provedli i pro umístění sedel. Ze vzorku jsme získali celkem 524 nukleů se sedlem, z toho 409 znělých, 115 neznělých. Jedná se

o výrazně menší vzorek, což může být jedním z důvodů větších rozptylů hodnot v následujících grafech. Zabývali jsme se již jen normalizovaným posunem.



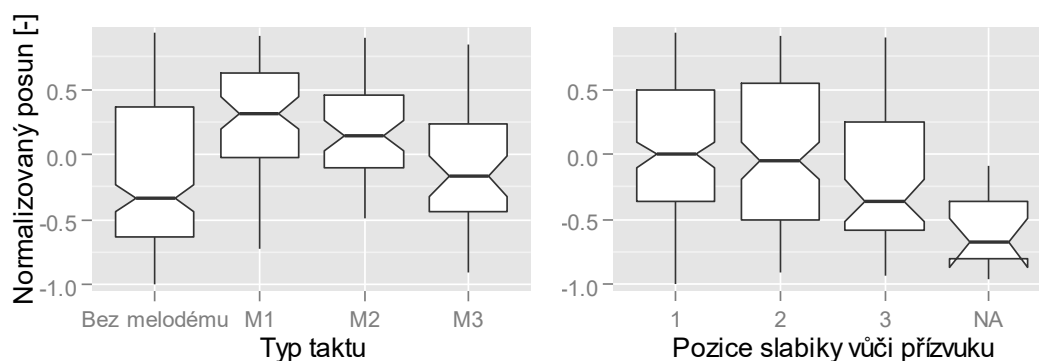
Graf 8: Histogramy normalizovaného posunu sedla F_0 dle znělosti slabičné přetury.

Graf 8 se výrazně liší od Grafu 2. Posun sedel na rozdíl od posunu vrcholů nevykazuje zřetelné maximum, ale spíše rovnoměrně rozdělený šum. V případě neznělých přetur se více případů logicky pohybuje v kladných hodnotách. Dále se opět zabýváme jen znělými případy.



Graf 9: Normalizovaný posun sedla F_0 dle repliky.

Podle Grafu 9 je posun sedel variabilnější než posun vrcholů, o čemž svědčí zvláště velikost IQR a CI. Například IQR repliky H6c_5 zabírá téměř tři čtvrtiny možného obsahu hodnot. Posun sedel dle mluvčích byl podobně rozptýlený, graf zde nezobrazujeme.



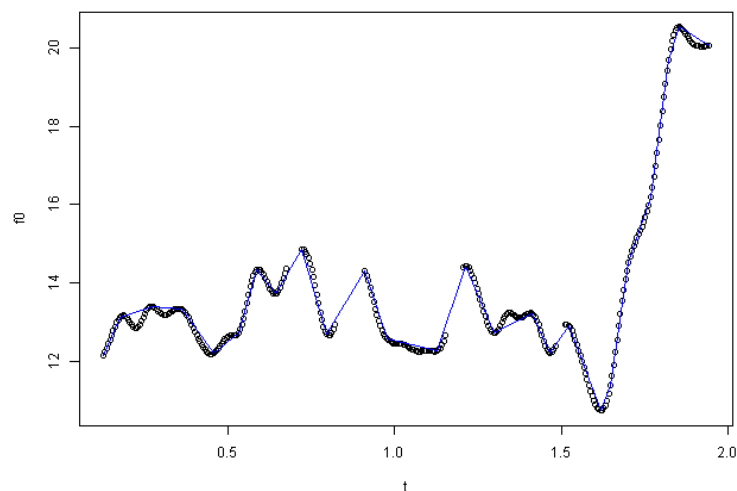
Graf 10 a 11: Normalizovaný posun sedla F_0 dle typu taktu a dle pozice slabiky vůči přízvuku.

V Grafu 10 vidíme, že téměř polovina případů pro takty s M1 se vyskytuje v kladných hodnotách. Z velikosti rozptylů v sadě Grafů 8–11 můžeme obecně soudit, že by bylo vhodné sedlové body definovat úžeji, například vyloučit případy, kde se ve vokálu vyskytuje vrchol i sedlo zároveň, a experiment zopakovat.

5.2.4. STYLIZACE KONTURY F_0 CELÉ REPLIKY

Na základě algoritmu k hledání extrémů kontury nukleů navrženého výše jsme navrhli algoritmus pro stylizaci kontury celé repliky. Jeho princip je následující: nejprve získá hodnoty prvního a posledního bodu původní kontury a jejich globálně extrémních bodů. Ty lineárně propojí a hledá časy, v nichž se stylizovaná kontura nejvíce liší od originální. Bod, v němž absolutní hodnota odchylky dosáhne maxima, a zároveň překročí stanovený práh, přidá do stylizované kontury, body obratu lineárně proloží a celou proceduru opakuje tolikrát, dokud již nenajde žádný bod s odchylkou vyšší prahu. Příklad stylizované kontury uvádíme na Obr. 4.

