

Univerzita Karlova, Filozofická fakulta

Fonetický ústav



DISERTAČNÍ PRÁCE

v oboru Fonetika

Mgr. Alžběta Houzar

Akustické parametry řečového signálu a jejich intraindividuální a interindividuální variabilita v češtině

Acoustic parameters of the speech signal and their
intra- and interindividual variability in Czech

Školitel: doc. Mgr. Radek Skarnitzl, Ph.D.

Praha 2023

Poděkování

Na tomto místě bych ráda poděkovala zejména svému školiteli doc. Mgr. Radku Skarnitzlovi, Ph.D., za jeho odborné vedení, četné rady a kromobyčejně vstřícný přístup. Díky patří také Ing. Tomáši Bořilovi, Ph.D., za jeho cenné rady v oblasti akustické i statistické analýzy a jeho neochvějnou trpělivost při jejich předávání. Na závěr chci také poděkovat oběma synům svého školitele, jejichž touha potkat pana prezidenta při příležitosti jmenování svého tatínka profesorem mi byla notnou motivací při dokončování této práce.

Prohlášení

Prohlašuji, že jsem disertační práci vypracovala samostatně, že jsem řádně citovala 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 21. června 2023

Mgr. Alžběta Houzar, v.r.

Abstrakt

Disertační práce se zaměřuje na variabilitu široké škály akustických parametrů řečového signálu v češtině. Shrnuje metodologické postupy aplikované v dosavadních studiích při zkoumání akustických charakteristik řeči a předkládá také souhrnnou analýzu intraindividuální a interindividuální variability obsáhlého souboru vybraných parametrů prostřednictvím vlastních experimentů provedených s využitím jednotného řečového materiálu. V rámci dílčích experimentů jsou parametrizovány základní frekvence, vokalické formanty, krátkodobé a dlouhodobé frekvenční spektrum a temporální rovina řeči, a to pomocí celkem 61 metrik. U všech parametrů byla sledována proměnlivost v řeči mluvčího v rámci jednoho typu projevu a také v závislosti na mluvním stylu; sledovány byly nahrávky čteného a spontánního projevu třiceti mluvčích. Na základě výsledků experimentů byly rovněž získány informace o distribuci hodnot sledovaných parametrů v obou mluvních stylech u dané populace. V rámci závěrečného experimentu byla provedena souhrnná analýza všech zkoumaných parametrů zaměřená na jejich vzájemné souvislosti a jejich podíl na celkové variabilitě řečového signálu. Výsledky této souhrnné analýzy ukazují, že pro akustickou variabilitu řečového signálu jsou určující zejména parametry krátkodobého i dlouhodobého spektrálního sklonu a vokalických formantů, zatímco ukazatele základní frekvence, charakteristiky frekvenčního spektra konsonantů či temporální parametry se na této variabilitě podílejí spíše v nižší míře.

Klíčová slova

akustická analýza, variabilita řečového signálu, intraindividuální variabilita, interindividuální variabilita, základní frekvence, vokalické formanty, frekvenční spektrum, spektrální sklon, spektrální momenty, LTAS, tempo řeči, globální temporální ukazatele

Abstract

The doctoral thesis focuses on the variability of a wide range of acoustic parameters of the speech signal in the Czech language. It summarizes the methodological approaches applied in previous studies on the acoustic characteristics of speech and presents a comprehensive analysis of intra-speaker and inter-speaker variability of an extensive set of selected parameters through original experiments conducted using a unified speech material. The individual experiments examined acoustic parameters that characterize fundamental frequency, vowel formants, short- and long-term frequency spectrum, and the temporal domain of speech, employing a total of 61 metrics. The variability of all parameters was observed within an individual's speech during one type of utterance as well as in relation to speaking style, with recordings of both read and spontaneous speech from thirty speakers being analyzed. Based on the experimental results, information regarding the distribution of parameter values in both speaking styles within the given population was obtained. In the final experiment, a comprehensive analysis of all investigated parameters was conducted, focusing on their mutual relationships and their contribution to the overall variability of the speech signal. The results of this comprehensive analysis indicate that the parameters of short- and long-term spectral slope and vowel formants play a decisive role in the acoustic variability of the speech signal, while indicators such as fundamental frequency, characteristics of consonant frequency spectra, and temporal parameters contribute to this variability to a lesser extent.

Key words

acoustic analysis, speech signal variability, intra-speaker variability, inter-speaker variability, fundamental frequency, vowel formants, frequency spectra, spectral slope, spectral moments, LTAS, speech rate, global temporal metrics

Obsah

Úvod	9
1 Statické a dynamické parametry	11
2 Východiska práce.....	15
2.1 Materiál.....	16
2.2 Analýzy.....	17
3 Základní frekvence	19
3.1 Střední hlasová frekvence	20
3.2 Proměnlivost základní frekvence.....	23
3.3 Kontury f_0	26
3.4 Experiment 1: Parametry základní frekvence	32
3.4.1 Metoda.....	32
3.4.2 Výsledky.....	34
3.4.3 Diskuse a závěr	43
4 Vokalické formanty	48
4.1 Hodnoty formantů v jednotlivých vokálech	48
4.2 Dlouhodobé formantové distribuce.....	54
4.3 Formantová dynamika.....	57
4.4 Formantové trajektorie	59
4.5 Experiment 2: Parametry vokalických formantů	62
4.5.1 Metoda.....	62
4.5.2 Výsledky.....	65
4.5.3 Diskuse a závěr	80
5 Frekvenční spektrum.....	83
5.1 Spektrální vlastnosti vokálů	83

5.2	Experiment 3: Parametry frekvenčního spektra vokálů	87
5.2.1	Metoda	87
5.2.2	Výsledky.....	88
5.2.3	Diskuse a závěr	94
5.3	Spektrální vlastnosti frikativ	96
5.4	Spektrální vlastnosti sonorních konsonantů	100
5.5	Experiment 4: Parametry frekvenčního spektra konsonantů	101
5.5.1	Metoda	101
5.5.2	Výsledky.....	103
5.5.3	Diskuse a závěr	111
5.6	Dlouhodobé průměrné spektrum	115
5.7	Experiment 5: Parametry dlouhodobého průměrného spektra.....	119
5.7.1	Metoda	119
5.7.2	Výsledky.....	121
5.7.3	Diskuse a závěr	124
6	Temporální jevy	127
6.1	Mluvní a artikulační tempo.....	127
6.2	Globální temporální ukazatele.....	129
6.3	Lokální temporální ukazatele	132
6.4	Experiment 6: Temporální parametry	134
6.4.1	Metoda	134
6.4.2	Výsledky.....	136
6.4.3	Diskuse a závěr	145
7	Intenzita	149
7.1	Proměnlivost intenzity	149
8	Experiment 7: Souhrnná analýza akustických parametrů řečového signálu	153
8.1	Metoda	153
8.2	Výsledky	155
8.3	Diskuse a závěr.....	162

9	Celková diskuse a závěr	166
	Seznam použité literatury	172
	Příloha A: Text čteného projevu.....	183
	Příloha B: Hodnoty akustických parametrů	184

Úvod

Řečový projev každého mluvčího je svým způsobem unikátní a lze v něm nalézt určité prvky, které jsou pro daného jedince charakteristické a zároveň jej odlišují od ostatních. Rozdíly mezi hlasy jednotlivců jsou různého typu a mají různé příčiny; některé z nich jsou organické, tedy zapříčiněné fyziologií vokálního traktu, jiné jsou osvojené, tedy dané např. sociálním či regionálním původem mluvčího a jeho individuálními artikulačními návyky. Oba tyto základní zdroje rozdílů se prolínají a v různé míře odrážejí v jednotlivých řečových parametrech (Nolan, 1999). Rozmanitý je ale také řečový projev jednotlivce, který variuje v závislosti na řadě faktorů, jakými jsou např. afektivní nastavení, mluvní styl či zdravotní stav.

Rozdílů v řečovém projevu mezi lidmi využíváme v běžných každodenních situacích, kdy od sebe rozpoznáváme jednotlivce podle hlasu. Přínosná je pro nás i variabilita hlasu jednoho mluvčího, která nám umožňuje odhadnout např. momentální rozpoložení komunikačního partnera. V situacích, s jakými se setkáváme v každodenním životě, zpracováváme zvuk pouze na základě poslechu, a to obvykle aniž bychom vědomě analyzovali jednotlivé aspekty řeči, protože se přirozeně soustředíme primárně na komunikované významy.

K analýze zvuku řeči je však možné přistupovat detailněji, a to jak pomocí poslechu, tak i s využitím instrumentálního měření vlastností akustického signálu. Výsledkem takové podrobné analýzy je rozložení řeči na jednotlivé aspekty a jejich oddělená charakterizace, případně i kvantifikace¹. Tímto způsobem lze do detailu postihnout jak variabilitu řeči napříč mluvčími (interindividuální), tak i v rámci jednoho mluvčího (intraindividuální). Podrobná poslechová a instrumentální

¹ Možnost kvantifikace je typická zejména pro akustickou analýzu, nicméně v některých případech je možná i u analýzy poslechové – např. při vyčíslení četnosti různých typů hezitačních prvků, výskytů určitých hláskových realizací atd.

analýza řečového signálu nachází využití ve specializovaných odvětvích, např. v diagnostické medicíně (některé charakteristiky mohou signalizovat určitá neurodegenerativní onemocnění, viz např. Whitfield & Goberman, 2014) či při fonetické identifikaci mluvčího ve forenzněfonetické praxi (viz např. Hollien, 2002; Nolan, 1999 nebo Skarnitzl, 2014).

Právě možnostem akustické analýzy řečového signálu a sledování interindividuální i intraindividuální variability různých akustických parametrů je věnována tato práce, která shrnuje dosavadní poznatky v této oblasti i metodologické postupy využívané v dostupných studiích a obsahuje také sérii vlastních experimentů. Podrobnější charakteristiku práce obsahuje kapitola 2.

1 Statické a dynamické parametry

Jak již bylo zmíněno, běžný posluchač obvykle vnímá řečový projev jakožto celek, při detailní analýze je ale možné v řečovém signálu nalézt širokou škálu znaků, které lze sledovat jednotlivě. Množství rysů charakteristických pro daného mluvčího, resp. danou promluvu (neboť charakteristiky projevu jednoho mluvčího se mohou lišit v závislosti na řadě faktorů) je možné vysledovat pomocí poslechové analýzy, např. výšku hlasu, otevřenost vokálů, tempo řeči, hezitační chování či řečové vady. Tato analýza je vždy do značné míry subjektivní a jednotlivé prvky mohou být obtížně kvantifikovatelné. Naproti tomu akustická (instrumentální) analýza umožňuje objektivní kvantifikaci akustických parametrů řečového signálu. Akustické proměnné, které je možné v signálu změřit, do značné míry korelují s určitými rysy zachytitelnými percepčně; je např. možné říci, že základní frekvence je akustickým korelátem vnímané výšky hlasu. Ačkoliv však mezi akustickou a percepční rovinou existuje souvislost, vztah mezi nimi není lineární (Fuchs, 2017). Mnohdy má také na výsledný percepční dojem vliv souhra více akustických faktorů; např. vnímanou výšku hlasu určuje nejen hodnota základní frekvence, ale do jisté míry také vokalických formantů.

Jednotlivé akustické proměnné, které jsou zastoupeny v řečovém signálu, lze analyzovat pomocí různých metod a získat tak řadu akustických parametrů. V současné době panuje určitá nejednoznačnost v názvosloví; při popisu řečových charakteristik autoři hovoří např. o „aspektech řeči“, „(idiosynkratických) rysech“, „vlastnostech řeči“ či „parametrech“. Není však vždy zcela zřejmé, ve kterých případech tyto pojmy odkazují obecně k akustické proměnné, jakou je např. f_0 , a ve kterých ke konkrétnímu ukazateli, jež je možné kvantifikovat a porovnat napříč vzorky, např. průměr nebo směrodatná odchylka f_0 (viz např. Nolan, 1999; Cambier-Langeveld, 2007; Skarnitzl, 2014). Jessen (2010) předkládá soubor charakteristik využitelných při srovnávání mluvčích, které dělí na organické,

idiolektální a habituální (viz tabulka 1.1). Tato tabulka obsahuje jak konkrétní měřitelné řečové parametry (např. průměrná základní frekvence), tak i obecné proměnné, na jejichž základě je zapotřebí jednotlivé parametry nejprve získat, a teprve ty je pak možné srovnávat mezi vzorky (např. formantové frekvence). V tomto případě tedy nejsou dvě výše popsané roviny, tj. obecná akustická proměnná a konkrétní ukazatel od této proměnné odvozený, rozlišovány.

ORGANICKÉ	IDIOLEKTÁLNÍ	HABITUÁLNÍ
1. průměrná základní frekvence (f_0)	4. individuální aspekty socio- lektu, dialektu a cizineckého přízvuku	7. artikulační tempo
2. (dlouhodobé) formantové frekvence	5. lingvisticko-fonetické detaily	8. variabilita f_0
3. kvalita hlasu	6. forenznělingvistické rysy	9. disfluentní ² chování

Tabulka 1.1. Charakteristiky využitelné při srovnávání mluvčích (podle: Jessen, 2010: 387).

Pro účely této práce je nicméně jednoznačné rozlišování uvedených rovin, a tedy i přesné definování užívaných pojmů, potřebné. V obecném smyslu budou popisovány jednotlivé proměnné (např. f_0), na jejichž základě mohou být pomocí daných metod kvantifikovány konkrétní parametry (např. průměr nebo směrodatná odchylka f_0). Celková podoba proměnných je charakteristická pro daného mluvčího, avšak srovnávání mezi vzorky je možné teprve u měřitelných parametrů od nich odvozených.

² Jessen (2010) ve své publikaci využívá označení „*dysfluent behaviour*“. V dostupných publikacích panuje určitá nejednotnost ohledně užívání forem „disfluence“ a „dysfluence“, které jsou mnohdy pojímány jako synonyma. Jak ale upozorňuje např. Quesal (1988), zatímco prefix *dis-* implikuje absenci, *dys-* poukazuje na abnormalitu; zatímco „disfluence“ (*disfluency*) podle něj tedy označuje jakoukoliv neplynulost v řeči, včetně těch, které se objevují běžně, pojem „dysfluence“ (*dysfluency*) již vnímá jako neplynulost vzniklou v důsledku řečové patologie, např. koktavosti. V českém prostředí používá toto rozlišení např. Lechta (2013). Varianta s „i“ je tedy využívána i v této práci.

Obecně lze akustické parametry rozdělit do dvou kategorií: na parametry statické a dynamické. Při výzkumu akustiky řeči byla tradičně věnována pozornost měření parametrů statických; ty jsou vyjádřeny jednotlivými hodnotami, a to jednak takovými, které charakterizují celkové chování dané proměnné v signálu (např. průměr či směrodatná odchylka), a jednak těmi, které jsou extrahovány z konkrétního časového bodu (např. hodnoty vokálních formantů ve středu časového průběhu vokálu). Výhoda statické analýzy spočívá v relativní časové úspornosti a poměrně snadné porovnatelnosti (včetně porovnatelnosti pomocí různých statistických postupů) takto získaných hodnot.

Jinou možností, jak přistupovat ke kvantifikaci proměnných v řečovém signálu, je sledování parametrů dynamických, tedy těch, které zachycují průběhy proměnných v čase. Čím dál více studií se přiklání k tomuto přístupu, a to z několika důvodů. Zaprvé, výběr konkrétních bodů v čase může být silně závislý na teoretickém konceptu a může být v některých případech cenný, ale nespolehlivý v případech jiných; řečový signál tvoří kontinuum, ve kterém nelze nalézt jednoznačně definované body, kde by bylo inherentně optimální měřit jednotlivé parametry (Fuchs, 2017: 11). Zadruhé, zatímco artikulační cíle jsou do jisté míry dány jazykovým systémem a nesou lingvistickou informaci, a tedy do jisté míry omezují možnost k vyjádření individuality daného mluvčího, přechody mezi artikulačními cíli jsou individuální a odrážejí charakteristiky dané fyziologií mluvčího (Nolan, 1999) i individuální osvojené artikulační strategie (viz např. Kühnert & Nolan, 1999). Na rozdíl od analýzy statických hodnot tedy dynamická analýza podává informace o pohybech artikulačních orgánů mluvčího, které charakterizují jednotlivce podobně jako jiné automatizované aktivity, např. chůze či běh (McDougall, 2006: 121).

Variabilitě proměnných přítomných v řečovém signálu a jednotlivých parametrů, které z nich vycházejí, se věnují následující oddíly. Studie, které jsou v souvislosti s jednotlivými parametry zmiňovány, jsou v tomto oddílu uváděny zejména kvůli

využívaným metodologickým postupům; tomu, nakolik jsou sledované parametry charakteristické pro jednotlivé mluvčí, se věnují pouze některé z nich. Je vhodné mít na paměti, že míra idiosynkratičnosti některých parametrů může být do určité míry jazykově specifická; také proto se tato práce zaměřuje na přínos jednotlivých řečových parametrů pro identifikaci mluvčího v češtině, nikoliv obecně. Studie představované v následujících oddílech využívaly materiál v různých jazycích, a ačkoliv lze předpokládat, že jejich závěry budou do značné míry platné i pro jazyky ostatní, není zcela na místě předpokládat jejich naprostou univerzálnost.

2 Východiska práce

Tato práce pojednává o široké škále akustických parametrů, jež lze v řečovém signálu sledovat. Shrnuje metodologické postupy aplikované při jejich analýze i dosavadní poznatky ohledně jejich intraindividuální a interindividuální variability a předkládá také pozorování této variability prostřednictvím vlastních experimentů. Ačkoliv existuje celá řada akustických parametrů, které nejsou v této práci zahrnuty, představuje soubor metrik zde obsažených vhodný základ pro důkladnou akustickou analýzu řečového signálu.

Jak je patrné z představení dosavadního výzkumu v následujících kapitolách, dostupné studie pojednávající o variabilitě akustických charakteristik řečového signálu se povětšinou zaměřují na úzký výběr parametrů, často z oblasti jediné proměnné; srovnatelnost jejich výsledků pak může být problematická vzhledem k jejich odlišnostem v metodologii, užitém řečovém materiálu i interpretaci výsledků. Předmětem této práce je analýza interindividuální i intraindividuální variability široké škály parametrů vycházejících z různých akustických proměnných, a to na jednotném řečovém materiálu. Charakteristiky řečového materiálu, který byl v rámci jednotlivých experimentů v této práci využit, představuje následující oddíl této kapitoly.

Další kapitoly jsou zaměřeny na jednotlivé proměnné řečového signálu – základní frekvenci, vokalické formanty, frekvenční spektrum, temporální rovinu řeči a intenzitu; obsahují výčet statických a dynamických parametrů, které byly doposud na základě daných proměnných analyzovány, a popisují metody jejich měření a výsledky zkoumání jejich variability napříč mluvčími i v rámci projevu jednotlivců. Následně jsou představena také vlastní zkoumání variability vybraných statických parametrů, a to v rámci dílčích experimentů věnovaných jednotlivým akustickým proměnným. Poslední experiment obsahuje zhodnocení relativního dopadu parametrů odvozených od všech akustických proměnných zkoumaných

v rámci experimentů předchozích na variabilitu akustického signálu, a to pomocí analýzy hlavních komponent. Práci uzavírá celková diskuse shrnující výsledky všech provedených experimentů a možná témata pro další výzkum.

2.1 Materiál

Všechny experimenty v této práci využívají tentýž řečový materiál, kterým jsou nahrávky 30 českých mluvčích mužského pohlaví. Mluvčí byli náhodně vybráni z referenční databáze obecné češtiny pro forenzní účely (Skarnitzl & Vaňková, 2017) obsahující vzorky hlasu od 100 mužů, rodilých mluvčích češtiny, mezi 19 a 50 lety (průměrný věk činí 25,6 roku se směrodatnou odchylkou 6,7 roku). Databáze obsahuje několik typů mluveného projevu a pro účely této práce byly využity dva z nich: čtený a spontánní.

Pro čtený projev byl využit text o 150 slovech, jehož přednes trvá zhruba 1 minutu (viz příloha A); tento text mimo jiné obsahuje přímou řeč dvou postav, nelze tedy vyloučit, že této skutečnosti mluvčí přizpůsobovali svůj projev a že charakteristiky přednesu textu o odlišném obsahu mohou být rozdílné – je tedy na místě podotknout, že pozorované chování řečových parametrů nemusí být zobecnitelné pro tento mluvní styl, tj. čtený text, všeobecně. Nahrávky přednesu tohoto textu byly využity přes jmenované potenciálně problematické aspekty vzhledem k tomu, že se jednalo o nejvhodnější typ mluveného projevu dostupný v dané databázi, přičemž tato databáze je pro češtinu v současnosti jediným zdrojem obsahujícím více nahrávek od dostatečného počtu mluvčích.

Vzorky spontánního projevu tvořily zhruba minutové úryvky řízeného rozhovoru na téma podle výběru mluvčího o celkovém trvání od 25 do 40 minut, přičemž před extrakcí minutových úryvků byly vyloučeny pětiminutové úseky na začátku a na konci nahrávek. U mluvčích, kteří při rozhovoru produkovali pouze repliky

trvající méně než jednu minutu, mohl být projev prokládán replikami výzkumníka (ty byly ovšem z nahrávek odstraněny); v případě některých minutových úseků se tedy nemusí jednat o celistvý projev daného mluvčího, ale o několik samostatných replik. Podrobnější údaje o obsahu využitých nahrávek relevantní pro analýzy konkrétních parametrů jsou představeny v metodických oddílech jednotlivých kapitol.

Nahrávky byly ve formátu WAV o vzorkovací frekvenci 32 kHz. Byly pořízeny pomocí profesionálního přenosného rekordéru Edirol HR-09, a to v tichých, nikoli však studiových podmínkách, určité akustické odchylky napříč vzorky od jednotlivých mluvčích tudíž nelze vyloučit.

Všechny řečové vzorky byly opatřeny segmentací na jednotlivé hlásky v programu Praat (Boersma & Weenink, 2023) – hranice segmentů byly nejprve určeny automaticky pomocí nástroje Prague Labeller (Pollák et al., 2007) a následně manuálně zarovnaný podle principů fonetické segmentace hlásek popsaných Macháčem a Skarnitzlem (2009). Byly také vyznačeny hranice souvislých promluvových úseků oddělených pauzami a počty v nich obsažených slabik, které byly v zájmu přesnosti spočítány manuálně.

2.2 Analýzy

Experimenty zaměřené na analýzu akustických parametrů, které jsou obsaženy v následujících kapitolách, využívají výše popsaný materiál. Extrakce hodnot je prováděna v programu Praat (Boersma & Weenink, 2023), případně v R (R Core Team, 2021) za využití knihovny rPraat (Bořil & Skarnitzl, 2016). Jedná se o explorační analýzu akustických charakteristik řečového signálu.

U jednotlivých parametrů je zjišťována významnost rozdílů jejich hodnot mezi sledovanými mluvními styly, tj. čteným a spontánním projevem, a to

prostřednictvím Wilcoxonova testu³, přičemž při určení hladiny významnosti je aplikována Bonferroniho korekce: $\alpha = 0,05/n$, kde n je počet sledovaných parametrů v daném experimentu.

Dále je v rámci jednotlivých experimentů sledována míra korelace sledovaných parametrů. Síla korelace je klasifikována na základě korelačního koeficientu r podle Evanse (1996):

- $|r|$ mezi 0,00 a 0,19 – velmi slabá korelace,
- $|r|$ mezi 0,20 a 0,39 – slabá korelace,
- $|r|$ mezi 0,40 a 0,59 – středně silná korelace,
- $|r|$ mezi 0,60 a 0,79 – silná korelace,
- $|r|$ mezi 0,80 a 1,00 – velmi silná korelace.

Vizualizace výsledků je realizována v R (R Core team, 2021) za využití knihoven ggplot2 (Wickham, 2016) a corrplot (Wei & Simko, 2021).

Populační statistiky všech sledovaných parametrů získané na základě využitého materiálu obsahuje příloha B.

³ Tj. neparametrického párového t-testu; podrobněji viz např. Volín (2007: 155–160).

3 Základní frekvence

Jednu z nejhojněji zkoumaných akustických proměnných v řečovém signálu představuje základní frekvence⁴, tj. frekvence kmitání hlasivek při fonaci. Identitu jedince odráží jak celková poloha základní frekvence, tak i její rozsah a také vzorce průběhu f_0 , které se v jeho projevu vyskytují. Jedná se ale o velice nestabilní proměnnou, kterou ovlivňuje celá řada faktorů, jakými jsou např. afektivní stavy, zdravotní stav mluvčího či mluvný styl, a její měření je značně náchylné ke zkreslení v důsledku kvalitativní degradace signálu (více viz např. Braun, 1995 či Skarnitzl & Hývlová, 2014).

Základní frekvenci, stejně jako další proměnné, lze parametrizovat prostřednictvím různých postupů. Při sledování intraindividuální i interindividuální variability f_0 lze využívat statické parametry, mezi které patří střední hodnota f_0 vyčíslená např. coby průměr, medián či základní hladina (viz níže), ale také rozsah hodnot, kterých f_0 v průběhu daného vzorku nabývá, vyjádřený nejčastěji jako směrodatná odchylka; dále je možné provádět i analýzu dynamického průběhu základní frekvence, tj. kontur f_0 , např. v rámci prozodické fráze.

⁴ Značení základní frekvence v dostupných studiích je do značné míry nekonzistentní: písmeno se objevuje jako verzálka i minuska, číslo pak někdy v dolním indexu, jindy nikoliv, v některých případech nejde o číslovku „0“, nýbrž o alfabetický znak „o“. Titze et al. (2015) navrhli sjednocení značení základní frekvence jako f_0 s využitím malého „f“ pro jasné odlišení od vokálních formantů, které jsou značeny pomocí velkého písmene „F“, a také s alfabetickým „o“ namísto číslovky v dolním indexu pro označení „oscilace“. V českém prostředí však bývá základní frekvence označována jakožto „ef nula“, písmeno „o“ se zde tedy jeví jako spíše nevhodné. V této práci tedy bude využíváno malé písmeno „f“ psané kurzívou, jak navrhuje Titze et al. (2015), avšak v dolním indexu bude následovat číslovka „0“ podle zavedeného českého úzu – viz např. Skarnitzl & Hývlová (2014) či Skarnitzl et al. (2016).

3.1 Střední hlasová frekvence

Způsobem parametrizace f_0 , který se předně nabízí, je změření určité hodnoty, která obecně charakterizuje běžnou polohu základní frekvence v daném řečovém vzorku, tedy tzv. střední hlasová frekvence. Jako nejpřímochařejší parametr tohoto typu se jeví aritmetický průměr f_0 ; právě ten využívá ve své studii např. Braun (1995) při porovnávání vlivu různých faktorů na střední hlasovou frekvenci. Existují však studie, podle nichž není aritmetický průměr optimálním způsobem, jak střední hlasovou frekvenci charakterizovat; např. Lindh (2006) se kloní spíše k využívání mediánu, který je méně náchylný ke zkreslení vlivem extrémních hodnot – ty se mohou objevit i kvůli chybné detekci f_0 využívaným softwarem, kdy může dojít ke vzniku oktávových skoků. Ke stejnému závěru docházejí i Lindh a Eriksson (2007) či Skarnitzl a Hývlová (2014).

Jiným parametrem, který postihuje jakousi hodnotu základní frekvence charakteristickou pro daný vzorek, je základní hladina. Její hodnota by měla odpovídat frekvenci běžné (nikoli třepené) fonace na konci prozodické fráze (Skarnitzl & Vaňková, 2017)⁵. Lindh a Eriksson (2007) charakterizují základní hladinu jako nosnou frekvenci řeči jedince, již lze chápat jako uvolněnou pozici, ke které se mluvčí vrací po intonačních úkonech, a která je tudíž ve většině podmínek mnohem lepším prediktorem „typické“ základní frekvence jedince než průměr nebo medián (Lindh & Eriksson, 2007: 2026; Skarnitzl & Hývlová, 2014). Lindh a Eriksson (2007) zmiňují, že nejnižší hodnota základní frekvence, která se v řeči daného mluvčího objevuje, je často shodná s nejnižší frekvencí, při níž je tento mluvčí schopen udržet fonaci, případně se jí alespoň velice blíží. Na základě této skutečnosti pak předpokládají, že zatímco rozsah, a tudíž také průměr f_0 bude

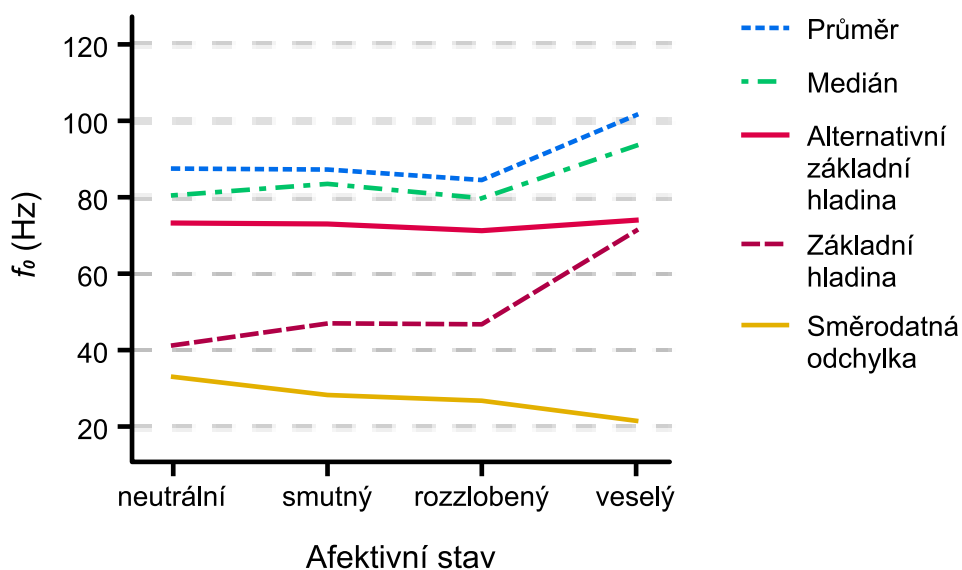
⁵ Skarnitzl a Hývlová (2014) podotýkají, že základní hladinu nelze považovat za ukazatel střední hodnoty f_0 v pravém slova smyslu, vzhledem k tomu, že její hodnota se blíží spíše spodní hranici než středu rozsahu f_0 v daném řečovém vzorku. Tento parametr však bývá srovnáván s průměrem a mediánem, je proto i zde zařazen právě do této kategorie.

proměnlivý v závislosti např. na afektivních stavech, nejnižší hodnota f_0 bude naopak zůstat neměnná.

Pro výpočet hodnoty základní hladiny uvádějí Traunmüller a Eriksson (1995) vzorec:

$$Fb = Fmean - k \cdot \sigma(F)$$

kde Fb je základní hladina f_0 , $Fmean$ její průměrná hodnota, $\sigma(F)$ její směrodatná odchylka a k představuje zvolenou škálovací konstantu. Jak upozorňují Lindh a Eriksson (2007), spolehlivost výsledné hodnoty základní hladiny vzhledem k výpočtu na základě průměru a směrodatné odchylky f_0 závisí na přesnosti těchto hodnot. Jejich spolehlivost přitom může být v případě méně kvalitních nahrávek sporná kvůli možným chybám v extrakci f_0 zmíněným již výše v souvislosti s průměrem. Autoři proto představují alternativní způsob výpočtu základní hladiny, který je méně ovlivněn odlehlými hodnotami. Jeho výsledkem jsou u kvalitních nahrávek (u kterých nedošlo k chybám v extrakci f_0 , jež by výrazně zkreslily hodnoty průměru a směrodatné odchylky f_0) přibližně tytéž hodnoty, které jsou získány pomocí původního vzorce. Jedná se o hodnotu f_0 , pod kterou leží 7,64 % všech hodnot f_0 v daném signálu. Takto vypočítaná základní hladina (autoři ji ve své studii nazývají alternativní základní hladinou) se v porovnání s průměrem, mediánem a základní hladinou vyčíslenou pomocí původního vzorce ukázala jako nejstabilnější parametr u daného mluvčího, a to napříč afektivními stavy i úrovněmi kvality řečového signálu, což ilustruje obrázek 3.1; zmíněná studie využívala švédský řečový materiál. Stabilitu tohoto parametru v rámci mluvčího sledovali také Da Silva et al. (2016) na řečovém materiálu v brazilské portugalštině, a to ve srovnání s průměrem a mediánem f_0 , ale také s parametry reflektujícími jiné aspekty f_0 než její střední hodnotu, např. směrodatnou odchylkou, modem či modální hustotou f_0 . Základní hladina byla i v tomto případě parametrem, který charakterizoval jednotlivce nejspolehlivěji.



Obrázek 3.1. Upraveno podle: Lindh & Eriksson (2007: 2027). Výsledky analýzy nahrávek, na kterých mluvčí imitoval různé afektivní stavy. Testované parametry – průměr, medián, alternativní základní hladina a základní hladina – jsou vyobrazeny shora dolů v tomto pořadí. Linie ve spodní části reprezentuje směrodatnou odchylku f_0 .

Variabilitu parametrů střední hodnoty f_0 sledovali také Skarnitzl a Hývlová (2014), a to na českém materiálu; zaměřili se na průměr, medián a základní hladinu (resp. alternativní základní hladinu, jak ji představili Lindh a Eriksson, 2007). Také oni dospěli k závěru, že základní hladina dobře charakterizuje chování základní frekvence u jednotlivce, a nadto uvádějí, že „u mluvčích s výraznějšími melodickými modulacemi se základní hladina stabilizuje výrazně dříve než průměr“ (Skarnitzl & Hývlová, 2014: 63) – tj. pro spolehlivou extrakci základní hladiny je zapotřebí menší množství řečového materiálu. K témuž závěru dospěli také Arantes a Eriksson (2014), kteří analyzovali parametry základní frekvence na materiálu ve 26 jazycích. Pro odlišení jednotlivců navzájem doporučují Skarnitzl a Hývlová (2014) kombinaci základní hladiny f_0 a mediánu, případně průměru.

Při porovnávání dvou vzorků za účelem zjištění, zda se jedná o téhož mluvčího, ovšem není klíčová pouze podobnost hodnot daných parametrů, ale také jejich

typičnost. K jejímu určení mohou sloužit statistiky sledovaných parametrů u dané populace – populační statistiky hodnot základní frekvence v současné době existují např. pro němčinu (Jessen et al., 2005), švédštinu (Lindh, 2006), britskou angličtinu (Hudson et al., 2007) a také pro češtinu (Skarnitzl & Vaňková, 2017).

3.2 Proměnlivost základní frekvence

Vedle střední hodnoty charakterizuje chování základní frekvence také její proměnlivost, tj. rozptyl, v jakém se f_0 v řeči daného mluvčího pohybuje, případně nakolik výrazně kolísá.

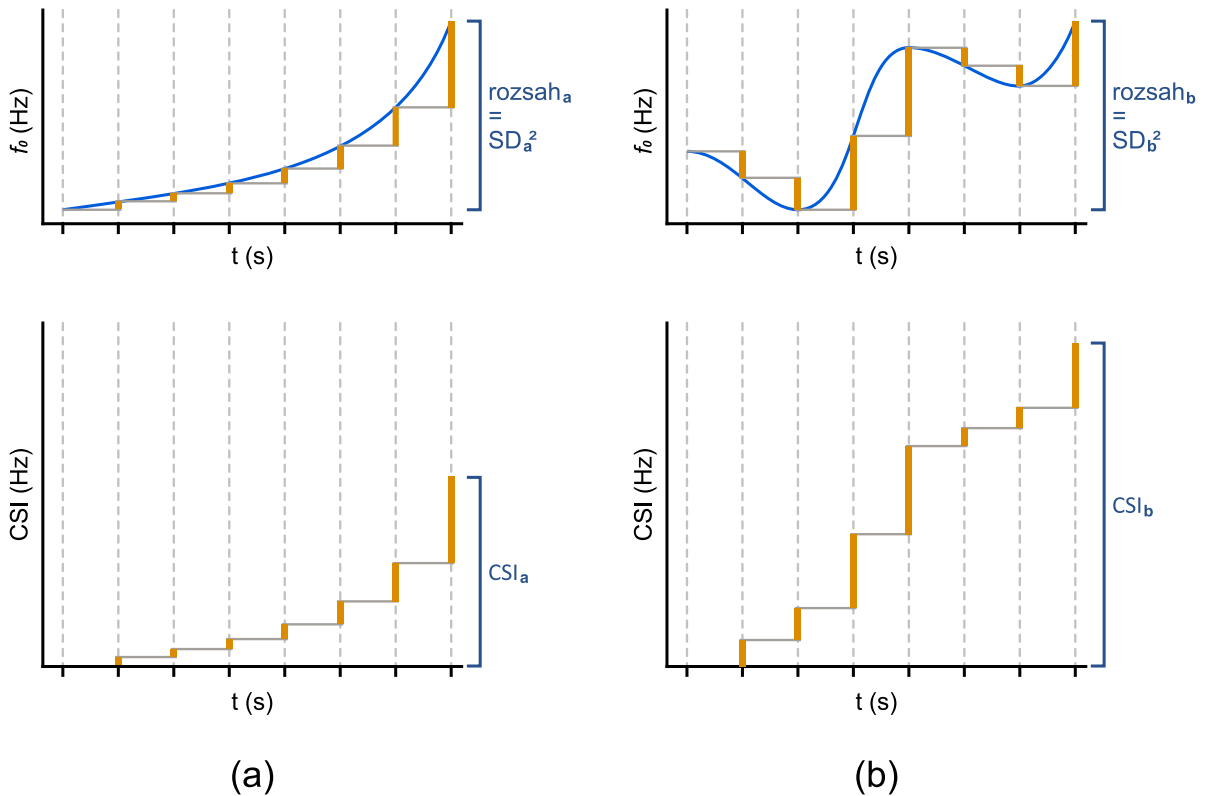
Rozptyl základní frekvence je možné měřit pomocí různých parametrů, např. směrodatné odchyly (*standard deviation*; SD), percentilového rozpětí (např. rozdílu mezi 95. a 5. percentilem či mezi 90. a 10. percentilem), celkového rozpětí (tj. rozdílu mezi maximální a minimální hodnotou f_0) či variačního koeficientu (*varco*; tj. normalizované směrodatné odchyly), viz např. Mennen et al. (2012) či Skarnitzl & Hývlová (2014). Patrně nejčastěji využívaným z těchto parametrů je směrodatná odchyly, která často doplňuje měření parametrů střední hodnoty základní frekvence (viz např. Lindh, 2006; Lindh & Eriksson, 2007; da Silva et al., 2016 či Skarnitzl & Vaňková, 2017).

Tři z parametrů zachycujících rozptyl f_0 byly porovnány v rámci experimentu Skarnitzla a Hývlové (2014), a to celkové rozpětí, směrodatná odchyly a rozdíl mezi 90. a 10. percentilem. Autoři upozorňují na značné rozdíly v chování celkového rozpětí a rozpětí mezi 10. a 90. percentilem u některých řečových vzorků a vysvětlují je častějším výskytem odlehlých hodnot vzniklých chybnou extrakcí f_0 ; užití daného percentilového rozpětí tedy považují za vhodnější. Jak již bylo zmíněno výše, podle autorů studie nejlépe charakterizuje mluvčího kombinace dvou parametrů střední hlasové frekvence, a to základní hladiny a mediánu (případně

průměru); o tom, že by zásadní přínos mělo také zařazení některého z parametrů rozptylu f_0 , např. právě rozpětí mezi 10. a 90. percentilem, se však nezmiňují.

Smirnova et al. (2007) sledovali několik parametrů spojených s variabilitou základní frekvence v jednoslabičných slovech rozdělených podle typu expresivity a typu věty, ve které se objevila, a ve vyplněných pauzách v tádžičtině. Mezi sledovanými faktory byly mimo jiné hodnoty f_0 na začátku a na konci sledovaného úseku, maximum, minimum a rozsah f_0 , zároveň pak také průměr (tedy ukazatel střední hodnoty f_0); ve vyplněných pauzách byly sledovány pouze minimum, maximum a průměr f_0 . Jak se ukázalo, tyto parametry byly u jednotlivých mluvčích poměrně konzistentní, zejména pak hodnota f_0 na konci sledovaných úseků (tento závěr koresponduje se skutečností, že jako nejstabilnější ukazatel střední hlasové frekvence se ukazuje základní hladina, viz výše). Jako nejvíce charakteristické pro mluvčího se ukázaly hodnoty f_0 extrahované z vyplněných pauz.

Výše popsané parametry kvantifikující variabilitu základní frekvence postihují celkový rozsah, v jakém se f_0 v daném vzorku pohybuje, neposkytují však informace o tom, jakým způsobem kolísá; u dvou mluvčích s podobným rozsahem f_0 , z nichž však u jednoho v rámci nádechového úseku dochází k jejím častým nárůstům a poklesům, zatímco u druhého f_0 pozvolna klesá, by tyto parametry byly velmi podobné, ačkoliv vzorce průběhu f_0 se u nich výrazně liší. O těchto rozdílech v pohybech f_0 vypovídá parametr, který ve své práci poprvé definoval Hruška (2016): tzv. index kumulativní strmosti (*cumulative slope index; CSI*). Ten zachycuje, jak rychle a do jaké míry kontura f_0 kolísá v průběhu řečového vzorku; jeho hodnota je „součtem absolutních hodnot rozdílů frekvence mezi po sobě jdoucími body, vyděleným trváním měřeného řečového segmentu“ (Hruška & Bořil, 2017: 37); princip jeho výpočtu ilustruje obrázek 3.2.



Obrázek 3.2. Schematické znázornění ukazatelů variability f_0 – rozsahu, resp. směrodatné odchylky, a indexu kumulativní strmosti: nahoře kontury f_0 s vyznačením rozdílů mezi po sobě jdoucími vzorky, vespod kaskádový graf zobrazující CSI, který je součtem absolutních hodnot rozdílů mezi po sobě jdoucími vzorky. Zatímco rozsah, a tedy i směrodatná odchylka obou kontur jsou shodné, CSI je u kolísající kontury (b) vyšší než u té, která pouze postupně narůstá (a).

Hruška a Bořil (2017) pozorovali, nakolik má proměnlivost melodie řeči vliv na vnímání expresivity dané promluvy v češtině, přičemž tuto proměnlivost kvantifikovali právě prostřednictvím indexu kumulativní strmosti f_0 . Vzhledem ke skutečnosti, že jejich studie byla zaměřena na řečovou percepci, byla kontura f_0 nejprve stylizována tak, aby obsahovala pouze percepčně relevantní pohyby (makrointonaci), zatímco drobné výkyvy f_0 , které na percepci neměly vliv (mikrointonace), byly odstraněny. CSI pak byl vyčíslen v půltónech za sekundu. Výsledky studie ukázaly, že hodnoty CSI (odrážející míru proměnlivosti melodie) se liší napříč mluvčími i stupni expresivity. Různí mluvčí však pro vyjádření expresivity využívají různé prostředky, nejen melodii řeči, ale také např. tempo či

barvu hlasu; změny hodnot CSI v závislosti na míře expresivity tudíž nejsou konzistentní napříč mluvčími.

Volín et al. (2017) se věnovali variabilitě CSI f_0 v závislosti na věku a pohlaví v češtině. Tato studie, na rozdíl od výzkumu Hrušky a Bořila (2017), nebyla zaměřena na vliv proměnlivosti základní frekvence na řečovou percepci, nýbrž na akustickou variabilitu prozodických charakteristik, z kontury f_0 tudíž nebyla vyřazena mikrointonace. Pro vyjádření CSI byly i v této studii využity půltóny; časovou jednotkou, pomocí které byl CSI kvantifikován, však v tomto případě nebyla sekunda, nýbrž slabika. Dále byl sledován také celkový rozsah f_0 jakožto druhý ukazatel vyjadřující variabilitu základní frekvence. Na základě výsledků experimentu dospěli autoři k závěru, že variabilita základní frekvence obecně narůstá s věkem, přičemž u žen byl zjištěn větší rozsah f_0 , hodnoty CSI naproti tomu poukázaly na výraznější kolísání f_0 u mužů.