Algoritmus jsme pak otestovali na základě resyntézy a percepční rovnosti bylo dosaženo při prahu nastaveném na 0,5 ST. Umístění výsledných bodů obratu však bylo podobně lingvisticky irelevantní jako v případě praatovské stylizace. Navíc námi definovaný práh se ukázal jako velmi neobratný parametr pro určování míry stylizace, protože při jeho manipulaci docházelo ve výsledné kontuře ke skokovým, nikoliv ke graduálním změnám. Pokud bychom se rozhodli algoritmus dále rozvíjet, určitě by bylo vhodné zapracovat do něj také zohlednění segmentálního složení kontury (a využít například dostupné TextGridy se segmentací).



Obrázek 4: Stylizace realizace repliky „Máte nějakou představu, jak to bude probíhat?“ s prahem 0,5 ST.

5.3. STRMOST KONTURY

Protože kromě přítomnosti výrazných vrcholů a sedel kontury F_0 člověk vnímá i její postupné změny, zaměřili jsme naše zkoumání i na směr a strmost kontury. K tomu jsme zvolili analýzu kontury stylizované v Prosogramu.

5.3.1. METODA

Prosogram si ve srovnání s algoritmy praatovské stylizace a MoMelu vedl výrazně lépe. Ač v módu plně automatické stylizace (tedy včetně vlastní detekce slabičných nukleů) vykazuje jisté chyby jako oktávové skoky a nespolehlivé umístování nukleů, v módu využití externí segmentace jsou jeho výsledky velmi uspokojivé. Při poslechovém porovnání několika replik resyntetizovaných s původní a se stylizovanou konturou jsme nepostřehli rozdíly.

Pro stylizaci v Prosogramu jsme nejprve v TextGridech replik pomocí skriptu TXT-sonzsegm.praat vytvořili vrstvu „segm“, kterou se Prosogram řídí při detekci slabičných nukleů. Prosogram jsme spustili s následujícím nastavením: perioda vzorkování 5 ms, rozsah detekce F_0 60–450 Hz, segmentace na základě vrstvy „segm“, adaptivní práh strmosti mezi $G = \frac{0,16}{T^2}$ a $G = \frac{0,32}{T^2}$, práh difference strmosti $DG = 20$, práh trvání $d_{min} = 50$ ms. V získaných konturách jsme pak detekovali dynamické tóny a měřili melodické skoky (frekvenční rozdíly) mezi sousedními slabikami.

Detekci dynamických tónů (tedy průběhů F_0 v nukleech, které překročí stanovený práh strmosti, viz 4.1.7) jsme provedli skriptem PTR+TXT-dynamicke.praat.¹⁴ K analýze melodických skoků jsme vytvořili skript PTR+TXT-skokyProso.praat.

5.3.2. DYNAMICKÉ TÓNY

V našem vzorku bylo dynamických tónů nalezeno poskrovnu. Z celého vzorku 2079 nukleů jich 74 (4 %) neslo tón klesavý, 37 (2 %) tón stoupavý a zbylých 1968 položek (95 %) neslo tón statický. Pouze ve 3 případech šlo o tón složený.

	Klesavý	Statický	Stoupavý	Celkem
H1a_5	9	90	1	100
H2d_1	3	274	8	285
H3a_1	8	251	1	260
H3b_5	0	145	15	160
H3c_1	9	231	0	240
H4b_1	11	228	1	240
H5c_1	9	167	0	176
H5d_5	2	160	9	171
H6c_5	8	191	1	200
H6d_1	15	231	1	247
Celkem	74	1968	37	2079

Tabulka 2: Počty dynamických tónů v závislosti na replikách.

V Tabulce 2 můžeme pozorovat podobnou tendenci jako v Tabulce 3, totiž že otázky (H2d_1, H3b_5 a H5d_5) obsahují nadprůměrný počet stoupavých tónů, zatímco v prostých sděleních se objevují spíše klesavé tóny.

	Klesavý	Statický	Stoupavý	Celkem
M1	58	329	3	390
M2	1	105	29	135
M3	6	208	1	215
Bez melodému	9	1326	4	1339
Celkem	74	1968	37	2079

Tabulka 3: Počty dynamických tónů v závislosti na typu taktu.

¹⁴ Vzhledem k tomu, že Prosogram někdy detekuje velmi krátké nebo žádné nukleky (nejspíš vlivem desonorizace, třepené fonace či jiných anomálií), není náš skript stoprocentně spolehlivý a detekuje nepatrně větší počet dynamických tónů než je počet skutečný. Složené dynamické tóny skript zpracovává stejně jako tóny jednoduché, neboť v celém vzorku se jich objevilo zanedbatelné množství.

Tabulka 3 ukazuje, že 15 % všech tónů v melodému 1 je klesavých a 21 % tónů v melodému 2 je stoupavých. Lze tedy říct, že dynamické tóny zde obvykle reflektují běžný tvar kadence, kterou je daný melodém realizován. Přízvukové takty bez melodému obsahují z 99 % tóny statické.

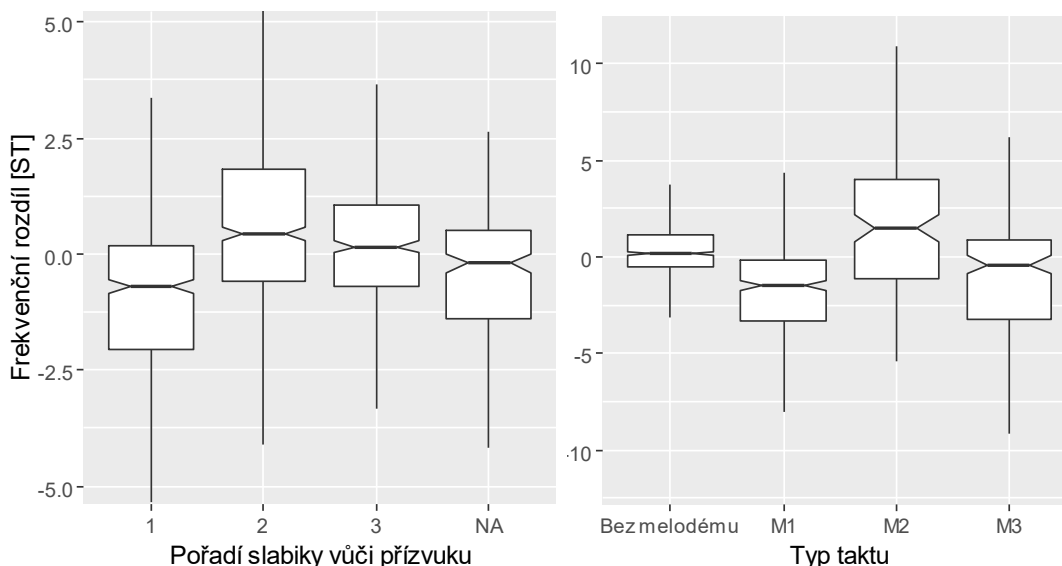
	Klesavý	Statický	Stoupavý	Celkem
Příz. slabiky	5	722	4	731
2. slabiky	38	645	27	710
3. slabiky	13	364	4	381
Nekl. slabiky	18	237	2	257
Celkem	74	1968	37	2079

Tabulka 4: Počty dynamických tónů v závislosti na pořadí slabiky vůči přízvuku.

V Tabulce 4 vidíme, že největší podíl (9 %) mají dynamické tóny na slabice, které následují po slabice přízvučné. Naopak v kategorii přízvučných slabik tvoří pouze 1 %. V případě neklasifikovaných slabik tvoří dynamické tóny takřka 8 %. Možná by bylo vhodné v dalším zkoumání slabiky označit také z hlediska jejich pozice vůči následujícímu přízvuku.

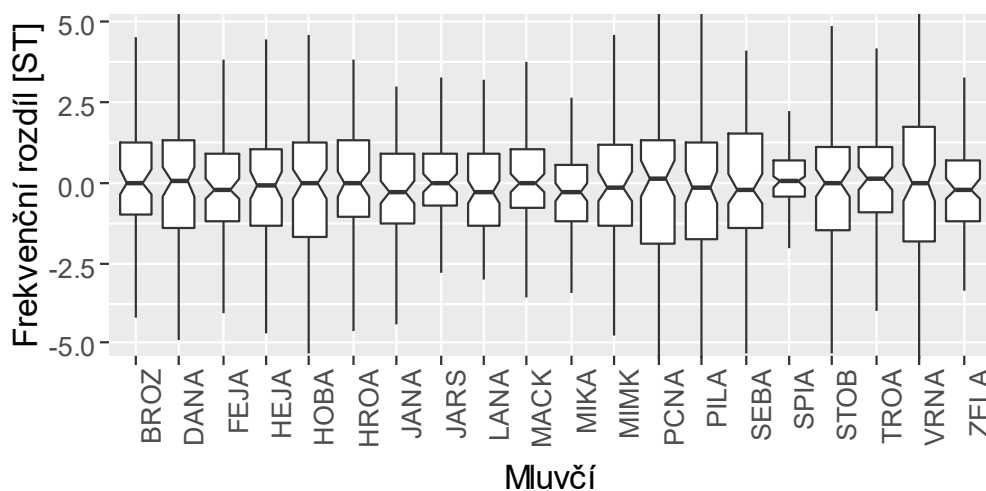
5.3.3. MELODICKÉ SKOKY

Z kontur stylizovaných v Prosogramu jsme skriptem PTR+TXT-skokyProso.praat získali frekvenci tónu na začátku každého slabičného jádra ve vzorku. Ta byla následně porovnána s frekvencí tónu na konci jádra předchozího a hodnota tohoto frekvenčního rozdílu byla zaznamenána. Pokud jde o skok směrem nahoru, rozdíl je kladný, pokud jde o skok dolů, rozdíl je záporný. První slabika každé promluvy byla ze vzorku vyřazena, neboť jí nepředchází slabika jiná, s níž by mohla být porovnána. Tím byl vzorek zmenšen na 1886 položek.