Na základě dostupných studií tedy lze usuzovat, že jak interindividuální, tak intraindividuální variabilita indexu kumulativní strmosti je značná. Jedná se však o poměrně nově zavedený parametr, jehož potenciál pro vzájemné odlišení mluvčích dosud nebyl podrobně prozkoumán.

3.3 Kontury f_0

Doposud popsané statické parametry základní frekvence zachycují všeobecné chování f_0 , nedokáží ale postihnout konkrétní podobu jejích vzorců, které se v řečovém signálu objevují. Ačkoliv index kumulativní strmosti poskytuje informace o rychlosti a velikosti změn základní frekvence, nelze jeho pomocí určit přesný tvar jejích pohybů; jak upozorňují Hruška a Bořil (2017), řečové vzorky s postupně klesající a postupně stoupající f_0 o stejném rozsahu budou

charakterizovány touž hodnotou CSI. V řečovém signálu je však vedle statických parametrů f_0 možné sledovat také dynamický průběh základní frekvence, tedy tvary kontur f_0 .

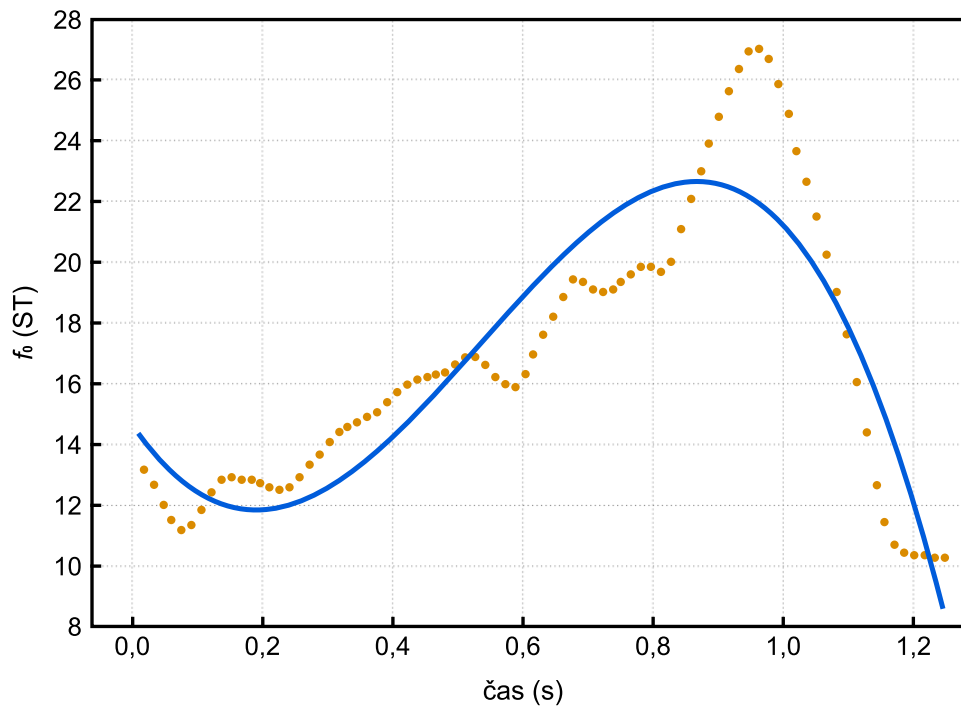
Melodie řeči, která je percepčním korelátem f_0 , se primárně odvíjí od významu, který má mluvčí v úmyslu sdělit. Vzorce f_0 ale zároveň obsahují také indexovou informaci, tzn. vypovídají i o individualitě jednotlivce: ukazuje se, že různí mluvčí při realizaci promluv se stejným lexikálním obsahem pro vyjádření stejného významu produkují různé vzorce f_0 , přičemž kontury, které využívají, jsou nenáhodné (Volín & Bořil, 2014). Nabízí se tedy možnost sledovat, nakolik tvary kontur f_0 variiují napříč mluvčími i v rámci projevu jednoho mluvčího.

Ke sledování dynamického průběhu základní frekvence lze přistupovat několika způsoby. Tím patrně nejjednodušším z nich je sledování obecného sklonu f_0 v daném úseku, tj. zda (a případně nakolik) základní frekvence postupem času stoupá, či klesá. Jeden ze způsobů aplikace tohoto přístupu představili ve své studii Adami et al. (2003), kteří sledovali sklon trajektorií f_0 a zároveň také intenzity, a to pouze v rámci jednotlivých znělých segmentů v americké angličtině. Autoři proložili konturu f_0 uvnitř každé znělé hlásky regresní přímkou a podle jejího sklonu následně k dané hlásce připojili označení + (stoupající f_0) nebo - (klesající f_0). Zároveň obdobným způsobem vyhodnocovali pohyb intenzity dané hlásky; takto byl každé znělé hlásce přiřazen bigram vyjadřující trend pohybu základní frekvence a intenzity: ++, +-, -+, --. (Vyhodnocováno bylo tedy pouze to, zda má daná kontura klesavý, nebo stoupavý charakter, přičemž velikost sklonu nebyla brána v úvahu; všechny kontury byly označeny jako stoupající či klesající, o rovných konturách se autoři nezmiňují.) Neznělé hlásky byly označeny jako „uv“ (*unvoiced*). Sledováno bylo také trvání jednotlivých stoupavých a klesavých úseků. Podle autorů studie mohou být právě takovéto vzorce spojené se základní frekvencí důležitým nositelem informace o individualitě mluvčího, neboť jedinci se navzájem výrazně liší ve způsobu realizací jednotlivých intonačních úkonů; idiosynkratický

může být např. podíl trvání stoupavých a klesavých úseků. Jak autoři usuzují na základě výsledků svého experimentu, toto modelování dynamiky f_0 (a intenzity) charakterizuje jednotlivce lépe než statické ukazatele základní frekvence.

Celkovou regresi kontury f_0 neboli gravitační trend, jak tento parametr autoři nazývají, sledovali také Volín a Bořil (2014), v tomto případě se však jednalo o zachycení průběhu f_0 v rámci celých promluvových úseků v češtině. Analyzovaným materiálem byly dvojice čtených vět o téměř shodném lexikálním obsahu, ze kterých byly extrahovány kontury f_0 . U těch byla následně provedena manuální korekce a interpolace hodnot f_0 v neznělých úsecích. Konturou f_0 pak byla proložena regresní přímka a byl vypočítán její sklon, vyjádřený v půltónech za sekundu. Tento gravitační trend byl shledán dobře charakterizujícím mluvčího, za předpokladu, že je srovnáván lingvisticky analogický materiál. Autoři však upozorňují, že kritéria pro takovouto srovnatelnost nejsou zcela jednoznačná a zároveň jsou v reálných kontextech srovnávání mluvčích (např. ve forenzní praxi) stěží dosažitelná.

Poněkud složitějším přístupem, který však poskytuje přesnější informace o tvaru kontury f_0 , je její proložení polynomiální křivkou. Tento postup v rámci své studie taktéž aplikovali Volín a Bořil (2014), a to u stejného materiálu jako ve výše zmíněném experimentu. Jednotlivé kontury f_0 byly proloženy polynomem 3. stupně (tj. kubickým polynomem; viz obr. 3.3), a byly tedy získány čtyři koeficienty určující tvar polynomiální křivky. Jako nejvíce charakteristický pro mluvčího se ukázal první koeficient, který odráží hodnotu f_0 na počátku kontury, což nasvědčuje tomu, že mluvčí mají tendenci začínat promluvu na stejné úrovni f_0 ; podle autorů studie je tato úroveň výsledkem interakce celkové polohy hlasu mluvčího a jeho individuálních intonačních návyků. Připomínají však, že nelze opomíjet také vliv jazykového systému, který podobu melodie řeči do značné míry předurčuje.



Obrázek 3.3. Upraveno podle: Volín & Bořil (2014: 14). Příklad proložení kontury f_0 (vyznačené žlutými body) polynomiální křivkou třetího řádu (vyznačenou modře).

K proložení kontury f_0 křivkou přistoupili také Grabe et al. (2007), kteří propojili měření akustických proměnných s intonační analýzou využívající autosegmentálně-metrický popis melodických kontur. Čtené anglické věty různého typu byly nejprve opatřeny anotací, vytvořenou na základě poslechové analýzy; takto bylo identifikováno celkem 7 různých typů melodické kadence, určených sekvencí stoupavých, klesavých a rovných úseků. Autoři studie sledovali, zda se tyto percepčně odlišné typy melodie různí také akusticky, tj. zda mají věty, u nichž byly poslechem určeny rozdílné melodické průběhy, také rozdílné kontury f_0 . Pro matematické modelování kontur základní frekvence Grabe et al. (2007) zvolili Legendreovy polynomy, a to proto, že mezi koeficienty, jimiž jsou definovány, existuje minimální korelace, na rozdíl např. od kubického polynomu (podrobněji viz např. Svatošová & Volín, 2022). Ze sledovaných vět byly extrahovány kontury f_0 , které byly následně normalizovány průměrnou základní frekvencí daného mluvčího (tj. byl vyloučen vliv celkové polohy hlasu mluvčích). Získané kontury byly proloženy Legendreovými polynomy; výsledkem analýzy pak byl

matematický model f_0 pro každý melodický průběh vyjádřený čtyřmi koeficienty. Výsledky ukázaly, že šest ze sedmi identifikovaných kadencí bylo skutečně charakterizováno statisticky významně odlišnými koeficienty; ačkoliv se tedy označování melodických průběhů na základě poslechu může jevit jako subjektivní, skutečně nachází oporu v reálných vzorcích základní frekvence. V rámci této studie bylo sledováno sedm různých dialektů angličtiny a autoři se také zaměřili na to, jak často se v jednotlivých dialektech objevují identifikované typy melodického průběhu; jak zjistili, některé z těchto typů se objevovaly pouze v jednom nebo dvou z analyzovaných dialektů, jediný z nich byl nalezen ve všech dialektech a každý dialekt se vyznačoval odlišnou distribucí jednotlivých typů. Ačkoli předmětem této studie nebyla otázka, jak se ve využívání daných typů melodických průběhů liší jednotliví mluvčí, mohly by tyto individuální intonační strategie poskytnout zajímavé informace o variabilitě základní frekvence mezi skupinami mluvčích a potenciálně také mezi jednotlivci. Další otázkou, která se v tomto kontextu nabízí, je, jaká by byla interindividuální a intraindividuální variabilita hodnot koeficientů Legendreových polynomů v rámci stejných typů melodických průběhů, které by byly v projevu mluvčích identifikovány.

Analýzu průběhu f_0 založenou na autosegmentálně-metrickém intonačním modelu provedl také Nolan (2002), využil však odlišný postup. Nejednalo se o modelování samotné křivky f_0 , nýbrž o sledování načasování vrcholů či propadů f_0 vůči segmentálnímu složení promluvy. Autosegmentálně-metrická teorie popisuje intonační kontury prostřednictvím intonačních vrcholů (H) a propadů (L), které jsou vázány na pozici intonačních jednotek uvnitř promluvy, zejména na přízvukné slabiky a hranice intonačních frází (podrobněji viz např. Van der Hulst & Smith, 1982). Vzhledem ke skutečnosti, že f_0 je akustickým korelátem intonace, považuje Nolan vrcholy a propady f_0 za akustickou reprezentaci zmíněných vrcholů a propadů v melodii řeči. Ve své studii sledoval právě synchronizaci těchto bodů a segmentálních prvků v britské angličtině, přičemž využíval nahrávky dvou replik s předem určenými intonačními vzorci, realizovaných osmi mluvčími.

Předpokládal, že to, s jakým načasováním vůči danému segmentálnímu úkonu se určitý zlom v kontuře f_0 objeví, se napříč mluvčími liší. Podle výsledků studie takto určené parametry vesměs nevykazují takovou interindividuální variabilitu a intraindividuální stabilitu, aby byly vhodnými ukazateli pro vzájemné rozlišení mluvčích. Ukázalo se však, že větší míru idiosynkratičnosti lze nalézt mezi zlomovými body kontury f_0 – tj. největší prostor pro individuální variabilitu patrně poskytují úseky mezi fonologickými intonačními cíli, nikoliv tyto cíle samotné – toto pozorování je v souladu s tezemi studií zmíněných v kapitole 1, podle nichž je podoba artikulačních cílů do značné míry určena jazykovým systémem, zatímco jsou to právě přechody mezi nimi, které poskytují prostor pro individuální artikulační strategie (Nolan, 1999; Kühnert & Nolan, 1999).

Dosud zmíněné experimenty, které provedli Adami et al. (2003) či Grabe et al. (2007), tedy využívaly přístupy nezávislé na lexikálním obsahu, zatímco Nolan (2002) a Volín a Bořil (2014) naopak srovnávali úseky o velmi podobném segmentálním a rytmickém složení. Další přístup, který představili Adami et al. (2003) v rámci již zmíněné studie, pak využívá přímo srovnávání totožných slov. Autoři navazují na systémy, které využívají tzv. dynamické borcení časové osy (*dynamic time warping*; DTW) a které jsou využívány k ověřování identity mluvčího prostřednictvím srovnávání nahrané šablony a cílového slova či fráze (více viz např. Pandit & Kittler, 1998). Adami et al. (2003) vybrali několik výrazů, které se ve sledovaném materiálu objevovaly často a zároveň nebyly příliš závislé na verbálním kontextu, a jejich podoba tedy byla v rámci projevu jednoho mluvčího poměrně stabilní. Jednalo se zejména o diskurzivní markery jako např. „right“, „okay“, „well“ apod. Autoři z řečového materiálu extrahovali realizace daných slov a pomocí DTW algoritmu srovnávali jejich kontury f_0 . Jak autoři zjistili, zahrnutí tohoto parametru zvýšilo úspěšnost automatického rozpoznávače mluvčích, a v oblasti vzájemného rozlišování jednotlivců jej tedy považují za slibný.

Je třeba podotknout, že využití dynamické analýzy f_0 je výrazně náročnější než aplikace analýzy statické, neboť vyžaduje podrobné zpracování sledovaného řečového materiálu. Statická analýza tedy umožňuje práci s výrazně větším množstvím řečového materiálu. Dynamická analýza na druhou stranu poskytuje přesnější informace o individuálních řečových strategiích jednotlivců. Jak uzavírají Adami et al. (2003), za nejvhodnější postup pro zachycení idiosynkratických vzorců základní frekvence lze považovat kombinaci obou přístupů.

3.4 Experiment 1: Parametry základní frekvence

3.4.1 Metoda

Za využití materiálu popsaného v kapitole 2 byly sledovány následující parametry základní frekvence: průměr, medián, základní hladina (ZH), směrodatná odchylka (SD), rozdíl mezi 95. a 5. percentilem, rozdíl mezi 90. a 10. percentilem a index kumulativní strmosti (CSI).

V programu Praat (Boersma & Weenink, 2023) byly z jednotlivých zvukových souborů pomocí příkazu *To Pitch (ac)* extrahovány kontury základní frekvence. Frekvenční rozmezí pro extrakci bylo stanoveno v souladu se zjištěními de Looze (2010), která se ve své disertační práci zaměřila na nalezení postupu, jehož pomocí lze co nejpřesněji určit reálný rozsah základní frekvence. Na základě výsledků svých pozorování de Looze (2010: 106) doporučuje nejprve extrahovat konturu f_0 s extrémními hodnotami hranic rozsahu, následně vypočítat 15. a 65. percentil získaných hodnot a provést extrakci kontury základní frekvence ze zvukového souboru podruhé, tentokrát o rozsahu odvozeném od daných percentilů, a to 0,83násobku 15. percentilu a 1,92násobku 65. percentilu. Popsaný postup byl tedy aplikován i v této práci; hodnoty základní frekvence byly

extrahovány nejprve s minimem 75 Hz a maximem 350 Hz⁶, poté byl vypočítán 15. a 65. percentil získaných hodnot, a opět byla extrahována kontura f_0 , tentokrát o rozsahu $0,83 \times 15.$ percentil až $1,92 \times 65.$ percentil kontury první. Mimo minima a maxima f_0 bylo při extrakci využito výchozí nastavení.

Na základě takto extrahovaných kontur základní frekvence byly pro každý promluvový úsek vyčísleny průměr, medián, základní hladina (tj. 7,64. percentil), směrodatná odchylka, rozdíl mezi 90. a 10. percentilem a rozdíl mezi 95. a 5. percentilem, přičemž jednotkou měření byl Hertz.

Pro měření indexu kumulativní strmosti bylo zapotřebí aplikovat korekci hodnot základní frekvence, neboť tento parametr se vyznačuje vysokou citlivostí na odlehle hodnoty. Před výpočtem CSI je žádoucí eliminovat chyby vzniklé při extrakci hodnot základní frekvence – rizikové jsou v tomto ohledu zejména oktávové skoky. Byly odstraněny hodnoty f_0 uvnitř znělých i neznělých obstruentů, v nichž je riziko chybné detekce f_0 relativně vysoké. Takto upravené kontury základní frekvence sloužily jako materiál pro první výpočet CSI. Nelze však zcela vyloučit, že i v těchto konturách byly nadále přítomny chybně detekované pohyby základní frekvence včetně oktávoových skoků; dále tedy byla aplikována funkce *smooth* (s výchozí šířkou pásma 10 Hz⁷), jejímž výsledkem je vyhlazená kontura základní frekvence. Na základě takto upravených kontur f_0 pak byl CSI změřen podruhé, aby bylo možné porovnat získané hodnoty CSI jak bez využití této korekce, tak s ním. Hodnoty CSI byly měřeny v rámci jednotlivých promluvových úseků, a to v půltónech pro zamezení vlivu celkové hladiny základní frekvence. Následně byly normalizovány celkovým trváním promluvových úseků a počtem slabik.

⁶ De Looze (2010) při extrakci první kontury základní frekvence využila výchozí nastavení Praatu, kterým je frekvenční rozsah pro extrakci 75 až 600 Hz. Vzhledem k využití materiálu pouze od mužských mluvčích a relativně ploché intonaci, která je typická pro češtinu (ve srovnání např. s francouzštinou či angličtinou, na něž se de Looze zaměřovala), byla v této práci zvolena nižší horní hranice.

⁷ Při aplikaci této funkce tedy dojde k vyhlazení kontury 10Hz filtrem, a tudíž k odstranění náhlých změn rychlejších než 100 ms.

Z každé nahrávky tudíž byly získány čtyři soubory hodnot CSI lišící se mírou korekce výchozí kontury f_0 a způsobem normalizace.

Pro každou nahrávku byly vypočítány průměrné hodnoty všech parametrů a také jejich konfidenční intervaly⁸. K vyhodnocení, zda se ve sledovaných mluvních stylech napříč mluvčími signifikantně liší průměrné hodnoty jednotlivých parametrů, byl využit Wilcoxonův test a při vyhodnocování významnosti rozdílů byla aplikována Bonferroniho korekce – hladina významnosti byla korigována počtem měřených parametrů: $\alpha = 0,05/10 = 0,005$. Byla také sledována míra korelace jednotlivých parametrů.

3.4.2 Výsledky

Jak je patrné z obrázku 3.4, medián a průměr základní frekvence se chovají velmi podobně. U mnoha mluvčích lze pozorovat výraznější rozdíly mezi jejich čtenými a spontánními promluvami než mezi jejich vlastním projevem a projevem jiných mluvčích (nejzřetelněji u mluvčího POLM); tyto parametry se tudíž nejeví jako příliš vhodné pro vzájemné rozlišování jednotlivců. Průměr představuje metriku, která je ve srovnání s mediánem náchylnější k vlivu odlehlých hodnot, bylo by tedy možné očekávat, že bude vykazovat odlišné chování než medián, mimo jiné v důsledku možných oktávových skoků v signálu; tento předpoklad se nicméně u většiny nahrávek nepotvrdil, vzájemné rozdíly mediánu i průměru napříč nahrávkami jsou velmi podobné – jejich korelační koeficient r je (při zaokrouhlení na dvě desetinná místa) roven 1, jak je ukazuje obrázek 3.11.

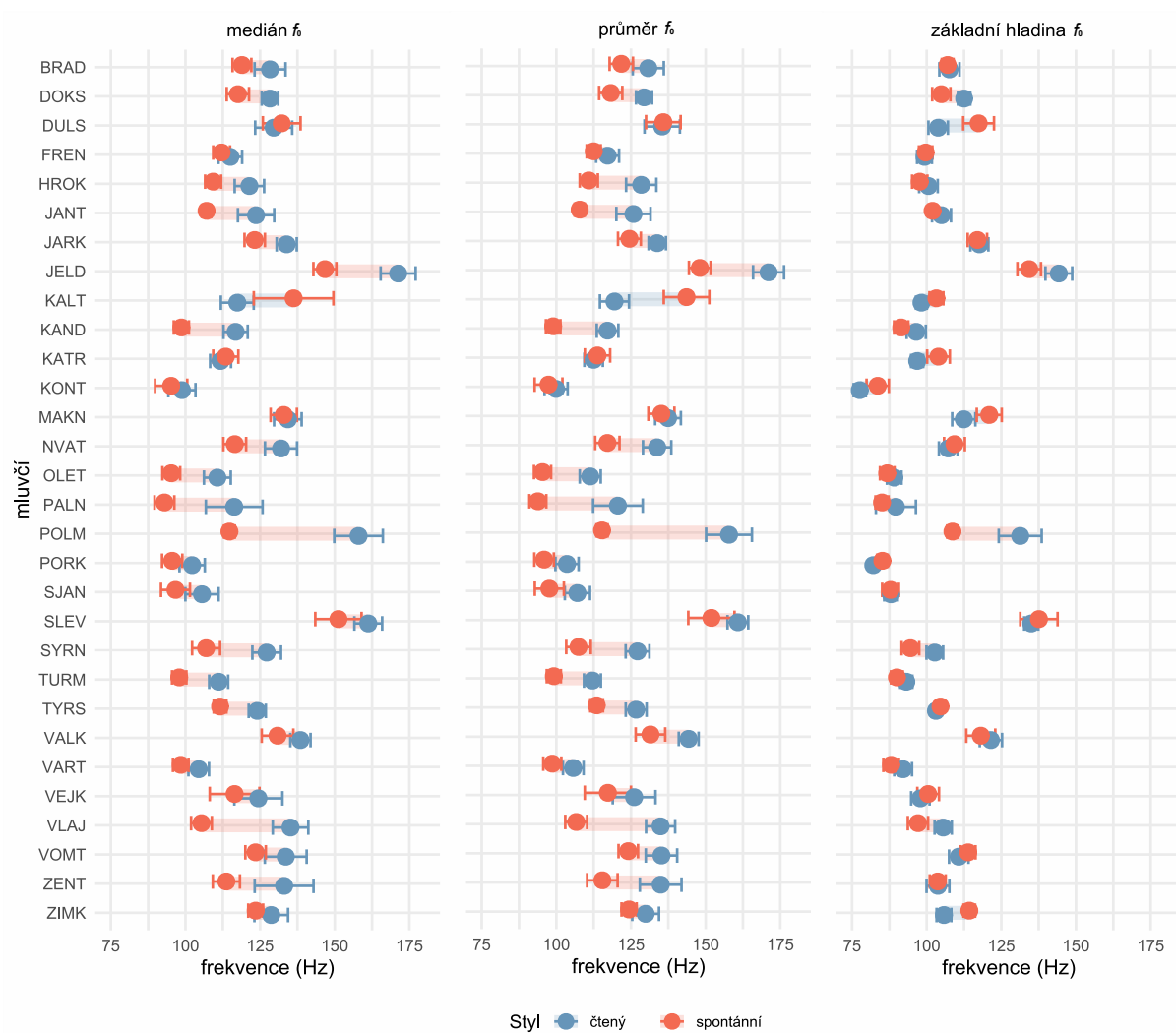
Zde je zapotřebí podotknout, že 95% konfidenční intervaly, které jsou vyznačeny na obrázcích 3.4, 3.6, 3.8 a 3.9, nebyly opatřeny korekcí pro opakovaná měření.

⁸ Za ucelenou intonační jednotku byl v tomto případě považován promluvý úsek, pro celkovou charakterizaci daného vzorku byl tudíž využit aritmetický průměr dílčích hodnot získaných z jednotlivých úseků, nikoli průměr vážený trváním úseků.

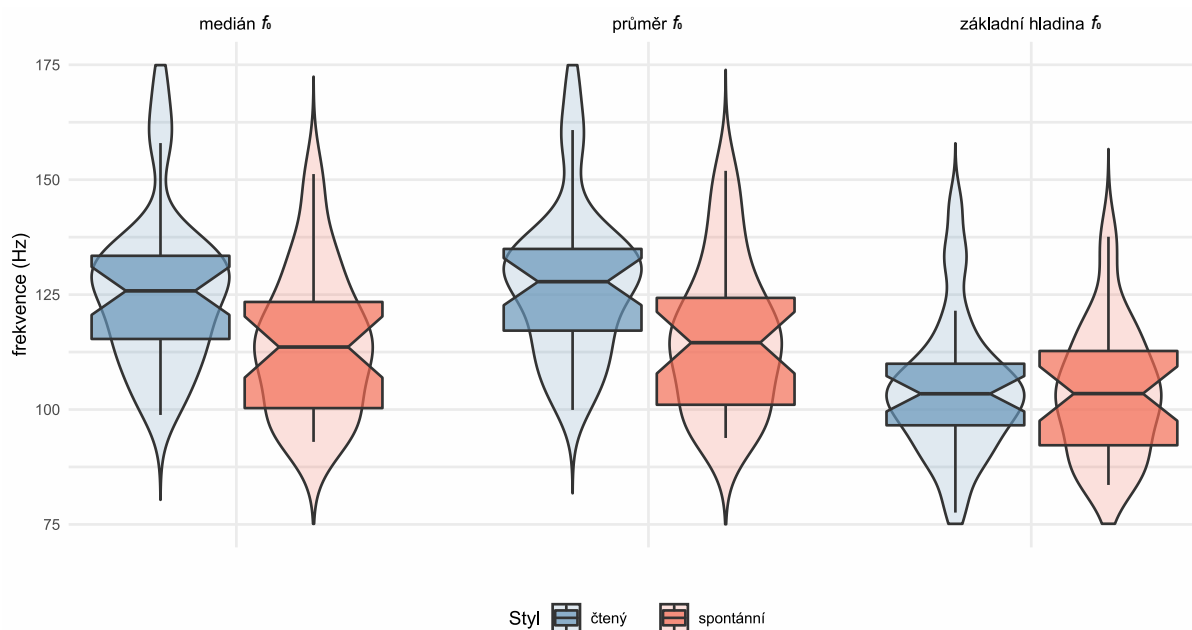
Jedná se o ilustrační zobrazení, na jehož základě nelze vynášet závěry ohledně statistické významnosti rozdílů mezi jednotlivými vzorky.

S průměrem i mediánem velmi silně koreluje také základní hladina f_0 ($r = 0,89$ a $0,91$, v tomto pořadí, viz obrázek 3.11), na obrázku 3.4 si však lze povšimnout, že právě základní hladina zůstává z těchto tří ukazatelů nejstabilnější v rámci projevu mluvčího. Zároveň jsou mezi některými mluvčími patrné výrazné rozdíly v základní hladině f_0 , jedná se tedy o parametr s poměrně nízkou intraindividuální variabilitou a zároveň značnou variabilitou interindividuální.

Celkové rozdíly mezi střední hlasovou frekvencí ve čteném a spontánním projevu shrnuje obrázek 3.5 (detailní informace o naměřených hodnotách těchto i dalších parametrů základní frekvence v obou mluvních stylech pak obsahuje příloha B). Zatímco průměr i medián vykazují signifikantně vyšší hodnoty ve čteném projevu (výsledky Wilcoxonova testu ukazují v obou případech $p < 0,001$), u základní hladiny podobný trend pozorovat nelze ($p > 0,59$) – je u ní však viditelný větší rozptyl hodnot napříč mluvčími ve spontánním projevu oproti projevu čtenému.

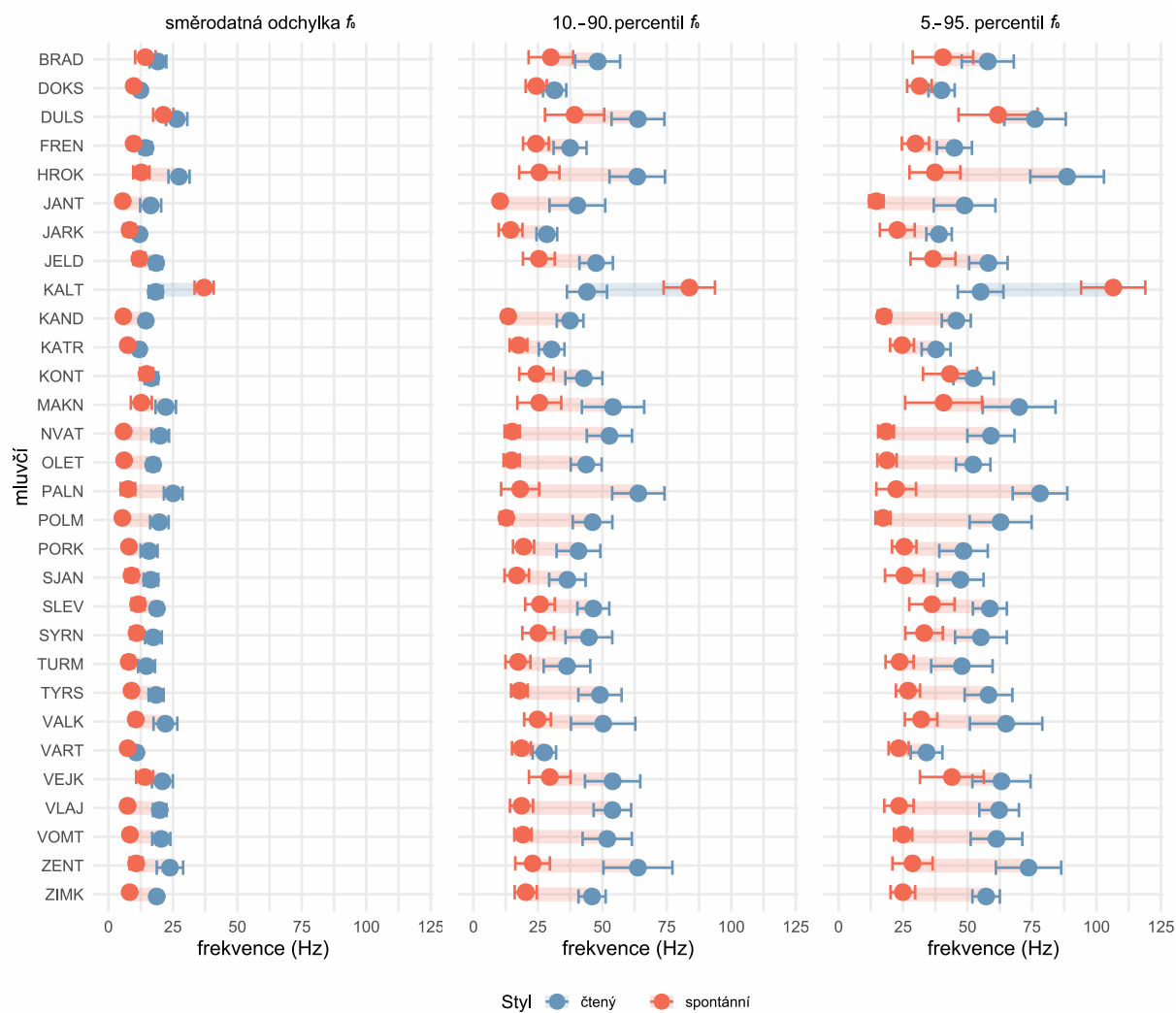


Obrázek 3.4. Medián, průměr a základní hladina f_0 jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluвовých úseků, linky znázorňují 95% konfidenční intervaly.

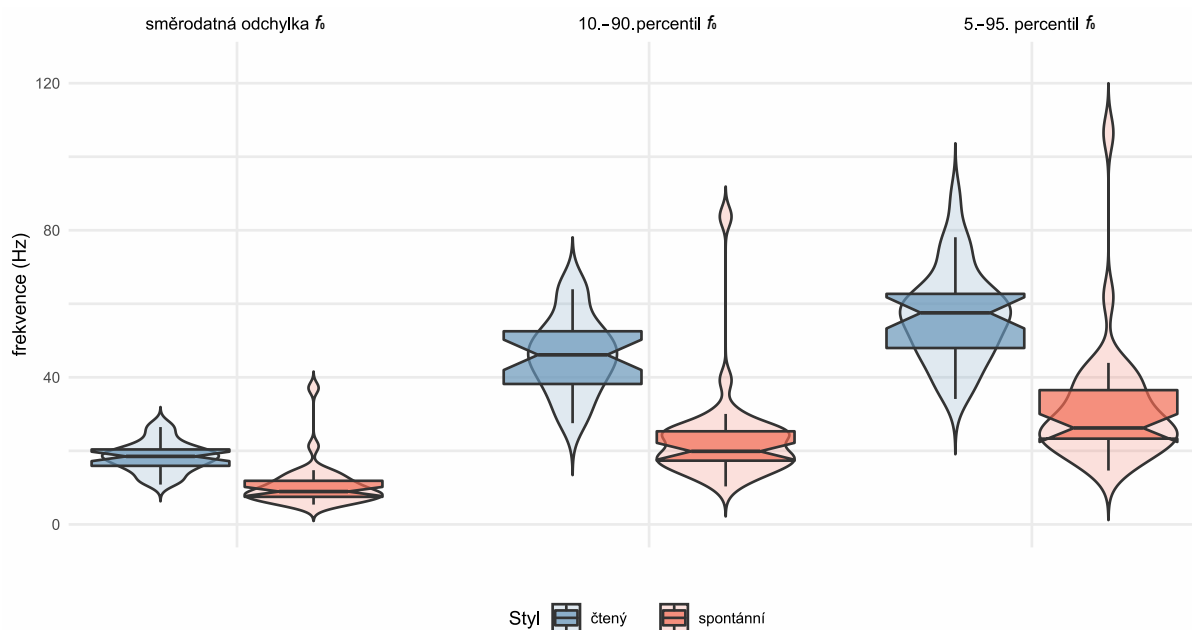


Obrázek 3.5. Srovnání průměrných hodnot mediánu, průměru a základní hladiny f_0 napříč mluvčími ve čteném a spontánním projevu.

Sledované ukazatele variability základní frekvence reflektující její rozptyl u mnoha mluvčích vykazují větší rozdíly intraindividuální než rozdíly oproti jiným mluvčím, jak ilustruje obrázek 3.6. Vzhledem k obdobným východiskům při jejich měření spolu všechny tři parametry velmi silně korelují (korelační koeficient je roven 0,98 až 0,99, viz obrázek 3.11). Obecně lze u všech těchto parametrů sledovat signifikantní nárůst ve čteném projevu ve srovnání s projevem spontánním, jak naznačuje obrázek 3.7 (podle výsledků Wilcoxonova testu je u všech tří parametrů variability f_0 $p < 0,0001$). Jak je patrné z obrázku 3.6, z tohoto trendu vybočuje pouze jeden ze třiceti sledovaných mluvčích, a to KALT.



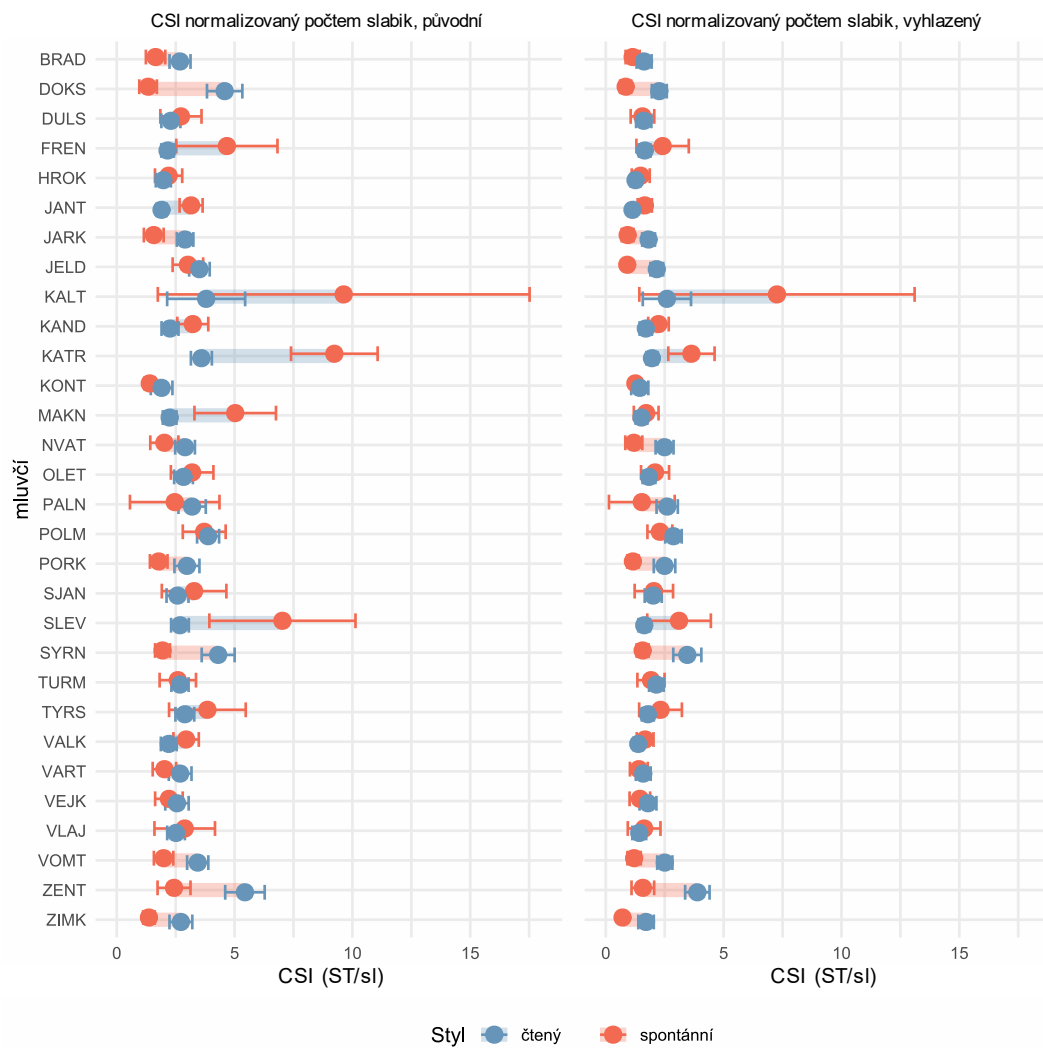
Obrázek 3.6. Směrodatná odchylka, rozpětí mezi 10. a 90. percentilem a rozpětí mezi 5. a 95. percentilem f_0 jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvočných úseků, linky znázorňují 95% konfidenční intervaly.



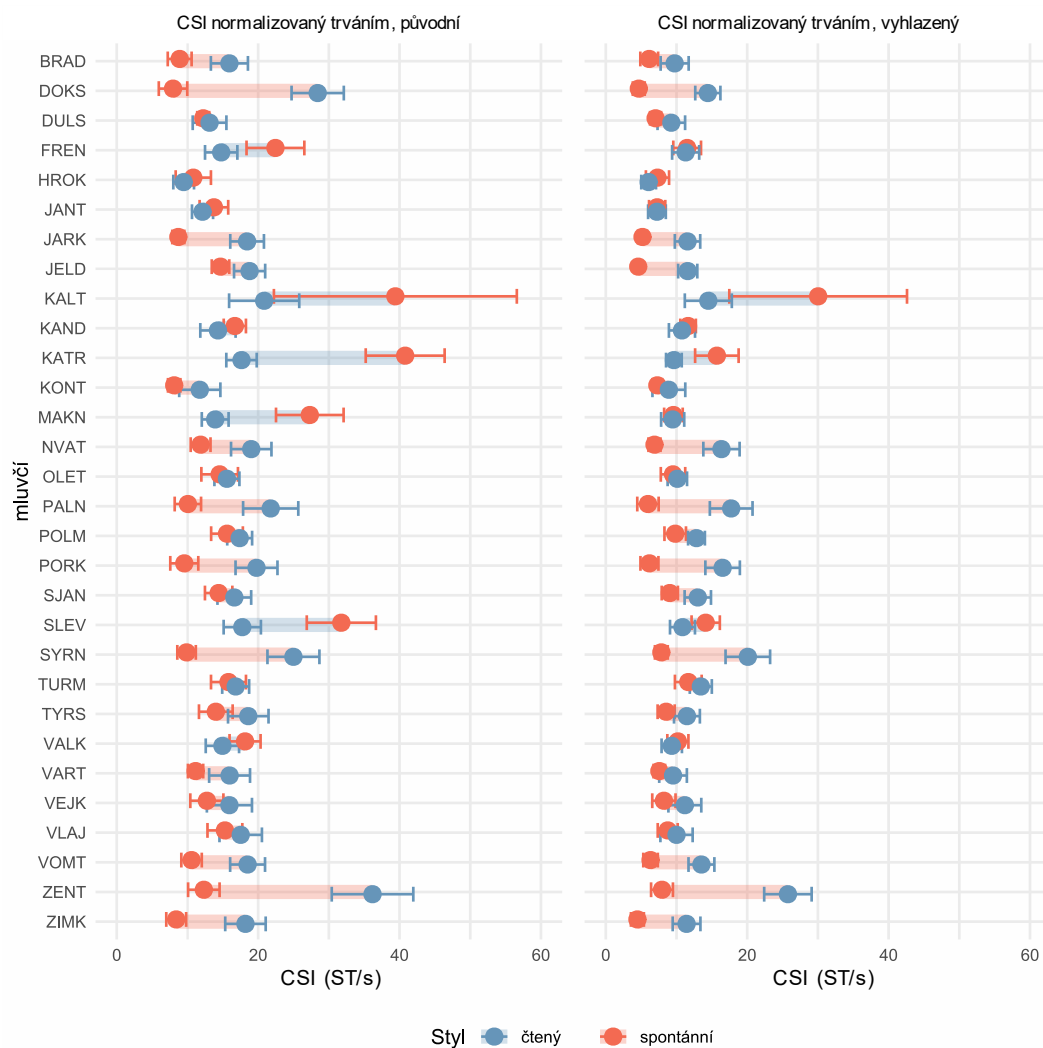
Obrázek 3.7. Srovnání směrodatné odchylky, rozdílu mezi 90. a 10. percentilem a rozdílu mezi 95. a 5. percentilem f_0 napříč mluvčími ve čteném a spontánním projevu.

Chování indexu kumulativní strmosti normalizovaného počtem slabik a trváním, které znázorňují obrázky 3.8 a 3.9, vykazuje obdobné tendence a jeho hodnoty jsou při obou způsobech normalizace velmi silně korelované (korelační koeficient je roven 0,93 u původních a 0,91 u vyhlazených kontur, viz obrázek 3.11). U všech způsobů extrakce CSI základní frekvence jsou zastoupeni jak mluvčí, jejichž hodnoty CSI se markantně liší v závislosti na mluvním stylu (přičemž relativně vyšší hodnoty se objevují ve čteném i ve spontánním projevu), tak ti, u kterých k výrazným posunům nedochází. Obecně lze říci, že u CSI f_0 , pro jehož extrakci byly využity popsání postupy, nelze pozorovat výraznější interindividuální rozdíly ve srovnání s těmi intraindividuálními, nejedná se tedy o parametr, na jehož základě by bylo možné vzájemně odlišit jednotlivce. Při využití popsání postupů při měření CSI také nelze vyloučit přítomnost chybně extrahovaných hodnot, které mohou zkreslovat výsledky – viz např. nápadně široký konfidenční interval u mluvčího KALT, který může být zapříčiněn právě touto skutečností (nicméně u něj zároveň byl pozorován větší rozsah hodnot f_0 ve spontánním projevu, vysoké

hodnoty CSI tak mohou být způsobeny reálným výrazným kolísáním základní frekvence).



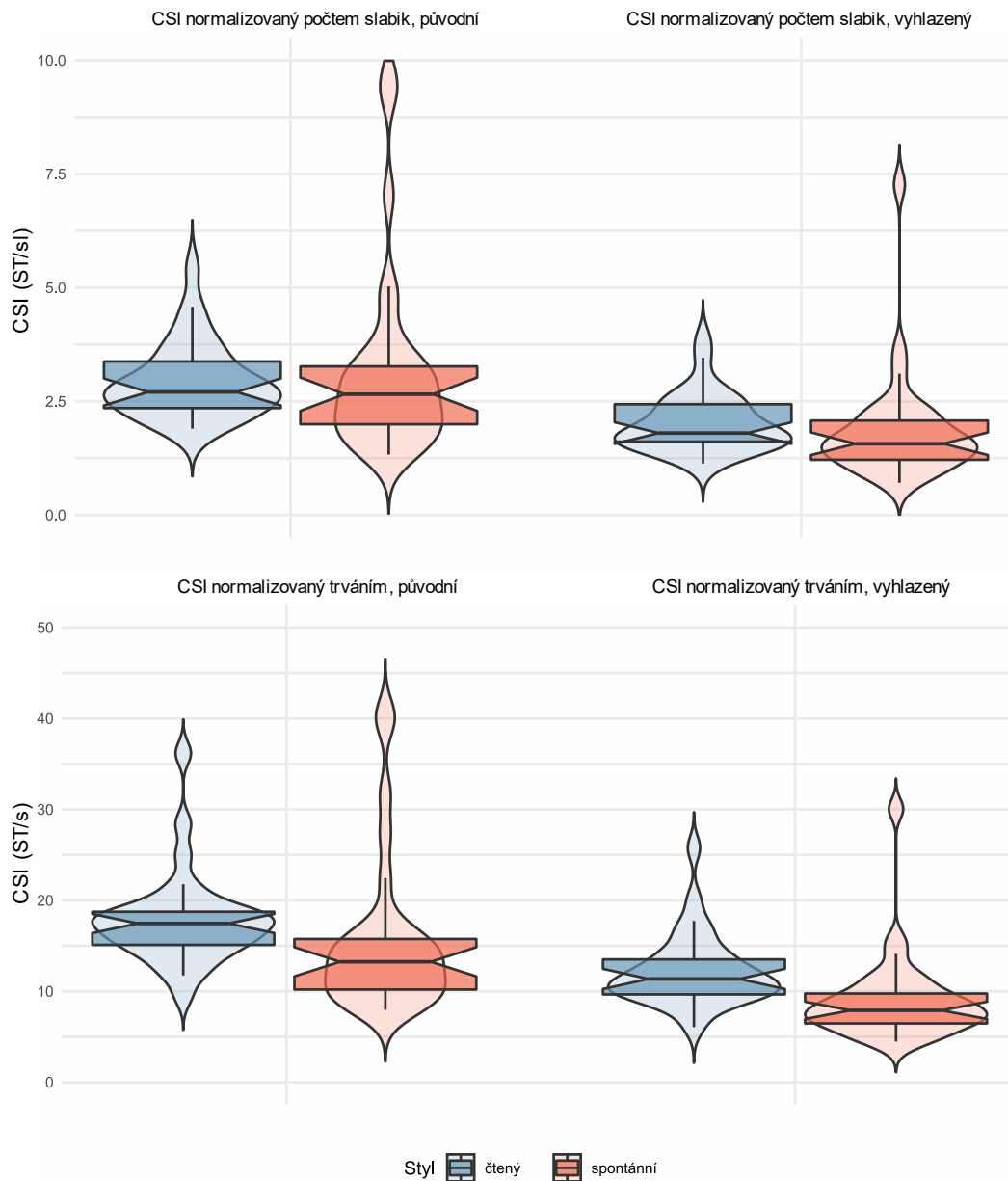
Obrázek 3.8. Index kumulativní strmosti f_0 jednotlivých mluvčích ve čteném a spontánním projevu před vyhlazením kontury f_0 a po něm, normalizovaný počtem slabik. Body reprezentují průměr hodnot z jednotlivých promluvočných úseků, linky znázorňují 95% konfidenční intervaly.



Obrázek 3.9. Index kumulativní strmosti f_0 jednotlivých mluvčích ve čteném a spontánním projevu před vyhlazením kontury f_0 a po něm, normalizovaný trváním. Body reprezentují průměr hodnot z jednotlivých promluvovacích úseků, linky znázorňují 95% konfidenční intervaly.

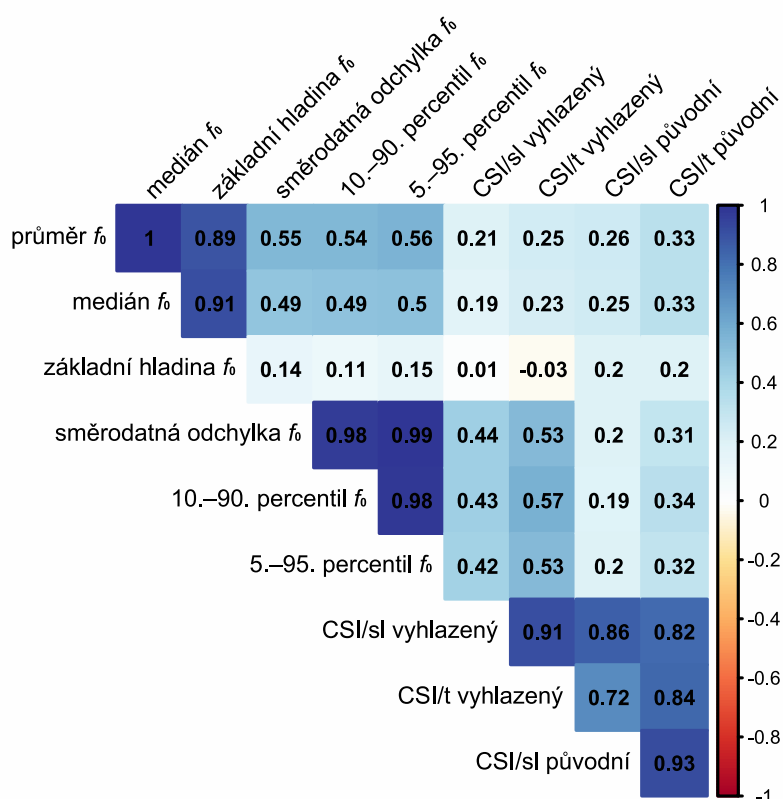
Jakkoliv jsou celkové vzorce CSI za využití normalizace počtem slabik a trváním podobné, u CSI normalizovaného trváním lze napříč mluvčími pozorovat výraznější rozdíly mezi čteným a spontánním projevem. Tyto rozdíly jsou výrazné např. u mluvčího PALN, jehož hodnoty CSI se při normalizaci slabikami podstatně neliší, zatímco při normalizaci trváním jsou výrazně vyšší ve čteném projevu. Průměrné hodnoty napříč mluvčími znázorňuje obrázek 3.10: CSI normalizovaný počtem slabik, extrahovaný z původní i vyhlazené kontury f_0 , nevykazuje v závislosti na mluvním stylu signifikantní rozdíly (výsledkem Wilcoxonova testu jsou

p-hodnoty vyšší než 0,9 a vyšší než 0,2, v tomto pořadí), u CSI normalizovaného trváním si však lze povšimnout tendence k výskytu vyšších hodnot ve čteném projevu. Tento rozdíl byl v rámci Wilcoxonova testu vyhodnocen jako signifikantní u CSI normalizovaného trváním získaného z vyhlazených kontur f_0 ($p = 0,002$), u jeho hodnoty extrahované z kontury původní nikoliv ($p = 0,1$).



Obrázek 3.10. Srovnání indexu kumulativní strmosti f_0 ve čteném a spontánním projevu před vyhlazením kontury f_0 a po něm, normalizovaný slabikami (nahore) a trváním (dole) napříč mluvčími.

Hodnoty CSI získané z vyhlazených kontur základní frekvence jsou relativně stabilnější v rámci projevu jednoho mluvčího než ty, které byly extrahovány z kontur původních; jak bylo možné očekávat, následkem vyhlazení kontury f_0 jsou hodnoty CSI celkově nižší a také méně proměnlivé, což reflektují užší konfidenční intervaly. Korelace mezi hodnotami CSI lišícími se užitím vyhlazené a nevyhlazené kontury f_0 je nižší než mezi těmi, které odlišuje způsob normalizace, nicméně korelační koeficient stále přesahuje 0,8 (viz obrázek 3.11).



Obrázek 3.11. Korelační matice sledovaných parametrů základní frekvence.

3.4.3 Diskuse a závěr

V rámci experimentu 1 byly analyzovány vybrané parametry střední hodnoty a variability základní frekvence: průměr, medián, základní hladina, směrodatná odchylka, rozdíl mezi 90. a 10. percentilem, rozdíl mezi 95. a 5. percentilem a

index kumulativní strmosti. Pozornost byla věnována jejich variabilitě napříč mluvčími a také mezi dvěma mluvními styly – čteným a spontánním.

Coby parametry postihující střední hlasovou frekvenci byly měřeny průměr a medián f_0 ; Lindh (2006) či Skarnitzl a Hývlová (2014) usuzují, že lépe typickou hodnotu základní frekvence reprezentuje medián, který je méně ovlivněn odlehlými hodnotami. Výsledek experimentu nicméně mezi těmito parametry u většiny mluvčích ukázal jen velmi malé rozdíly. Jedním z možných vysvětlení je, že v důsledku individuálního určení maximální a minimální hodnoty základní frekvence při její extrakci ze zvukového signálu podle postupu, který na základě výsledků své studie doporučuje de Looze (2010), obsahovaly výsledné kontury základní frekvence méně odlehlých hodnot vzniklých kvůli chybné extrakci než za využití jednotného nastavení pro všechny mluvčí. U mediánu i průměru f_0 byla pozorována intraindividuální variabilita mezi sledovanými mluvními styly, která u řady mluvčích přesahovala variabilitu interindividuální.

Třetí metrikou odrážející typickou hodnotu základní frekvence byla základní hladina, tedy hranice, pod níž leží 7,64 % hodnot f_0 v daném vzorku. Právě základní hladina se z těchto tří parametrů ukázala jako nejstabilnější pro mluvčího, v souladu se závěry, ke kterým dospěli Lindh a Eriksson (2007), Skarnitzl a Hývlová (2014) či Da Silva et al. (2016). Nadto byly u řady mluvčích pozorovány značné interindividuální rozdíly, přesahující rozdíly intraindividuální; lze tedy konstatovat, že právě tento ukazatel poměrně dobře vzájemně odlišuje různé mluvčí, pakliže u nich lze pozorovat výrazně rozdílné hodnoty, a to nezávisle na mluvním stylu.

Ukazatele variability f_0 odvozené od rozptylu základní frekvence, tedy směrodatná odchylka, rozpětí mezi 10. a 90. percentilem a rozpětí mezi 5. a 95. percentilem, podle očekávání vykazaly velmi silnou vzájemnou korelaci. Skarnitzl a Hývlová (2014) se přiklánějí spíše k využití rozpětí mezi 10. a 90. percentilem než celkového rozpětí vzhledem k možnému výskytu chybně detekovaných

odlehlych hodnot, které jsou při odstranění 10 % nejnižších a 10 % nejvyšších extrahovaných hodnot eliminovány. V provedeném experimentu figurovalo také rozpětí mezi 5. a 95. percentilem a na základě jeho výsledků je možné pozorovat, nakolik se tyto metriky vzájemně odlišují. Prezentované výsledky ukázaly velmi silnou korelaci směrodatné odchylky (tj. odmocniny celkového rozptylu) i obou percentilových rozpětí f_0 , množství odlehlych hodnot v extrahovaných konturách základní frekvence tudíž patrně nebylo natolik výrazné, aby zřetelně zkreslilo výsledné hodnoty sledovaných parametrů. Nelze tedy vyvozovat, že by při parametrizaci rozsahu základní frekvence bylo možné v zájmu odstranění chybně detekovaných odlehlych hodnot doporučit oříznutí extrahovaných hodnot f_0 mimo některé ze sledovaných percentilových rozpětí. Také tento závěr patrně může naznačovat, že postup při extrakci kontury f_0 podle de Looze (2010) vykazuje relativně nízkou chybovost, resp. produkuje poměrně nízké množství odlehlych hodnot včetně oktávových skoků. Výsledky tohoto experimentu tudíž podporují využití tohoto postupu při extrakci kontur základní frekvence.

Jmenované ukazatele variability základní frekvence vykázaly značnou intraindividuální variabilitu v závislosti na mluveném projevu; u všech tří byly zjištěny signifikantně vyšší hodnoty ve čteném textu než ve spontánním hovoru. Při konfrontaci této skutečnosti s výsledky analýzy ukazatelů střední hodnoty základní frekvence lze pozorovat následující trend: v projevech s větším rozptylem základní frekvence se obecně zvyšují také její průměr a medián, základní hladina, která se blíží spíše spodní hranici rozsahu hlasové frekvence, však zůstává bez signifikantních změn. Při využívání vyššího rozsahu f_0 se u mluvčích tudíž má tendenci zvyšovat horní hranice jejich rozsahu, zatímco hranice spodní zřetelných změn nedoznává. Tento poznatek koresponduje s předpokladem, který ve své studii zmiňují Lindh a Eriksson (2007).

Poměrně nově využívaným ukazatelem variability základní frekvence, který byl také sledován v tomto experimentu, je index kumulativní strmosti. Vzhledem

k jeho vysoké náchylnosti ke zkreslení vlivem chybně detekovaných odlehlých hodnot bylo před jeho extrakcí provedeno odstranění hodnot f_0 uvnitř obstruentů, ve kterých bývá základní frekvence častěji chybně detekována, a to i tehdy, když v nich není přítomna, a bylo provedeno také automatické vyhlazení kontury f_0 – CSI pak byl změřen v jednotlivých promluvových úsecích jak za využití kontur před tímto vyhlazením, tak po něm, aby bylo možné provést srovnání výsledných hodnot. CSI pak byl normalizován slabikami a trváním; výsledné hodnoty získané pomocí těchto dvou způsobů normalizace velmi silně korelovaly, nicméně zatímco u CSI normalizovaného počtem slabik nebyl pozorován signifikantní rozdíl mezi sledovanými mluvními styly, při normalizaci trváním vykázal tendenci k vyšším hodnotám ve čteném projevu, přičemž jako signifikantní byl tento rozdíl vyhodnocen u hodnot CSI extrahovaných z vyhlazených kontur f_0 . Tato diskrepance mezi popsányými způsoby normalizace souvisí s rozdíly v artikulačním tempu; čím vyšší je artikulační tempo, tím je hodnota CSI normalizovaného slabikami (měřeného ve stejném časovém úseku) nižší. Při srovnání průměrných hodnot artikulačního tempa napříč mluvčími se ukázalo, že ačkoli ve čteném projevu jeho hodnoty nebyly na dané hladině významnosti vyhodnoceny jako signifikantně vyšší, určitou tendenci k tomuto rozdílu přesto bylo možné zaznamenat (více viz kapitolu 6, oddíl 6.4). Zatímco při normalizaci CSI trváním se tedy objevují vyšší hodnoty ve čteném projevu, v důsledku popsaného trendu u artikulačního tempa při normalizaci počtem slabik zřejmě dochází k setření těchto rozdílů. Jak je patrné, způsob normalizace má výrazný dopad na výsledné hodnoty CSI a jejich vzájemný poměr mezi řečovými vzorky; v rámci budoucího výzkumu by tedy mohlo být přínosné zaměřit se na otázku, který způsob normalizace přináší hodnoty CSI stabilnější pro mluvčího v různých mluvních stylech, přičemž nejspolehlivější výsledky by poskytla analýza s využitím manuálně korigovaných kontur základní frekvence (viz níže).

Hodnoty CSI získané před vyhlazením kontur f_0 a po něm vykázaly vysokou korelaci, nicméně u některých mluvčích bylo při aplikaci těchto dvou postupů možné

pozorovat markantní odlišnosti; po korekci kontur základní frekvence u nich pak byly pozorovány menší rozdíly v CSI ve sledovaných mluvních stylech. Tento výsledek by mohl napovídat, že po odstranění oktávových skoků (tj. po eliminaci vlivu náhodně se vyskytujících odlehlých hodnot) je CSI daného mluvčího méně proměnlivé. Jako potenciálně přínosné se jeví otestování dalších metod odstraňování oktávových skoků a sledování, zda bude jejich výsledkem stabilizace CSI v rámci projevu jednotlivce nezávisle na mluvním stylu. Je ovšem na místě upozornit, že při automatických opravách kontur f_0 hrozí, že veškeré chyby se přesto nepodaří eliminovat, zatímco mohou být zanedbány některé její skutečné pohyby, případně že dojde k jejímu celkovému zkreslení. Postupem, jehož výsledkem by byly nejpřesnější hodnoty CSI, je tudíž manuální korekce kontur základní frekvence.

Podobně jako v případě ukazatelů variability základní frekvence vycházejících z jejího rozptylu, také u CSI normalizovaného trváním získaného z vyhlazených kontur f_0 byly signifikantně vyšší hodnoty pozorovány ve čteném projevu ve srovnání s projevem spontánním (i u hodnot extrahovaných z nevyhlazených kontur základní frekvence byla pozorována podobná tendence, nicméně nebyla vyhodnocena jako statisticky signifikantní). Čtený projev se tudíž vyznačoval nejen větším rozsahem f_0 , který mluvčí využívali, ale také jejím výraznějším kolísáním. Je však potřeba podotknout, že tyto výsledky nejsou zobecnitelné pro čtený a spontánní projev všeobecně; akustickou podobu řečového signálu může ovlivňovat řada dalších faktorů, jakými jsou např. obsah promluv. Pro získání přesnějších informací o chování sledovaných parametrů v rámci jednotlivých mluvních stylů by bylo prospěšné v rámci každého stylu využít materiál z více řečových vzorků.

4 Vokalické formanty

Vokalické formanty⁹, tj. rezonanční frekvence vokálního traktu, se tradičně považují za jeden z parametrů, které nejlépe charakterizují jednotlivce (Hollien, 2002), neboť do jisté míry odrážejí jak jeho fyziologické vlastnosti a individuální osvojené artikulační strategie, tak i různé faktory sociolingvistické povahy (regionální původ, věk aj.). Hodnoty F_1 , F_2 a částečně také F_3 , respektive zejména jejich vzájemná poloha, závisí především na kvalitě vokálu; F_1 roste s větší otevřeností vokálu, F_2 je tím vyšší, čím je artikulace vokálu přednější. Hodnoty nižších formantů jsou tedy v rámci projevu jednotlivce proměnlivé; přesto však podoba vzorců F_1 – F_3 do značné míry odráží i jeho individualitu a napříč mluvčími se různí. Vyšší formanty jsou spíše charakteristické pro daného mluvčího a zůstávají v rámci jeho projevu relativně stabilní (Reetz & Jongman, 2009: 184, citováno v: Skarnitzl et al., 2014: 21).

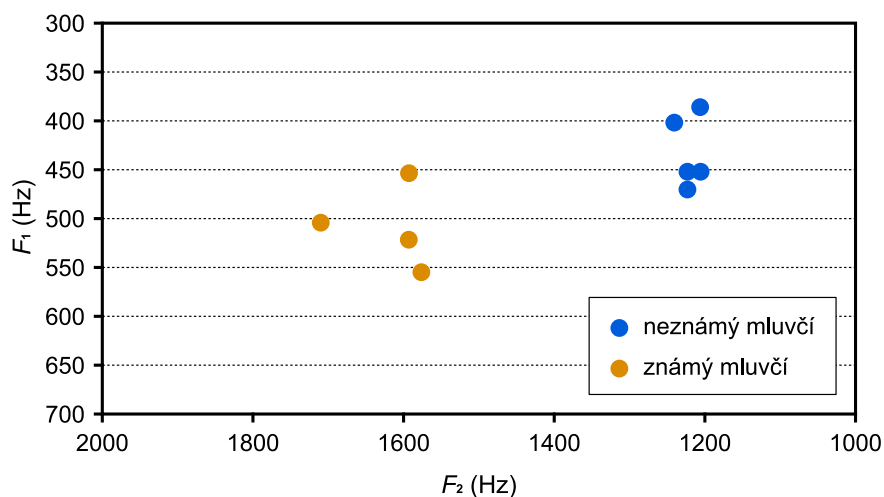
4.1 Hodnoty formantů v jednotlivých vokálech

Jednou z nejčastějších a také zřejmě nejjednodušších metod analýzy vokalických formantů je sledování jejich hodnot v jednotlivých vokálech, tj. jeden vokál je charakterizován jednou hodnotou každého formantu. Postupy při získávání hodnot formantů pro jednotlivé vokály se různí; jednou možností je extrahovat hodnotu z jednoho bodu v průběhu vokálu, a to z jeho temporálního středu či artikulačního cíle, tj. místa, kde formanty dosáhly stabilního průběhu – jak ale

⁹ Pro vokalické formanty se užívá značení pomocí velkého „F“ a číslovky označující pořadí. Tato číslovka bývá v některých dostupných studiích v úrovni běžného textu, v jiných v dolním indexu. Titze et al. (2015) navrhli sjednocení značení některých proměnných, mezi nimi i vokalických formantů, u kterých je „F“ psáno kurzívou a číslovka v dolním indexu. V této práci bude v souladu s jejich doporučením využíváno označování formantů s číslovkou v dolním indexu, nicméně velké „F“ bude v zájmu zachování konzistence se značením dalších parametrů (např. dlouhodobých formantových hodnot či harmonických složek řečového signálu) psáno bez kurzívy.

upozorňuje Rose (2015), takový bod, ve kterém by všechny sledované formanty dosáhly stabilního průběhu, nemusí být možné v rámci vokálu nalézt, neboť jednotlivé formanty nemusejí očekávaných vrcholů dosahovat zároveň. Druhou možnost představuje stanovení hodnoty formantu na základě více bodů v průběhu vokálu.

Cavalcanti et al. (2021) analyzovali průměrné hodnoty vokalických formantů F_1 – F_4 a sledovali, nakolik se liší napříč jednotlivci. Hodnoty formantů byly v rámci této studie extrahovány z bodů uprostřed trvání jednotlivých vokálů. Jako parametr, kterým se od sebe jednotliví mluvčí nejvýrazněji lišili, se ukázal F_4 , následoval jej F_3 . Hodnoty formantů z temporálního středu vokálů extrahovali také Nolan a Grigoras (2005) při srovnávání řečových vzorků a určování, zda se jedná o téhož mluvčího v rámci reálného forenzního případu; analyzovali však pouze F_1 a F_2 (tj. formanty, jejichž hodnoty se odvíjejí od vokalické kvality), a to v určitých vybraných vokálech. Při srovnávání daných řečových vzorků se ukázalo, že hodnoty F_1 a F_2 ve sledovaných vokálech se liší natolik, že je možné s vysokou pravděpodobností vyloučit, že se jedná o téhož mluvčího – výsledky jejich měření ukazuje obrázek 4.1; v tomto konkrétním případě se tedy hodnoty F_1 a F_2 extrahované z temporálního středu vokálu ukázaly jako spolehlivý parametr pro rozlišení dvou mluvčích. (Jednalo se však o případovou studii, jejíž součástí nebylo vyhodnocení, nakolik je tento parametr idiosynkratický všeobecně.)



Obrázek 4.1. Upraveno podle: Nolan & Grigoras (2005: 147). Hodnoty F_1 a F_2 v realizacích vokálu /i/ naměřené v nahrávkách neznámého a známého mluvčího.

K jinému způsobu extrakce hodnot vokalických formantů přistoupili Fletcher et al. (2015); předpokládali, že artikulačního cíle může být dosahováno v jiné fázi trvání vokálu než v jeho temporálním středu. Hodnoty formantů tedy sledovali nejen v bodě uprostřed daného vokálu, ale také v artikulačním cíli, tj. tam, kde artikulace dosáhla stabilního průběhu – za artikulační cíl bylo považováno místo mezi 20 a 80 % trvání vokálu, ve kterém formanty nabyly krajních hodnot: u předního zavřeného vokálu [i:] byl sledován vrchol F_2 , pro otevřený vokál [e:] vrchol F_1 a pro zadní [o:] největší propad F_2 . (Tato studie využívala materiál od mluvčích australské angličtiny.) Výsledky studie v souladu s jejich předpokladem ukázaly signifikantní rozdíly mezi hodnotami formantů extrahovanými z temporálního středu vokálu a z místa dosažení artikulačního cíle. Podobný přístup uplatnili také Eichhorn et al. (2018), kteří na základě svých pozorování uvádějí, že „temporální střed vokálu se ne vždy shoduje s místem stálého průběhu formantů“ (Eichhorn et al., 2018: 644.e2). Ve svém experimentu měřili hodnoty prvních čtyř formantů vokálů v americké angličtině: u vokálu [i] bylo extrakčním bodem místo vrcholu F_2 (stejně jako v případě studie Fletcher et al., 2015), u [u] největší propad F_2 (analogicky k zadnímu vokálu [o:] v předchozí studii), u /a/

pak místo největšího přiblížení F_1 a F_2 a u /æ/ místo nejrovnoměrnějšího rozložení formantů ve spektru.

Skarnitzl a Volín (2012) měřili hodnoty prvních dvou formantů u mluvcích standardní češtiny pro referenční účely. Hodnota formantu byla vypočtena jakožto aritmetický průměr ze sedmi ekvidistantních bodů v prostřední třetině průběhu daného vokálu. Jejich studie předkládá referenční hodnoty vokalických formantů dospělých českých mluvcích mužského pohlaví, formanty mluvcích ženského pohlaví jsou pak podle závěrů studie přibližně o 15 až 20 % vyšší. Tato studie není zaměřena na to, nakolik jsou F_1 a F_2 idiosynkratické, nýbrž na všeobecnou podobu vokalického systému češtiny. Populační statistiky však, jak upozorňují např. Skarnitzl a Vaňková (2017), do značné míry napomáhají určení, nakolik podobnost vokalických formantů (jakož i jiných parametrů) mezi dvěma vzorky nasvědčuje tomu, že se jedná o téhož mluvčího. Kromě míry podobnosti je totiž na základě populačních statistik možné určit také míru typičnosti dané hodnoty.

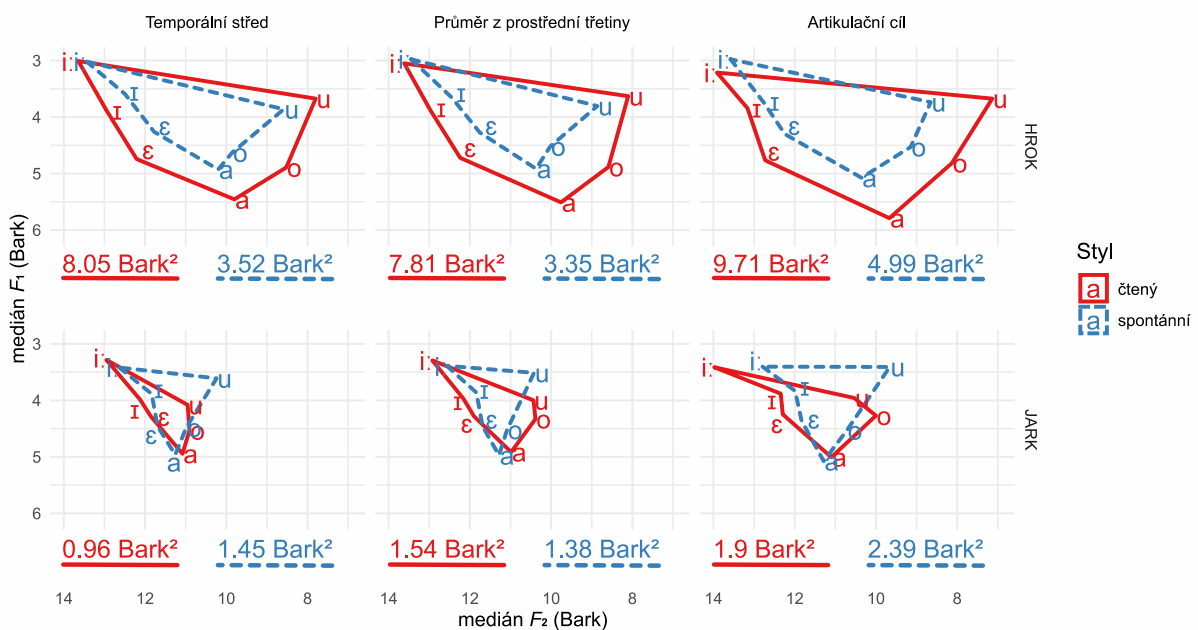
Různé způsoby extrakce vokalických formantů porovnali Houzar a Skarnitzl (2022), kteří měřili F_1 a F_2 u českých monoftongů. Hodnoty formantů měřili pomocí tří různých metod: v temporálním středu vokálu, jakožto průměr z prostřední třetiny vokálu a v akusticky definovaném artikulačním cíli mezi 20 a 80 % trvání vokálu, přičemž replikovali postup, který využili Fletcher et al. (2015): u předních vokálů /i i: ε ε:/ byl za artikulační cíl považován vrchol F_2 , u zadních vokálů /u u: o o:/ naopak největší propad F_2 a u otevřených vokálů /a a:/ vrchol F_1 . Na základě výsledků svého experimentu uvádějí, že zatímco výsledky měření formantů v temporálním středu a v prostřední třetině vokálu se podstatně neliší, extrakcí v (akusticky definovaném) artikulačním cíli jsou získávány výrazně extrémnější hodnoty poukazující na perifernější artikulaci, a upozorňují na možnost, že algoritmus, který je nastaven tak, aby identifikoval místa vrcholů a propadů vokalických formantů, může inklinovat k extrémním hodnotám, objevujícím se v důsledku chybné identifikace formantových hodnot daným softwarem.

Podotýkají také, že zejména ve spontánním projevu jsou jednotlivé vokály často realizovány redukovane či jinak odchylně od kanonické podoby, a tudíž pro ně extrakce formantů pomocí takto sestaveného algoritmu nemusí být zcela vhodná.

Jednu z metod analýzy hodnot vokalických formantů v jednotlivých vokálech představuje zobrazení velikosti vokalického prostoru (*vowel space area; VSA*), tj. rozsahu, ve kterém se pohybují hodnoty formantů F_1 a F_2 , což koreluje s artikulačním nastavením; při centralizovanější artikulaci vokálů bude vokalický prostor menší, při hyperartikulaci naopak rozsáhlejší. Velikost vokalického prostoru bývá typicky vymezena hodnotami formantů u krajních vokálů (tj. takových, které mají nejvíce otevřenou/zavřenou a nejvíce přední/zadní artikulaci), a to minimálně tří, nicméně lze zahrnout i více vokalických kvalit. (Např. v češtině by pro změření velikosti vokalického prostoru bylo třeba využít vokály [i:, a/a:, u/u:], ale bylo by možné zahrnout i [ɪ, ε/ε:, o/o:]). Hodnoty vokalických formantů charakterizující velikost vokalického prostoru lze extrahovat právě pomocí výše zmíněných postupů.

Analýze velikosti vokalického prostoru se věnovali např. Fletcher et al. (2015; viz výše), kteří zkoumali závislost velikosti vokalického prostoru na věku a tempu řeči. Tito autoři, jak již bylo uvedeno, měřili hodnoty formantů jednak v temporálním středu vokálu a jednak v artikulačním cíli, a to ve třech vokálech australské angličtiny ([e:, i:, o:]). Následně sestavili dvě varianty vokalického prostoru: jednu na základě hodnot formantů extrahovaných v temporálním středu vokálů (pro každý vokál byla stanovena průměrná hodnota F_1 a F_2), druhou podle jejich hodnot v artikulačním cíli. Podle očekávání autorů byl obecně rozsah vokalického prostoru naměřený na základě hodnot formantů v artikulačními cíli vokálů signifikantně větší. Velikostí vokalického prostoru se zabývali také Weirich a Simpson (2013), kteří sledovali její souvislost se základní frekvencí; ve své studii se primárně věnovali otázce, proč se u žen (i přes využití psychoakustické jednotky Bark) obecně objevuje větší vokalický prostor než u mužů, resp. zda jeho velikost

koreluje se střední hodnotou základní frekvence (té se podrobněji věnuje kapitola 3, oddíl 3.1); hodnoty vokalických formantů byly extrahovány z artikulačních cílů pěti sledovaných německých vokálů ([i:, ε, a:, o, u:]), pro každý vokál byly následně vypočítány průměrné hodnoty F_1 a F_2 a na jejich základě byla následně vyčíslena velikost vokalického prostoru. Idiosynkratičnosti tohoto parametru se věnovali Houzar a Skarnitzl (2022), kteří stanovili hodnotu VSA jakožto velikost mnohoúhelníku vytyčeného mediány prvních dvou formantů jednotlivých českých monoftongů ve dvourozměrném prostoru, jehož dimenze tvoří F_1 a F_2 , a sledovali závislost výsledných hodnot na mluvčím a mluvním stylu; jejich výsledky, které ilustruje obrázek 4.2, poukazují na značnou variabilitu VSA mezi mluvčími, ale i v rámci mluvčího mezi dvěma sledovanými mluvními styly.



Obrázek 4.2. Převzato z: Houzar & Skarnitzl (2022: 90). Velikost vokalického prostoru ve čteném a spontánním projevu získaná za využití tří metod extrakce vokalických formantů (v temporálním středu vokálu, jakožto průměr z prostřední třetiny vokálu a v artikulačním cíli vokálu) u dvou mluvčích (ilustrační výňatek; studie obsahuje zobrazení VSA u 4 mluvčích).

4.2 Dlouhodobé formantové distribuce

Další možností využití vokalických formantů jsou dlouhodobé formantové distribuce (*long-term formant distributions*; LTF), které ve své studii představili Nolan a Grigoras (2005). Tento ukazatel odráží, jakých hodnot nejčastěji dosahují frekvence každého sledovaného formantu v celém řečovém signálu, nezávisle na kvalitě vokálu. Pozice vrcholů v distribuci frekvencí v rámci jednotlivých formantů odrážejí jak rozměry vokálního traktu mluvčího, tak také jeho artikulační návyky, např. tendenci k palatalizaci nebo zaokrouhlování rtů (Nolan & Grigoras, 2005). Analýzu dlouhodobých formantových distribucí v současnosti využívá i řada dalších studií, způsoby jejich extrakce se však různí (viz níže).

Nolan a Grigoras (2005) nejprve vyčlenili pouze znělé intervaly z řečového signálu, v nichž následně automaticky detekovali hodnoty formantů F_1 – F_4 , a to v průběhu celého vzorku¹⁰. Výsledné distribuce byly získány výpočtem, kolikrát byly jednotlivé frekvence detekovány jako hodnoty některého formantu. Podle autorů zmíněné studie se jedná o parametr, který má výrazný potenciál pro rozlišování mluvčích, a na základě výsledků svého výzkumu konstatují, že ačkoliv vokalické formanty do značné míry v rámci projevu jednotlivce variiují a mohou podléhat zkreslení vlivem nahrávacích podmínek, jsou vůči obojímu výrazně odolnější než parametry, které odrážejí činnost hlasivek, jakými jsou např. střední hodnoty základní frekvence nebo parametry sklonu dlouhodobého průměrného spektra.

Skarnitzl et al. (2014) provedli analýzu dlouhodobých formantových distribucí na českém řečovém materiálu. Pro účely analýzy vyčlenili pouze řetězce vokálů, nenazálních sonor [j l r] a vokalických hezitačních zvuků, přičemž hranice segmentů

¹⁰ Ve zmiňované studii byly využívány řečové vzorky v angličtině. Vzhledem k tomu, že fonologicky znělé obstruenty bývají v angličtině na rozdíl od češtiny obvykle realizovány jako foneticky neznělé, byly patrně v analyzovaném signálu zahrnuty pouze vokály a sonory, tj. hlásky, ve kterých jsou typicky přítomny vokalické formanty. Při obdobném postupu využívajícím český materiál by bylo vhodné pro účely analýzy vyřadit nejen neznělé intervaly, ale také znělé obstruenty.

byly manuálně zarovnány. Z takto získaného vzorku byly v 10ms intervalech extrahovány formanty F_1 – F_3 . Autoři této studie zjišťovali, nejen nakolik se navzájem liší dlouhodobé formantové distribuce různých mluvčích, ale i do jaké míry jsou proměnlivé v rámci mluvčího. Podle jejich závěrů jsou formantové distribuce intraindividuálně stabilní; pro sledování proměnlivosti LTF v rámci daného mluvčího však využili pouze jednu nahrávku jeho projevu, je otázkou, nakolik by LTF zůstávaly stabilní i při srovnávání různých nahrávek. Zároveň byly LTF významně odlišné napříč mluvčími, autoři však upozorňují, že jsou značně náchylné ke zkreslení vlivem okolního šumu, a to navíc nikoli konzistentně. Obecně se jako odolnější vůči zkreslení ukázaly vyšší formanty. Autoři uzavírají, že jako nejrobustnější vůči zkreslení a zároveň nejužitečnější pro rozpoznání mluvčího se jeví LTF_3 .

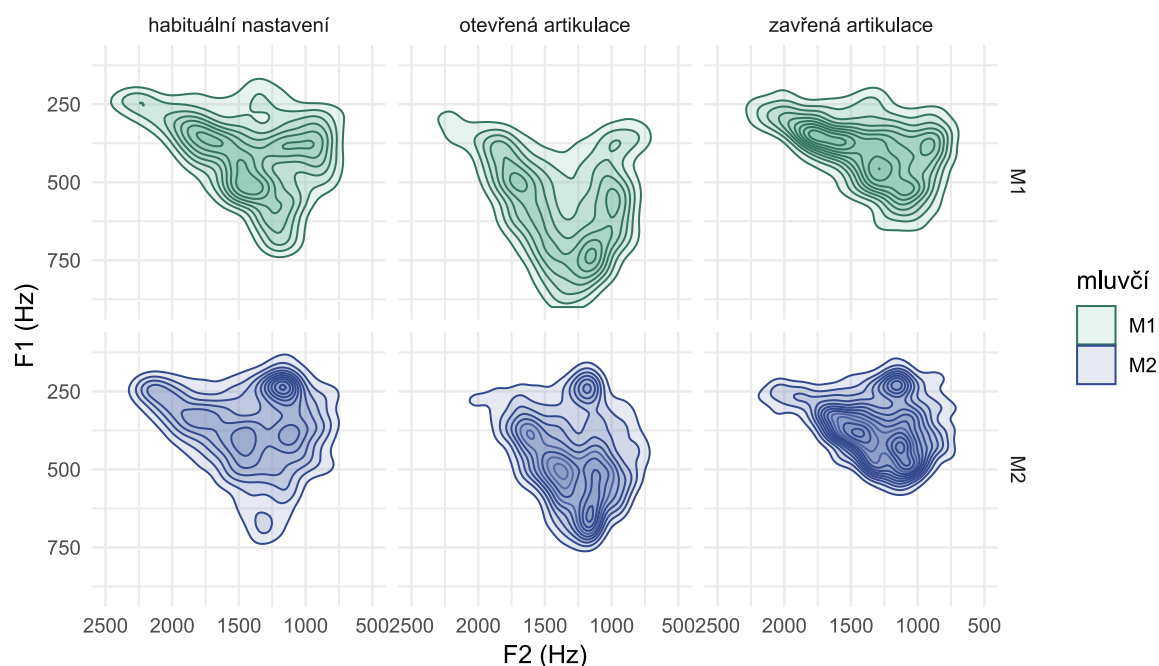
Podobně jako je tomu u zobrazení hodnot F_1 a F_2 jednotlivých vokálů pomocí velikosti vokalického prostoru, existuje možnost prostorového vynesení hodnot i u LTF_1 a LTF_2 . Jedná se o zobrazení hustoty vokalického prostoru (*vowel space density*; VSD), které zachycuje, nejen jaké vokalické kvality lze v rámci daného řečového projevu nalézt, ale také jak jsou časté. Tuto metodu ve své studii představili Story a Bunton (2017). Zatímco výpočet velikosti vokalického prostoru obvykle vychází z hodnot formantů v konkrétních cílových hláskách a ukazuje tudíž jakési mezní hodnoty, hustota vokalického prostoru poskytuje informace o vokalické produkci během promluvy celkově, neboť hodnoty formantů jsou extrahovány nezávisle na kvalitě jednotlivých vokálů, a to v celém jejich průběhu (nikoli pouze v prostřední části či v artikulačním cíli). Ve dvourozměrném prostoru určeném dimenzemi F_1 a F_2 jsou vyneseny body definované extrahovanými hodnotami vokalických formantů. Třetí rozměr vokalického prostoru pak představuje právě hustota těchto bodů.

Story a Bunton (2017) pro tvorbu zobrazení VSD využili automaticky vyčleněné znělé úseky¹¹ a následnou detekci vokalických formantů. Formanty byly extrahovány v intervalech 5 milisekund. Pro účely této metody není zapotřebí segmentace na jednotlivé hlásky, v této studii nebyly ani manuálně korigovány hranice znělých intervalů či extrahovaných formantových hodnot. Autoři se ve své publikaci zaměřovali na rozdíl hustoty vokalického prostoru ve třech typech projevu jednoho mluvčího: (1) opakování určitých slabik, (2) hyperartikulace a (3) hypoartikulace. Slabiky, které mluvčí opakoval v rámci prvního typu projevu, obsahovaly čtyři krajní vokály: /i æ a u/; hustota vokalického prostoru získaná na základě těchto vokálů představuje reprezentaci určitých extrémních hodnot vokalických formantů, kterých může daný mluvčí dosahovat. Graf VSD získaný na základě hyperartikulovaného projevu se podobal čtyřem krajním vokálům; oblasti s vysokou hustotou se nacházely blíže hranicím vokalického prostoru. Naproti tomu v případě hypoartikulace byla hustota vokalického prostoru nejvyšší v oblasti mediánu sledovaných formantů. Jak se tedy ukázalo, může se hustota vokalického prostoru v rámci projevu jednoho mluvčího značně lišit v závislosti na typu projevu, intraindividuální variabilita VSD je tedy výrazná. Vzhledem k tomu, že tato studie zahrnovala řečový materiál jediného mluvčího, není možné usuzovat, nakolik je VSD variabilní interindividuálně.

Zobrazení hustoty vokalického prostoru u českých mluvčích využili Houzar et al. (2022), kteří sledovali posuny vokalických formantů při cílených manipulacích artikulačního nastavení. Analyzovaný materiál tvořily nahrávky pěti mluvčích, kteří předčítali krátký text nejprve se svým habituálním artikulačním nastavením a poté se zaostřenými a zaokrouhlenými rty, se zavřenější a otevřenější artikulací, palatalizovaně a faryngalizovaně. Zobrazení hustoty vokalického prostoru (viz obrázek 4.3) zachycuje jak určité rozdíly mezi mluvčími, tak mezi jednotlivými modifikacemi. Ve jmenované studii se autoři zabývali otázkou, nakolik se

¹¹ Jednalo se o analýzu anglického materiálu, znělé úseky tedy nezahrnovaly znělé obstruenty; viz výše.

sledované parametry včetně vokálních formantů mění v důsledku jednotlivých artikulačních manipulací, a pozorovali signifikantní posun F_2 při zaostření a zao-krouhlení rtů a při faryngalizaci, zatímco u F_1 žádný signifikantní posun zaznamenaný nebyl. Co se týče hustoty vokálního prostoru, rozdíly v jejím rozložení statisticky testovány nebyly; její zobrazení sloužilo pouze k vizualizaci chování vokálních formantů.