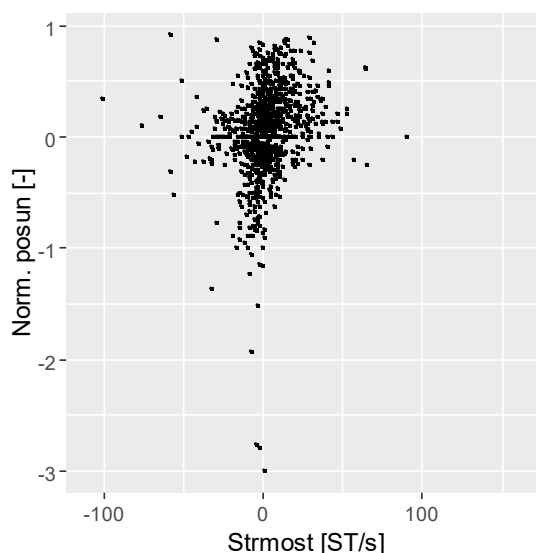
Grafy 12 a 13: Melodické skoky v závislosti na pozici slabiky vůči přízvuku a na typu taktu.

Graf 12 je v souladu s běžným průběhem melodie přízvukového taktu (Palková & Volín, 2003). Patrné je zde především to, že před přízvuknou slabikou bývá pokles a po ní melodické stoupnutí. V Grafu 13 je nápadné především malé rozpětí hodnot (IQR cca 2 ST) u taktů bez melodému ve srovnání s takty s melodémem. Největší IQR (cca 5 ST) mají takty s M2. Více než polovina hodnot naměřených v taktech s M1 je záporných, což odpovídá jeho klesavým kadencím.



Graf14: Melodické skoky v závislosti na mluvčím.

V Grafu 14 je opět nápadné rozpětí hodnot, které některé mluvčí charakterizuje. Například SPIA využívá v 50 % případů skoky menší než 1 ST. Zároveň hodnoty u každého mluvčího jsou vystředěny podél 0, případně nepatrně pod ní, což pravděpodobně svědčí o tom, že využívají skoky nahoru i dolů přibližně stejnou měrou.



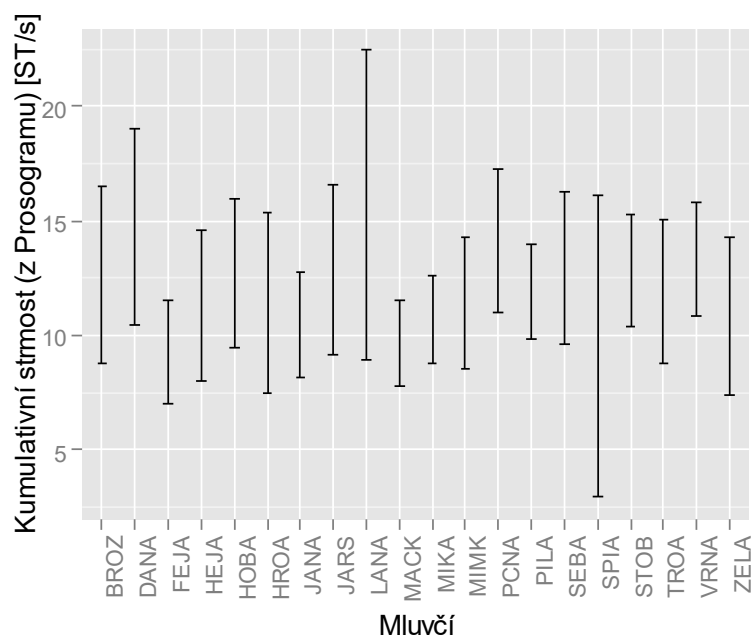
Graf 15: Závislost mezislabičné strmosti a norm. posunu vrcholu F_0 .

Grafem 15 jsme se pokusili zachytit závislost mezislabičného skoku a normalizovaného posunu vrcholu F_0 . Skoky jsme zde podělili jejich trváním (analogicky k normalizaci, kterou jsme prováděli v případě posunu vrcholů) a získali jsme tím parametr mezislabičné strmosti [ST/s]. Oba parametry spolu zjevně do jisté míry korelují.

5.4. VARIABILITA PROMLUVY

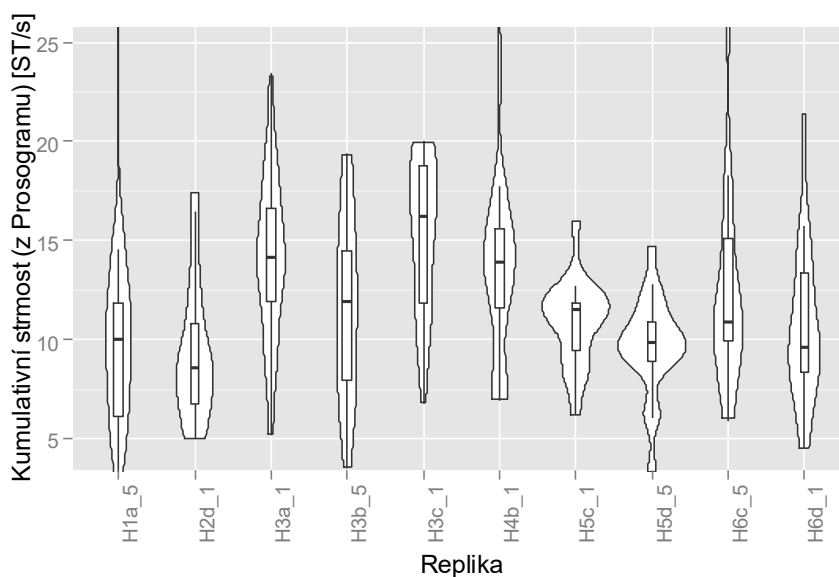
Četnost a míra intonačních událostí v čase – ať už mluvíme o bodech obratu, melodických skocích nebo nápadných plynulých změnách – by mohla souviset s tím, zda řeč vnímáme jako monotónní nebo variabilní. Předpokládáme, že řeč s vyšším počtem melodických změn, které využívají větší frekvenční rozsah, bude obecně vnímána jako variabilnější a živější. Naopak řeč s málem melodických změn, které navíc nejsou příliš melodicky prominentní, budeme spíše hodnotit jako monotónní.