Obrázek 4.3. Převzato z: Houzar et al. (2022). Zobrazení hustoty vokálního prostoru u dvou mluvčích při jejich habituálním artikulačním nastavení a projevu s otevřenější a zavřenější artikulací (ilustrační výňatek; studie obsahuje zobrazení VSD u 5 mluvčích a 7 artikulačních nastavení).

4.3 Formantová dynamika

He et al. (2019) představili způsob parametrizace prvního formantu zahrnující temporální aspekt, kterým je tzv. formantová dynamika. Citovaná studie navazuje na publikaci zaměřenou na analogický způsob analýzy intenzity řečového signálu (He & Dellwo, 2017; podrobněji viz kapitolu 7, oddíl 7.1). Autoři vycházejí z předpokladu, že podobně jako změny intenzity, také pohyby prvního formantu, který je akustickým korelátem otevřenosti vokálu, bude korespondovat

s artikulačními úkony, konkrétně otevíráním a zavíráním vokálního traktu. Dynamika prvního formantu odráží jeho kolísání v rámci vokalických intervalů, přičemž autoři je dělí na dynamiku pozitivní a negativní. Tyto parametry odrážejí rychlost a strmost nárůstů a poklesů F_1 ; vzdálenosti mezi frekvenčními minimy a maximy jsou měřeny ve dvourozměrném prostoru definovaném časem a hodnotou frekvence prvního formantu. Pozitivní dynamika odpovídá velikosti vektoru vyneseno od bodu minimální dosažené frekvence k frekvenci maximální, zatímco dynamika negativní je naopak velikostí vektoru mezi frekvenčním vrcholem F_1 a jeho následujícím propadem.

V rámci své studie autoři využívali korpus čtených promluv ve švýcarské němčině. Hodnoty F_1 nejprve normalizovali pro mluvčího a následně vyčíslili průměr, směrodatnou odchylku a index párové variability¹² pozitivní i negativní dynamiky. Výsledky jejich experimentu odhalily výraznější idiosynkratičnost u parametrů dynamiky negativní, což autoři dávají do souvislosti s průběhem artikulace: jak již bylo uvedeno, autoři předpokládají, že pohyby kontury F_1 kovariují s otevíracími a zavíracími artikulačními úkony během produkce vokálu, a usuzují, že ve fázi otevírací, tj. při dosahování vokalického artikulačního cíle, mluvčí může aktivněji ovládat artikulační pohyby, zatímco při fázi zavírací, tedy po překonání artikulačního cíle, může být ovládání artikulačních orgánů již méně precizní, a tudíž v něm může existovat větší prostor pro individuální variaci závisící na fyziologii vokálního traktu a artikulačních návycích jednotlivce. Výsledky této studie jsou v souladu se závěry publikace věnované intenzitní dynamice (He & Dellwo, 2017) – i zde byla nalezena větší idiosynkratičnost u dynamiky negativní (podrobněji viz kapitolu 7, oddíl 7.1).

Heeren a He (2021) provedli analýzu formantové dynamiky v angličtině, holandštině a švýcarské němčině, avšak namísto čteného projevu, který analyzovali He

¹² Index párové variability (*pairwise variability index*; PVI) odráží průměrný rozdíl mezi dvěma po sobě jdoucími hodnotami v rámci vzorku.

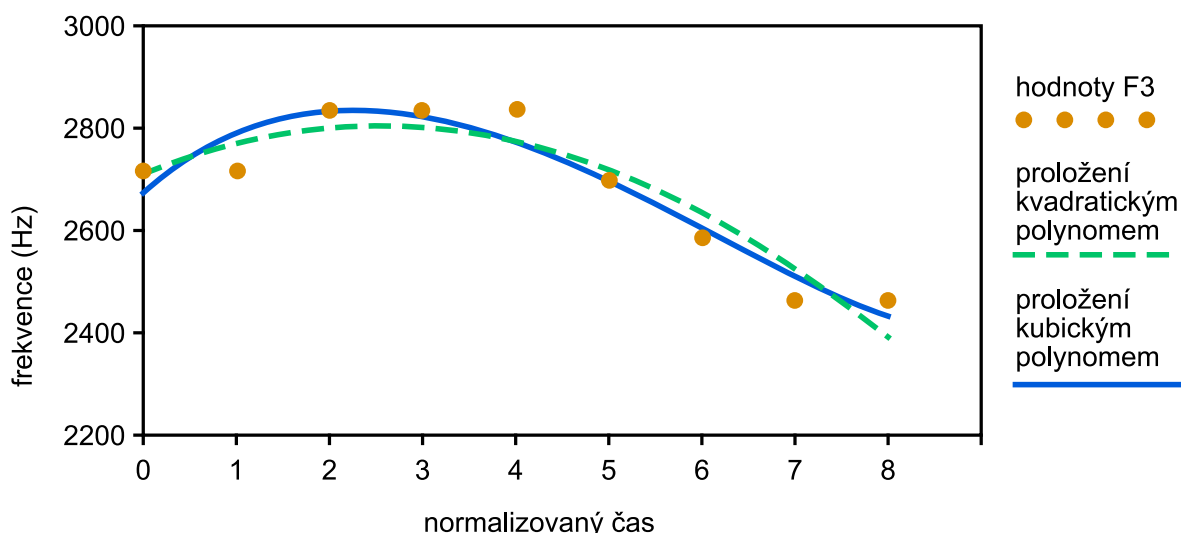
et al. (2019), využili projev spontánní, ve kterém, jak uvádějí, lze očekávat celkově méně pečlivé ovládní artikulačních gest mluvčími než ve čtených promluvách, a tudíž i méně zřejmou převahu idiosynkratického potenciálu u negativní dynamiky ve srovnání s dynamikou pozitivní. Jejich výsledky poukazují na jazykovou specifičnost míry idiosynkratickosti u pozitivní a negativní formantové dynamiky: zatímco výsledky pro švýcarskou němčinu v podstatě odpovídají závěrům předchozí studie zaměřené na materiál z téhož jazyka, v angličtině ani holandštině se stejný trend neukázal – v holandštině přínosnější informací naopak patrně nese dynamika pozitivní, v angličtině pak nebyla zaznamenána převaha ani jedné z nich. V rámci svého experimentu autoři zohlednili také kvalitu sledovaných vokálů, které rozdělili na otevřené/zavřené a zaokrouhlené/nezaokrouhlené; formantová dynamika napříč jazyky vykazala nejvyšší idiosynkratický potenciál v otevřených nezaokrouhlených vokálech. (Uvedená publikace je nicméně jen abstraktem a nezahrnuje interpretaci výsledků.)

4.4 Formantové trajektorie

Je také možné nenahlížet na formanty pouze jako na jednorozměrné hodnoty, ale sledovat jejich pohyb v čase, tj. extrahovat několik hodnot v průběhu vokálu a zaměřit se na tvar formantových trajektorií. Zatímco analýza hodnot formantů bez časového rozměru, jako např. jejich měření v artikulačním cíli či sledování dlouhodobých formantových distribucí, vypovídá o rozměrech vokálního traktu mluvčího či jeho artikulačních návycích, analýza zohledňující průběh změn hodnot formantů v čase poskytuje také informace o pohybech artikulačních orgánů jedince. Jak už bylo zmíněno, právě charakteristické pohyby mezi artikulačními cíli mohou velmi dobře odrážet individualitu mluvčího (McDougall, 2006).

McDougall (2006) se ve své studii zaměřila na formantové trajektorie v realizacích několika konkrétních řetězců fonémů v angličtině. Ve sledovaných segmentech byly manuálně změřeny hodnoty formantů v několika bodech. Následně byly aplikovány dva typy postupů pro charakterizaci průběhu formantových trajektorií: využití jednotlivých naměřených hodnot a jejich proložení polynomy¹³ – ilustraci poskytuje obrázek 4.4. Autorka si v rámci této studie kladla otázku, zda by bylo vhodnější využití koeficientů polynomů oproti analýze zohledňující všechny naměřené hodnoty formantů při rozlišování mluvčích. Podle jejích závěrů je využití polynomů pro parametrizaci formantových trajektorií nadějným postupem, neboť je tak zapotřebí méně porovnávaných hodnot, a to při relativně malém snížení přesnosti klasifikace pomocí lineární diskriminační analýzy. Takto charakterizované formantové trajektorie podle závěrů studie poskytují zásadní informace o individualitě mluvčího, a to výrazně více než jednotlivé formantové hodnoty změřené v temporálním středu vokálů. V rámci této studie byly testovány různé metody parametrizace průběhu formantů, jejím cílem ale nebylo zvolit jednu z nich coby nejefektivnější. Autorka uzavírá, že vývoj sofistikovaných modelů pro popis kontur zůstává předmětem dalšího výzkumu.

¹³ Jednalo se o proložení třemi typy polynomů: polynom 1. stupně (lineární polynom) definovaný dvěma koeficienty (proložení přímkou), polynom 2. stupně (kvadratický polynom) definovaný třemi koeficienty (proložení parabolou) a polynom 3. stupně (kubický polynom) definovaný čtyřmi koeficienty (proložení kubickou parabolou).



Obrázek 4.4. Upraveno podle: McDougall (2006: 107). Příklad proložení kontury F_3 polynomiální křivkou s využitím kvadratického a kubického polynomu (ilustrační výňatek; studie obsahuje zobrazení polynomiálních křivek prokládajících F_1 , F_2 a F_3).

Formantovým trajektoriím se věnoval také Rose (2015). V rámci jeho studie byly sledovány průběhy formantů F_1 – F_3 v realizacích vokálu /ɜ/ v australské angličtině; jejich hodnoty byly extrahovány v průběhu celých segmentů a následně proloženy polynomy třetího stupně (viz výše), jejichž koeficienty pak byly předmětem srovnávání. Také on dospěl na základě svých výsledků k závěru, že formantové trajektorie charakterizují mluvčího lépe než hodnoty formantů změřené v jednotlivých bodech uprostřed vokálů.

Tvary formantových trajektorií charakteristické pro jednotlivé vokalické kvality reflektuje ukazatel označovaný jako inherentní proměna spektra vokálu (*vowel inherent spectral change*; VISC; Nearey & Assmann, 1986). Obvykle je tento faktor dáván do souvislosti s percepčním rozlišováním vokalických kvalit (viz např. Jin & Liu, 2013, Morrison, 2009 či Nearey, 2013), nicméně existují i studie zaměřené na jeho variabilitu mezi jazyky (viz např. Williams et al., 2015), dialekty (viz např. Jacewicz & Fox, 2013) i jednotlivými mluvčími. Právě posledním jmenovaným se zabýval Morrison, 2013, který vycházel z předpokladu, že zatímco pro percepci se jako určující ukazují spíše jednoduché parametrizace VISC jako např.

sklon mezi výchozími a koncovými hodnotami formantů (Gottfried et al., 1993; Nearey & Assmann, 1986; obojí citováno v: Morrison, 2013: 266), zůstává v detailnějších pohybech formantových trajektorií prostor pro individuální variaci (jeho poznatky ohledně percepce se tedy podobají zjištěním McDougall, 2006 v oblasti artikulace, podle kterých – jak již bylo zmíněno – existuje mezi artikulačními cíli, které jsou pevněji určeny jazykovým systémem, větší volnost pro idiosynkratické artikulační pohyby). Také Morrison (2013) dospěl na základě rešerše dostupných studií k závěru, že parametry vypovídající o celkovém tvaru formantových trajektorií (zejména hodnoty koeficientů křivek, jimiž lze trajektorie proložit) mohou být výhodným ukazatelem pro vzájemné rozlišení mluvčích.

4.5 Experiment 2: Parametry vokálních formantů

4.5.1 Metoda

Za využití materiálu popsaného v kapitole 2 byly sledovány následující parametry vokálních formantů: dlouhodobé hodnoty F_1 – F_3 (LTF_1 – LTF_3), hodnoty F_1 – F_3 v jednotlivých vokálech a velikost vokálního prostoru (VSA). Jednotkou měření byl ve všech sledovaných metrikách Hertz (resp. Hertz² v případě VSA).

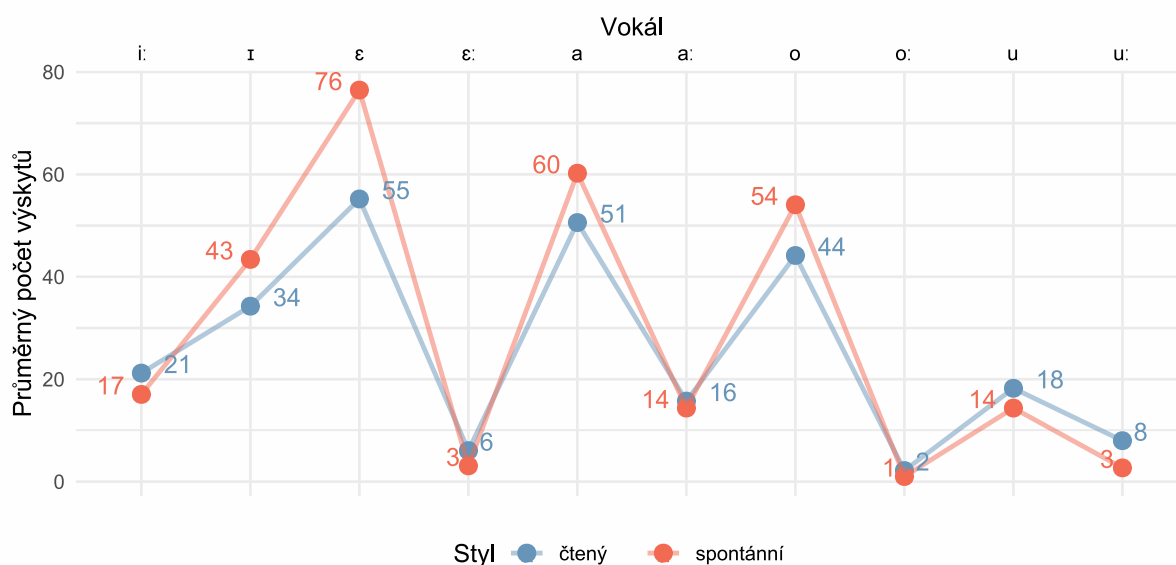
V programu Praat byly extrahovány hodnoty vokálních formantů pomocí příkazu *To Formant (robust)*, a to s využitím nastavení, jež doporučují Skarnitzl et al. (2015) pro mluvčí mužského pohlaví, tj. maximální počet formantů byl stanoven na 5,5 ve frekvenčním rozsahu do 5,5 kHz; autoři uvedené studie na základě svých výsledků dále konstatují, že lze doporučit jednotné nastavení nezávisle na kvalitě vokálu, v němž jsou hodnoty formantů měřeny (Skarnitzl et al., 2015: 180).

Nejprve byly extrahovány hodnoty vokálních formantů v průběhu celé promluvy nezávisle na identitě vokálu, a to v 10ms intervalech uvnitř vokálů v průběhu celého jejich trvání (jak již bylo zmíněno v kapitole 2, každá nahrávka byla opatřena segmentací na jednotlivé hlásky), a pro každou nahrávku byl vypočítán průměr a konfidenční intervaly jednotlivých formantů.

Byla také provedena extrakce hodnot formantů v jednotlivých vokálech. Jeden z nejčastějších způsobů parametrizace vokálních formantů představuje extrakce jejich hodnot z temporálního středu vokálu (viz oddíl 4.1), nicméně jak upozorňují Houzar a Skarnitzl (2022), u tohoto postupu existuje riziko detekce chybných hodnot. Optimální extrakční metodu podle zmíněných autorů nepředstavuje ani měření formantů v automaticky detekovaném artikulačním cíli vokálu vzhledem k potenciálnímu riziku, že by algoritmus mohl zaznamenávat chybně rozpoznané odlehle hodnoty. V tomto experimentu byl tedy sledován průměr vokálních formantů z prostřední třetiny vokálů. Pro jednotlivé nahrávky byl z takto získaných hodnot následně vypočítán průměr a 95% konfidenční interval pro každou vokální kvalitu: /i:/, /ɪ/, /ɛ ɛ:/, /a a:/, /o o:/ a /u u:/ – ačkoliv v současné době v češtině patrně dochází k diferenciaci vokální kvality krátkého centrálnějšího /u/ a dlouhého perifernějšího /u:/ (viz Volín & Skarnitzl, 2018: 18), v této studii byly zařazeny do společné kategorie, a to kvůli nízkému počtu výskytů obou zmíněných vokálů, které ilustruje obrázek 4.4; čtený text obsahoval průměrně 18 výskytů krátkého /u/ a 8 dlouhého /u:/¹⁴, v nahrávkách spontánního projevu se pak objevovalo v průměru 14 realizací /u/ a pouze 3 realizace /u:/ (pro srovnání: nejfrekventovanější vokál /ɛ/ byl ve čteném projevu zastoupen v průměru 55krát, ve spontánním pak dokonce průměrně zaznívalo 76 jeho realizací).

¹⁴ Text, který mluvčí předčítali, byl cíleně formulován tak, aby obsahoval i v běžné řeči málo frekventované hlásky. Obsah textu byl jednotný, nicméně u některých mluvčích se objevovaly fonetické redukce či chybně přečtená slova; celkové počty výskytů jednotlivých hlásek v nahrávkách čteného projevu se tudíž mohly mírně lišit.

Generalizace výsledků na základě takto nízkého počtu realizací sledovaných vokálů by tudíž byla problematická.



Obrázek 4.4. Průměrný počet výskytů jednotlivých vokálů ve zkoumaných nahrávkách.

Velikost vokalického prostoru, tj. plocha ve dvourozměrném prostoru definovaném dimenzemi F_1 a F_2 , byla vyčíslena na základě průměrů formantů všech vokalických kvalit pomocí následujícího vzorce (podle Houzar & Skarnitzl, 2022):

$$VSA = \left| \frac{(x_1 y_2 - y_1 x_2) + (x_2 y_3 - y_2 x_3) \dots + (x_6 y_1 - y_6 x_1)}{2} \right|$$

Kde x a y představují průměrné hodnoty F_1 a F_2 a čísla 1–6 reprezentují jednotlivé vokalické kvality v tomto pořadí: /i: ɪ ε a o u/.

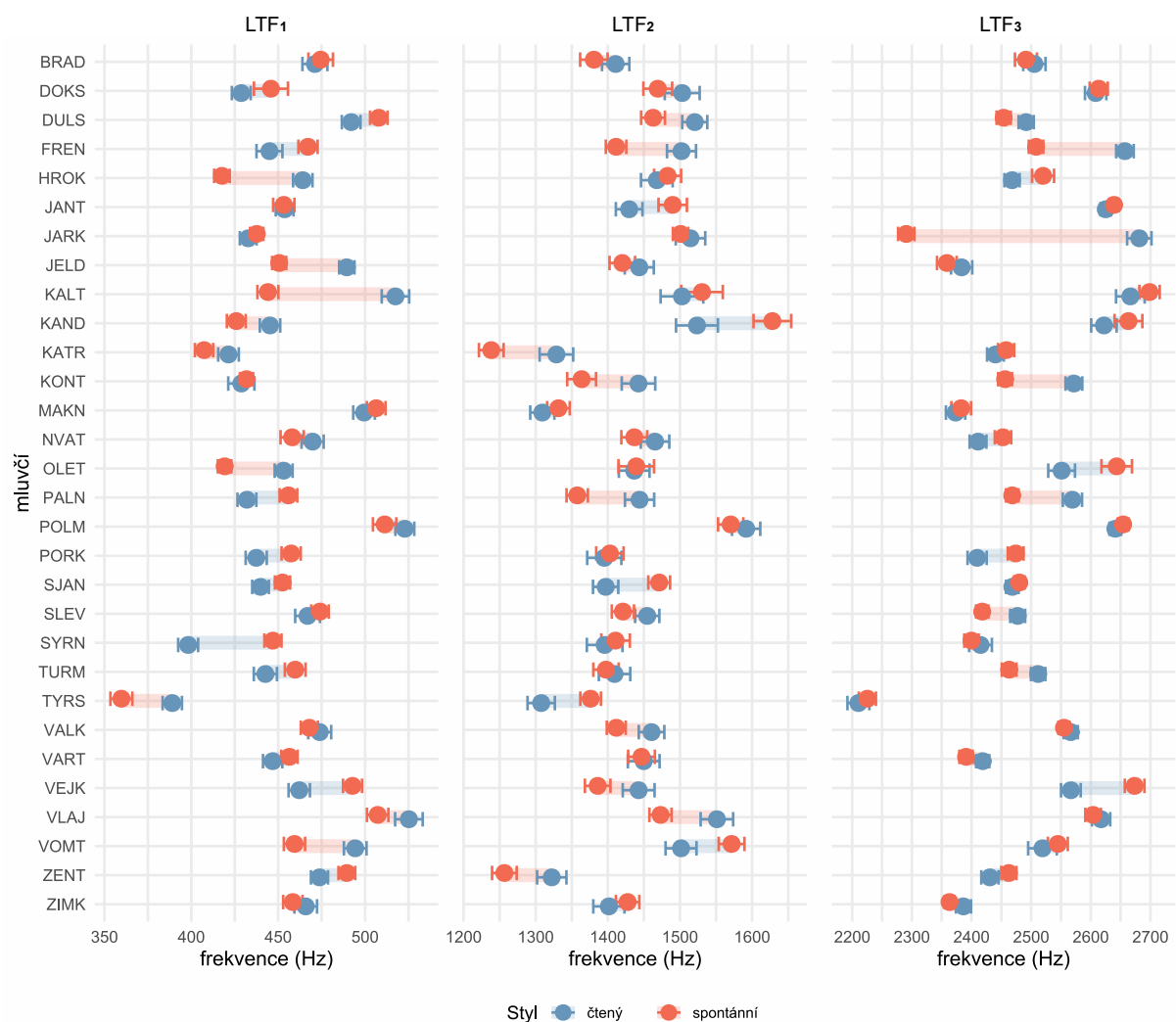
K vyhodnocení, zda se hodnoty sledovaných parametrů liší ve čteném a spontánním projevu bez ohledu na mluvčího, byl využit Wilcoxonův test. Při vyhodnocování významnosti rozdílů byla aplikována Bonferroniho korekce – hladina významnosti byla korigována počtem měřených parametrů: $\alpha = 0,05/22 = 0,0023$.

4.5.2 Výsledky

Dlouhodobé hodnoty F_1 , které korelují s celkovou otevřeností artikulace (otevřenější vokály jsou charakterizovány vyššími hodnotami F_1)¹⁵, u většiny mluvčích doznávají změn v závislosti na mluvním stylu, jak je patrné z obrázku 4.5, nicméně u řady mluvčích intraindividuální variabilita nepřevyšuje variabilitu interindividuální, tudíž by patrně bylo možné je od sebe na základě tohoto parametru odlišit – viz např. mluvčí TYRS, jehož hodnoty LTF_1 jsou celkově nižší oproti ostatním mluvčím, a na druhé straně POLM, jehož LTF_1 se naopak pohybuje ve vyšších frekvencích. Obdobná situace nastává také u LTF_2 .

Třetí formant, který bývá považován za relativně stabilní v rámci projevu jednotlivce, u řady mluvčích skutečně nevykazuje výrazné rozdíly mezi mluvními styly, avšak u jiných jednotlivců jsou tyto rozdíly naopak znatelné; nejvýraznější posun LTF_3 lze pozorovat u mluvčího JARK (jeho příčiny nicméně nejsou na základě dostupných dat zřejmé). I v případě tohoto formantu lze nicméně konstatovat, že jeho hodnoty mohou vzájemně odlišit některé jednotlivce nezávisle na mluvním stylu (např. JELD a KALT či TYRS a VALK), a to včetně těch, u kterých nejsou přítomny výrazné rozdíly mezi hodnotami prvního a druhého formantu (např. VALK a VART).

¹⁵ Na hodnoty vokalických formantů má ovšem vliv také anatomie vokálního traktu, nelze tudíž přímo odvozovat, že jejich rozdíly mezi dvěma mluvčími implikují rozdíly v jejich artikulaci. Liší-li se však tyto hodnoty v různých projevech jednoho mluvčího, pak tyto rozdíly na odlišnou artikulaci poukazují.



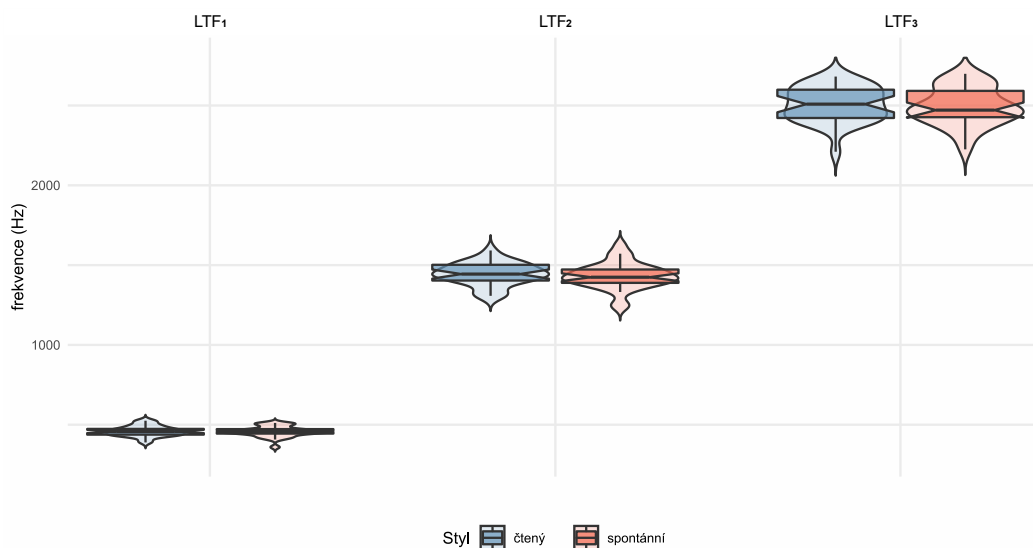
Obrázek 4.5. LTF_1 , LTF_2 a LTF_3 jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvovacích úseků, linky znázorňují 95% konfidenční intervaly.

Pokud jde o vzájemné rozlišování mluvčích, značené rozdíly lze mezi některými jedinci pozorovat zejména při kombinaci všech tří sledovaných formantů – např. již jmenovaní mluvčí TYRS a POLM se všeobecně vyznačují jejich nižšími, resp. vyššími hodnotami (v tomto pořadí). Právě vzorce chování více vokalických formantů tedy mohou poskytovat bližší informace o individualitě mluvčího.

Zde je zapotřebí podotknout, že konfidenční intervaly, které jsou vyznačeny na obrázku 4.5 a dále 4.7, 4.8 a 4.9, slouží k ilustraci, avšak nevypovídají

o statistické významnosti rozdílů mezi jednotlivými vzorky, neboť jsou zde zobrazeny jejich hodnoty bez korekce pro opakovaná měření.

Srovnání celkového chování jednotlivých formantů mezi sledovanými mluvními styly poskytuje obrázek 4.6 (pro podrobné informace o naměřených hodnotách těchto i ostatních parametrů vokálních formantů v obou typech projevu viz přílohu B). Jak lze na jeho základě pozorovat, mezi čteným a spontánním projevem nejsou přítomny výrazné posuny hodnot LTF_1 , LTF_2 ani LTF_3 ; ani výsledky Wilcoxonova testu nepotvrzují jejich statistickou významnost (ve všech případech je $p > 0,1$). LTF_1 v obou mluvních stylech vykazuje poměrně malý rozptyl napříč mluvčími, u LTF_2 je tento rozptyl výraznější a nejvyšších hodnot dosahuje u LTF_3 ; je nicméně třeba poukázat na dopad využití hertzové stupnice – při využití některé škály vycházející z psychoakustického základu (např. Bark či ERB) by byla relativní velikost těchto rozdílů odlišná.

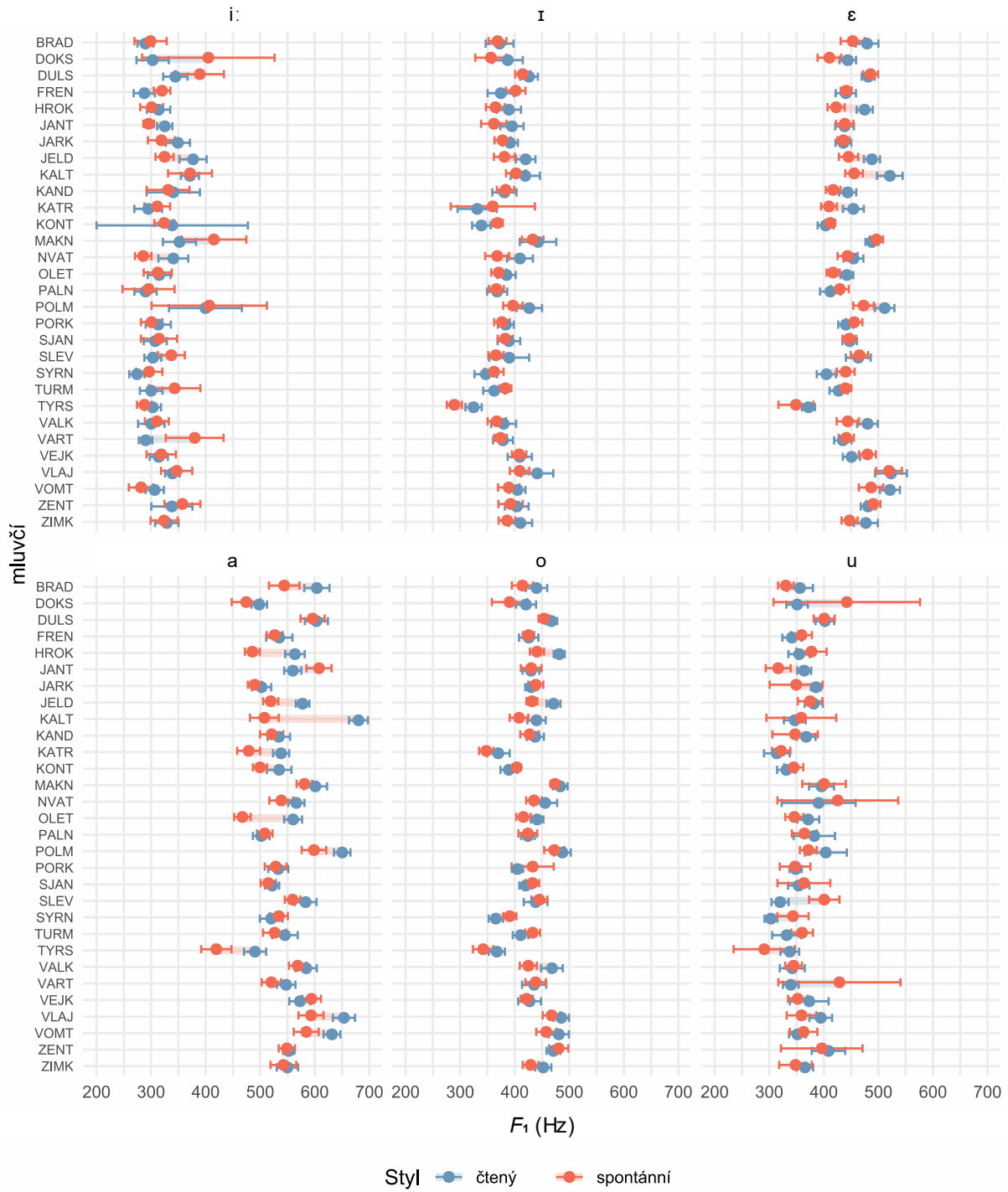


Obrázek 4.6. Srovnání LTF_1 , LTF_2 a LTF_3 napříč mluvčími ve čteném a spontánním projevu.

Podrobnější informace o chování vokálních formantů podává jejich analýza podle jednotlivých vokálních kvalit, jejíž výsledky zobrazují obrázky 4.7, 4.8 a

4.9. Jak je z obrázku 4.7 patrné, rozdíly mezi hodnotami prvního formantu ve čteném a spontánním projevu nejsou konzistentní napříč mluvčími; největšího množství nepřekrývajících se konfidenčních intervalů mezi mluvními styly daného mluvčího si lze povšimnout u vokálu /a/, přičemž jeho vyšší hodnoty jsou zastoupeny v rámci čteného projevu, což poukazuje na obecnou tendenci k jeho větší otevřenosti (tj. extrémnější artikulaci) v tomto mluvním stylu.

Na vyobrazení hodnot vokalických formantů v jednotlivých vokálech pro každého mluvčího zvlášť (obrázky 4.7, 4.8 a 4.9) je možné si povšimnout širších konfidenčních intervalů u zavřenějších vokálů, což by mohlo nabádat k dojmu, že jsou hodnoty formantů v těchto vokálech v rámci řečových vzorků celkově proměnlivější. Je nicméně třeba opět upozornit na různou četnost zastoupení jednotlivých vokalických kvalit v použitých nahrávkách, kde se vokály /i:/ a /u/ objevovaly v menším zastoupení než otevřené /a/ (viz obrázek 4.4). Velikost konfidenčních intervalů může být ovlivněna právě touto skutečností.



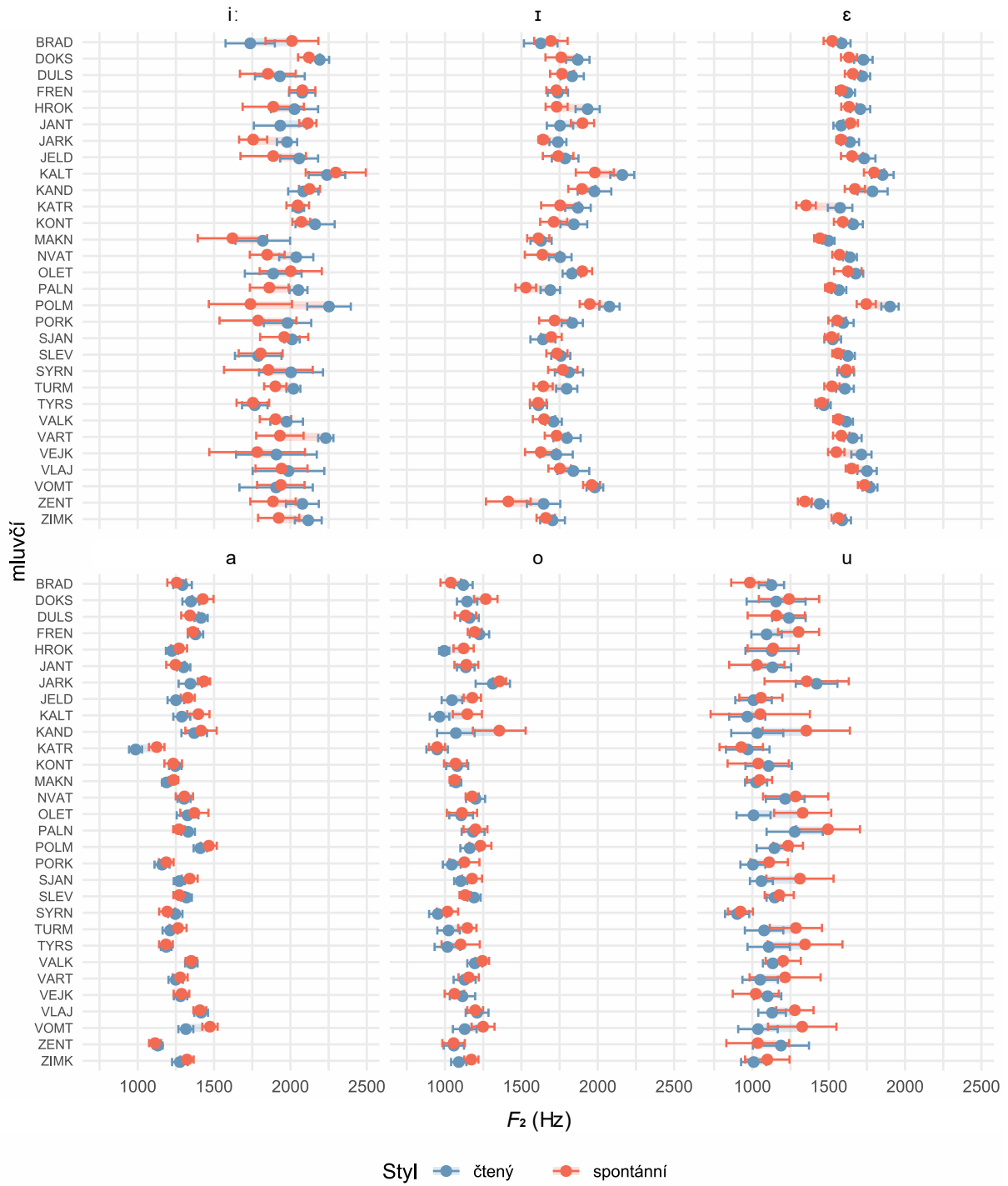
Obrázek 4.7. F_1 podle vokalické kvality u jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluвовých úseků, linky znázorňují 95% konfidenční intervaly.

Na druhou stranu ovšem hodnoty F_1 u některých mluvčích zejména v realizacích vokálu /a/ vykazují menší rozdíly mezi mluvnickými styly než oproti jiným mluvčím, a na jejich základě by tudíž patrně bylo možné odlišit některé jedince, kteří se znatelně vymykají populačnímu standardu (např. DOKS a DULS).

Při celkovém srovnání čteného a spontánního projevu, které ilustruje obrázek 4.10, nelze u většiny vokálů pozorovat výrazné rozdíly prvního formantu v žádném ze sledovaných vokálů, s výjimkou /a/; ačkoliv se zde konfidenční intervaly překrývají, podle výsledků Wilcoxonova testu je rozdíl mezi hodnotami F_1 vokálu /a/ signifikantní ($p < 0,0001$). Z hlediska otevřenosti vokálů tedy ve čteném projevu ve srovnání s projevem spontánním existuje tendence k perifernější artikulaci, konkrétně k otevřenější artikulaci otevřených vokálů.

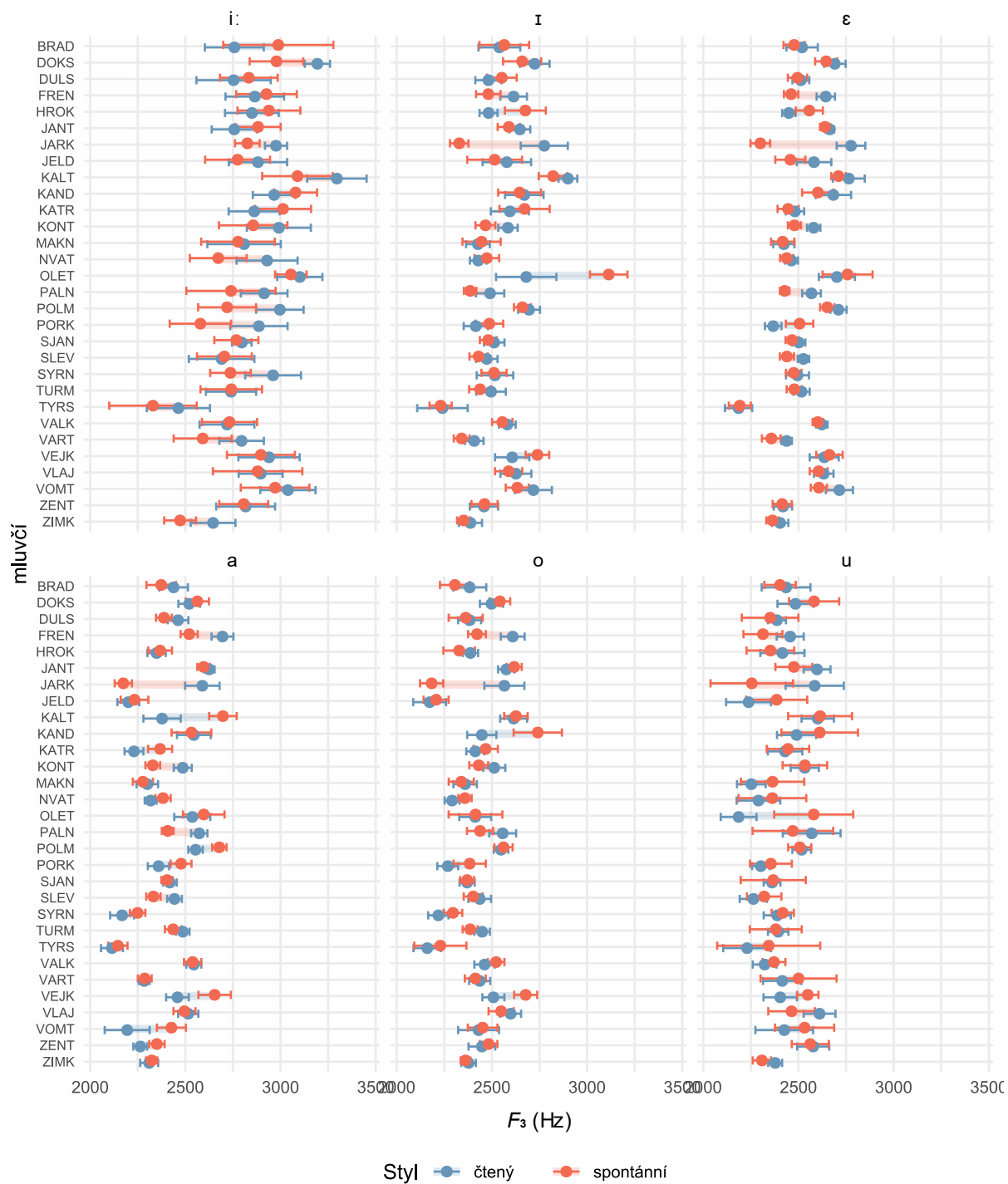
V rámci druhého formantu jednotlivých vokálů lze u většiny mluvčích pozorovat překrývající se konfidenční intervaly, jak ukazuje obrázek 4.8, nicméně u některých jednotlivců jsou hodnoty F_2 natolik vzdálené, že by zřejmě mohly představovat přínosnou informaci při jejich vzájemném rozlišení (např. VOMT a ZENT) – jako nejpřínosnější se v tomto směru jeví hodnoty extrahované z vokálů /ε/ a /a/ díky nejužším konfidenčním intervalům, které nasvědčují nejmenší fluktuaci hodnot sledovaného formantu v této vokalické kvalitě, ovšem jsou také důsledkem skutečnosti, že se realizace těchto dvou vokalických fonémů v řečovém materiálu vyskytují relativně nejčastěji (což potvrzuje obrázek 4.4).

Ačkoliv nejsou posuny druhého formantu mezi mluvnickými styly jednotné napříč mluvčími, je patrné, že tento formant ve čteném projevu obecně vykazuje tendenci k vyšším hodnotám u předních vokálů a naopak nižší v realizacích vokálů zadních; ve spontánním projevu tedy mluvčí patrně směřují k horizontálně centralizovanější artikulaci než v projevu čteném. Wilcoxonův test potvrdil významnost této tendence u vokálů /i:/ ($p < 0,0021$), /ɪ/ ($p < 0,0002$) a /ε/ ($p < 0,0001$).



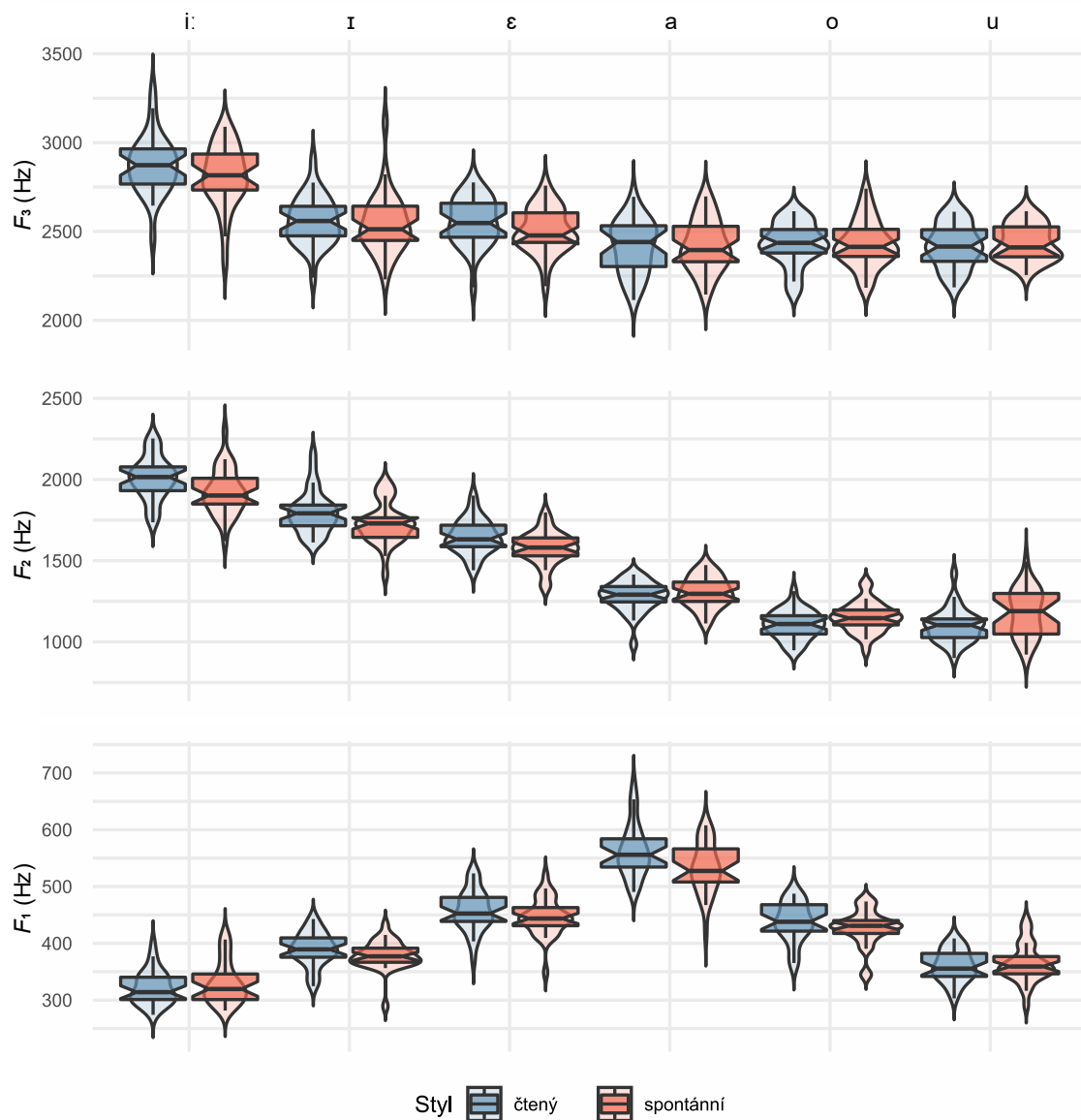
Obrazek 4.8. F_2 podle vokalické kvality u jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvočných úseků, linky znázorňují 95% konfidenční intervaly.

Individuální vzorce F_3 v jednotlivých vokálech zachycuje obrázek 4.9. Zatímco LTF_3 poskytuje souhrnné informace o středních hodnotách třetího formantu ve zkoumaných řečových vzorcích (viz obrázek 4.5), při jeho detailní analýze s rozdělením podle kvality vokálu lze pozorovat určité rozdíly získaných hodnot v jednotlivých vokálech. Např. u mluvího OLET lze v rámci LTF_3 pozorovat nepřekrývající se konfidenční intervaly (viz obrázek 4.5), ovšem při pohledu na hodnoty tohoto formantu v jednotlivých vokálech je tento trend patrný pouze ve vokálech /i/ a /u/. Naproti tomu u mluvího JARK, u kterého je rozdíl mezi hodnotami LTF_3 ve sledovaných mluvních stylech nejmarkantnější, se konfidenční intervaly nepřekrývají u žádného ze sledovaných vokálů s výjimkou /u/, a také v rámci /i:/ se od sebe hodnoty F_3 ve čteném a spontánním projevu vzdalují méně než u zbývajících vokálů. I vzorce chování třetího formantu tedy mohou poskytovat informace o individualitě mluvího.



Obrazek 4.9. F_3 podle vokalické kvality u jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvočných úseků, linky znázorňují 95% konfidenční intervaly.

Ze srovnání hodnot třetího formantu ve čteném a mluveném projevu vyplývá, že podobně jako v případě LTF_3 , ani u hodnot třetího formantu v jednotlivých vokálech celkově nedochází k výrazným posunům v závislosti na mluvním stylu, jak ukazuje obrázek 4.10. Wilcoxonův test nepotvrdil signifikantní rozdíl F_3 mezi čteným a spontánním projevem ($p > 0,02$) s výjimkou / ϵ / ($p < 0,002^{16}$).

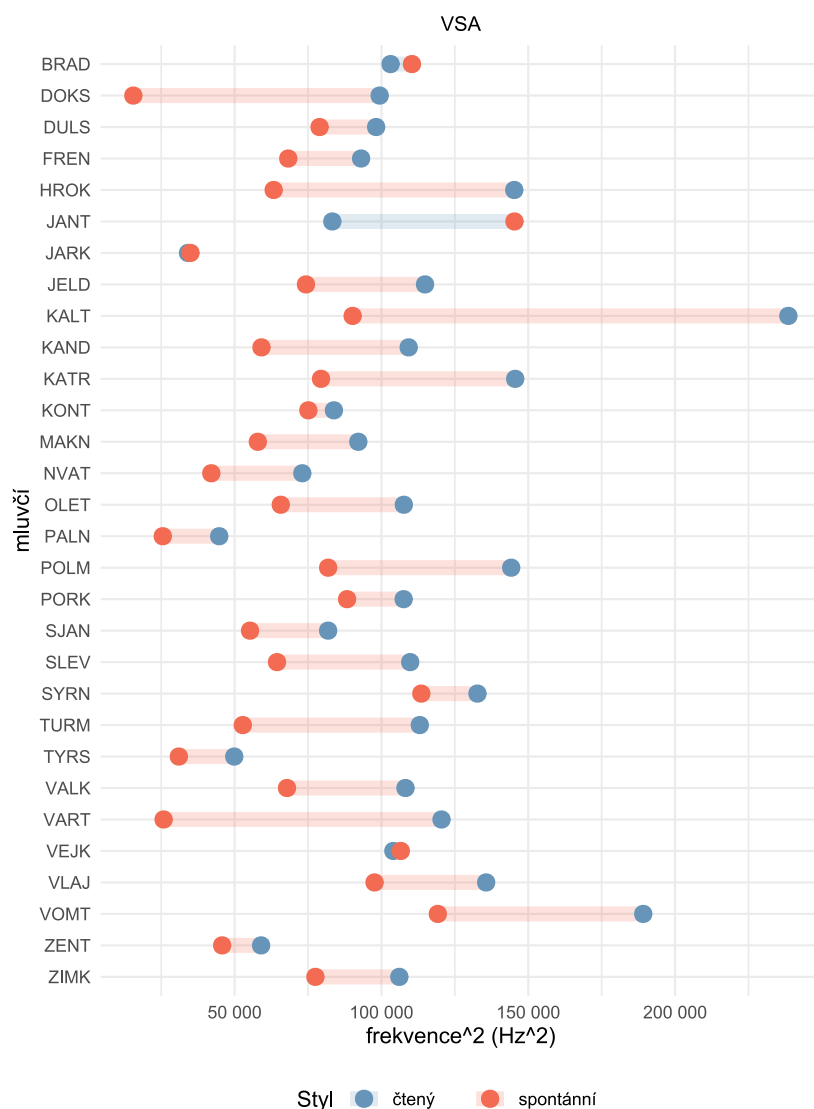


Obrázek 4.10. Srovnání F_1 , F_2 a F_3 v jednotlivých vokálech napříč mluvčími ve čteném a spontánním projevu.

¹⁶ V tomto případě se ovšem p-hodnota blíží hladině významnosti, která je rovna 0,0023 (viz výše).

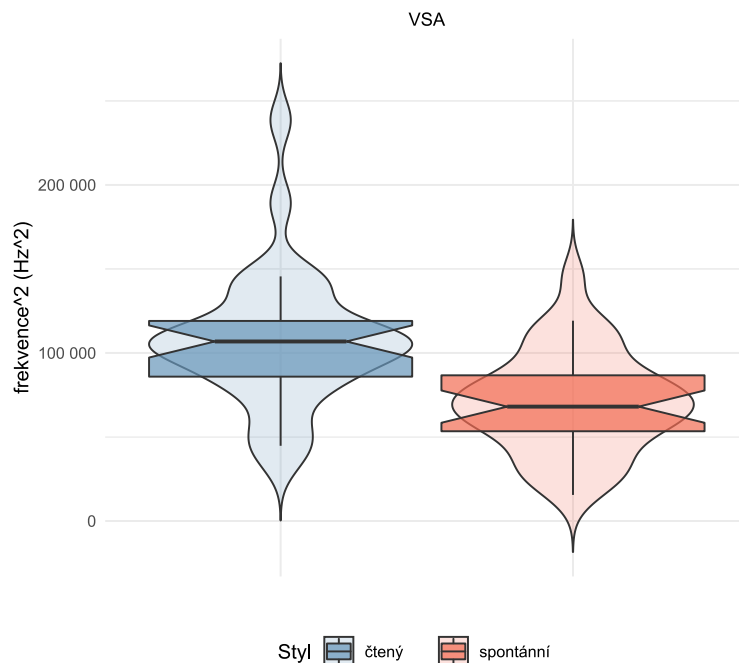
Na obrázku 4.10 si dále lze povšimnout výrazného rozdílu mezi F_3 v realizacích předního zavřeného /i:/ a ostatních vokálů; zde je namístě podotknout, že nelze zcela vyloučit vliv chybně detekovaných hodnot třetího formantu v řečovém signálu, které by mohly ovlivňovat výslednou podobu dat; vzhledem k relativně vyšším hodnotám F_2 v předních vokálech existuje riziko natolik výrazného přiblížení hodnot F_2 a F_3 , že využitý algoritmus v daném pásmu zachytí pouze F_2 a hodnotu F_3 chybně detekuje ve vyšší frekvenční oblasti (jak je z obrázku 4.10 patrné, F_2 v realizacích /i:/ mnohdy nabývá hodnot nad 2000 Hz, přičemž F_3 se v ostatních vokálech nezdá pohybuje pod 2500 Hz.). Na témže obrázku je patrná tendence k poklesu F_3 směrem od předních vokálů k zadním, ačkoliv není přítomna natolik nápadně jako u F_2 . Mezi těmito dvěma ukazateli také byla zjištěna středně silná až silná korelace pohybující se mezi 0,53 a 0,66, s výjimkou hodnot v zadních vokálech /o/ a /u/, což znázorňuje obrázek 4.14.

Od hodnot prvního a druhého formantu v jednotlivých vokálech se odvíjí velikost vokalického prostoru (VSA), jejíž hodnoty v analyzovaných řečových vzorcích zachycuje obrázek 4.11. Tento parametr u řady mluvčích vykazuje značné intraindividuální rozdíly; všeobecně je patrný posun k vyšším hodnotám ve čteném projevu (výraznou výjimku tvoří mluvčí JANT), přičemž velikost tohoto rozdílu se napříč jednotlivci různí; zatímco např. u mluvčího JARK byla v obou typech projevu zaznamenána relativně nízká hodnota VSA a jako poměrně stabilní se tento parametr ukázal také např. u mluvčích PORK či ZENT, u dalších jednotlivců byly nalezeny rozdíly vysoce přesahující odlišnosti mezi jejich projevy a projevy jiných mluvčích, viz zejména mluvčího KALT, ale i mnohé další. Je ovšem třeba upozornit, že vzhledem k vyčíslení hodnot VSA na základě průměrů F_1 a F_2 jednotlivých vokálů v celých řečových vzorcích u tohoto parametru nejsou k dispozici konfidenční intervaly, a tudíž ani bližší informace o stabilitě získaných hodnot uvnitř řečových vzorků.



Obrázek 4.11. Velikost vokálního prostoru jednotlivých mluvčích ve čteném a spontánním projevu.

V souladu s výsledky měření hodnot F_1 a F_2 v jednotlivých vokálech vypovídají i získané hodnoty VSA o celkové tendenci k perifernější artikulaci ve čteném projevu oproti projevu spontánnímu; při porovnání těchto mluvních stylů napříč mluvčími, které znázorňuje obrázek 4.12, lze pozorovat určitý rozdíl, jenž je podle Wilcoxonova testu statisticky signifikantní ($p < 0,0001$).

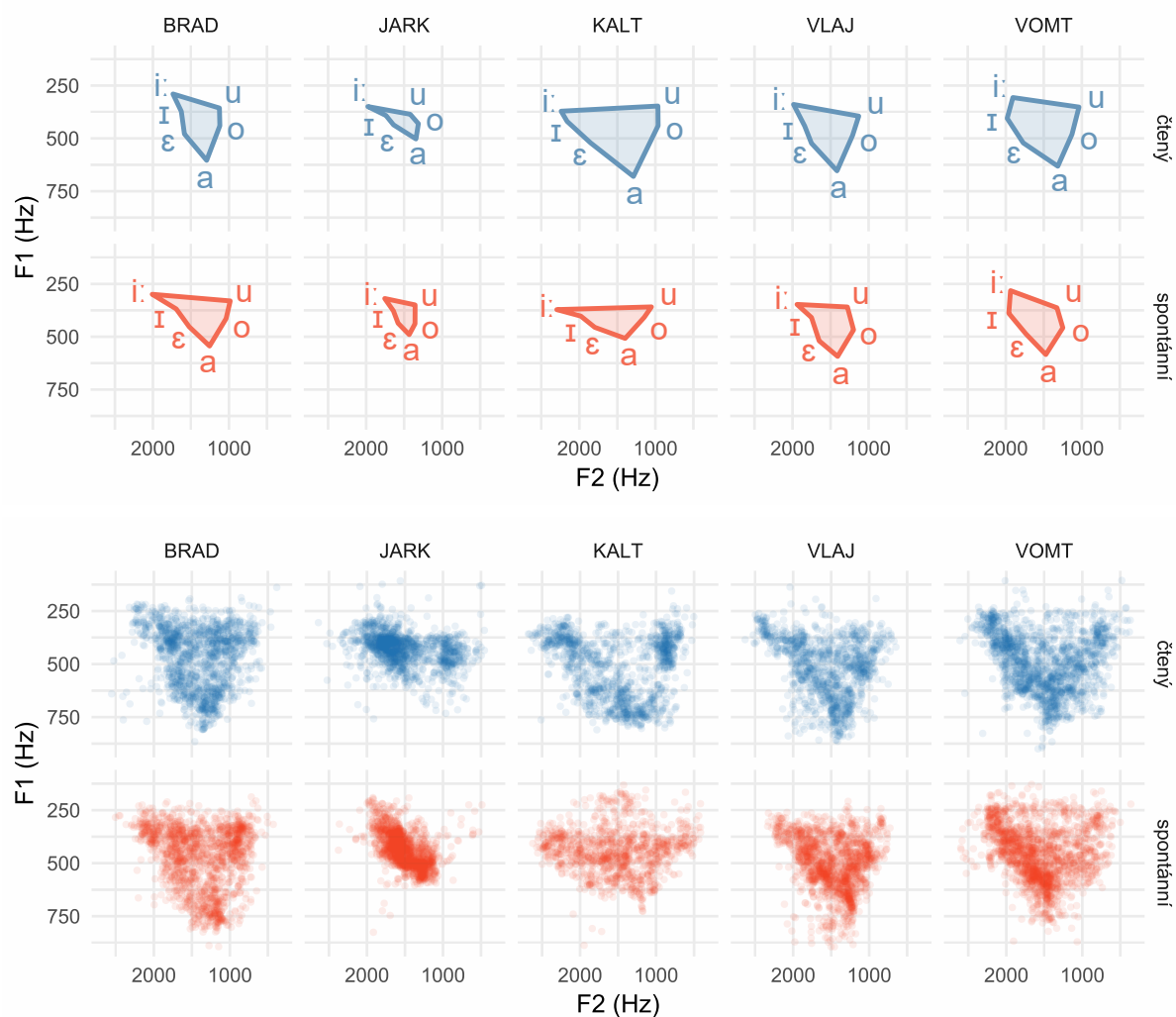


Obrázek 4.12. Velikost vokálního prostoru napříč mluvčími ve čteném a spontánním projevu.

Ilustraci rozložení vokálního prostoru u vybraných mluvčích nabízí obrázek 4.13, který obsahuje vykreslení vokálního prostoru na základě naměřených hodnot F_1 a F_2 v jednotlivých vokálech, jehož ploše odpovídá právě parametr VSA, a pro srovnání je doplněn také zobrazením hustoty vokálního prostoru využívajícím formantové hodnoty extrahované průběhu artikulace všech vokálů (viz výše).

U mluvčích VLAJ i VOMT je patrné zmenšení vokálního prostoru ve spontánním projevu, které koresponduje s obecnou tendencí k centralizovanější artikulaci v tomto mluvním stylu. Spontánní projev mluvčího KALT taktéž vykazuje zmenšení VSA ve vertikální dimenzi v důsledku zvýšení F_1 u vokálu /a/, nicméně v horizontální dimenzi (F_2) centralizace patrná není. Právě u mluvčího KALT byl zjištěn nejvýraznější absolutní rozdíl velikosti vokálního projevu ve čteném a spontánním projevu (viz obrázek 4.11), a tento rozdíl je patrný i na obrázku 4.13. Mluvčí BRAD se vyznačuje extrémnějšími hodnotami F_2 a naopak – podobně jako KALT – centralizací F_1 ve spontánním projevu, nicméně na rozdíl právě od

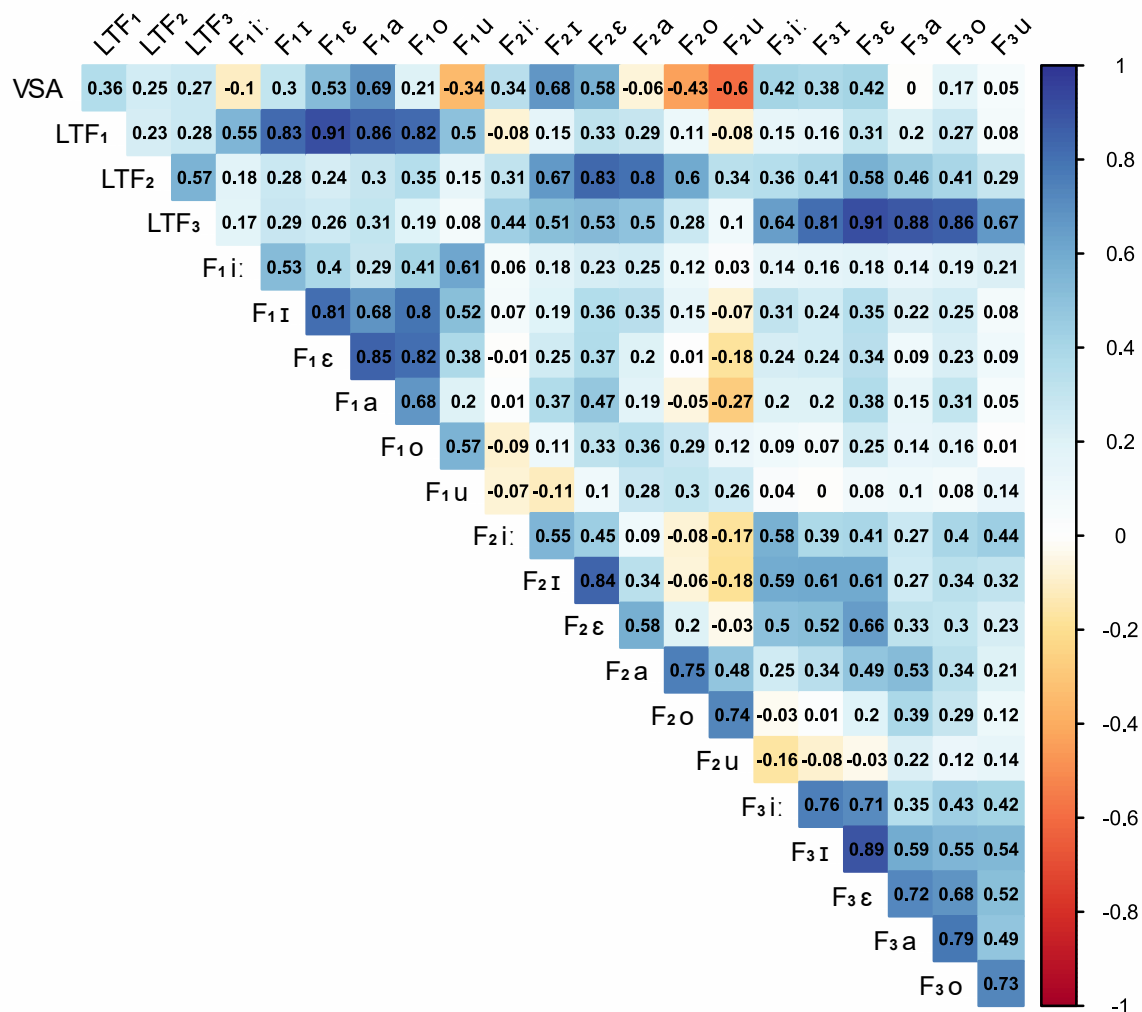
mluvčího KALT u něj byly v obou mluvních stylech naměřeny podobné hodnoty VSA. Jak bylo zmíněno výše, u mluvčího JARK byly v obou mluvních stylech zaznamenány relativně nízké hodnoty VSA, a i na zobrazení jeho vokálního prostoru je znatelný relativně malý rozestup hodnot formantů v jednotlivých vokálech. (U téhož mluvčího byl zároveň zaznamenán značný rozdíl mezi mluvními styly u hodnoty LTF_3 , viz obrázek 4.5.)



Obrázek 4.13. Zobrazení velikosti vokálního prostoru (nahore) na základě průměrných hodnot F_1 a F_2 a hustoty vokálního prostoru pomocí hodnot F_1 a F_2 extrahovaných v 10ms intervalech v průběhu artikulace vokálů (dole) u vybraných mluvčích.

U jednotlivých ukazatelů je možné sledovat také vzájemnou korelaci, kterou zachycuje obrázek 4.14; jak bylo výše uvedeno, např. F_3 středně silně až silně

koreluje s F_2 ve vokálech /i: ɪ ɛ a/ ($r \in \langle 0,53; 0,66 \rangle$). Velmi silná korelace s LTF_1 byla nalezena u F_1 všech vokálů s výjimkou zavřených /i:/ a /u/ ($r \in \langle 0,82; 0,91 \rangle$), parametr LTF_2 pak velmi silně koreloval s F_2 vokálů /ɛ/ ($r=0,83$) a /a/ ($r = 0,8$). LTF_3 vykázal velmi silnou korelaci s F_3 všech vokalických kvalit s výjimkou vokálů /i:/ a /u/ ($r \in \langle 0,81; 0,91 \rangle$). Silná korelace s velikostí vokalického prostoru se objevila u F_1 otevřeného vokálu /a/ ($r = 0,69$), F_2 předního /ɪ/ ($r = 0,68$ a také F_2 zadního /u/ (v jehož případě se jednalo o korelaci negativní: $r = -0,6$).



Obrázek 4.14. Korelační matice sledovaných parametrů vokalických formantů F_1 - F_3 .

4.5.3 Diskuse a závěr

V rámci experimentu 2 byly analyzovány vybrané parametry vokalických formantů F_1 , F_2 a F_3 : jejich dlouhodobé průměrné hodnoty (LTF_1 , LTF_2 , LTF_3), hodnoty F_1 – F_3 podle kvality vokálu a na jejich základě vyčíslená velikost vokalického prostoru. Pozornost byla věnována variabilitě těchto ukazatelů napříč mluvčími a také mezi dvěma mluvními styly – čteným a spontánním.

Hodnoty vokalických formantů, které jsou rezonančními frekvencemi vokálního traktu, obecně odrážejí tvar nadhrtanových dutin, přičemž první dva formanty a částečně také formant třetí závisejí na artikulačním nastavení. Dlouhodobé hodnoty F_1 – F_3 tedy závisejí jak na anatomii nadhrtanových dutin, tak na celkovém habituálním artikulačním nastavení daného mluvčího – tedy na jeho případné tendenci např. k otevřenější artikulaci, labializaci a podobně. Výsledky experimentu ukazují, že v rámci všech tří dlouhodobých formantových ukazatelů byly u části mluvčích pozorovány rozdíly mezi čteným a spontánním projevem, tyto rozdíly však napříč mluvčími nebyly konzistentní. Při celkovém srovnání obou mluvních stylů nezávisle na mluvčím nebyl pozorován významný rozdíl ani v jednom ze sledovaných formantů, což naznačuje, že zřejmě neexistuje všeobecná tendence k horizontálnímu či vertikálnímu posunu artikulace vokálů mezi sledovanými mluvními styly.

Přestože se dlouhodobé hodnoty vokalických formantů obecně neukázaly jako zcela stabilní pro jednotlivce nezávisle na mluvním stylu, některé mluvčí by patrně na jejich základě bylo možné odlišit, a to tím spíše, čím výrazněji se odchyľují od populačního standardu. F_3 , který bývá ze sledovaných formantů považován za nejstabilnější v rámci mluvčího, se v tomto směru jeví jako potenciálně nejpřínosnější, a to i v případě jednotlivců, u kterých první dva formanty nevykazují markantní rozdíly – je však na místě upozornit, že ani chování třetího formantu není konzistentní napříč mluvčími, a zatímco u některých zůstává stabilní nezávisle na mluvním stylu, u jiných byly zaznamenány jeho výrazné posuny.

Podrobnější náhled na chování vokalických formantů poskytují jejich hodnoty v realizacích jednotlivých vokalických fonémů. Zatímco dlouhodobé formantové hodnoty vypovídají, jak již bylo uvedeno, mimo jiné o celkovém habituálním artikulačním nastavení mluvčího, tyto parametry odrážejí i rozdíly mezi jednotlivými vokály, jež jsou v rámci dlouhodobých formantových hodnot zanedbány. Výsledky experimentu poukázaly i v tomto případě na individuální tendence napříč mluvčími, u některých z nich ovšem byly zjištěny větší interindividuální rozdíly ve srovnání s rozdíly intraindividuálními, a bylo by tedy patrně možné je na základě těchto ukazatelů rozlišit. Ačkoliv posuny mezi sledovanými mluvními styly byly značně individuální, výsledky měření F_1 a F_2 vypovídají o celkové tendenci k centralizovanější artikulaci vokálů ve spontánním projevu v porovnání s projevem čteným: ve čteném projevu se obecně vyskytovaly vyšší hodnoty F_1 v otevřených vokálech, druhý formant se pak ve čteném projevu pohyboval ve vyšších hodnotách u předních vokálů a naopak v nižších hodnotách u vokálů zadních, což poukazuje na celkově perifernější artikulaci ve čteném projevu. Právě tyto rozdíly byly v rámci dlouhodobých formantových hodnot setřeny v důsledku zprůměrování hodnot formantů extrahovaných ze všech vokálů.

Vzorci třetího formantu taktéž variovaly mezi jednotlivci – u většiny mluvčích se F_3 projevil jakožto stabilní parametr nezávislý na mluvním stylu, u některých však došlo k jeho výrazným posunům. Obecně u F_3 nebyly pozorovány markantní rozdíly mezi čteným a spontánním projevem (s výjimkou vokálu / ϵ /, důvody tohoto rozdílu však nejsou zcela zřejmé). Jak již bylo zmíněno, třetí formant bývá považován za do jisté míry závislý na kvalitě vokálu; ze získaných výsledků je patrná jeho mírná tendence k vyšším hodnotám u přednějších vokálů, nelze však hovořit o signifikantních rozdílech – výjimku nicméně tvoří F_3 předního zavřeného /i:/, u něž je patrné výrazné zvýšení oproti ostatním vokálům. Tento výsledek však patrně vyplývá nikoli ze skutečně natolik vyššího F_3 ve zmiňovaném vokálu, nýbrž ze skutečnosti, že pro tento přední vokál jsou charakteristické relativně vysoké hodnoty F_2 , jež se mohou pohybovat ve frekvenčním pásmu

blízkém třetímu formantu; při extrakci vokalických formantů pak může dojít ke splynutí reálného F_3 s F_2 a chybné detekci hodnot třetího formantu ve vyšším frekvenčním pásmu, které již může odpovídat F_4 . Na jakých okolnostech závisí hodnoty F_3 a do jaké míry kovariují s prvními dvěma formanty, zůstává otázkou pro další výzkum.

Naměřené hodnoty velikosti vokalického prostoru je vzhledem ke způsobu jejího vyčíslení (viz výše) problematické srovnávat mezi jednotlivými řečovými vzorky, nicméně projevila se u ní zjevná tendence k nabývání vyšších hodnot ve čteném projevu v porovnání s projevem spontánním – tyto závěry korespondují se vzorci F_1 a F_2 v jednotlivých vokalických kvalitách, které nasvědčují celkové inklinaci mluvcích k perifernější artikulaci ve čteném projevu, resp. k centralizaci vokálů v projevu spontánním. Pro vzájemné odlišování mluvcích se tento parametr jeví jako spíše nevhodný, neboť při tomto způsobu jeho extrakce nejsou k dispozici informace o jeho proměnlivosti uvnitř řečových vzorků a patrně se jedná o parametr náchylný ke změnám v závislosti na mluvním stylu.

5 Frekvenční spektrum

Mimo ukazatele, které lze vysledovat ve frekvenčním spektru v průběhu řeči, jakými jsou vokální formanty a základní frekvence, je možné se zaměřit také na tvar spektra jako takového. Frekvenční spektrum signálu odráží, nakolik jsou jednotlivé frekvence zastoupené v signálu prominentní, a percepčním korelátem této charakteristiky je barva hlasu. Parametry vycházející z frekvenčního spektra, u nichž je možné sledovat idiosynkratický potenciál, jsou dvojího charakteru: krátkodobé, které charakterizují jednotlivé hlásky, případně jejich části, a dlouhodobé, které odrážejí průměrné spektrum celého zkoumaného vzorku.

Je třeba podotknout, že celkový tvar frekvenčního spektra může být do značné míry ovlivněn kvalitou nahrávky, akustickými vlastnostmi prostředí či přítomností ruchů v pozadí, a možnost charakterizovat jeho pomocí daného mluvčího tedy může být problematická (Weingartová et al., 2014).

5.1 Spektrální vlastnosti vokálů

Kapitola 4 pojednávala o vokálních formantech, které lze nalézt ve frekvenčním spektru vokálů – patří tedy také mezi jejich spektrální vlastnosti, nicméně vzhledem k zásadní roli, kterou hrají při zkoumání variability řečového signálu, bylo na místě věnovat jim samostatnou kapitolu. Existují ovšem i další prvky, které je možné ve frekvenčním spektru vokálů měřit a sledovat jejich interindividuální i intraindividuální variabilitu.

Proměnnou, na kterou se zaměřuje řada studií věnovaných spektrálním vlastnostem vokálních segmentů, je spektrální sklon. Jedná se o strmost poklesu či nárůstu amplitudy s rostoucí hodnotou frekvence ve spektru; Weingartová et al. (2014) definují spektrální sklon jakožto „spád křivky interpolující závislost

amplitudy na frekvenci v řečovém spektru“ (Weingartová et al., 2014: 77; viz také oddíl 5.6). Tento ukazatel je akustickým korelátem barvy hlasu, resp. fonačních charakteristik – při zvýšeném mluvním úsilí, třepeně a tlačené fonaci či falzetu, které se zároveň vyznačují nižším poměrem otevření hlasivek, dochází ke snížení spektrálního sklonu (viz např. Monsen & Engebretson, 1977: 988 či Doddington, 1985: 1659, obojí citováno v: Weingartová et al., 2014: 83), zatímco např. dyšná fonace se pojí s výraznějším úbytkem energie ve vyšších frekvencích, tj. se strmějším spektrálním sklonem (viz např. Hammarberg et al., 1980: 446).

Spektrální sklon vokalických segmentů bývá charakterizován pomocí rozdílů určitých harmonických složek, a to definovaných jejich řádem (označované písmenem H a pořadovým číslem) nebo formantem, v němž představují nejsilnější harmonickou složku (označované písmenem A a pořadovým číslem daného formantu) – např. H1-H2, H2-H4, H1-A1, H1-A2 či H1-A3 (viz např. Hanson, 1997 či Tavi, 2018), případně také frekvenční oblastí, ke které leží nejbližší (v tom případě jsou označovány písmenem H a hodnotou dané frekvence), např. H2kHz–H5kHz (viz např. Garellek et al., 2013). Hodnoty amplitud jednotlivých harmonických složek řečového spektra odrážejí jednak činnost hlasivek, ale také fyziologii a momentální nastavení vokálního traktu¹⁷ – promítají se do nich i vokalické formanty. Idiosynkratickými faktory v rámci vokalických segmentů jsou jak fonace, tak i artikulační nastavení a individuální fyziologie vokálního traktu, a je žádoucí napříč mluvčími srovnávat akustické ukazatele, které jsou určovány všemi těmito aspekty. Nastavením nadhrtanových dutin v případě vokalických segmentů jsou určovány hodnoty vokalických formantů, jež jsou často sledovanou proměnnou. V případě spektrálního sklonu jakožto další proměnné, jejíž interindividuální variabilitu lze zkoumat, tedy může být výhodné vliv vokalických formantů (a tedy individuálních rozdílů v anatomii nadhrtanových dutin i

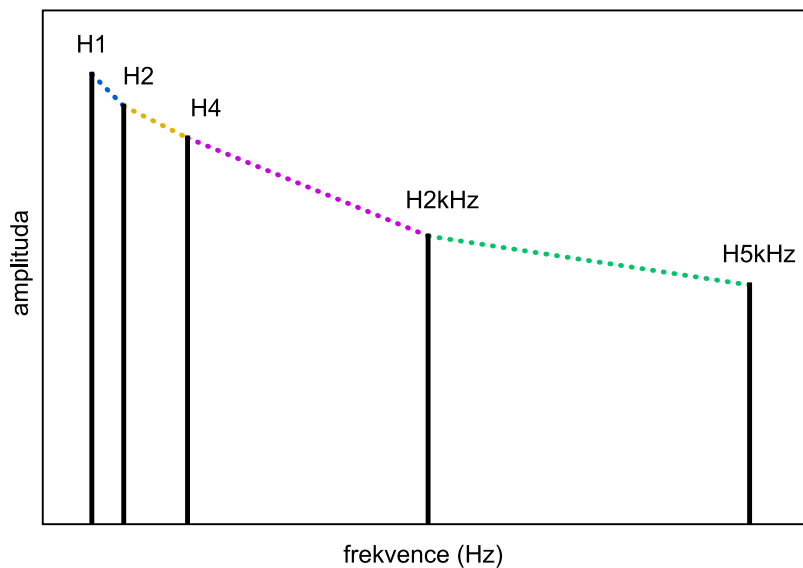
¹⁷ Podle filtrového modelu produkce řeči, jak jej popsal Fant (1960), je řečový signál formován dvěma činiteli, a to zdrojem, tedy hlasivkami tvořícími primární tón, případně artikulátory produkujícími turbulentní šum, a filtrem, který je představován nadhrtanovými dutinami a ve kterém výchozí signál nabývá výslednou podobu.

momentálního artikulačního nastavení) odfiltrovat a získat tak spektrální charakteristiku hlasivkové činnosti; výsledkem jsou tedy vzájemně nezávislé proměnné: vokální formanty a normalizovaný spektrální sklon. Způsob normalizace amplitud jednotlivých frekvencí popisují např. Iseli et al. (2007), takto upravené hodnoty jsou pak označovány hvězdičkou, tedy např. $H1^*$ či $A1^*$. Normalizovaný spektrální sklon lze tedy parametrizovat např. jako $H1^*-H2^*$, $H2^*-H4^*$, $H1-A1^*$, $H1^*-A3^*$ atd. (viz např. Seyfarth & Garellek, 2015; Nara, 2017; Tylečková & Skarnitzl, 2019 či Esposito et al., 2021).

Inter- i intraindividuální variabilitu spektrálního sklonu vokálních segmentů pomocí takto normalizovaných parametrů spektrálního sklonu zkoumali např. Vaňková a Skarnitzl (2014), kteří měřili $H1^*-H2^*$, $H2^*-H4^*$, $H1^*-A1^*$, $H1^*-A2^*$ a $H1^*-A3^*$ v českých monoftonzích /a/ a /a:/ a porovnávali jejich stabilitu v rámci mluvčího a variabilitu napříč mluvčími (jinými slovy tedy jejich idiosynkratický potenciál) s vokálními formanty F_1-F_4 . Jakožto nejstabilnější parametr v rámci mluvčího napříč mluvčími styly se projevil parametr $H1^*-A2^*$, následovaný $H1^*-H2^*$ a $H1^*-A1^*$, zatímco největší variabilitu mezi mluvčími vykazoval parametr $H1^*-H2^*$, o něco méně pak také $H1^*-A1^*$. Co se týká srovnání s vokálními formanty, konstatují autoři, že u parametrů spektrálního sklonu byla v rámci jejich experimentu zjištěna o něco větší idiosynkratičnost.

Garellek et al. (2013) ve své studii zaměřené na percepci kvality hlasu sledovali hodnoty $H1-H2$, $H2-H4$, $H4-H2\text{kHz}$ (tj. rozdíl čtvrté harmonické složky a harmonické složky nejbližší 2 kHz) a $H2\text{kHz}-H5\text{kHz}$ v izolovaném drženém vokálu /a:/; vycházeli přitom z psychoakustického modelu postihujícího spektrum hlasivkového tónu, který využívá členění právě pomocí těchto harmonických složek a je definován jmenovanými parametry spektrálního sklonu (Kreiman & Gerratt, 2010; Kreiman et al., 2011; Kreiman & Gerratt, 2012, vše citováno v: Garellek et al., 2013: 2); schematické zobrazení uvedeného modelu ukazuje obrázek 5.1. Podle výsledků Garelleka et al. (2013) jsou spektrální parametry ovlivněny f_0 –

vyšší hodnoty f_0 se většinou pojí se strmějším sklonem mezi 0 Hz a H₄, ale plošším sklonem mezi H₄ a 2kHz.



Obrázek 5.1. Upraveno podle: Garellek et al. (2013: 2). Schematické zobrazení psychoakustického modelu postihujícího spektrum hlasivkového tónu.

Efektivitu parametrů spektrálního sklonu při srovnávání mluvčích zkoumali Cardoso et al. (2019); zaměřili se na rozdíly amplitud H1-H2, H2-H4, H1-A1, H1-A2 a H1-A3 v automaticky extrahovaných vokalických intervalech¹⁸. Na základě výsledků svého experimentu uzavírají, že tyto ukazatele extrahované z vokálů zachycují zásadní idiosynkratické informace a při porovnání s výsledky studie zaměřené na idiosynkratický potenciál vokalických formantů, kterou provedli Hughes et al. (2019; citováno v: Cardoso et al., 2019: 4), parametry spektrálního sklonu vykazují vyšší idiosynkratický potenciál než vokalické formanty měřené v témže vokalickém materiálu.

¹⁸ Využívali materiál od 97 mužů, kteří byli mluvčími standardní britské angličtiny.

5.2 Experiment 3: Parametry frekvenčního spektra vokálů

5.2.1 Metoda

Za využití materiálu popsaného v kapitole 2 byly sledovány následující parametry frekvenčního spektra, resp. spektrálního sklonu vokálů: $H1^*-H2^*$, $H2^*-H4^*$, $H1^*-A1^*$, $H1^*-A2^*$, $H1^*-A3^*$, $H4^*-H2kHz^*$ a $H2kHz^*-H5kHz^*$.

Hodnoty jednotlivých parametrů byly získány automaticky pomocí softwaru VoiceSauce (Shue et al., 2011; Shue, 2013) za využití segmentace na hlásky (podrobněji viz kapitolu 2). V průběhu vokalických segmentů byly v 1ms intervalech extrahovány hodnoty sledovaných parametrů a následně byl v R pro každý vokál vyčíslen průměr získaných hodnot. Jak bylo uvedeno v oddíle 5.1, značení hvězdičkou je využíváno pro normalizované parametry spektrálního sklonu, přičemž účelem normalizace je odfiltrování dopadu vokalických formantů na amplitudu harmonických složek; cílem normalizace je tedy potlačení vlivu filtru vokálního traktu a získání informací pouze o chování hlasivkového zdroje.

Pro každý řečový vzorek byly vypočítány průměrné hodnoty jednotlivých parametrů a jejich konfidenční intervaly, u kterých byla následně sledována intra- i interindividuální variabilita. (U konfidenčních intervalů nebyla aplikována korekce pro opakovaná měření, jejich zobrazení tedy slouží k ilustraci výsledků, nikoli k vynášení závěrů ohledně statistické významnosti rozdílů mezi vzorky.) Za využití průměrných hodnot jednotlivých parametrů byly vyhodnoceny také rozdíly mezi čteným a spontánním projevem; v rámci statistického testu byla hladina významnosti korigována počtem sledovaných parametrů: $\alpha = 0,05/7 = 0,0071$.