Teoreticky nejjednodušší monotónní konturu můžeme modelovat jednou pomalu deklinující úsečkou. Čím živější řeč bude, tím více změn a výškových skoků se v kontuře objeví. Na základě toho jsme navrhli parametr tzv. **kumulativní strmosti kontury**. Jde vlastně o součet absolutních hodnot všech výškových skoků v kontuře podělený jejím celkovým trváním. Tento parametr jsme spočítali pro všech 193 kontur stylizovaných Prosogramem pomocí skriptu PTR+TXT-prumStrmostProso.praat.



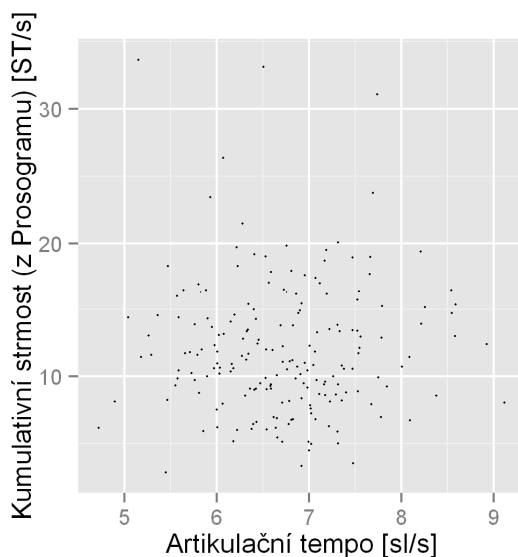
Graf 16: Průměrná kumulativní strmost kontury řeči jednotlivých mluvčích.
(hladina významnosti intervalu spolehlivosti $\alpha=0,05$)

Pokud byl náš předpoklad správný, pak by vyšší kumulativní strmost měla znamenat variabilnější řeč s většími a četnějšími melodickými skoky. Graf 16 ukazuje, že se střední hodnoty kumulativní strmosti většiny mluvčích pohybují v rozmezí 10–15 ST/s. Smysluplnost a výpovědní hodnotu kumulativní strmosti by bylo třeba ověřit poslechem.



Graf 17: Kumulativní strmost kontury jednotlivých replik.

Graf 17 zobrazuje poměrně velký rozptyl v realizacích jednotlivých replik. Viditelnou výjimkou jsou repliky H5c_1 a H5d_5. Na základě toho by se dalo soudit, že některé repliky jsou mezi mluvčími realizovány s podobnou variabilitou, zatímco jiné si každý realizuje po svém. Pravděpodobnější však je, že kumulativní strmost není dobrý parametr pro vyhodnocování variability kontury základní frekvence.



Graf 18: Vzájemná závislost kumulativní strmosti a artikulačního tempa.

Grafem 18 jsme se pokusili zjistit, zda není námi navržený parametr korelátem artikulačního tempa, ale zobrazený výsledek tomu nijak nenasvědčuje.

6. Závěr

V bakalářské práci jsme se seznámili s obecnými principy a funkcemi prozodie a intonace a uvedli si některé z přístupů k jejich pojetí. Definovali jsme si konturu základní frekvence (F_0) jako základní akustický korelát intonace a popsali jsme si některé její obecné vlastnosti spojené s její produkcí a percepcí. Podívali jsme se na intonační a prozodické aspekty češtiny a uvedli jsme si, jak se intonační průběh výpovědi řídí jejím typem a umístěním přízvuků.

Dále jsme si přiblížili proces zpracování kontury F_0 od akustického signálu až po fonologickou vrstvu. Popsali jsme si proces stylizace v pojetí tří různých algoritmů a způsob zpracování stylizované kontury v podání čtyř různých fonetických modelů. Fonetické intonační modely jsme si rozdělili do kategorií podle toho, jak zpracovávají konturu základní frekvence a které její prvky jsou pro ně klíčové. V závěru teoretické části jsme se podívali na samotné aspekty popisu intonační kontury a na její nejmenší jednotky ve vztahu k segmentální vrstvě řeči.

V rámci praktické části jsme se podíleli na segmentaci nahrávek, které jsou součástí korpusu *Minidialogy* vyvíjenému na Fonetickém ústavu FF UK. Poté jsme na vybraném vzorku z korpusu provedli výzkum posunu bodů obratu (resp. lokálních extrémů kontury F_0) vůči hranici konsonant-vokál a pokusili jsme se navrhnout jednoduchou stylizaci právě na principu detekce extrémů. V dalším bodě jsme využili stylizaci kontury programem *Prosogram* a analyzovali jsme v našem vzorku dynamické tóny slabičných nukleů a melodické skoky mezi slabikami. Na závěr jsme navrhli parametr kumulativní strmost kontury pro posuzování variability intonace a zkusili jím popsat náš vzorek.