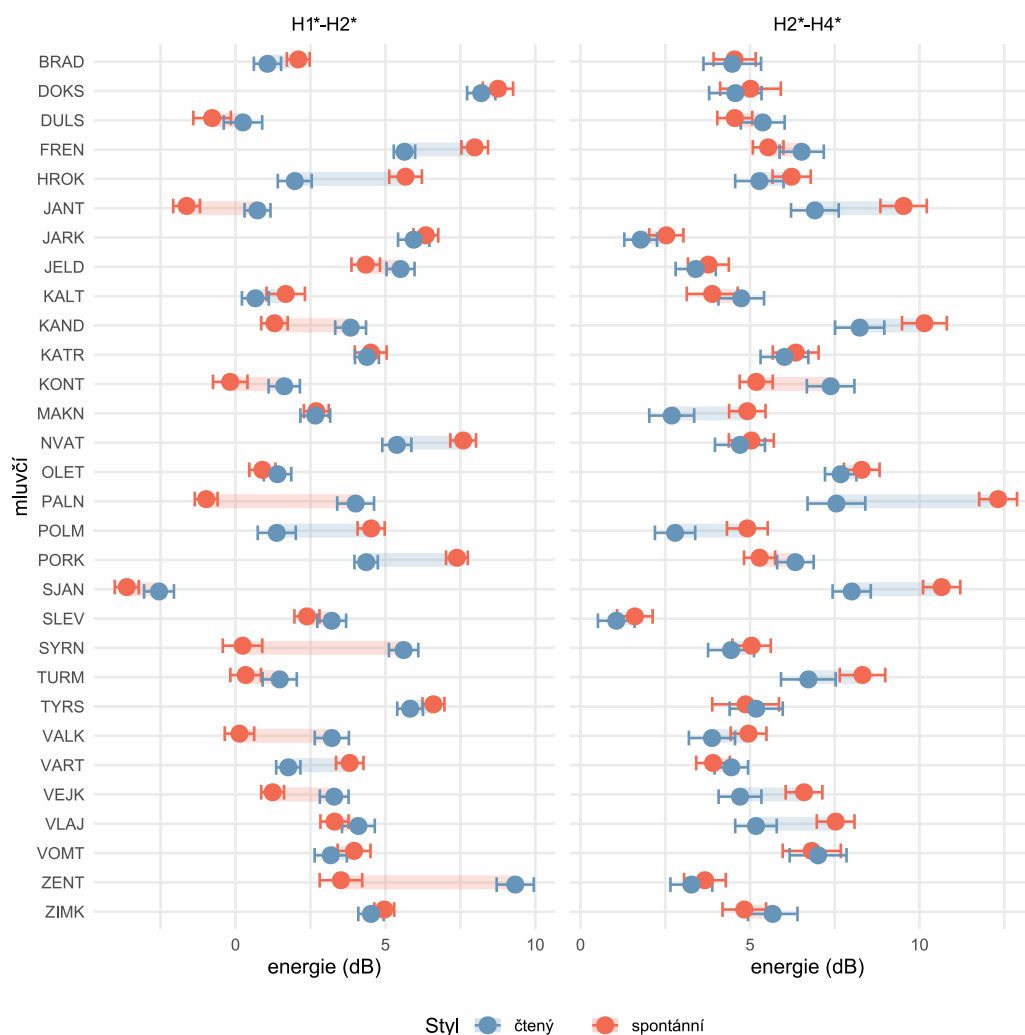
5.2.2 Výsledky

Obrázky 5.2, 5.3 a 5.4 předkládají hodnoty krátkodobých parametrů spektrálního sklonu u jednotlivých mluvčích. Jak je z těchto zobrazení patrné, vzorce sledovaných parametrů se u jednotlivců výrazně rozcházejí. U většiny mluvčích hodnoty parametrů zůstávají stabilní nezávisle na mluvním stylu, nicméně tento závěr není zobecnitelný pro všechny jednotlivce; nadto není případná přítomnost či absence rozdílu mezi mluvními styly jednotná u všech parametrů, což naznačuje, že mezi jednotlivými parametry neexistuje přímočarý vztah. Např. u mluvčího ZIMK zůstávají bez výrazných rozdílů hodnoty $H1^*-H2^*$, $H2^*-H4^*$ i $H1^*-A1^*$, na druhou stranu u $H1^*-A2^*$ a $H1^*-A3^*$ lze u tohoto mluvčího pozorovat nápadné posuny. U mluvčího PALN jsou patrné rozdíly mezi čteným a spontánním projevem u $H1^*-H2^*$ a $H2^*-H4^*$, u ostatních parametrů nikoliv. V případě mluvčích VOMT a TYRS patrně není přítomen rozdíl mezi mluvními styly u žádného z parametrů, zatímco u mluvčího VALK jisté posuny pozorovat lze, s výjimkou parametru $H2^*-H4^*$.

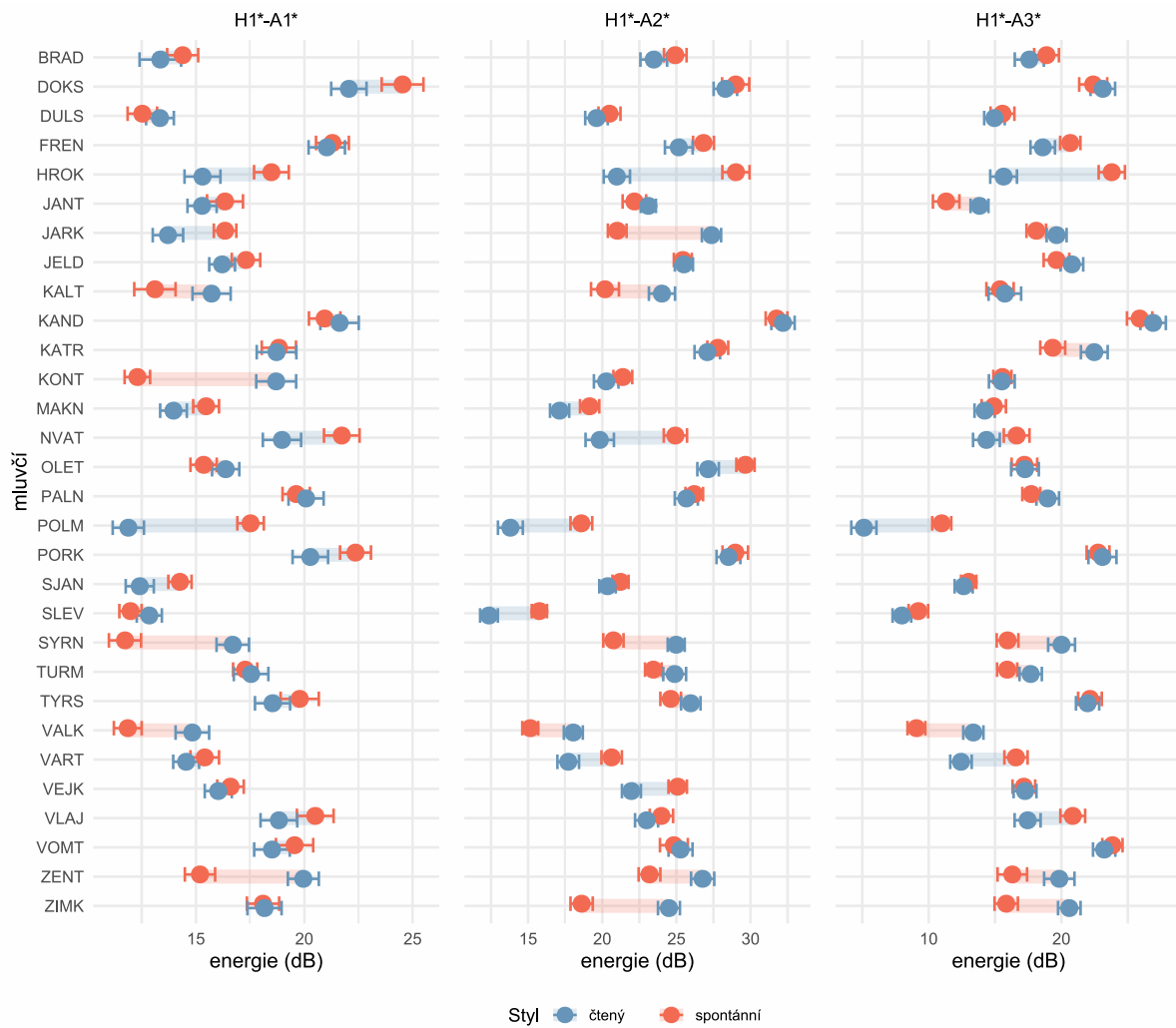
Obecně je možné usuzovat, že na základě určitých krátkodobých parametrů spektrálního sklonu by bylo možné rozlišit některé mluvčí, neboť jsou u nich přítomny značné interindividuální rozdíly, které výrazně přesahují rozdíly intraindividuální. Krátkodobý spektrální sklon by tedy mohl podávat určité informace o individualitě mluvčího, a to zejména v případě, že by se jeho hodnoty nápadně vymykaly populačnímu standardu. Je nicméně třeba nespouštět ze zřetele případný vliv mluvního stylu, který lze u některých mluvčích pozorovat, a zmíněnou nejednotnost rozdílů mezi mluvními styly napříč parametry: např. mluvčí BRAD a DOKS vykazují značně odlišné hodnoty parametru $H1^*-H2^*$, což ovšem již neplatí pro parametr $H2^*-H4^*$, zatímco u mluvčích KALT a KAND nastává opačná situace.

Na základě získaných výsledků nelze jednoznačně určit, který z krátkodobých parametrů spektrálního sklonu nese největší idiosynkratický potenciál.

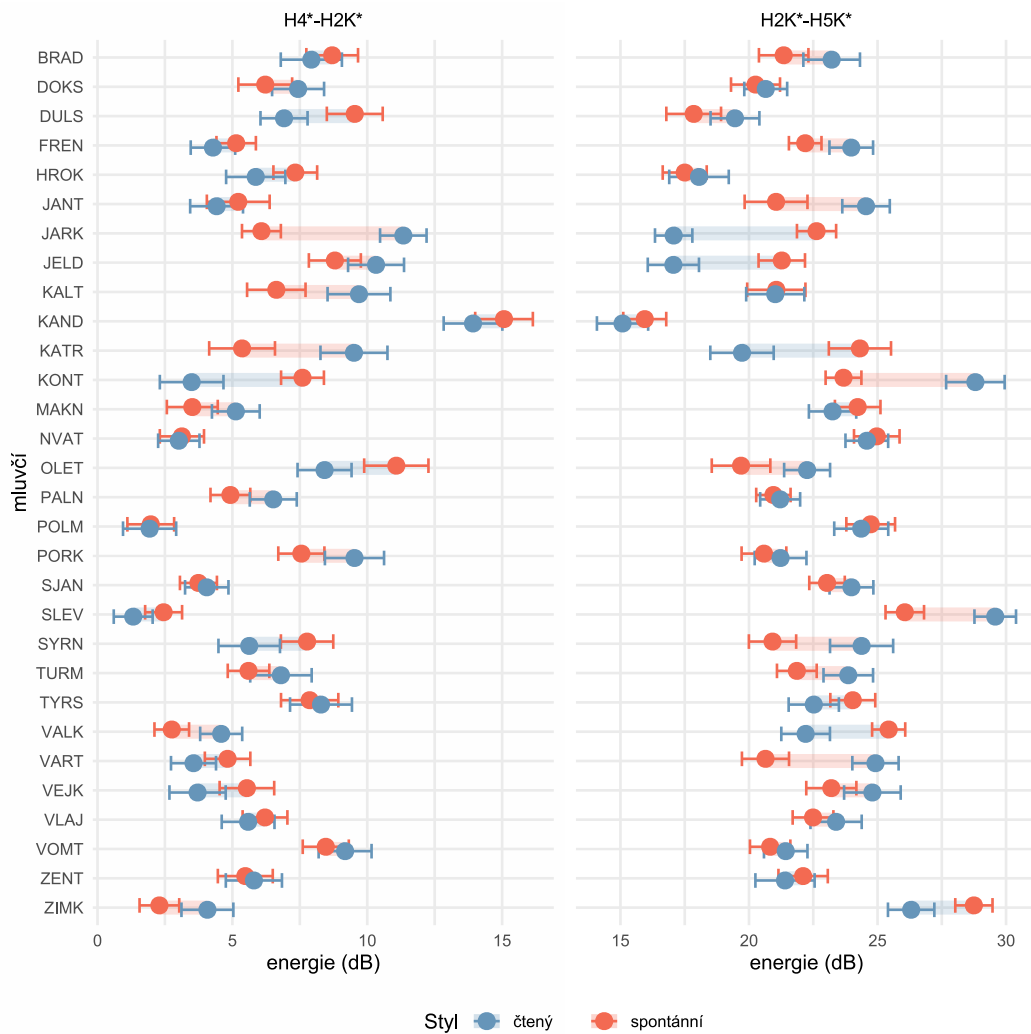
Nejnápadnější překryvy hodnot napříč mluvčími jsou pozorovatelné u parametru $H4^*-H2kHz^*$ a $H2kHz^*-H5kHz^*$ (viz obr. 5.4), pro vzájemné odlišení jednotlivců se tedy jako výhodnější jeví ostatní sledované ukazatele.



Obrázek 5.2. Krátkodobé ukazatele spektrálního sklonu založené na rozdílech mezi energiemi dvou harmonických složek daného řádu $H1^*-H2^*$ a $H2^*-H4^*$ u jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvových úseků, linky znázorňují 95% konfidenční intervaly.



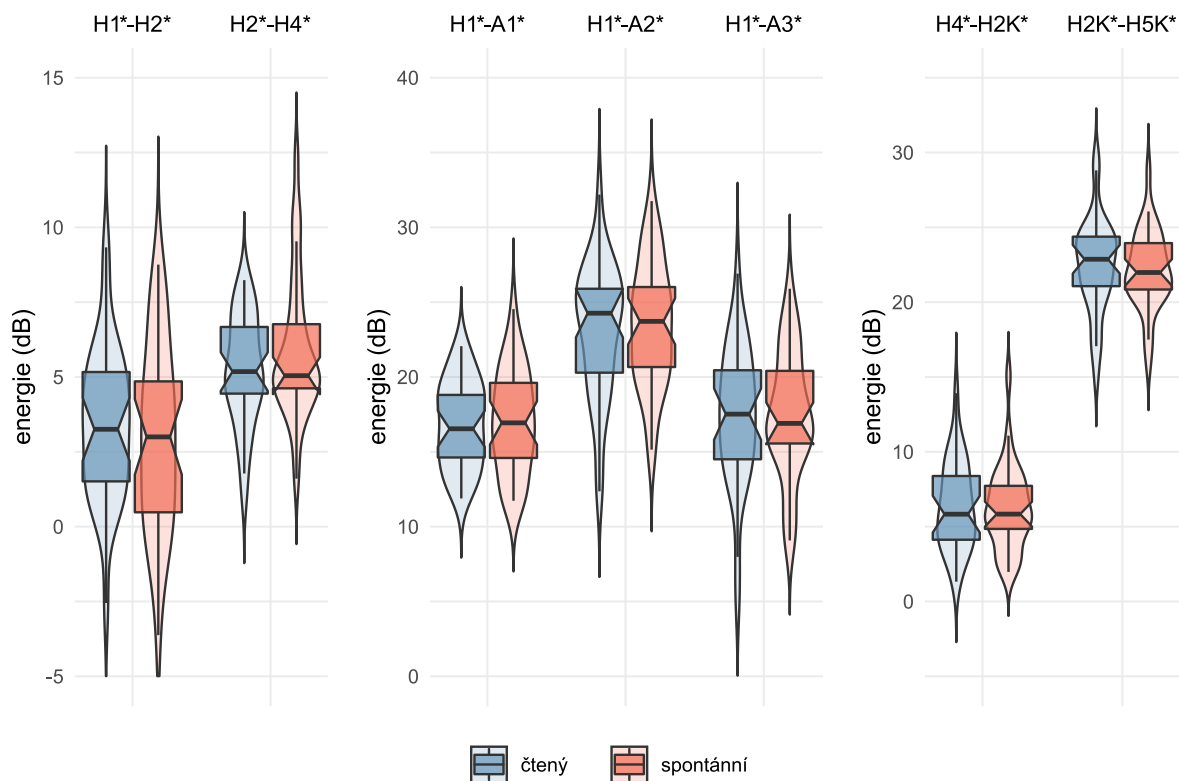
Obrázek 5.3. Krátkodobé ukazatele spektrálního sklonu založené na rozdílech mezi energiemi první harmonické složky a harmonické složky v oblasti určitého formantu $H1^*-A1^*$, $H1^*-A2^*$ a $H1^*-A3^*$ u jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvových úseků, linky znázorňují 95% konfidenční intervaly.



Obrázek 5.4. Krátkodobé ukazatele spektrálního sklonu založené na rozdílu mezi energiemi čtvrté harmonické složky a harmonické složky nejbližší 2 kHz ($H4^*-H2K^*$) a na rozdílu harmonických složek nejbližší 2 kHz a 5 kHz ($H2K^*-H5K^*$) u jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvových úseků, linky znázorňují 95% konfidenční intervaly. (Pro účely zobrazení je kvůli zjednodušení využito označení „K“ namísto „kHz“.)

V případech, kdy jsou znatelné rozdíly daného parametru mezi čteným a spontánním projevem, nelze hovořit o konzistentní tendenci napříč mluvčími k vyšším hodnotám v jednom mluvním stylu ve srovnání s druhým. Při srovnání průměrných hodnot napříč mluvčími nebyly u žádného ze sledovaných parametrů pozorovány významné rozdíly mezi čteným a spontánním projevem, což ilustruje obrázek 5.5 a dokladem jsou také výsledky Wilcoxonova testu ($p \in \langle 0,021; 0,777 \rangle$).

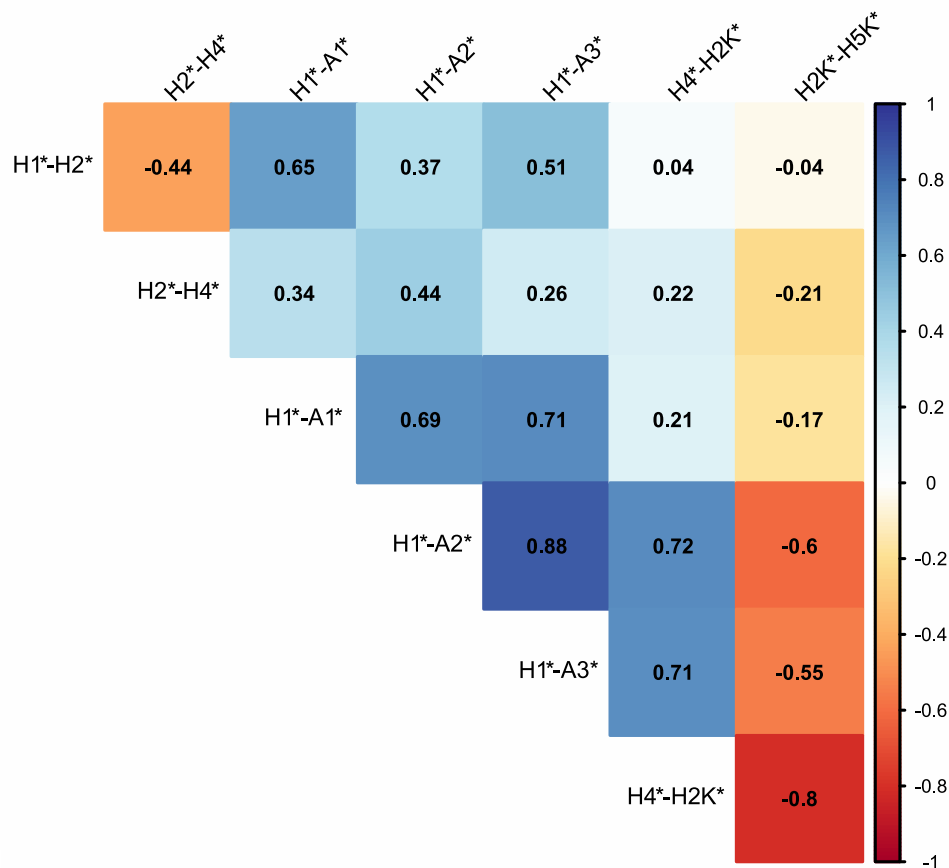
Konkrétní hodnoty sledovaných parametrů frekvenčního spektra naměřené v obou mluvních stylech obsahuje příloha B.



Obrázek 5.5. Srovnání průměrných hodnot všech sledovaných krátkodobých ukazatelů spektrálního sklonu napříč mluvčími ve čteném a spontánním projevu. (Pro účely zobrazení je kvůli zjednodušení využito označení „K“ namísto „kHz“.)

Jak již bylo řečeno, při porovnání krátkodobých parametrů spektrálního sklonu u jednotlivých mluvčích není patrná konzistentní spojitost mezi hodnotami různých ukazatelů. Korelační matice, kterou předkládá obrázek 5.6, nicméně vypovídá o silné korelaci mezi některými ze sledovaných parametrů. Nejvyšší korelační koeficient odpovídající velmi silné korelaci byl nalezen u parametrů $H1^*-A2^*$ a $H1^*-A3^*$ ($r = 0,88$) a $H4^*-H2kHz^*$ a $H2kHz^*-H5kHz^*$ ($r = -0,8$), silná korelace se pak objevila také u řady dalších dvojic parametrů (jediným ukazatelem, který nevykázal silnou korelaci s žádným z ostatních, je $H2^*-H4^*$). Jakási spojitost mezi hodnotami těchto parametrů tedy patrně existuje, jakkoliv není při porovnání naměřených hodnot napříč řečovými vzorky zcela zřetelná.

Podle Garelleka et al. (2013), kteří měřili ukazatele H1-H2, H2-H4, H4-H2kHz a H2kHz-H5kHz, existuje určitý vztah mezi hodnotami spektrálního sklonu v sousedících frekvenčních pásmech: pakliže je v daném pásmu nalezen strmější spektrální sklon, v pásmu následujícím se s vyšší pravděpodobností objeví sklon mírnější než opět plošší a naopak (Garellek et al., 2013: 4). S těmito závěry koreponduje velmi silná negativní korelace ukazatelů H4*-H2kHz* a H2kHz*-H5kHz* ($r = -0,8$). Mezi ukazateli H1*-H2* a H2*-H4* byla nalezena taktéž negativní, avšak pouze středně silná korelace ($r = -0,44$). U zbývajících dvojic, tedy H2*-H4* a H4*-H2kHz* tento trend pozorován nebyl, korelace v tomto případě byla slabě pozitivní ($r = 0,22$).



Obrázek 5.6. Korelační matice sledovaných krátkodobých parametrů spektrálního sklonu.

5.2.3 Diskuse a závěr

V rámci experimentu 3 byly analyzovány vybrané parametry spektrálního sklonu vokálních segmentů: $H1^*-H2^*$, $H2^*-H4^*$, $H1^*-A1^*$, $H1^*-A2^*$, $H1^*-A3^*$, $H4^*-H2kHz^*$ a $H2kHz^*-H5kHz^*$. Byla zkoumána míra jejich variability napříč mluvčími a také jejich případné rozdíly mezi čteným a spontánním projevem.

Výsledky provedeného experimentu ukázaly, že hodnoty jednotlivých krátkodobých parametrů spektrálního sklonu zůstávají u řady mluvčích stabilní nezávisle na mluvním stylu a zejména v případě jedinců, kteří se výrazně vzdalují hodnotám standardním pro danou populaci, je lze považovat za vhodné ukazatele pro rozlišování mluvčích. Využitý materiál ovšem obsahoval i nahrávky jednotlivců, u kterých byly pozorovány zřetelné posuny některých sledovaných parametrů mezi mluvními styly, přičemž tyto rozdíly nebyly napříč mluvčími konzistentní – u různých mluvčích docházelo k posunům různých parametrů. Při využití krátkodobých parametrů spektrálního sklonu pro charakterizaci mluvčího je tudíž třeba vzít v úvahu případný vliv mluvního stylu, který patrně v případě velké části mluvčích není příliš markantní, avšak nelze jej zcela vyloučit.

Na základě získaných výsledků nelze s jistotou stanovit, které ze sledovaných parametrů nesou největší idiosynkratický potenciál; Vaňková a Skarnitzl (2014) jej spatřují u ukazatelů $H1^*-H2^*$ a $H1^*-A1^*$, které se v rámci jejich studie projevíly jakožto nejstabilnější pro mluvčího a zároveň nejvíce variující napříč jednotlivci. Také na základě zde předložených výsledků lze konstatovat, že tyto parametry podávají důležité informace o identitě mluvčího, nicméně jejich převahu nad zbývajícím parametry není možné jednoznačně potvrdit. Výjimku představují parametry odrážející spektrální sklon ve vyšších frekvenčních pásmech, $H4^*-H2kHz^*$ a $H2kHz^*-H5kHz^*$, u nichž byly pozorovány značné překryvy hodnot napříč mluvčími, a pro jejich vzájemné rozlišování se tudíž jeví jako méně přínosné.

U žádného ze sledovaných parametrů nebyl zjištěn významný rozdíl mezi čteným a spontánním projevem nezávisle na mluvčím; ačkoliv se tedy u určitých jednotlivců objevily posuny některých parametrů v závislosti na mluvním stylu, nejednalo se o posuny konzistentní – byly pozorovány jen u určitých mluvčích a v obou směrech.

U většiny parametrů nebyly pozorovány přímočaré vzájemné vztahy, tj. hodnoty spektrálního sklonu v různých frekvenčních pásmech jsou na sobě z větší části nezávislé. Výjimku představovaly ukazatele $H_4^*-H_{2kHz}^*$ a $H_{2kHz}^*-H_{5kHz}^*$, mezi kterými byla nalezena velmi silná negativní korelace – tento výsledek koresponduje se závěry Garelleka et al. (2013), kteří sledovali parametry H_1-H_2 , H_2-H_4 , H_4-H_{2kHz} a $H_{2kHz}-H_{5kHz}$ (tj. nenormalizované varianty parametrů měřených v rámci tohoto experimentu); podle výsledků jejich studie implikuje strmější spektrální sklon v určitém frekvenčním pásmu vyšší pravděpodobnost ploššího spektrálního sklonu v sousedním frekvenčním pásmu a naopak. Méně výrazný byl tento efekt v případě dvojice parametrů $H_1^*-H_2^*$ a $H_2^*-H_4^*$, jejichž negativní korelace byla pouze středně silná; u dvojice $H_2^*-H_4^*$ a $H_4^*-H_{2kHz}^*$ pak tento trend pozorován nebyl.

Jak již bylo zmíněno na začátku kapitoly, spektrální sklon je akustickým korelátorem barvy hlasu, resp. způsobu fonace. Jednotlivé parametry spektrálního sklonu ovšem odrážejí různé aspekty fonačního nastavení; např. parametr $H_1^*-H_2^*$, který odráží rozložení energie v nižších frekvencích, souvisí s poměrem otevření hlasivek (Hanson & Chuang, 1999: 1066), $H_1^*-A_3^*$ je ovlivňován mluvním úsilím (Mooshammer, 2010), ukazatel H_2-H_4 pak společně s H_1-H_2 ¹⁹ koresponduje s vnímanou dyšností (Kreiman & Garellek, 2011). Lze předpokládat, že krátkodobé parametry spektrálního sklonu, které vzájemně silně, případně velmi silně

¹⁹ V odkazované publikaci byly využity tyto ukazatele bez normalizace, nejsou tudíž označeny hvězdičkou.

korelují, mají obdobnou spojitost s fonačním nastavením při řečové produkci i percepčními charakteristikami řečového signálu.

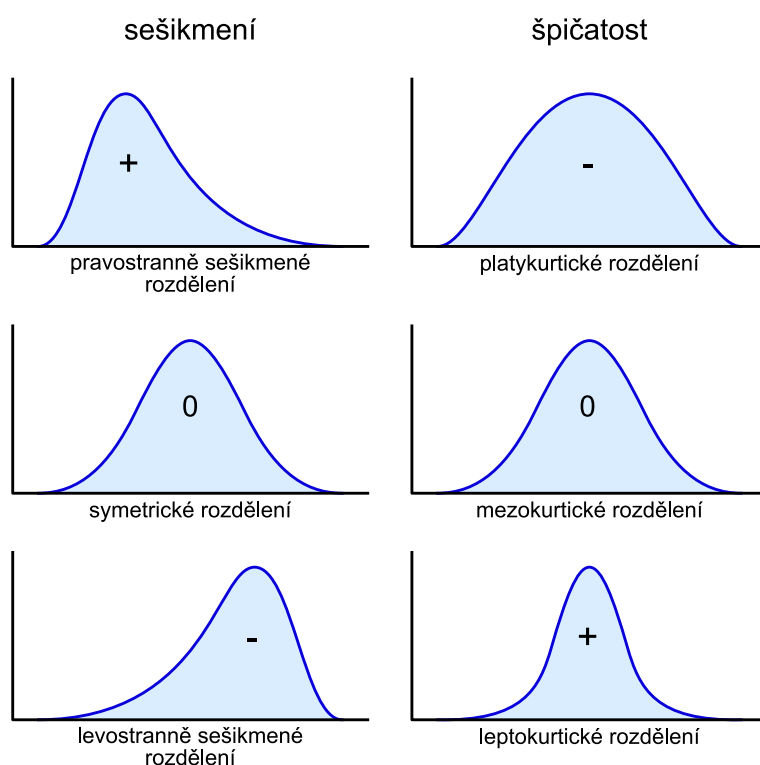
Výsledky experimentu 3 poukazují na značnou vzájemnou nezávislost různých krátkodobých parametrů spektrálního sklonu, a také studie zaměřené na jejich spojitost s fonačním nastavením i řečovou percepcí popisují, že jednotlivé parametry korespondují s různými produkčními i percepčními faktory (viz výše). Při charakterizaci spektrálního sklonu jakož i jeho vztahu k fonačním charakteristikám a percepčním vlastnostem hlasu je tudíž vhodné upřesnit frekvenční oblast, v jejímž rámci je spektrální sklon sledován, neboť pojetí spektrálního sklonu jakožto homogenní skupiny parametrů může být značně zavádějící.

5.3 Spektrální vlastnosti frikativ

Akustická podoba frikativních hlásek se vyznačuje přítomností turbulentního šumu, jehož rozložení ve frekvenčním spektru závisí na místě artikulace daného segmentu, a v případě znělých frikativ je součástí signálu také základní frekvence (podrobněji viz např. Stevens, 1998, citováno v: Jongman et al., 2000: 1253).

Studie zaměřující se na tvar spektra frikativních segmentů využívají k jeho parametrizaci řadu ukazatelů. Mezi nejčastější z nich patří skupina parametrů označovaných souhrnně jako spektrální momenty: jedná se o spektrální těžiště (*center of gravity*; CoG), směrodatnou odchylku (*standard deviation*; SD), sešikmení a špičatost. Jedná se o parametry obecně využívané k popisu tvaru distribucí. Hodnota spektrálního těžiště odpovídá frekvenci, která rozděluje spektrum (resp. rozdělení) na dvě části se stejnou souhrnnou energií, směrodatná odchylka vypočítává o míře rozptylu hodnot kolem průměru, sešikmení reflektuje míru asymetrie rozložení energie ve spektru, tj. četnější výskyt odlehlých hodnot směrem k vyšším frekvencím (pozitivní sešikmení) či nižším frekvencím (negativní

sešikmení) směrem od střední hodnoty, a špičatost vyjadřuje, nakolik je energie ve spektru nakumulována kolem střední hodnoty (kladná špičatost) či naopak rozprostřena napříč celým spektrem (záporná špičatost) ve srovnání s normálním rozdělením – ilustraci sešikmení a špičatosti poskytuje obrázek 5.7.



Obrázek 5.7. Tvary rozdělení – sešikmení a špičatost.

Právě na potenciál spektrálních parametrů frikativních segmentů při rozlišování mluvcích se zaměřila Kavanagh (2012), která sledovala jejich hodnoty v realizacích fonému /s/ v britské angličtině. Pro jejich extrakci využila frekvenční pásmo 0–8 kHz; zvolená horní hranice je poměrně vysoká, neboť sykavky se vyznačují šumem ve vysokých frekvencích, a jejich odfiltrováním by tedy mohlo dojít ke ztrátě důležitých informací. K ještě vyšší hranici pásmové propusti se právě z tohoto důvodu uchýlili např. Czaplicki et al. (2016), kteří analyzovali spektrální momenty ve frekvenčním pásmu do 11 kHz, či Jannedy a Weirich (2017), v jejichž experimentu zaměřeném rovněž na spektrální momenty bylo využito frekvenční

pásmo až do 12 kHz. Podle výsledků experimentu Kavanagh (2012) se faktor mluvčího projevil jako statisticky signifikantní u všech měřených spektrálních momentů, považuje je tedy za potenciálně užitečné parametry pro rozlišování mluvčích.

Jannedy a Weirich (2017) sledovaly rozdíly ve spektru realizací fonémů /ç/ a /ʃ/ ve třech dialektech němčiny. Jak již bylo řečeno, umístily autorky horní hranici analyzovaného frekvenčního pásma do 12 kHz. Zároveň aplikovaly horní propust od 0,5 kHz, a to za účelem vyloučení vlivu případného přeznívání hlasivkového tónu ze sousedních hlásek. Vedle spektrálních momentů využily pro parametrizaci frekvenčního spektra další soubor parametrů odrážejících celkový tvar spektrální obálky, tzv. DCT koeficienty (*discrete cosine transformation*; DCT), přičemž navazovaly na metodologii aplikovanou ve studiích Guzika a Harringtona (Guzik & Harrington, 2007; Harrington, 2010; obojí citováno v: Jannedy & Weirich, 2017: 396), kteří jako první popsali metodu parametrizace tvaru frikativního spektra tímto způsobem. Stejně jako spektrální momenty i DCT koeficienty postihují tvar spektra pomocí několika hodnot: DCT₁ odráží celkový sklon spektra, DCT₂ jeho zakřivení a DCT₃ vypovídá zejména o amplitudách ve vyšších frekvencích (DCT₀ odpovídá průměrné amplitudě spektra, nelze jej tedy považovat za relevantní ukazatel při srovnávání nahrávek, neboť jeho hodnota silně závisí na okolnostech nahrávání). Jmenovaná studie se týkala rozdílů mezi sledovanými frikativami v daných dialektech a tomu, nakolik dané parametry korespondují s percepcí frikativ i vizuální podobou jejich frekvenčního spektra; autorky dospěly k závěru, že ve srovnání se spektrálními momenty je pro tyto účely vhodnější využití DCT koeficientů. Na druhou stranu výzkum zaměřený na intraindividuální ani interindividuální variabilitu těchto parametrů patrně dosud nebyl proveden.

Poměrně zhusta se v dostupné literatuře vyskytujícím parametrem je hodnota frekvence, která se ve spektru vyznačuje nejvyšší amplitudou. Do svých

experimentů jej zařadili např. Jesus a Shadle (2002), Van der Harst et al. (2007) či Czaplicki et al. (2016). I tento parametr bývá měřen v rámci různých frekvenčních pásem. Czaplicki et al. (2016) analyzovali polské sykavky, a to znělé i neznělé, a extrahovali spektrální vrchol ve frekvenčním pásmu mezi 20 Hz a 11 kHz; Jesus a Shadle (2002) se pak zaměřili na nejsilnější frekvenci ve spektru portugalských frikativ dokonce v rozsahu do 20 kHz, ovšem s vyloučením f_0 a jejich harmonických složek u znělých segmentů. Naproti tomu Van der Harst et al. (2007) měřili frekvenci s maximální amplitudou v realizacích holandského /ʁ/ pouze v pásmu 0,8–4 kHz (ovšem tato velární frikativa se obecně nevyznačuje frikativním šumem v tak vysokých frekvencích jako např. frikativy alveolární). Jesus a Shadle (2002) dále měřili také parametr označovaný jako dynamická amplituda, odpovídající rozdílu mezi maximální amplitudou v pásmu 0,5–20 kHz (spodní hranice byla takto opět stanovena kvůli vyloučení vlivu f_0) a minimální amplitudou v pásmu 0–2 kHz. Studie jmenované v tomto odstavci se ovšem nezaměřovaly na variabilitu sledovaných parametrů napříč mluvčími či v rámci projevu jednoho mluvčího, a nelze tedy na základě jejich poznatků formulovat závěry ohledně potenciálu oněch parametrů v této oblasti.

Při charakterizaci tvaru frekvenčního spektra frikativ se uplatňuje také spektrální sklon, tedy proměnná, která je typická spíše pro analýzu spektra vokálů, případně dlouhodobého spektra v průběhu řeči (viz oddíly 5.1 a 5.6). Měření spektrálního sklonu uplatnili např. Jesus a Shadle (2002), kteří zkoumali portugalské labiodentální, alveolární a postalveolární frikativy, a to jak znělé, tak neznělé. Pro účely měření spektrálního sklonu sledovali frekvenční pásmo od 0,5 do 20 kHz, které rozdělili pivotem, stanoveným pro jednotlivé frikativy podle místa artikulace následujícím způsobem: ve všech sledovaných segmentech byla změřena frekvence s nejvyšší amplitudou, načež byla pro každé místo artikulace vypočítána její průměrná hodnota a následně zaokrouhlena na nejbližší celý kHz. Pivotem pro realizace fonémů /f v/ byla frekvence 5 kHz, pro /s z/ 6 kHz a pro /ʃ ʒ/ 4 kHz. Pro každé z takto rozdělených frekvenčních pásem pak byl stanoven

spektrální sklon, a to proložením regresní přímkou. Jejich postup pak napodobili např. Czaplicki et al. (2016) při analýze polských sykavek. Ani u tohoto způsobu parametrizace frekvenčního spektra frikativ, na rozdíl od vokálů, dosud zřejmě nebyl prozkoumán jeho idiosynkratický potenciál.

5.4 Spektrální vlastnosti sonorních konsonantů

Sonorní konsonanty, které – na rozdíl od obstruentů – neobsahují šumovou složku, se vyznačují akustickými vlastnostmi podobnými vokálům. Nabízí se u nich tedy podobné způsoby parametrizace jako právě u vokálů, nicméně dostupné studie popisují spíše analýzu jejich spektra prostřednictvím spektrálních momentů, které obvykleji nacházejí využití při analýze spektra frikativ (viz výše) – pro parametrizaci spektrálních vlastností sonorních konsonantů je využili např. Gottsmann a Harwardt (2011) či Themistocleous et al. (2022).

Kavanagh (2012) se ve své disertační práci zaměřené na akustické parametry využívané při forenzním srovnávání mluvčích mimo jiné věnovala parametrizaci spektra anglických sonor [m n ŋ l], a to prostřednictvím spektrálního těžiště, směrodatné odchylky a hodnot frekvencí s maximální a minimální amplitudou. V případě laterální aproximanty upozorňuje na její artikulační variabilitu, a to zejména v závislosti na kontextu – podle pozice v rámci slabiky se v angličtině vyskytuje alveolární aproximanta či její velarizovaná varianta. Kavanagh (2012) ve své práci potvrzuje Nolanův (1997) předpoklad, že jsou to nazální hlásky, které by mohly nést značný idiosynkratický potenciál, neboť při jejich artikulaci vzniká rezonance nejen v ústní, ale také v nosní dutině, a ta je při řeči ve srovnání s dutinou ústní relativně nehybná, tudíž by akustické vlastnosti nazál také mohly zůstat poměrně stabilní (Nolan, 1997: 750–751). Výsledky experimentu Kavanagh (2012) skutečně ukazují, že nazální konsonanty [m] a [n] společně s frikativou [s] vykazují značný potenciál coby ukazatele při rozlišování mluvčích,

zatímco spektrální vlastnosti [l] se ukázaly jako nejméně idiosynkratické. Z jednotlivých měřených parametrů napříč segmenty hodnotí autorka jako nejpřínosnější z hlediska možnosti využití při identifikaci mluvčího spektrální těžiště a směrodatnou odchylku.

Je namístě podotknout, že zatímco publikací věnovaných spektrálním charakteristikám vokálů či frikativ a jejich idiosynkratičnosti je dostupné poměrně hojné množství, problematika podoby frekvenčního spektra sonorních konsonantů a jeho vztahu k identitě mluvčího je zkoumána spíše okrajově.

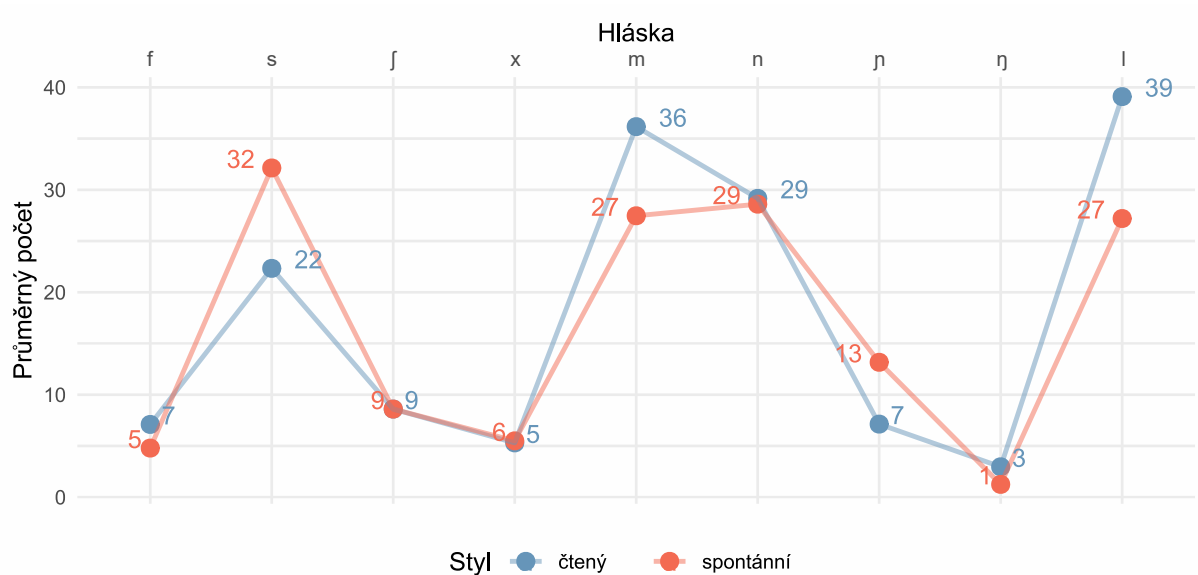
5.5 Experiment 4: Parametry frekvenčního spektra konsonantů

5.5.1 Metoda

Za využití materiálu popsaného v kapitole 2 byly sledovány spektrální momenty, tj. spektrální těžiště, směrodatná odchylka (SD), sešikmení a špičatost, u vybraných konsonantů. První sledovanou skupinu konsonantů tvořily neznělé frikativy, u kterých je zkoumání tvaru frekvenčního spektra hojně zastoupeno v dostupné literatuře (viz oddíl 5.3), druhou skupinu pak představovaly nazály společně s laterální aproximantou [l], jejichž idiosynkratickým potenciálem se zabývala např. Kavanagh (2012); v případě sonorních segmentů ovšem byly v souladu se jmenovanou studií sledovány pouze spektrální těžiště a směrodatná odchylka.

Pro analýzu ovšem nebyly využity všechny segmenty spadající do jmenovaných tříd, které se v češtině vyskytují, a to vzhledem k jejich četnosti ve využitém materiálu; jak ukazuje obrázek 5.8, realizace určitých segmentů se v nahrávkách vyskytovaly v příliš nízkých počtech, a generalizace výsledků získaných na jejich základě by tudíž byla problematická. Analýze frekvenčního spektra tedy byly

podrobeny pouze hlásky, jejichž průměrný počet v řečových vzorcích přesahoval 20 výskytů; jednalo se o jedinou neznělou frikativu [s], dvě nazály [m] a [n] a laterální aproximantu [l].



Obrázek 5.8. Průměrný počet jednotlivých neznělých frikativ, nazál a laterálních aproximant ve využitých řečových vzorcích.