6.1.1. POZNATKY Z PRAKTICKÉ ČÁSTI

Ve zkoumání bodů obrátů jsme zjistili, že pokud se ve slabičném jádře vyskytuje melodické maximum, bývá umístěno na hranici konsonant-vokál. Nenalezli jsme však žádné tendence, které by napovídaly tomu, že je posun vrcholu vůči hranici využíván

distinktivně. Pouze v případě slabik se znělou préturou je mírná tendence ke zpoždění vrcholu (ve srovnání s případy s préturou neznělou). Oba tyto pozorované jevy budou motivované spíše mikrintonačně než jazykově. Tomu nasvědčuje i nalezený náznak korelace posunu vrcholového bodu se strmostí kontury, která mu bezprostředně předchází. Posun vrcholu může tedy být pouhým důsledkem velikosti melodického skoku. Určitě by bylo vhodné provést podobnou analýzu znovu, tentokrát i se zohledněním melodické (výškové) prominence vrcholu.

Analýza posunů sedlových bodů vykazala velké rozptyly dat (například i ve srovnání s posuny vrcholů), proto nebylo možné žádnou tendenci ve vztahu k segmentální vrstvě nalézt. Experiment by bylo možné zopakovat s úžeji definovaným sedlem (vymezit ho jeho frekvenčními a časovými rozměry, stanovit přesnější práh, co všechno chápeme jako sedlo, to nejen na základě „geometrie“ kontury, ale zohlednit také zákonitosti percepce zvuku a řeči).

Při zkoumání nahrávek s využitím stylizace dle tonálně percepčního modelu jsme objevili, že průběhy F_0 uvnitř vokálů nejsou příliš variabilní, neboť jen malá část z nich překročí i relativně nízký práh strmosti, aby byly algoritmem Prosogram identifikovány jako dynamické tóny. V případě, že se v replice dynamický tón vyskytne, obvykle to bývá v té části, jejíž průběh je určen melodémem, a jeho směr se řídí směrem kadence, tj. že klesavé tóny se objevují v klesavých kadencích a stoupavé tóny v kadencích stoupavých.

Výsledky analýzy melodických skoků, tedy výškových melodických rozdílů mezi sousedními slabikami, potvrzují dřívější poznatky o průběhu větné melodie v češtině. Přízvukové takty bez melodému obsahují převážně menší skoky než takty s melodémem. Takty s melodémem M1 a M3 obsahují skoky spíše sestupné, takty s M2 oproti nim častěji skoky stoupavé.

Námi navrhovaná kumulativní strmost se neprokázala jako efektivní parametr pro posuzování variability promluv, neboť ve většině zkoumaných skupin se její hodnoty pohybovaly v podobném rozpětí. Přesto však by bylo vhodné parametr před úplným zamítnutím posoudit na základě poslechu replik.

7. Zdroje

- BAGSHAW, P. C. 1993. An investigation of acoustic events related to sentential stress and pitch accents. In *Speech Communication*, 1993, vol. 13, 333–342.
- BARNES, J., VEILLEUX, N., BRUGOS, A. & SHATTUCK-HUFNAGEL, S. 2010. The effect of global F_0 contour shape on the perception of tonal timing contrasts in American English intonation. In *Speech Prosody*, 2010, 100445: 1-4.
- BOERSMA, P. & WEENINK, D. 2015. *Praat: doing phonetics by computer*. [Počítačový program]. Verze 5.3.51, staženo 25. 11. 2015 z <<http://www.praat.org>>
- BRAUN, B., KOCHANSKI, G., GRABE, E. & ROSNER, B. S. 2006. Evidence for attractors in English intonation. In *Journal of the Acoustical Society of America*, 2006, 119(6): 4006-15.
- DILLEY, L. C. & BROWN, M. 2007. Effects of pitch range variation on F_0 extrema in an imitation task. In *Journal of Phonetics*, 35/ 4: 523-551.
- DUBĚDA, T. 2014. Czech intonation: A tonal approach. *Slovo a slovesnost*, vydání č. 75, s. 83–98.
- FUJISAKI, H. & OHNO, S. 1997. Comparison and assessment of models in the study of fundamental frequency contours of speech. In: INT – 1997, 131-134.
- GERHARD, D. 2003. *Pitch Extraction and Fundamental Frequency: History and Current Techniques*. [PDF, online].
Dostupné z: <<http://www.cs.uregina.ca/Research/Techreports/2003-06.pdf>>
- GRABE, E., KOCHANSKI, G. & COLEMAN, J. 2007. Connecting intonation labels to mathematical descriptions of fundamental frequency.
In *Language and Speech*, 50 (3), 281–310.
- HÁLA, B. 1962. *Uvedení do fonetiky češtiny na obecně fonetickém základě*.
1. vydání. Praha: Nakladatelství Československé akademie věd.
- HERMES, D. J. 2006. Stylization of Pitch Contours. In W. de Gruyter (ed.) *Methods in Empirical Prosody Research*. ISBN 978-311018856-1.
- HIRSCHBERG, J. & BECKMAN, M. 1994. *The ToBI Annotation Conventions*.
Ohio State University, 1994. Dostupné z <www.cs.columbia.edu/~julia/files/conv.pdf>
- HIRST, D. & ESPESER, R. 1993. Automatic Modelling Of Fundamental Frequency Using A Quadratic Spline Function. In *Travaux de l'Institut de Phonétique d'Aix*, Vol. 15, 75–85.
- HIRST, D. & DI CRISTO, A. 1998. *Intonation Systems. A survey of Twenty Languages*.
Cambridge: Cambridge University Press. ISBN 978-052139550-2.