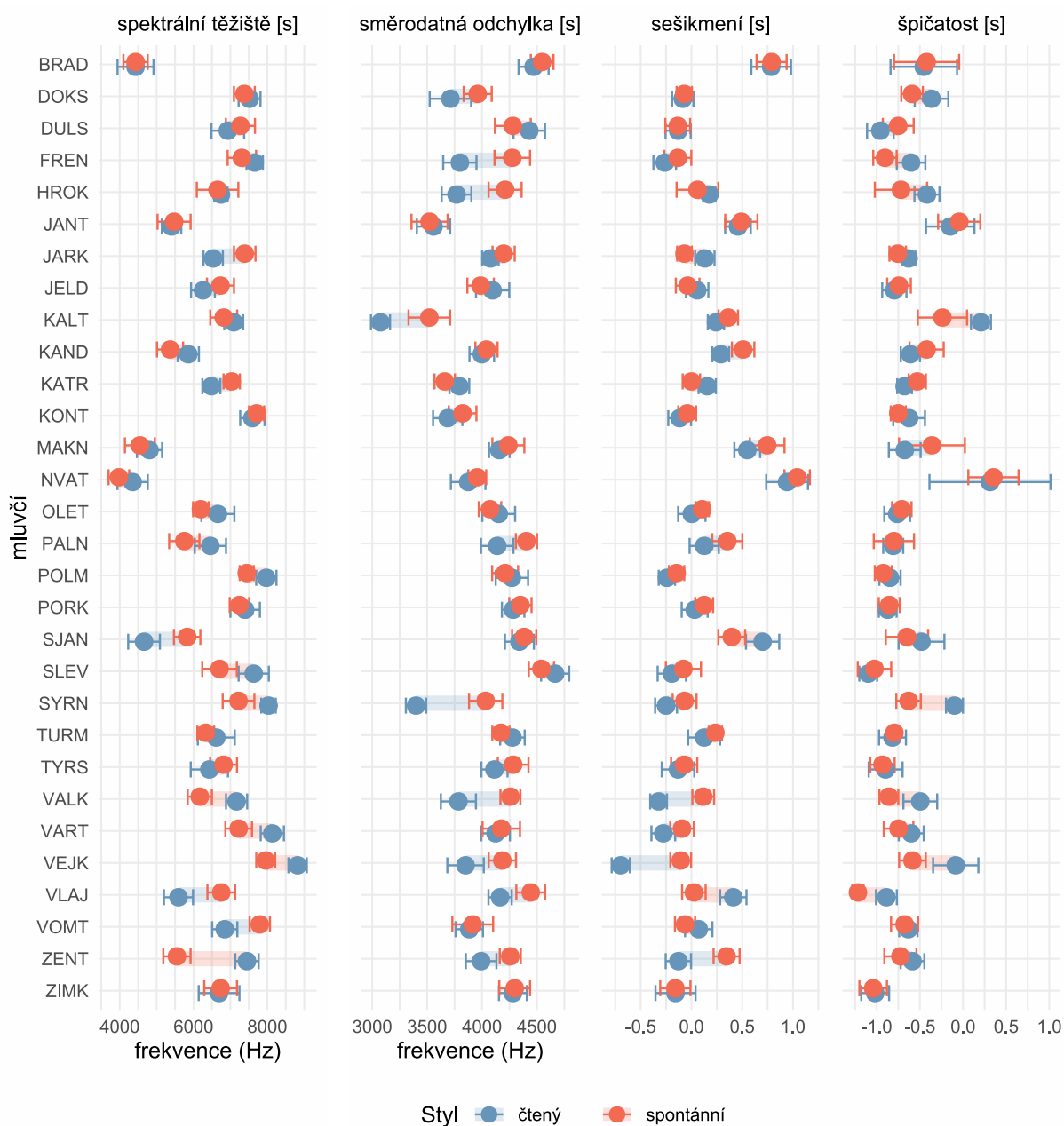
V programu Praat bylo na základě segmentace na hlásky (podrobněji viz kapitolu 2) v každém sledovaném konsonantu identifikováno prostředních 20 ms a jimi určené hranice pak byly posunuty na nejbližší průchod zvukové vlny nulou směrem k temporálnímu středu segmentu. Takto vymezený úsek byl následně vyříznut pomocí obdélníkového okna a prostřednictvím funkce *To Spectrum (fft)* pro něj bylo vygenerováno frekvenční spektrum, v němž byly následně vypočítány spektrální momenty (viz výše), a to v celém spektru, v němž byly obsaženy frekvence do 16 kHz (všechny nahrávky byly pořízeny se vzorkovací frekvencí 32 kHz; viz kapitolu 2).

Pro každý řečový vzorek byly vypočítány průměrné hodnoty jednotlivých parametrů (tj. všech spektrálních momentů [s] a spektrálního těžiště a směrodatné odchylky [m], [n] a [l]) a jejich konfidenční intervaly, a následně u nich byla

sledována míra variability mezi různými mluvčími a dvěma sledovanými mluvčími styly – spontánním a čteným. (U konfidenčních intervalů nebyla provedena korekce pro opakovaná měření, odpovídají hladině významnosti 0,05 – jejich zobrazení tedy poskytuje orientační informace o spolehlivosti naměřených hodnot v jednotlivých řečových vzorcích, nicméně nepředstavuje základ pro usuzování, zda jsou rozdíly mezi vzorky statisticky signifikantní.) Průměrné hodnoty jednotlivých parametrů byly porovnávány ve čteném a spontánním projevu pomocí Wilcoxonova testu (podrobněji viz kapitolu 2), v němž byla hladina významnosti korigována počtem sledovaných parametrů: $\alpha = 0,05/10 = 0,005$.

5.5.2 Výsledky

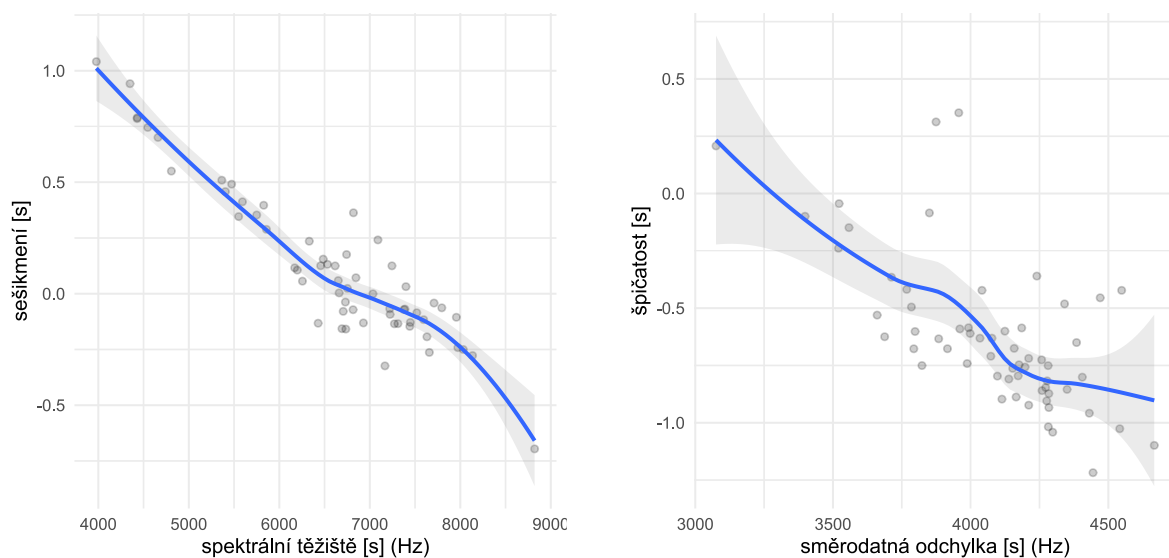
Jak je patrné z obrázku 5.9, spektrální momenty hlásky [s] u většiny mluvčích nedoznávají výrazných posunů v závislosti na mluvním stylu a mezi určitými jednotlivci existují natolik markantní rozdíly, že by na základě spektrálních momentů patrně bylo možné je od sebe vzájemně odlišit. Jako nejslibnější se v tomto směru jeví spektrální těžiště a sešikmení [s] – např. u mluvčích BRAD a DOKS jsou hodnoty těchto parametrů značně odlišné, přičemž zůstávají bez markantních změn v závislosti na mluvním stylu.



Obrázek 5.9. Spektrální momenty naměřené v konsonantu [s] u jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvovacích úseků, linky znázorňují 95% konfidenční intervaly.

Mezi spektrálním těžištěm a sešikmením [s] existuje velmi silná negativní korelace, jejich korelační koeficient je roven -0,93 (viz obrázek 5.15) – tj. spektrální těžiště o vyšší frekvenci se pojí s nižším koeficientem sešikmení. Podrobnější informace poskytuje obrázek 5.10, ze kterého je patrné, že nulový koeficient

sešikmení, který implikuje symetrické rozložení hodnot kolem těžiště, koresponduje s hodnotami spektrálního těžiště kolem 7000 Hz²⁰, přičemž s narůstající hodnotou spektrálního těžiště klesá koeficient špičatosti (podrobněji viz oddíl 5.3). Vzhledem k tomuto vztahu mezi spektrálním těžištěm a sešikmením spektra [s] lze konstatovat, že při využití spektrálních momentů pro vzájemné rozlišování mluvčích se zahrnutí obou parametrů jeví jako redundantní. Silně pak koreluje směrodatná odchylka a špičatost [s] ($r = -0,67$; viz obrázek 5.15), které taktéž zachycuje obrázek 5.10.

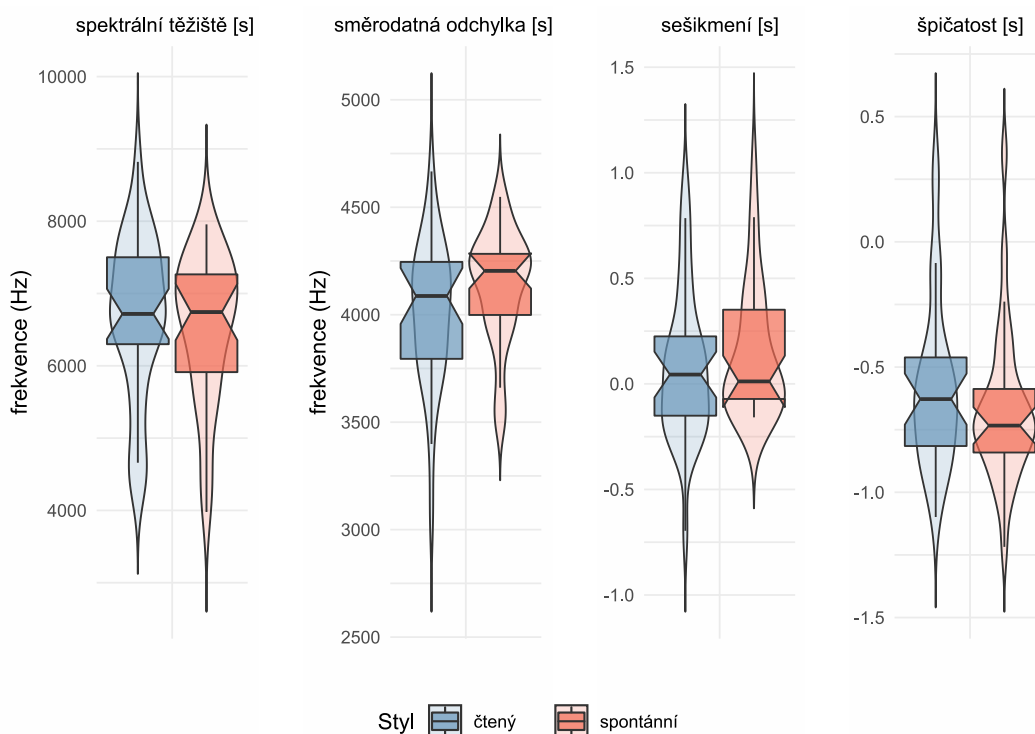


Obrázek 5.10. Závislost koeficientu sešikmení na spektrálním těžišti [s] (vlevo) a závislost koeficientu špičatosti na směrodatné odchylce [s] (vpravo). Body reprezentují průměrné hodnoty v jednotlivých řečových vzorcích, modrá křivka charakterizuje výsledný trend, přičemž šedivá oblast představuje její 95% konfidenční interval.

Z celkového srovnání průměrných hodnot spektrálních momentů [s] napříč řečovými vzorky vyplývá, že mezi čteným a spontánním projevem u těchto parametrů neexistují signifikantní rozdíly; vizuální srovnání nabízí obrázek 5.11 a tento závěr podporují také výsledky Wilcoxonova testu ($p > 0,05$, ovšem s výjimkou směrodatné odchylky, u které byla zjištěna p-hodnota 0,00501 – nejedná se tedy

²⁰ Průměrná hodnota spektrálního těžiště napříč řečovými vzorky činila 6600 Hz.

o statisticky signifikantní rozdíl, nicméně p-hodnota se blíží hladině významnosti²¹).

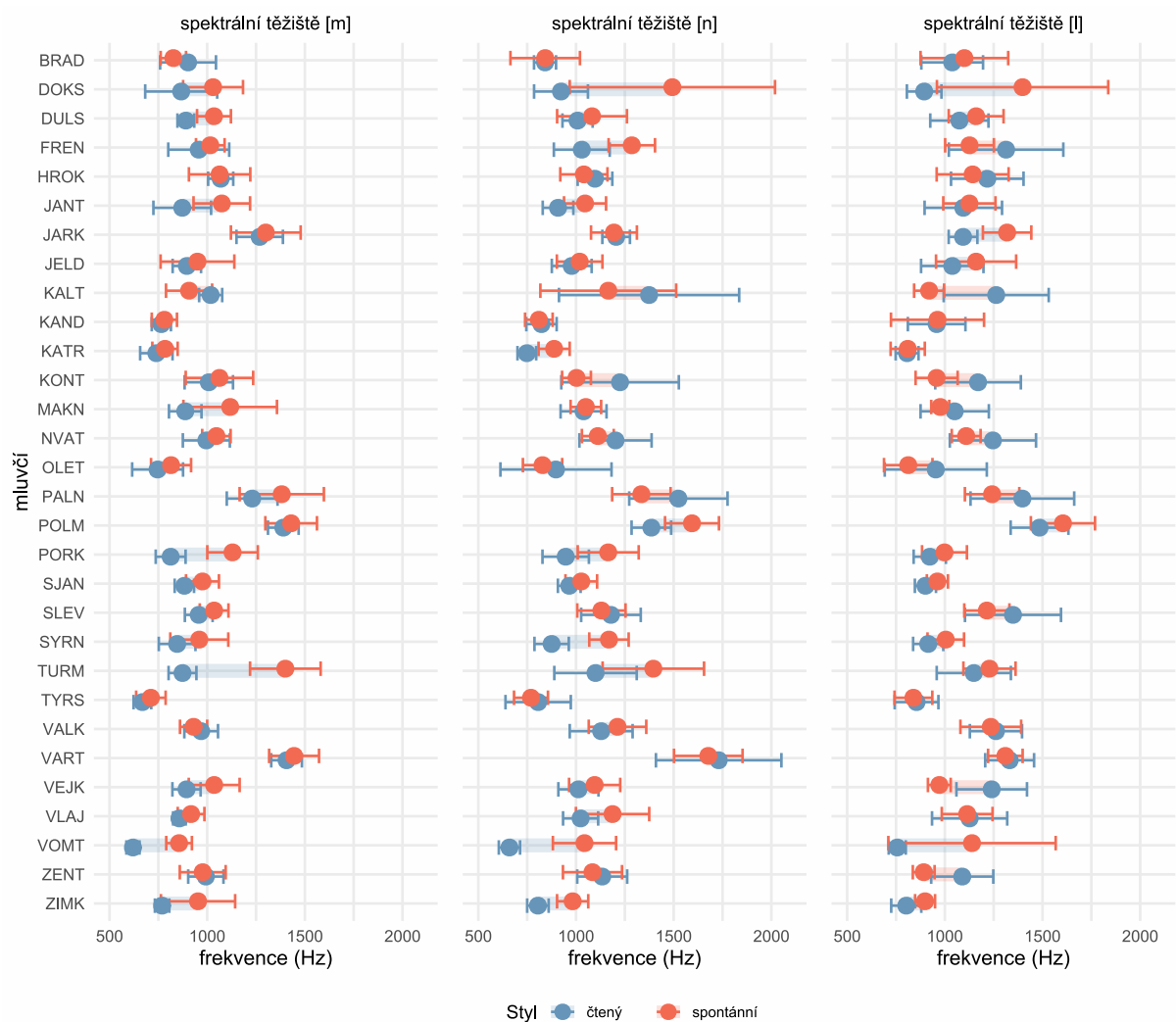


Obrázek 5.11. Průměrné hodnoty spektrálních momentů naměřených v konsonantu [s] napříč mluvčími ve čteném a spontánním projevu.

Hodnoty spektrálního těžiště sonorních konsonantů [m n l] ukazuje obrázek 5.12, ze kterého je patrné, že stejně jako v případě spektrálních momentů [s], ani tento parametr u většiny mluvčích nevykazuje výrazné rozdíly mezi čteným a spontánním projevem. Ve srovnání se spektrálními momenty [s] jsou však přítomny zřetelnější překryvy hodnot napříč mluvčími, a to zejména u spektrálního těžiště laterální aproximanty [l]. Obecně se tedy patrně jedná o parametry s nižším

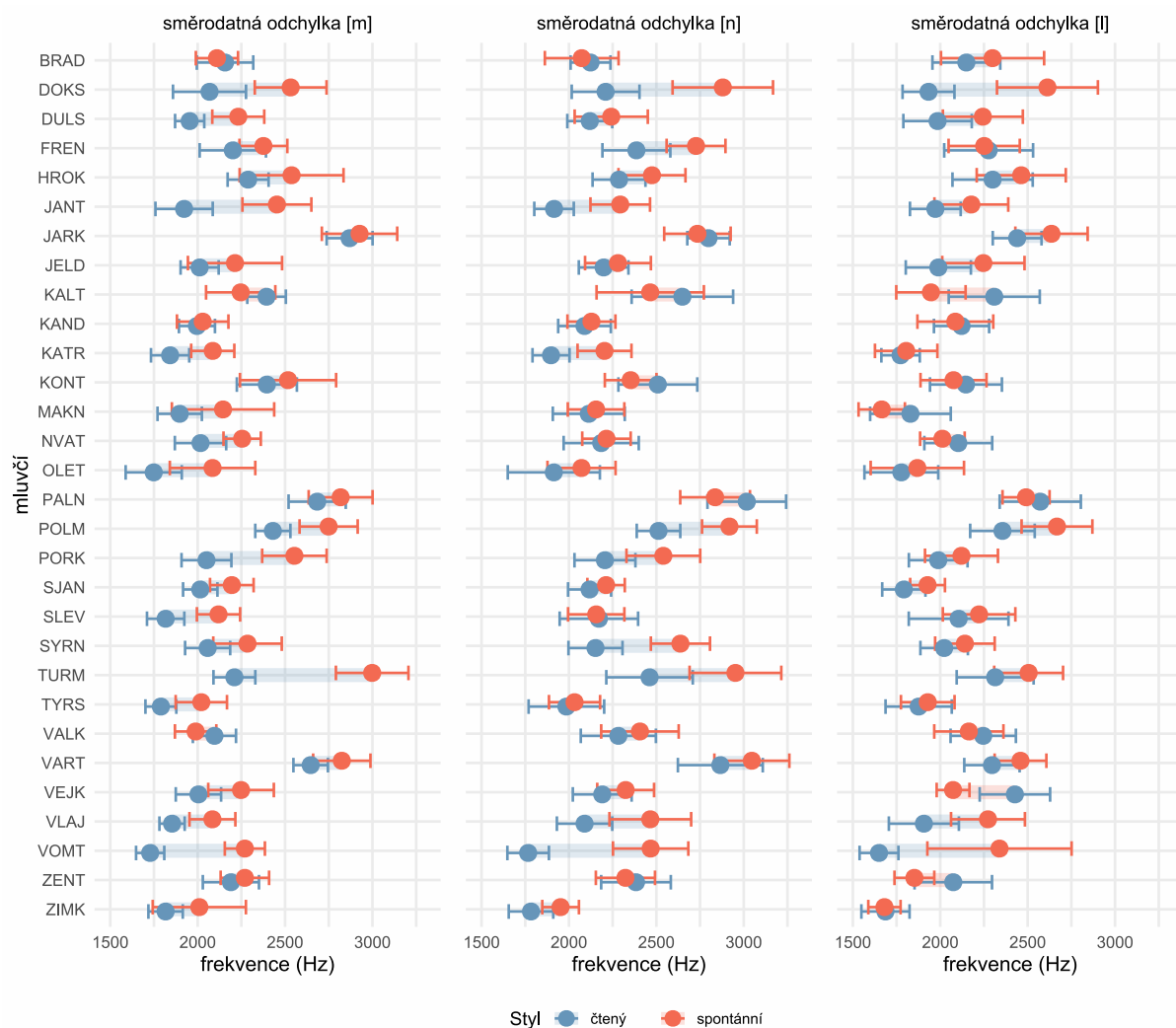
²¹ V některých studiích je pro p-hodnoty mírně přesahující hladinu významnosti využíváno označení „marginální“ či „okrajová“ signifikance (*marginal significance*) – jedná se o v určitých oblastech výzkumu zhusta zavedenou, avšak nikoli nekontroverzní praxi – viz např. Pritschet et al. (2016). Na dané hladině významnosti takový výsledek v každém případě signifikantní není, nicméně takto hraniční p-hodnota nastoluje otázku, zda výsledek nedosahuje statistické signifikance proto, že je rozdíl skutečně pravděpodobně náhodný, či kvůli nedostatku dat, na jejichž základě byl získán – a může tedy být na místě zvážit další výzkum daného jevu s využitím rozsáhlejšího materiálu.

idiosynkratickým potenciálem, nicméně určité mluvčí by na jejich základě patrně bylo možné rozlišit; např. hodnoty spektrálního těžiště všech sledovaných sonor-
ních konsonantů se poměrně výrazně odlišují mezi mluvčími TYRS a VART. U ně-
kterých mluvčích jsou však přítomny natolik široké konfidenční intervaly, že by
tyto parametry k jejich identifikaci pravděpodobně znatelně nepřispěly – viz
např. hodnoty spektrálního těžiště [n] u mluvčího KALT či ve spontánním projevu
mluvčího DOKS. (Široké konfidenční intervaly mohou být způsobeny velkým roz-
ptylem naměřených hodnot, ale také nízkým počtem výskytů sledovaných konso-
nantů.)



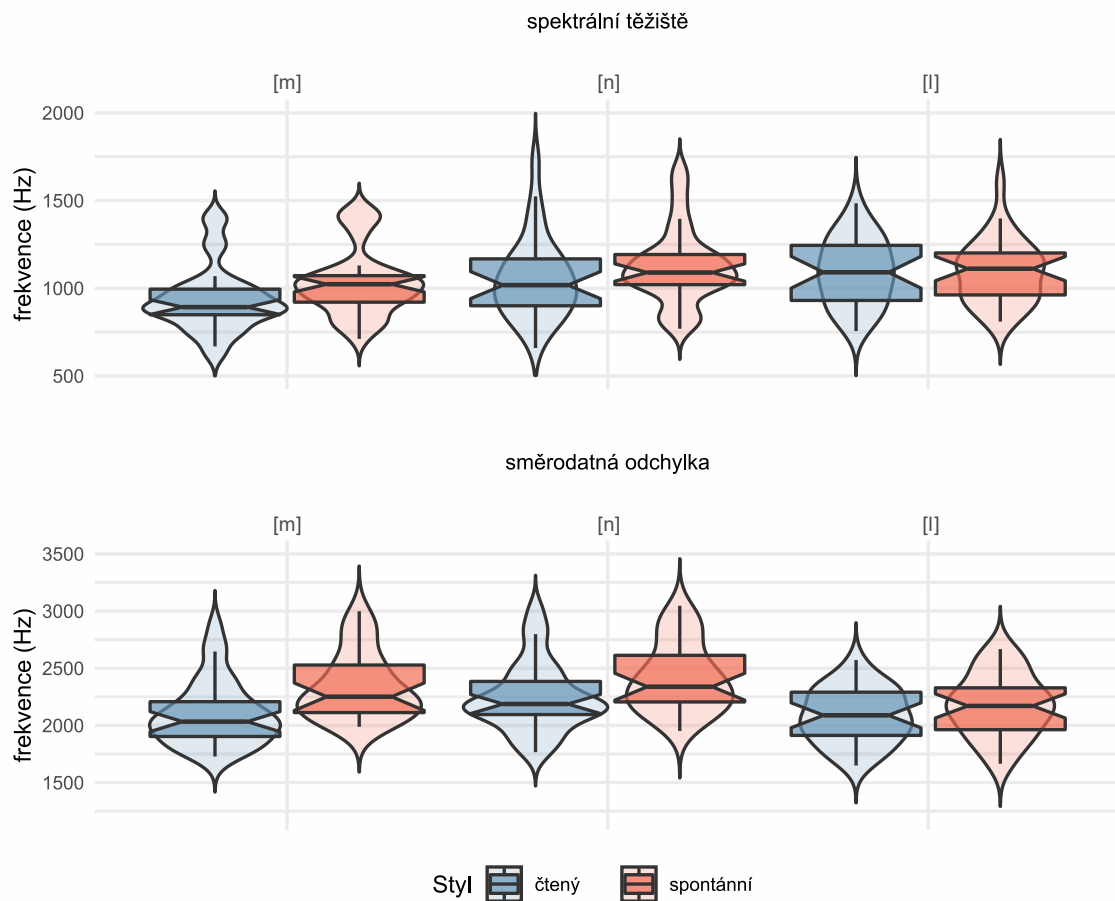
Obrázek 5.12. Spektrální těžiště naměřené v konsonantech [m n l] u jednotlivých mluvčích ve čte-
ném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvo-
vých úseků, linky znázorňují 95% konfidenční intervaly.

Směrodatná odchylka sonorních konsonantů vykazuje značné překryvy hodnot napříč mluvčími, a to do vyšší míry než jejich spektrální těžiště, jak demonstruje obrázek 5.13. I v tomto případě by však patrně bylo možné rozlišit určité jednotlivce, např., stejně jako v případě spektrálního těžiště, mluvčí TYRS a VART. Spektrální momenty sonorních konsonantů by zřejmě mohly přispět i k rozlišení některých jednotlivců, u kterých nejsou přítomny výrazné rozdíly mezi spektrálními momenty frikativy [s] – např. mezi mluvčími OLET a PALN lze zaznamenat viditelný rozdíl mezi hodnotami spektrálního těžiště [m] a SD všech sledovaných sonor (viz obr. 5.12 a 5.13), zatímco mezi hodnotami spektrálních momentů [s], které byly v nahrávkách těchto mluvčích změřeny, existuje značný překryv (viz obr. 5.9). U některých mluvčích ovšem dochází také ke zdatelným posunům mezi čteným a spontánním projevem, a to zejména v případě SD nazálních konsonantů [m] a [n]; např. u mluvčího VOMT je směrodatná odchylka spektra všech sledovaných sonor výrazně vyšší ve spontánním projevu než v projevu čteném – tento trend byl u téhož mluvčího pozorován také v rámci spektrálního těžiště [m] a [n].



Obrázek 5.13. Směrodatná odchyška frekvenčního spektra naměřená v konsonantech [m n l] u jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvových úseků, linky znázorňují 95% konfidenční intervaly.

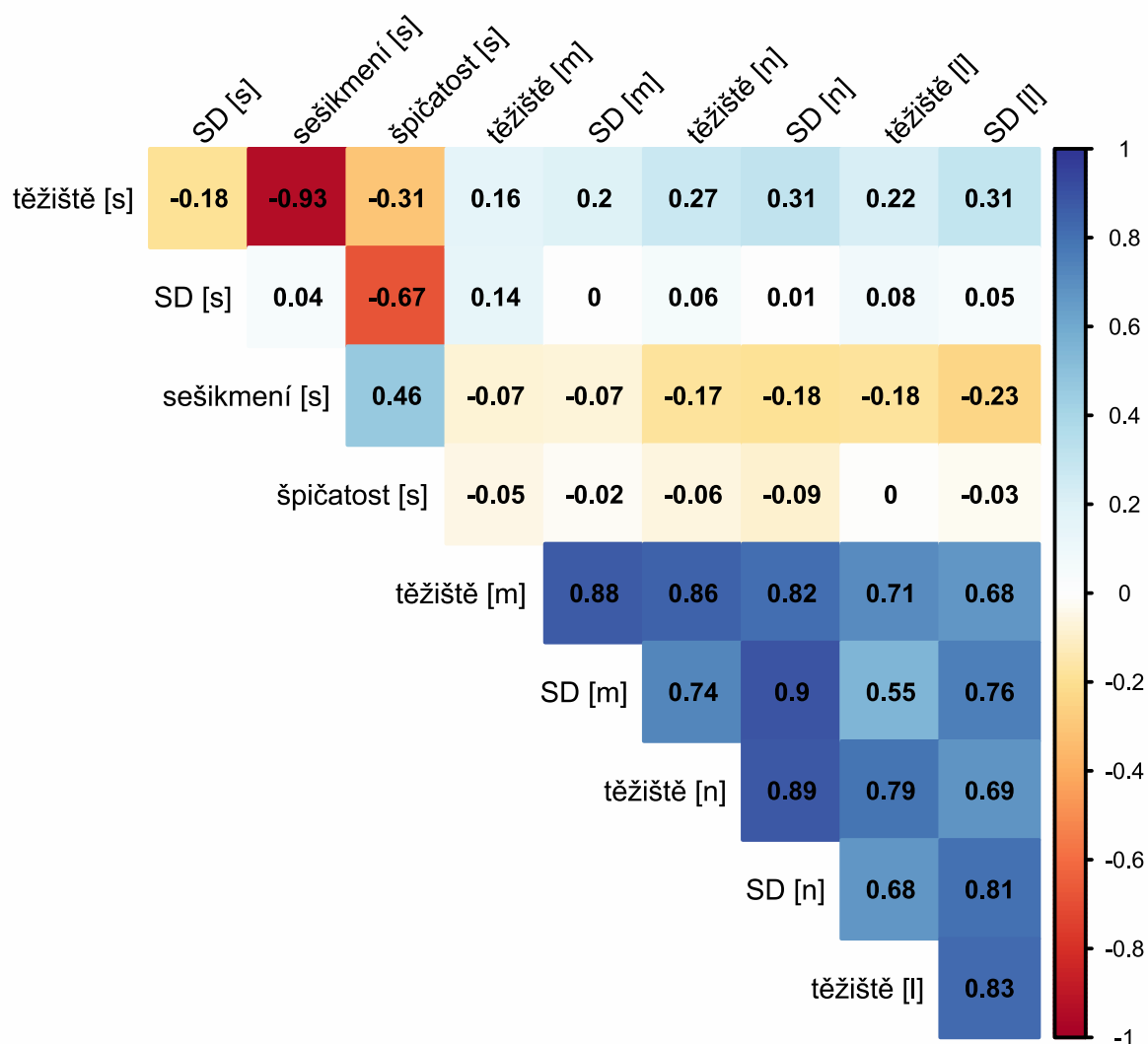
Obrázek 5.14 předkládá srovnání průměrných hodnot parametrů napříč využitými řečovými vzorky. Zatímco spektrální těžiště ani směrodatná odchyška [l] nedoznávají významných změn v závislosti na mluvním stylu (což potvrzují i výsledky Wilcoxonova testu: $p = 0,76$ a $0,03$, v tomto pořadí), v případě nazálních konsonantů je patrná tendence k vyšším hodnotám obou parametrů ve spontánním projevu; tento rozdíl je významný u spektrálního těžiště i SD [m] ($p > 0,0001$ u obou parametrů) a u směrodatné odchyšky [n] ($p = 0,0006$; v případě spektrálního těžiště [n] rozdíl signifikantní není, $p = 0,06$).



Obrázek 5.14: Průměrné hodnoty spektrálního těžiště a směrodatné odchylky frekvenčního spektra naměřených v konsonantech [m n l] napříč mluvčími ve čteném a spontánním projevu.

U všech sledovaných sonorních konsonantů spolu spektrální těžiště a směrodatná odchylka velmi silně korelují, viz obrázek 5.15 ($r \in \langle 0,83; 0,89 \rangle$), a silnou až velmi silnou korelaci také vykazují jednotlivé parametry napříč sonorními konsonanty (spektrální těžiště: $r \in \langle 0,71; 0,86 \rangle$, SD: $r \in \langle 0,76; 0,9 \rangle$).

Silnou či středně silnou korelaci lze pozorovat u většiny dvojic parametrů v rámci jedné skupiny tvořené spektrálními momenty [s] a druhé skupiny sestávající ze spektrálních momentů sonorních konsonantů. Naproti tomu parametry napříč těmito skupinami obecně vykazují pouze slabou či velmi slabou korelaci (viz obr. 5.15).



Obrázek 5.15. Korelační matice sledovaných parametrů frekvenčního spektra konsonantů.

5.5.3 Diskuse a závěr

V rámci experimentu 4 byly analyzovány spektrální momenty (spektrální těžiště, směrodatná odchylka, sešikmení a špičatost) frikativy [s] a spektrální těžiště a směrodatná odchylka sonorních konsonantů [m], [n] a [l]. Jejich hodnoty byly sledovány v rámci jednotlivých řečových vzorků a porovnávány napříč mluvčími a také mezi čteným a spontánním projevem.

Jak výsledky experimentu ukázaly, relativně nejpřínosnější informace pro vzájemné rozlišování mluvcích by mohly poskytovat spektrální momenty [s], jejichž hodnoty zůstávaly u většiny mluvcích bez znatelných změn v závislosti na mluvním stylu, přičemž u některých dvojic mluvcích bylo možné pozorovat výrazné rozdíly. Tyto závěry jsou v souladu se zjištěními Kavanagh (2012), která právě spektrální momenty [s] hodnotí jakožto přínosné pro identifikaci mluvcího v angličtině. Nejvýraznější variabilitu mezi mluvcími a zároveň ve většině případů stabilitu pro mluvcího vykazaly spektrální těžiště a sešikmení [s]; tyto dva parametry spolu zároveň velmi silně negativně korelovaly – ukázala se tedy tendence k většímu výskytu odlehklých hodnot směrem k vyšším frekvencím při níže položeném spektrálním těžišti a naopak. Spektrální charakteristiky [s] závisejí nejen na konkrétním umístění a míře zúžení v ústní dutině, ale do velké míry také na individuální fyziologii ústní dutiny mluvcího, např. tvaru horního patra či podobě chrupu (variabilitě struktur, které jsou součástí vokálního traktu, a jejichž anatomie má tudíž dopad na podobu řečového signálu, se podrobně věnuje např. Beck, 2010), a vzhledem k absenci znělosti nejsou ovlivňovány momentálním fonačním nastavením. V případě tohoto segmentu lze tedy předpokládat relativní stabilitu v rámci projevu daného mluvcího a zároveň variabilitu napříč jednotlivci. Opět je nicméně na místě upozornit na možný vliv akustických podmínek při nahrávání (např. přítomnosti šumu, jak upozorňují např. Garellek et al., 2013 – podrobněji viz oddíl 5.6) na podobu frekvenčního spektra, a tudíž také na hodnoty spektrálních momentů konsonantů; při vzájemném srovnávání mluvcích je tudíž žádoucí využívat materiál srovnatelné kvality, jak doporučuje French (1994: 175; citováno v: Weingartová et al., 2014: 79) pro analýzu dlouhodobého průměrného spektra (viz oddíl 5.7).

Sledovaným spektrálním momentům sonorních konsonantů lze na základě získaných výsledků přisoudit spíše nižší idiosynkratický potenciál, a to i kvůli poměrně širokým konfidenčním intervalům; ty mohou být důsledkem velkého rozptylu hodnot extrahovaných z daného řečového vzorku či nízkého počtu těchto hodnot

– v takovém případě by zúžení konfidenčních intervalů, a tedy zpřesnění údajů o spolehlivosti naměřených dat, napomohl větší objem zkoumaného materiálu. V případě určitých dvojic mluvčích se ovšem i tyto parametry (zejména spektrální těžiště) projeví jako potenciálně přínosné pro jejich rozlišení, neboť mezi nimi bylo možné pozorovat značné rozdíly, přičemž hodnoty daných parametrů zůstávaly stabilní v obou mluvnických stylech – a to včetně některých dvojic mluvčích, jejichž spektrální momenty frikativy [s] se podstatně nelišily. I spektrální těžiště a směrodatná odchylka sledovaných sonorních konsonantů by tudíž mohly v některých případech přispět k rozlišení mluvčích, ačkoli celkově se pro identifikaci mluvčího ukázaly jako spíše méně přínosné.

Kavanagh (2012) v rámci svého výzkumu potvrzuje Nolanův (1997) předpoklad, že vysoký idiosynkratický potenciál by mohl být připsán nazálním konsonantům, vzhledem k přítomnosti rezonancí nosní dutiny, která – na rozdíl od dutiny ústní – při řeči zůstává v podstatě nehybná. Výsledky provedeného experimentu však tomuto závěru neodpovídají, neboť se u spektrálních momentů nazálních konsonantů, konkrétně spektrálního těžiště [m] a směrodatné odchylky [m] a [n], projevil významný vliv mluvnického stylu: ve spontánním projevu ve srovnání s projevem čteným byly zjištěny vyšší hodnoty těchto parametrů. Příčiny těchto rozdílů nejsou na základě získaných výsledků zcela zřejmé – podmínky nahrávání, které mohou značně ovlivňovat podobu frekvenčního spektra (podrobněji viz kapitolu 5, str. 83), byly při nahrávání čteného a spontánního projevu daného mluvčího totožné, tento faktor by tedy neměl hrát roli. Součástí spektra sonorních konsonantů je také základní frekvence, která tudíž může ovlivňovat spektrální parametry; při vyšší průměrné f_0 by tedy bylo možné předpokládat také zvýšení spektrálního těžiště. U sledovaného materiálu však byla zjištěna tendence ke zvyšování spektrálního těžiště nazálních konsonantů ve spontánním projevu, zatímco vyšší průměrná f_0 byla pozorována v projevu čteném (viz kapitolu 3, oddíl 3.4), a mezi hodnotami spektrálního těžiště sonorních konsonantů a průměrem, resp. mediánem základní frekvence byla nalezena pouze slabá korelace ($r < 0,2$).

Tento vztah mezi jmenovanými parametry tudíž nelze na základě získaných výsledků potvrdit. Vliv na podobu frekvenčního spektra byl popsán u hláskového okolí; např. Shadle a Scully (1995) pozorovaly snížení spektrálního těžiště [s] v důsledku koartikulační labializace v sekvencích se sousedícím vokálem [u] u rodilých mluvčích americké angličtiny. Vzhledem ke specifické podobě textu, který mluvčí předčítali v rámci čteného projevu (podrobněji viz kapitolu 2), v něm nelze vyloučit výskyt určitých sekvencí, které se ve spontánním projevu vyskytovaly s nižší četností. Pro zjištění případného vlivu segmentálního okolí na parametry frekvenčního spektra nazálních konsonantů by byl zapotřebí větší objem vhodného materiálu (v materiálu využitém v rámci této práce nelze zaručit dostatečné množství cílových sekvencí pro zkoumání vlivu konkrétního segmentálního okolí na dané hlásky). Je na místě podotknout, že ve srovnání se sykavkami lze zásadní informace o akustických vlastnostech sonorních konsonantů nalézt v nižších frekvencích, zatímco ve vyšších frekvencích je pak možný výskyt šumu, který by mohl mít zkreslující dopad na výsledné hodnoty. Pro jejich analýzu by tedy mohlo být vhodné využití užšího frekvenčního pásma.

U všech sledovaných sonorních konsonantů byla zjištěna velmi silná korelace spektrálního těžiště se směrodatnou odchylkou. Jak spektrální těžiště, tak směrodatná odchylka také vykazaly velmi silnou korelaci napříč těmito konsonanty. Na druhou stranu byla zjištěna pouze slabá či velmi slabá korelace u spektrálního těžiště a směrodatné odchylky mezi sonorními konsonanty a frikativou [s]; tato skutečnost naznačuje, že hodnoty všech analyzovaných parametrů zřejmě nebyly zásadním způsobem ovlivněny podmínkami nahrávání, neboť pokud by jejich podstatný dopad na podobu frekvenčního spektra řečových vzorků byl přítomen, bylo by možné očekávat vyšší korelaci mezi parametry [s] a sonorních konsonantů v rámci řečových vzorků.

Na základě výsledků experimentu 4 lze konstatovat, že jako nejpřínosnější pro vzájemné rozlišování mluvčích se jeví spektrální momenty frikativy [s], zejména

její spektrální těžiště a sešikmení – vzhledem k jejich velmi silné (negativní) korelaci by pro účely identifikace mluvího patrně stačil jen jeden z těchto parametrů. Spektrální těžiště a směrodatná odchylka sonorních konsonantů se ukázaly jako spíše méně přínosné, avšak určité dvojice jedinců by patrně s jejich pomocí bylo možné rozlišit, a to i v některých případech, ve kterých se výrazně neliší hodnoty spektrálních momentů [s]. U spektrálních momentů nazálních konsonantů (na rozdíl od ostatních sledovaných parametrů) ovšem byla také pozorována obecná tendence k dosahování vyšších hodnot ve spontánním projevu – příčiny tohoto trendu však zůstávají otázkou pro další výzkum.

5.6 Dlouhodobé průměrné spektrum

Dlouhodobé ukazatele spektrálního sklonu, podobně jako ty krátkodobé (viz oddíl 5.1), charakterizují tvar frekvenčního spektra, využívají však dlouhodobé průměrné spektrum (*long-term average spectrum*; LTAS) řečového signálu. LTAS, jak jej charakterizují Nordenberg a Sundberg (2004: 93), „zobrazuje průměrnou hladinu zvuku v různých frekvenčních pásmech a odráží charakteristiky hlasivek i vokálního traktu“. Spektrální vlastnosti jednotlivých hlásek jsou v rámci LTAS zprůměrovány, a segmentální rozdíly jej tudíž neovlivňují (Weingartová et al., 2014); podle Fritzella et al. (1974; citováno v: Nordenberg & Sundberg, 2004: 93) se jeho podoba typicky stabilizuje asi po 40 sekundách. Úskalí při možnosti využití LTAS pro rozlišování mluvích může představovat jeho značná citlivost na hlasitost řeči, jak upozorňují Nordenberg a Sundberg (2004) a podotýkají, že zvýšení amplitudy při hlasitější řeči není konzistentní napříč frekvencemi. Garellek et al. (2013) také konstatují, že tvar spektra je ovlivněn přítomností šumu; při vyšší hladině šumu lze pozorovat plošší spektrum ve vyšších frekvencích (2–5 kHz).

Jedním ze způsobů parametrizace dlouhodobého spektrálního sklonu je vyčíslení jeho strmosti prostřednictvím rozdílů amplitud v jednotlivých frekvenčních pásmech. Mezi takové parametry patří např. tzv. Hammarberg index, (Hammarberg et al., 1980), index α (Frøkjær-Jensen & Prytz, 1976; tento index využili také např. Bahmanbiglu et al., 2017 či Ng et al., 2012), spektrální emfáze (Eriksson et al., 2001) a řada dalších – jejich přehled nabízí např. studie Weingartové et al. (2014: 91), kteří je ovšem využili pro parametrizaci krátkodobého spektrálního sklonu v českých vokálech:

- 1) Hammarberg index: rozdíl mezi amplitudovými vrcholy ve frekvenčních pásmech 0–2 kHz a 2–5 kHz (jak autoři uvádějí, třetí pásmo, tj. 5–8 kHz, nebylo využito vzhledem k jeho zřejmé náchylnosti k vlivu nahrávacího zařízení);
- 2) index α : rozdíl mezi součty energií v pásmech 0–1 kHz a 1–5 kHz;
- 3) lineární regrese: proložení přímkou v pásmu 0,5–3 kHz (podle Kochanski et al., 2005);
- 4) proložení celkové energie v pásmech 0–0,5 kHz, 0,5–1 kHz, 1–2 kHz a 2–4 kHz regresní přímkou;
- 5) rozdíl energií v pásmech 0,35–1,1 kHz a 2,3–5,5 kHz;
- 6) rozdíl energií v pásmech 0–1,1 kHz a 2,3–5,5 kHz;
- 7) rozdíl energií v pásmech oddělených pohyblivým pivotem na druhém formantu: 0,35– F_2 a F_2 –5,5 kHz (podrobněji viz níže);
- 8) rozdíl energií v pásmech 0– F_2 a F_2 –5,5 kHz (oddělených druhým formantem, stejně jako v č. 7);
- 9) spektrální emfáze: rozdíl mezi celkovou energií ve spektru do 16 kHz a energií v pásmu do 1,43násobku f_0 .

Tamarit et al. (2008) vznášejí určité námitky k parametrům spektrálního sklonu s pevně stanoveným pivotem: argumentují, že tyto metriky pomíjejí průběh základní frekvence u daného mluvčího. Dále těmto parametrům vytýkají, že vzhledem k hrubé aproximaci LTAS může dojít k zanedbání důležitých prvků ve tvaru frekvenčního spektra. Namísto pevného pivotu v konkrétní frekvenci proto navrhují pohyblivý pivot charakteristický pro daného mluvčího, přičemž nabízejí tři způsoby jeho určení: (1) ve vybraném násobku, tj. vyšší harmonické složce průměrné f_0 (autoři pro svůj experiment zvolili 5., 10. a 15. harmonickou složku), (2) v hodnotě druhého formantu (tuto pozici pivotu aplikovali např. Weingartová a Volín, 2014a, při analýze spektra vokálů – viz výše), přičemž autoři uvádějí využití F_2 vokálu [a]²², a (3) v takové hodnotě frekvence, která rozděluje LTAS do dvou částí s určitým poměrem energií.