- LADD, R. 2008. *Intonational Phonology*. Second Edition. Cambridge: Cambridge University Press. ISBN 978-052167836-0.
- DE LOOZE, C. & HIRST, D. 2010. The OMe (Octave-Median) scale: a natural scale for speech melody. In *Proceedings of XXVIIIème Journées d'Etude sur la Parole (JEP 2010)*, Mons, Belgium, 4 pages.
- MACHAČ, P. & SKARNITZL, R. 2009. *Fonetická segmentace hlásek*. Praha: Nakladatelství Epoque.
- MERTENS, P. 2004. The Prosogram: Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. In B. Bel & I. Marlien (eds.) *Proceedings of Speech Prosody*, 2004, Nara, Japan.
- MIXDORFF, H. 2000. A novel approach to the fully automatic extraction of Fujisaki model parameters. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000*.
- NIEBUHR, OLIVER. 2014. *Kiel Intonation Model*. [Přednáška]. Praha: FF UK, 7. října 2014.
- PALCOVÁ, Z. 1994. *Fonetika a fonologie češtiny*. 1. vydání. Praha: Univerzita Karlova v Praze. 366 s. ISBN 80-7066-843-1.
- PALCOVÁ, Z. 1997. Modelling intonation in Czech: neutral vs. marked TTS F_0 -patterns. In *Intonation: Theory, Models, and Applications*. Athens, Greece. 1998.
- PALCOVÁ, Z. & VOLÍN, J. 2003. The role of F_0 contours in determining foot boundaries in Czech. In *Proceedings of the 15th ICPhS*, Vol. II, pp. 1783–1786. Barcelona: Organizing Committee.
- POLLÁK, P., VOLÍN, J. & SKARNITZL, R. 2007. HMM-Based Phonetic Segmentation in Praat Environment. In *Proceedings of the XIIth International Conference "Speech and computer – SPECOM 2007"*, 537–541, Moscow.
- R CORE TEAM. 2015. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Dostupné z: <<https://www.R-project.org/>>.
- ROSSI, P. S, PALMIERI, F. & CUTUGNO, F. 2002. Method for Automatic Extraction of Fujisaki-Model Parameters. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, Volume 1, 509–512.
- TAYLOR, P. 1994. The Rise/Fall/Connection Model of Intonation. In *Speech Communication*, 10/1994; 15(1-2):169-186.

TAYLOR, P. 1999. Analysis and synthesis of intonation using the Tilt model.
In *Journal of the Acoustical Society of America*, Vol. 107, No. 3, March 2000.

VOLÍN, J. & BOŘIL, T. 2014. General and speaker-specific properties of F₀ contours in short utterances. In *AUC Philologica 1/2014*, *Phonetica Pragensia XIII*, pp. 101-112.

8. Přílohy

Vzhledem k rozsahu tabulek s naměřenými a zpracovávanými daty je všechny příkládáme pouze v elektronické formě.

8.1. OBSAH ELEKTRONICKÉ PŘÍLOHY

8.1.1. NAMĚŘENÁ DATA

Naměřená data využitá v této práci i data z nich dopočítaná jsou zkompletována a seříděna v tabulkovém souboru KONTURYdata.xlsx. Soubory s daty použitými k vykreslení grafů se nachází spolu se skripty, které je zpracovávají, ve složce R_grafy.

8.1.2. SKRIPTY K PROGRAMU PRAAT

Skripty vytvořené v rámci této práce jsou dostupné v příloze v podsložce PRAAT_skripty. Názvy souborů jsou tvořeny prefixem a samotným názvem. Prefix napovídá, jaký typ objektů daný skript zpracovává: DPT je diferenciální PitchTier, PITCH je objekt Pitch, PTR je běžný PitchTier, SND je zvuk a TXT značí TextGrid. Všechny skripty slouží k hromadnému zpracování objektů.

Zde je abecední seznam skriptů dle názvu souborů a stručný popis jejich účelu:

- 1) **DPT+TXT-prumStrmost.praat** počítá absolutní strmost kontury celé nahrávky, a to z absolutní diferenciální melodické vrstvy.
- 2) **DPT+TXT-strmostNastupu.praat** počítá strmost kontury na hranici souhlásky a vokálu z diferenciální melodické vrstvy.
- 3) **DPT+TXT-taktyAbsStrm.praat** vypíše průměrné absolutní strmosti všech přízvukových taktů.
- 4) **DPT+TXT-vokalyPrumStrm.praat** vypíše průměrnou strmost vokálů z dif. melodické vrstvy.
- 5) **DPT-2AbsDifPitchTier.praat** vytvoří z DPT absolutní diferenciální vrstvu.
- 6) **DPT-2PitchTier.praat** převede DPT zpět na obyčejný PitchTier.

- 7) **PITCH+TXT-unvoice.praat** převede vybrané intervaly objektu Pitch na neznělé.
- 8) **PTR+TXT-dynamicke.praat** vypíše strmosti dynamických tónů z kontury stylizované Prosogramem.
- 9) **PTR+TXT-prumStrmostProso.praat** spočítá kumulativní strmost kontury.
- 10) **PTR+TXT-skokyProso.praat** vypíše velikosti a strmosti mezislabičných skoků.
- 11) **PTR+TXT-vokalKontury.praat** navzorkuje a vypíše průběhy F_0 všech vokálů a uloží je jako tabulku.
- 12) **PTR-2DifPitchTier.praat** vytvoří diferenciální PitchTier.
- 13) **PTR-resample.praat** převzorkuje libovolný PitchTier.
- 14) **PTR-ulozitTabulku.praat** uloží PitchTier jako tabulku.
- 15) **TXT-artTempo.praat** spočítá artikulační tempo z TextGridu.
- 16) **TXT-son2segm.praat** vytvoří z vrstvy sonority vrstvu „segm“.
- 17) **TXT-sonorita.praat** vytvoří vrstvu sonority z vrstvy hlásek.
- 18) **TXT-znelostPretury.praat** vypíše znělost prétury každé slabiky.

8.1.3. SKRIPTY V JAZYCE R

V jazyce R jsme vytvořili tři sady skriptů:

- 1) **Skripty pro detekci extrémních bodů obratu** v konturách jednotlivých vokálů se nachází ve složce R_extremy.
- 2) **Skript pro stylizaci PitchTierů** se nachází ve složce R_stylizace.
- 3) **Skripty pro vykreslení grafů** se nachází ve složce R_grafy.

8.2. DODATEK: VRSTVA SONORITY

Většina skriptů, které jsme během této práce vytvořili, pracuje s tzv. vrstvou sonority. Přínejmenším kvůli této závislosti bychom měli osvětlit její podstatu. Vrstva sonority je námi navržená vrstva TextGridu zarovnaná s vrstvou hláskovou, ovšem místo fonetického přepisu jednotlivých segmentů obsahuje čísla odkazující na jejich *třídou sonority*. Sonoritu chápeme jako schopnost hlásky tvořit jádro slabiky. Třída sonority

seskupuje hlásky s podobnou sonoritou. Nejvyšší třída má číslo 0 a zahrnuje vokály, nejnižší třída má v češtině číslo 7 a zahrnuje neznělé plozivy. Motivací pro vytvoření vrstvy sonority byla jednak jednoduchá fonetická klasifikace hlásek (jednotlivé třídy zahrnují hlásky se stejným či blízkým způsobem artikulace a stejnou třídou znělosti) a jednak využití pro detekci nukleů, slabičných hranic a pobočných slabik. Detekci pobočných slabik jsme původně plánovali v rámci práce využít, ale pro náš vzorek dat to nebylo třeba. Nakonec jsme v této práci vrstvu sonority použili pouze k rozlišení konsonantů od vokálů, tedy jako ekvivalent vrstvy CV užívané v korpusu.

K vytvoření vrstvy sonority z vrstvy hlásek jsme navrhli skript `TXT-sonzsegm.praat`.

8.3. DODATEK: DIFERENCIÁLNÍ MELODICKÁ KONTURA

V rámci analýzy strmosti kontury F_0 jsme navrhli koncept diferenciální melodické kontury, ale vzhledem k rozsahu práce jsme jej nakonec nevyužili. Přesto je zde uvádíme, protože věříme, že by mohl být využit například pro účely stylizace kontury.

Podstatou diferenciální melodické vrstvy (*differential PitchTier*, DPT) je interpretace strmosti kontury jako změna základní frekvence v čase. Křivka diferenční melodické kontury je tak první časovou derivací křivky melodické kontury. Skript `PTR-2DifPitchTier.praat`, který jsme pro výpočet DPT navrhli, vzorkuje originální `PitchTier` se zadaným vzorkovacím krokem (standardně 10 ms) a do nové vrstvy vkládá rozdíly mezi každými dvěma sousedními vzorky, dokud nedospěje ke konci nahrávky. Důležité je podotknout, že zatímco obyčejný `PitchTier` zaznamenává hodnoty v jednotkách Hz, náš diferenciální `PitchTier` používá jednotky ST/s, neboť logaritmická stupnice je pro popis změn názornější. K převodu obyčejného `PitchTieru` na diferenciální jsme sestavili skript `PTR-2DifPitchTier.praat`.

K získání absolutní hodnoty z diferenciální melodické kontury (kterou by bylo možné využít k alternativnímu výpočtu kumulativní strmosti) jsme připravili skript `DPT-2AbsDifPitchTier.praat`, který převrátí všechny její záporné hodnoty na kladné.