Po tomto předestření možností, jakým způsobem určit pohyblivý pivot, autoři popisují vlastní způsoby parametrizace spektrálního sklonu, přičemž zmiňují, že vedle pivotu je zásadní také stanovení horní frekvenční hranice pro analýzu. Na rozdíl od Hammarberg indexu, pro který je relevantní rozložení energie ve frekvenční oblasti do 5 kHz, se Tamarit et al. (2008) rozhodli přistoupit k horní hranici 8 kHz, neboť usuzují, že i rozložení energie ve vyšších frekvencích může obsahovat důležité informace. Představují pak tři metody parametrizace spektrálního sklonu: (1) modifikovaný Hammarberg index, který spočívá, stejně jako původní Hammarberg index (Hammarberg et al., 1980), ve vyčíslení rozdílů mezi amplitudových maximy v pásmech 0–2 kHz a 2–5 kHz, avšak 2kHz pivot je nahrazen pohyblivým pivotem definovaným podle postupů popsanych výše a horní hranici tvoří 8 kHz, (2) proložení spektrální obálky mezi pohyblivým pivotem a

²² Tamarit et al. (2008) využili jazykový korpus obsahující materiál od 10 francouzských mluvčích, a to jednak nahrávky souvislé řeči s hranými afektivními stavy a jednak záznam dlouze drženého vokálu [a] – ten využili při měření hodnot F_2 pro jednotlivé mluvčí, které následně využili coby pohyblivý pivot.

8 kHz regresní přímkou a (3) proložení téhož úseku exponenciální křivkou. Výsledky jejich experimentu zaměřeného na rozlišování afektivních stavů poukazují na přínos využití pohyblivého pivotu a exponenciálních koeficientů ve srovnání s koeficienty lineárními. Na závěr ovšem uvádějí, že aproximace spektrálního sklonu pomocí exponenciální křivky může být spíše než s afektivními stavy propojena s identitou mluvčího; při různě afektivně zabarvených promluvách se podle nich mění zejména hrubé charakteristiky spektra, zatímco jeho detailnější tvar, který nejlépe ze sledovaných parametrů odráží právě exponenciální křivka, je typičtější pro daného jednotlivce (Tamarit et al., 2008: 4).

Míru idiosynkratičnosti vybraných ukazatelů dlouhodobého spektrálního sklonu včetně Hammarberg indexu s pivotem v oblasti druhého formantu sledovali Nechanský et al. (2023), kteří zkoumali vliv cílených modifikací artikulačního a fonačního nastavení u českých mluvčích na charakteristiky akustického signálu. Při analýze dlouhodobého spektrálního sklonu využili automaticky extrahované vokalické proudy (tj. dlouhodobé spektrum bylo získáno z řečového materiálu, z něžž byly odstraněny automaticky detekované pauzy a konsonantické segmenty) a hodnota pohyblivého pivotu byla stanovena coby dlouhodobý medián F_2 . Jejich výsledky ovšem vyšší idiosynkratičnosti Hammarberg indexu s pohyblivým pivotem oproti jeho původní variantě s konstantním 2kHz pivotem nenasvědčují: při srovnávání 15 sledovaných modifikací nastavení vokálního traktu s nastavením modálním byly u standardního Hammarberg indexu zjištěny signifikantní posuny u 4 z nich, zatímco u ukazatele s pohyblivým pivotem se jednalo o 6 modifikací.

5.7 Experiment 5: Parametry dlouhodobého průměrného spektra

5.7.1 Metoda

Za využití materiálu popsaného v kapitole 2 byly sledovány následující parametry spektrálního sklonu dlouhodobého průměrného spektra: Hammarberg index (HI; tj. rozdíl mezi amplitudovými maximy ve frekvenčních pásmech 0–2 kHz a 2–5 kHz), index α (poměr energií ve frekvenčních pásmech 1–5 kHz a 0–1 kHz), poměr energií ve frekvenčních pásmech 2,3–5,5 kHz a 0,35–1,1 kHz (BgNoFo²³) a poměr energií ve frekvenčních pásmech 2,3–5,5 kHz a 0–1,1 kHz (Bg²⁴). (Měření prvních tří parametrů představuje replikaci metodologie, která byla využita v rámci studie Houzar et al., 2023; čtvrtý parametr označovaný jako Bg byl zahrnut pro srovnání s BgNoFo, od nějž se odlišuje zahrnutím pásma obsahujícího základní frekvenci.)

V rámci analýzy dlouhodobého spektrálního sklonu byly využity pouze znělé intervaly (v souladu např. s Hammarberg et al., 1980, Ng et al., 2012 či Bahmanbiglu et al., 2017)²⁵, které byly ze signálu extrahovány na základě segmentace na hlásky (podrobněji viz kapitolu 2), a to po prvotním zarovnání hranic hlásek na nejbližší průchod zvukové vlny nulou. (Spektrální sklon vypovídá jak o charakteristikách nadhrtanových dutin, tak i hlasivkového signálu, který při produkci

²³ Podle Weingartové a Volína, kteří tento parametr označují jako „BgNoFo“ – „*band difference with gap and excluded Fo*“ (Weingartová & Volín, 2014: 7).

²⁴ Obdobně jako BgNoFo, avšak se zahrnutým pásmem f_0 (podle Weingartové a Volína, 2014).

²⁵ Např. Nordenberg a Sundberg (2004) pro získání LTAS využili materiál včetně neznělých konsonantů, neboť analyzovali LTAS pouze do 4 kHz a podle jejich pozorování je podoba frekvenčního spektra v tomto pásmu prakticky identická jak při zahrnutí neznělých segmentů, tak při jejich vyřazení. Nad touto hranicí však již přítomnost neznělých segmentů podobu frekvenčního spektra ovlivňuje, zřetelné rozdíly se pak objevují nad 5 kHz (Nordenberg & Sundberg, 2004: 95).

neznělých segmentů není přítomen – jejich zahrnutí do analýzy dlouhodobého frekvenčního spektra by tudíž v naměřených datech mohlo tvořit nežádoucí šum.)

Pro výpočet Hammarberg indexu bylo ze získaných proudů znělých segmentů v programu Praat vytvořeno dlouhodobé průměrné spektrum (LTAS) pomocí příkazu *To LTAS*, v němž byla následně detekována amplitudová maxima (v dB) v daných frekvenčních pásmech (viz výše). Pro vyčíslení ostatních parametrů bylo z proudů znělých segmentů (taktéž v programu Praat) vytvořeno frekvenční spektrum pomocí příkazu *To Spectrum*, v němž byly následně získány celkové hodnoty energií (v Pa²·s) ve sledovaných pásmech (viz výše). (Způsob parametrizace Hammarberg indexu se tedy od ostatních sledovaných parametrů podstatně liší – jedná se o rozdíl mezi hodnotami dvou amplitudových maxim, zatímco ostatní parametry představují podíl celkových energií ve dvou frekvenčních pásmech. Hammarberg index je ovšem vyčíslen v decibelech, které jsou logaritmickou jednotkou – výpočet rozdílu tedy odpovídá vyčíslení podílu při konverzi na jednotku lineární. De facto tedy všechny sledované parametry vyjadřují poměr mezi dvěma hodnotami akustického tlaku²⁶.)

Pro ustálení podoby LTAS je zapotřebí řečový materiál o trvání v řádu desítek sekund; Fritzell et al. (1974; citováno v: Nordenberg & Sundberg, 2004: 93) doporučují asi 40 sekund, podle Sergeanta a Welche (2007: 660) je v rámci většiny studií přijímáno jakožto adekvátní trvání mezi 20 až 30 sekundami. Trvání řečových vzorků využitých v tomto experimentu se pohybovalo kolem 1 minuty, přičemž proudy znělých intervalů, které byly při analýze dlouhodobého spektrálního sklonu využity, trvaly průměrně 31,3 s (SD = 3,5 s). Z tohoto důvodu byla pro každý řečový vzorek extrahována jediná hodnota každého sledovaného parametru; získané výsledky jsou tedy spíše orientační vzhledem k tomu, že nelze

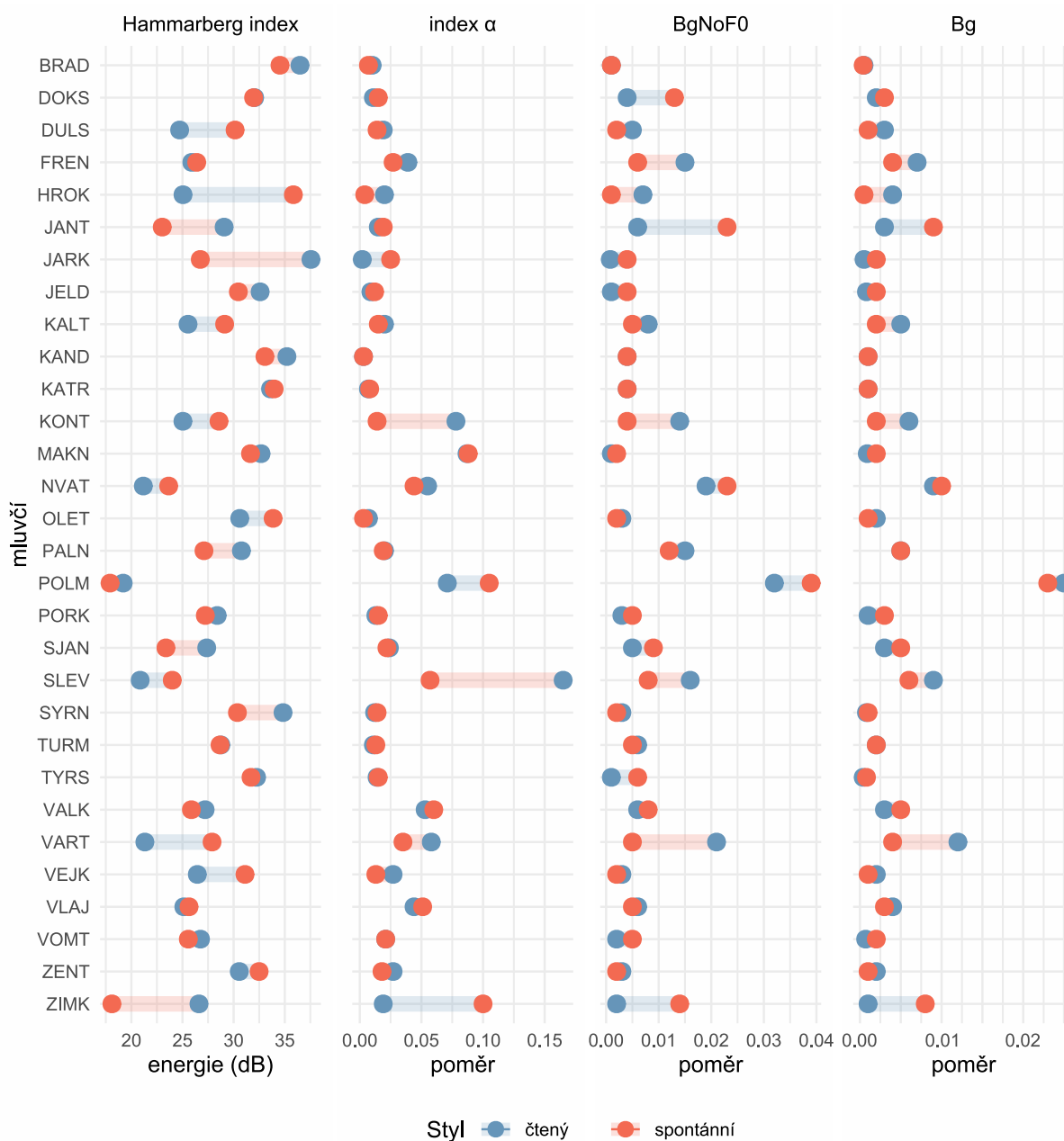
²⁶ Vzhledem k této skutečnosti nebyla zapotřebí normalizace amplitudy využitých řečových vzorků.

vyhodnotit spolehlivost získaných hodnot dlouhodobého spektrálního sklonu v rámci vzorků, neboť nejsou k dispozici jejich konfidenční intervaly.

Hodnoty sledovaných parametrů byly porovnávány mezi čteným a spontánním projevem; v rámci statistického testu byla hladina významnosti korigována počtem sledovaných parametrů: $\alpha = 0,05/4 = 0,0125$.

5.7.2 Výsledky

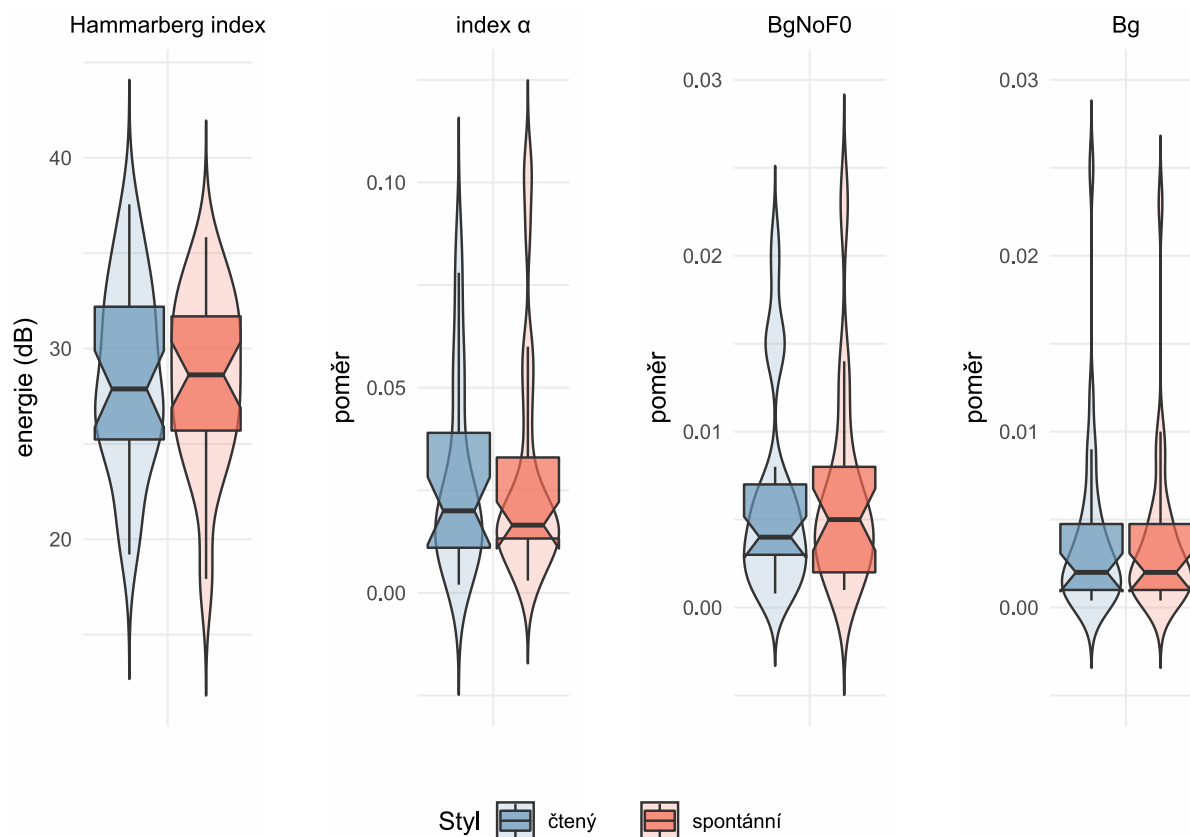
Výsledné hodnoty sledovaných parametrů v jednotlivých řečových vzorcích ukazuje obrázek 5.16. Jak je z něj patrné, nejrovnoměrnější rozmístění naměřených hodnot mezi minimem a maximem zjištěným u sledované populace lze pozorovat u Hammarberg indexu, který odráží rozdíly mezi amplitudovými maximy ve dvou frekvenčních pásmech. Ostatní parametry, které vyjadřují poměr energií v určitých frekvenčních pásmech, se vyznačují několika výraznými odchylkami směrem k vyšším hodnotám (nejvýraznějšími případy jsou čtený i spontánní projev mluvčího POLM), zatímco většina ostatních hodnot se pohybuje relativně blízko zjištěnému minimu. Vzhledem k absenci konfidenčních intervalů ovšem nelze vyhodnotit spolehlivost získaných výsledků.



Obrázek 5.16. Dlouhodobé ukazatele spektrálního sklonu u jednotlivých mluvčích ve čteném a spontánním projevu.

Z celkového srovnání naměřených hodnot dlouhodobých ukazatelů spektrálního sklonu vyplývá, že mezi spontánním a čteným projevem neexistuje významný rozdíl, což ilustruje obrázek 5.17 a tento závěr dokládají také výsledky Wilcoxonova testu ($p \in \langle 0,42; 0,79 \rangle$). V případě HI je opět pozorovatelný rovnoměrnější

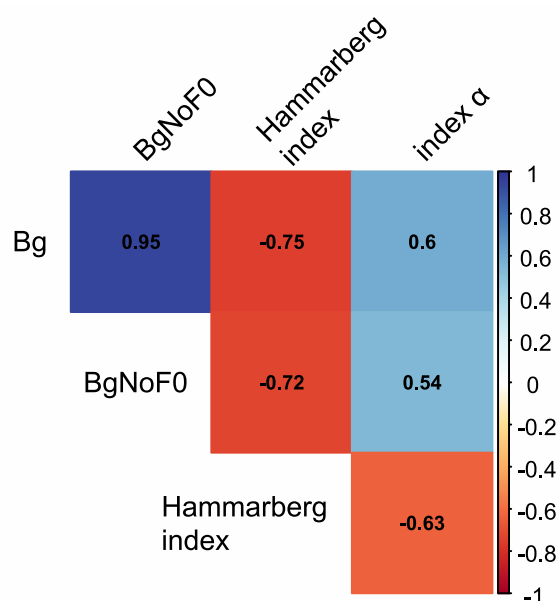
rozptyl hodnot kolem průměru, zatímco rozdělení u ostatních parametrů je zřetelně asymetrické.



Obrázek 5.17. Srovnání hodnot dlouhodobých ukazatelů spektrálního sklonu napříč mluvčími ve čteném a spontánním projevu.

Všechny sledované dlouhodobé parametry spektrálního sklonu vykázaly vzájemnou středně silnou, silnou či velmi silnou korelaci (absolutní hodnota korelačního koeficientu se pohybovala mezi 0,54 a 0,95; viz obr. 5.18); velmi silně korelovaly parametry Bg a BgNoF0 ($r = 0,95$), které odrážejí poměry energií v obdobných frekvenčních pásmech, přičemž u prvního jmenovaného je zahrnuto pásmo obsahující základní frekvenci, zatímco u druhého z nich nikoli. Zápornou silnou korelaci ($r \in \{-0,75; -0,63\}$) se všemi ostatními parametry vykázal Hammarberg index, zatímco mezi ostatními parametry byly zjištěné hodnoty korelačního koeficientu kladné. (Tento výsledek vyplývá ze způsobu výpočtu: Hammarberg index je

rozdílem²⁷ hodnot dvou amplitudových maxim, přičemž menšencem je hodnota maxima v nižším pásmu a menšitelem hodnota v pásmu vyšším; ostatní parametry představují podíl energií ve dvou frekvenčních pásmech, přičemž dělencem je souhrn energií ve vyšším pásmu a dělitelem souhrn energií v pásmu nižším – činitelé jsou tedy v opačném pořadí než v případě Hammarberg indexu.)



Obrázek 5.18. Korelační matice sledovaných dlouhodobých parametrů spektrálního sklonu.

5.7.3 Diskuse a závěr

V rámci experimentu 5 byly analyzovány vybrané dlouhodobé parametry spektrálního sklonu vyčíslené na základě dlouhodobého průměrného spektra proudů znělých intervalů (tj. s vyřazením pauz a neznělých segmentů z analyzovaných řečových vzorků): Hammarberg index (rozdíl mezi amplitudovými maximy ve frekvenčních pásmech 0–2 kHz a 2–5 kHz), index α (poměr energií ve frekvenčních pásmech 1–5 kHz a 0–1 kHz), BgNoFo (poměr energií ve frekvenčních pásmech 2,3–5,5 kHz a 0,35–1,1 kHz; tj. bez frekvenčních pásem zahrnujících

²⁷ Ovšem jak již bylo uvedeno, jedná se o logaritmickou jednotku, tj. výpočet rozdílu je zde ekvivalentní operací k výpočtu podílu u lineárních jednotek.

přibližně základní frekvenci a oblast druhého formantu) a Bg (poměr energií ve frekvenčních pásmech 2,3–5,5 kHz a 0–1,1 kHz; tj. parametr odlišující se od BgNoFo zahrnutím pásma obsahujícího základní frekvenci).

Pro vzájemné odlišování mluvčích se tyto parametry nejeví jako příliš přínosné, neboť jejich extrakce vyžaduje poměrně velké množství řečového materiálu – ke stabilizaci dlouhodobého průměrného spektra je podle dostupných studií zapotřebí materiál o trvání v řádu desítek sekund (viz např. Fritzell et al., 1974; citováno v: Nordenberg & Sundberg, 2004: 93 či Sergeant & Welch, 2007). Pro přesnější představu o variabilitě dlouhodobého spektrálního sklonu u daného mluvčího (jak v rámci jednoho řečového vzorku, tak mezi vzorky lišícími se např. mluvním stylem) by pak bylo vhodné využití podstatně většího objemu řečového materiálu. Vedle toho je při porovnávání řečových vzorků třeba brát v úvahu případný vliv okolností na podobu řečového spektra, např. akustických podmínek při nahrávání (Weingartová et al., 2004: 77), hlasitosti řeči (Nordenberg & Sundberg, 2004) či přítomnosti šumu (Garellek et al., 2013). French (1994: 175; citováno v: Weingartová et al., 2014: 79) doporučuje poměřovat parametry dlouhodobého frekvenčního spektra pouze u nahrávek srovnatelné kvality, případně jednu z nahrávek kvalitativně degradovat tak, aby bylo této srovnatelnosti dosaženo.

Výsledné hodnoty sledovaných parametrů nevykázaly významné rozdíly mezi čteným a spontánním projevem. Tyto parametry představují akustický korelát kvality hlasu, resp. způsobu fonace (viz např. Hammarberg et al., 1980 či Houzar et al., 2023); nelze tedy usuzovat, že by existovaly konzistentní rozdíly mezi fonačním nastavením ve sledovaných mluvních stylech napříč mluvčími. (Nejedná se o překvapivý závěr, o existenci konzistentních rozdílů kvality hlasu ve čteném a spontánním projevu se dostupné studie nezmiňují.)

Mezi všemi sledovanými parametry byla zjištěna středně silná, silná až velmi silná korelace, při využití parametrizace dlouhodobého spektrálního sklonu pro

postižení akustických charakteristik řečového signálu se tedy měření všech čtyř jeví jako nadbytečné. Je ovšem celkově otázkou, nakolik by v takovém případě bylo zahrnutí parametrizace dlouhodobého spektrálního sklonu relevantní, vzhledem k jeho citlivosti na jiné okolnosti, než je samotný řečový projev mluvčího (viz výše).

6 Temporální jevy

Dosud popsané akustické parametry vycházely z frekvenční oblasti, tj. ze zastoupení jednotlivých frekvencí ve spektru řečového signálu. I ty obsahovaly časovou složku a byly ovlivňovány probíhajícími proměnami frekvenčního spektra (tato skutečnost byla patrná zejména u parametrů dynamických). Další skupina parametrů je pak odvozována přímo od temporálních jevů, pro něž je časová doména primární: jsou založeny na (absolutním či relativním) trvání různých řečových úkonů. Oproti ukazatelům z frekvenční oblasti mají temporální charakteristiky z pohledu akustické analýzy zásadní výhodu, která spočívá v jejich odolnosti vůči kvalitativní degradaci signálu (Leemann et al., 2014) – zatímco podoba frekvenčního spektra může být vlivem řady faktorů (např. telefonního přenosu či ruchu na pozadí) zkreslena, časový průběh zpravidla zůstává nedotčen.

6.1 Mluvní a artikulační tempo

Patrně nejpřímočařejším způsobem parametrizace časového průběhu promluvy je změření tempa řeči, tj. počet realizovaných zvukových jednotek (hlásek, slabik či slov) za jednotku časovou (sekundu či minutu); ve fonetickém výzkumu bývá obvykle vyčíslováno ve slabikách za sekundu. Zastřešující termín „tempo řeči“ zahrnuje dvě podobné, ne však zaměnitelné metriky, a to jednak tempo mluvní a jednak tempo artikulační. Zatímco mluvní tempo zahrnuje veškerou řečovou produkci včetně pauz a disfluenčních prvků, artikulační tempo odráží pouze rychlost vlastní artikulace, přičemž pauzy a disfluence jsou z měření vyřazeny. Při srovnání dvou mluvčích, z nichž první rychle artikuluje a jeho řeč se zároveň vyznačuje řadou disfluencí, zatímco druhý hovoří pomaleji a plynuleji, může mít tudíž první mluvčí ve srovnání s druhým vyšší artikulační, avšak nižší mluvní tempo.

Künzel (1997) sledoval míru idiosynkratičnosti mluvního i artikulačního tempa v němčině napříč různými mluvními styly, jmenovitě ve čteném, polospontánním a spontánním projevu (přičemž za polospontánní projev je považován řízený rozhovor). Výsledky ukázaly značené rozdíly v mluvním tempu mezi spontánními a polospontánními projevy na jedné straně a čtenými promluvami na straně druhé, autor tedy usuzuje, že mluvní tempo daného mluvčího značně variuje v závislosti na mluvním stylu. Artikulační tempo naproti tomu na základě svých výsledků považuje spíše za stabilní, a tedy charakterističtější pro jednotlivce. Připouští nicméně, že rozlišovací potenciál artikulačního tempa je ve srovnání s některými parametry ze spektrální oblasti spíše slabý – ovšem jeho výhodu spatřuje v odolnosti vůči kvalitativní degradaci signálu.

Künzel (1997) měřil průměrné mluvní i artikulační tempo v celých řečových vzorcích, získal tedy jednu hodnotu zastupující každý z nich. Odlišný přístup uplatnil, taktéž za použití německého materiálu, Jessen (2007) při měření artikulačního tempa – vzhledem k možným proměnám tempa v průběhu delší promluvy považuje za výhodné rozdělit řečový vzorek na více úseků o několika slabikách a získat hodnotu artikulačního tempa pro každý z nich. Na základě těchto lokálně extrahovaných hodnot artikulačního tempa pak pro každý řečový vzorek vyčíslil průměrné artikulační tempo²⁸ a také jeho směrodatnou odchylku. Na rozdíl od Künzelových (1997) závěrů Jessen (2007) seznává, že artikulační tempo je signifikantně vyšší (a jeho směrodatná odchylka naopak nižší) při čteném projevu než při projevu spontánním, a ani to tedy nelze považovat za stabilní charakteristiku v rámci projevu jednoho mluvčího. Jessenovu (2007) metodologii pak ve své studii zaměřené na britskou angličtinu replikovala Gold (2018), která rovněž konstatuje, že artikulační tempo nevykazuje významný idiosynkratický potenciál, neboť jeho intraindividuální variabilita přesahuje variabilitu interindividuální. Podotýká ale, že u některých jednotlivců, jejichž artikulační tempo se výrazně

²⁸ Autor ovšem nezmiňuje, zda se jedná o prostý aritmetický průměr, nebo zda byl vypočítán průměr vážený s ohledem na trvání jednotlivých úseků.

odchyluje od populačního standardu, může být i tento parametr dobrým rozlišovacím prvkem.

Artikulační tempo v češtině měřili také Weingartová a Volín (2014b), kteří sledovali jeho idiosynkratický potenciál za využití nahrávek čteného projevu a zjistili, že zatímco jeho variabilita mezi mluvčími je výrazná, v rámci mluvčího zůstává spíše stabilní, a považují jej tedy za parametr poměrně dobře odrážející individualitu jednotlivce. Je ovšem potřeba zmínit, že intraindividuální variabilitu artikulačního tempa autoři testovali porovnáním hodnot mezi dvěma polovinami, na něž byla nahrávka daného mluvčího rozdělena, nejednalo se tedy o srovnání napříč mluvčími styly, ba dokonce ani dvěma různými nahrávkami téhož mluvčího stylu. Podobně jako v němčině či angličtině (viz výše) však artikulační (i mluvní) tempo v závislosti na mluvním stylu výrazně variuje, jak na základě své studie využívající nahrávky čteného textu, popisu obrázku a řízeného dialogu uvádí Balkó (2005).

6.2 Globální temporální ukazatele

Vedle samotného měření rychlosti produkce všech zvukových jednotek (tj. tempa řeči) lze sledovat také podíl a variabilitu trvání zvukových jednotek určitého typu (např. vokálů a konsonantů či znělých a neznělých segmentů) v řečovém signálu²⁹. Metrikami, které takové charakteristiky odrážejí, jsou tzv. globální temporální ukazatele, někdy též nazývané rytmické, viz např. Dellwo et al. (2015), či Asadi et al. (2018) – tento termín však lze považovat za problematický, neboť evokuje spojitost s rytmem, jenž je však velmi komplexním jevem zakořeněným

²⁹ Mezi globální temporální ukazatele lze zařadit i mluvní a artikulační tempo, nicméně těm je v rámci této práce věnována samostatná kapitola a mezi globální temporální ukazatele jsou zde zahrnuty parametry založené na relativním načasování akustických prvků v řeči.

v řečové percepci a takto pojaté temporální charakteristiky jej nemohou spolehlivě postihnout, jak objasňuje Volín (2017).

Weingartová a Volín (2014b) uvádějí výčet nejčastěji používaných ukazatelů odrážejících načasování zvukových jednotek v řeči:

- 1) %V: podíl trvání vokalických intervalů vůči celkovému trvání promluvy (do celkového trvání nejsou zahrnuty pauzy, tj. promluva sestává pouze z vokalických a konsonantických intervalů; Ramus et al., 1999);
- 2) $\Delta V/\Delta C$: směrodatná odchylka trvání vokalických a konsonantických intervalů v řeči (nejedná se tedy o trvání jednotlivých segmentů – za jeden konsonantický interval je považován celý konsonantický klastr, obdobně je tomu též v případě vokálů);
- 3) $\text{VarcoV}/\text{VarcoC}$: variační koeficient ΔV a ΔC , tedy podíl směrodatné odchylky a průměru trvání vokalických/konsonantických intervalů; jak uvádí White a Mattys (2007), jedná se o $\Delta V/\Delta C$ normalizované tempem řeči, na němž jsou samotné parametry ΔV a ΔC výrazně závislé³⁰;
- 4) $r\text{PVI-V}/r\text{PVI-C}$ ³¹: index párové variability pro vokalické a konsonantické intervaly; tento parametr (na rozdíl od $\Delta V/\Delta C$ či $\text{VarcoV}/\text{VarcoC}$) odráží rozdíly v trvání každé dvojice po sobě jdoucích intervalů a je vyčíslením průměru těchto rozdílů v řeči (Grabe & Low, 2002);
- 5) $n\text{PVI-V}/n\text{PVI-C}$ ³²: hodnota $r\text{PVI-V}/r\text{PVI-C}$ normalizovaná průměrným trváním dané dvojice intervalů (tzn. jedná se de facto o normalizaci tempem řeči);

³⁰ Jak uvádí Volín (2007: 63), variační koeficienty jsou oproti nenormalizovaným ΔV a ΔC vhodnější pro srovnávání promluv o rozdílných tempech; „[b]ez této normalizace by totiž pomalejší tempa nejspíše vykazovala větší hodnoty ΔC , což by bylo zavádějící.“

³¹ rPVI – raw pairwise variability index (Grabe & Low, 2002)

³² nPVI – normalised pairwise variability index (Grabe & Low, 2002)

- 6) %VO: podíl trvání znělých intervalů vůči celkovému trvání promluvy; Weingartová a Volín (2014b) poukazují na možnost automatické extrakce tohoto parametru, na rozdíl od těch, které reflektují temporální vlastnosti vokalic-
kých a konsonantických intervalů a které zpravidla vyžadují manuální segmentaci signálu. V současné době jsou ovšem dostupné i nástroje pro automatickou extrakci vokalic-
kých a konsonantických intervalů jako např. Vocal Toolkit (Corretge, 2023), který je doplňkem softwaru Praat (Boersma & Weenink, 2023). Leemann et al. (2014) pak za využití automatické extrakce vyčíslují také ΔVO , VarcoVO a nPVI-VO (analogicky k ΔV , VarcoV a nPVI-V popsaným výše).

Leemann et al. (2014) sledovali, nakolik globální temporální ukazatele odrážejí individualitu mluvčího ve švýcarské němčině, přičemž využili nahrávky čteného a spontánního hovorů. Ze signálu extrahovali pět parametrů odrážejících načasování vokalic-
kých a konsonantických intervalů, jmenovitě %V, VarcoV, nPVI-V, VarcoC a nPVI-C, dále tři parametry založené na znělých a neznělých intervalech, a to %VO, VarcoVO a nPVI-VO, a nakonec dvě metriky odvozené od intervalů mezi intenzitními vrcholy jednotlivých slabik, které označují jako VarcoPeak a nPVI-Peak. Výsledky jejich měření ukázaly značnou variabilitu sledovaných parametrů mezi mluvčími, a to zejména ve čteném projevu. Jako nejlépe rozlišující mluvčí a zároveň nejstabilnější napříč mluvčími styly se projeví parametry %V a %VO.

Užitečnost globálních temporálních ukazatelů pro rozlišování mluvčích ve své studii testovali také Weingartová a Volín (2014b), kteří ve čteném projevu českých mluvčích (jednalo se o týž materiál jako v případě experimentu zaměřeného na artikulační tempo, viz výše) měřili %V, ΔV , ΔC , VarcoV, VarcoC, rPVI-V, rPVI-C, nPVI-V a nPVI-C. Coby nejefektivnější ukazatel z hlediska rozlišování mluvčích se projevílo %V. Na idiosynkratický potenciál globálních temporálních ukazatelů v češtině se zaměřila také Weingartová (2014). V rámci jejího

experimentu však žádný ze sledovaných parametrů (týchž jako v případě studie Weingartové a Volína, 2014b) neumožnil vzájemné odlišení všech tří mluvčích, od nichž pocházel analyzovaný řečový materiál. Autorka uzavírá, že analyzované ukazatele patrně lépe odlišují různé jazyky než přímo jednotlivé mluvčí, a větší idiosynkratický potenciál spatřuje u dalšího měřeného parametru, kterým bylo lokální artikulační tempo (viz níže).

6.3 Lokální temporální ukazatele

Tempo řeči nezůstává zcela konstantní v průběhu celé promluvy, ale kolísá v závislosti na řadě faktorů, z nichž jedním z nejprominentnějších a nejvíce prozkoumaných je závěrové zpomalování, tj. snižování tempa na konci fráze (pro češtinu viz např. Dankovičová, 1997 či Volín & Skarnitzl, 2007). Při měření tempa řeči či globálních temporálních ukazatelů v průběhu celé promluvy se tyto lokální rozdíly stírají – i ty však mohou být charakteristické pro mluvčího. Hypotézu o deklinaci artikulačního tempa v průběhu intonační fráze testovala Dankovičová (1999), a to za využití čteného projevu mluvčích češtiny a angličtiny. Hodnotu artikulačního tempa měřila v jednotlivých fonologických slovech uvnitř fráze. Výsledky jejího experimentu prokázaly postupné snižování tempa řeči v průběhu promluvy (neboli *rallentando*), autorka upozorňuje zejména na výrazně pomalejší realizaci posledního fonologického slova v rámci intonační fráze (tj. závěrové zpomalování). Konstatuje také, že artikulační tempo souvisí také s délkou a významovou kategorií daného slova: jak v češtině, tak v angličtině bylo zjištěno vyšší artikulační tempo u slov sestávajících z více slabik, a také u celků, které sestávala z neplnovýznamových lexikálních jednotek.

Volín (2009; citováno v: Volín & Weingartová, 2012) zavedl ukazatel nazvaný lokální artikulační tempo (*local articulation rate*; LAR), který odráží změny artikulačního tempa v čase; hodnotu LAR představuje převrácená hodnota časové

vzdálenosti dvou po sobě jdoucích slabičných vrcholů. Volín a Weingartová (2012) provedli měření LAR ve čteném projevu mluvčích češtiny; byly srovnávány hodnoty LAR v týchž krátkých větách předčítaných jednotlivými mluvčími. Zatímco v některých typech úseků se jeho hodnoty mezi jednotlivci výrazně nelišily, v určitých slovech či morfémech napříč mluvčími lokální artikulační tempo zdatelně variovalo. Jak autoři uzavírají, v češtině patrně existují slova, morfémy a jejich kontexty, v nichž mohou temporální vzorce poukazovat na individuální rytmické preference.

Zatímco Volín a Weingartová (2012) analyzovali pouze hodnoty LAR v jednotlivých slabikách, Weingartová (2014) charakterizovala změny artikulačního tempa v čase pomocí lineární regrese. Ve dvourozměrném prostoru definovaném hodnotou LAR a pořadovým číslem slabiky v rámci dané promluvy vynesla kontury LAR, které následně proložila regresní přímkou. I v její studii byly využity nahrávky čteného projevu, a to od tří českých mluvčích, jednalo se však o odlišný text než v případě studie Volína a Weingartové (2012). Kontury LAR byly sledovány pouze v posledních šesti slabikách jednotlivých nádechových úseků; na jejich základě byly sledovány individuální vzorce závěrového zpomalování. U všech sledovaných mluvčích bylo zpozorováno závěrové zpomalování, nicméně výsledky experimentu potvrdily určité individuální vzorce; ačkoliv byla zjištěna určitá intraindividuální variabilita, všechny tři mluvčí bylo možné podle přesné podoby závěrového zpomalování vzájemně odlišit. Autorka na základě výsledků svého experimentu usuzuje, že lokální artikulační tempo nese značný idiosynkratický potenciál, a považuje jej za slibnější pro vzájemné rozlišování mluvčích než globální temporální ukazatele, které ve své studii taktéž zkoumala.

6.4 Experiment 6: Temporální parametry

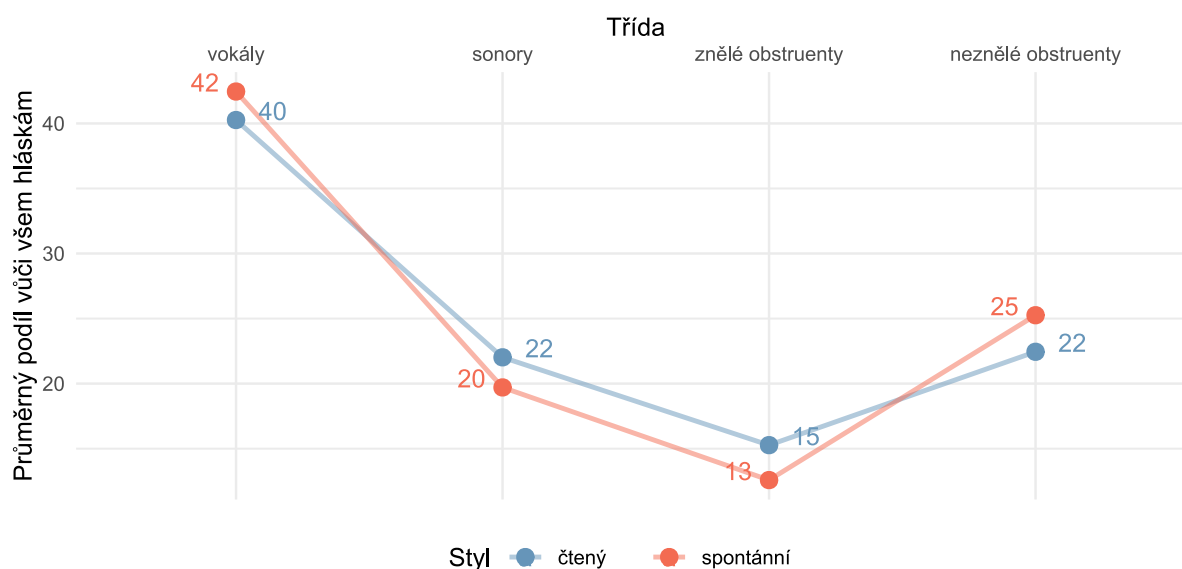
6.4.1 Metoda

Za využití materiálu popsaného v kapitole 2 byly sledovány následující temporální parametry: artikulační tempo (AT), podíl vokalických intervalů v řeči (%V), podíl znělých intervalů v řeči (%VO), směrodatná odchylka trvání vokalických a konsonantických intervalů v řeči (ΔV a ΔC) a variační koeficient ΔV a ΔC (VarcoV a VarcoC), tj. ΔV a ΔC normalizované průměrným trváním vokalických, resp. konsonantických intervalů. Hodnoty jednotlivých parametrů byly vyčísleny na základě segmentace na hlásky a promluvové úseky s uvedením počtu slabik (podrobněji viz kapitolu 2).

Pro výpočet artikulačního tempa byly v každé nahrávce v rámci jednotlivých promluvových úseků sečteny počty slabik a vyděleny celkovým trváním daného úseku. Z hodnot AT v promluvových úsecích pak byly vypočítány průměrné hodnoty³³ a 95% konfidenční intervaly artikulačního tempa v dané nahrávce. Byla také sledována variabilita artikulačního tempa, která byla vyjádřena prostřednictvím směrodatné odchylky. Druhý způsob parametrizace tempa řeči, kterým je mluvní tempo, jehož výpočet zahrnuje i trvání pauz a disfluenčních prvků, zde zahrnut nebyl, a to vzhledem k podobě využitých nahrávek, z nichž mohly být některé původně obsažené pauzy či disfluenční prvky odstraněny. Při nahrávání spontánního projevu do hovoru v některých případech také mohli vstupovat nahrávající výzkumníci, jejichž repliky byly z výsledných nahrávek vyňaty; v takovýchto případech se tedy nemuselo jednat o plynulý projev daného mluvčího a trvání pauz by tudíž mohlo být zkresleno.

³³ Za jednotku byl v tomto případě považován promluvový úsek, při výpočtu průměru tedy nebyla jednotlivým hodnotám přiřazována váha podle trvání úseku.

%V a %VO byly vyčísleny jakožto procentuální zastoupení celkového trvání vokalických a znělých intervalů v rámci promluvových úseků. Tyto parametry ovšem závisí na celkovém počtu daných segmentů; zatímco ve spontánním projevu se výskyt jednotlivých hlásek odvíjel od individuálních strategií mluvčích, text pro čtený projev byl u všech mluvčích týž – jednalo se o text záměrně formulovaný tak, aby obsahoval i běžně málo frekventované hlásky a hláskové kombinace. Poměry, v jakých v něm byly jednotlivé hlásky, resp. třídy hlásek zastoupeny, se tudíž mohly lišit od jejich standardní četnosti v běžné řeči³⁴. Průměrný podíl jednotlivých tříd hlásek (vokálů, sonor a znělých a neznělých obstruentů) předkládá obrázek 6.1.



Obrázek 6.1. Průměrný podíl jednotlivých tříd hlásek ve využitém materiálu.

Pro získání hodnot ΔV , ΔC , VarcoV a VarcoC byla nejprve od segmentace na jednotlivé hlásky odvozena segmentace na vokalické a konsonantické intervaly (tj. sousedící vokály či konsonanty byly sloučeny do jednoho intervalu). V každé nahrávce byly vypočítány směrodatné odchylky trvání vokalických a

³⁴ Pro podrobné informace o četnosti jednotlivých hlásek v češtině viz např. Bartoň et al. (2009: 25–39).

konsonantických intervalů, tj. ΔV a ΔC , které byly následně vyděleny průměrným trváním těchto intervalů, výsledkem čehož byly hodnoty VarcoV a VarcoC.

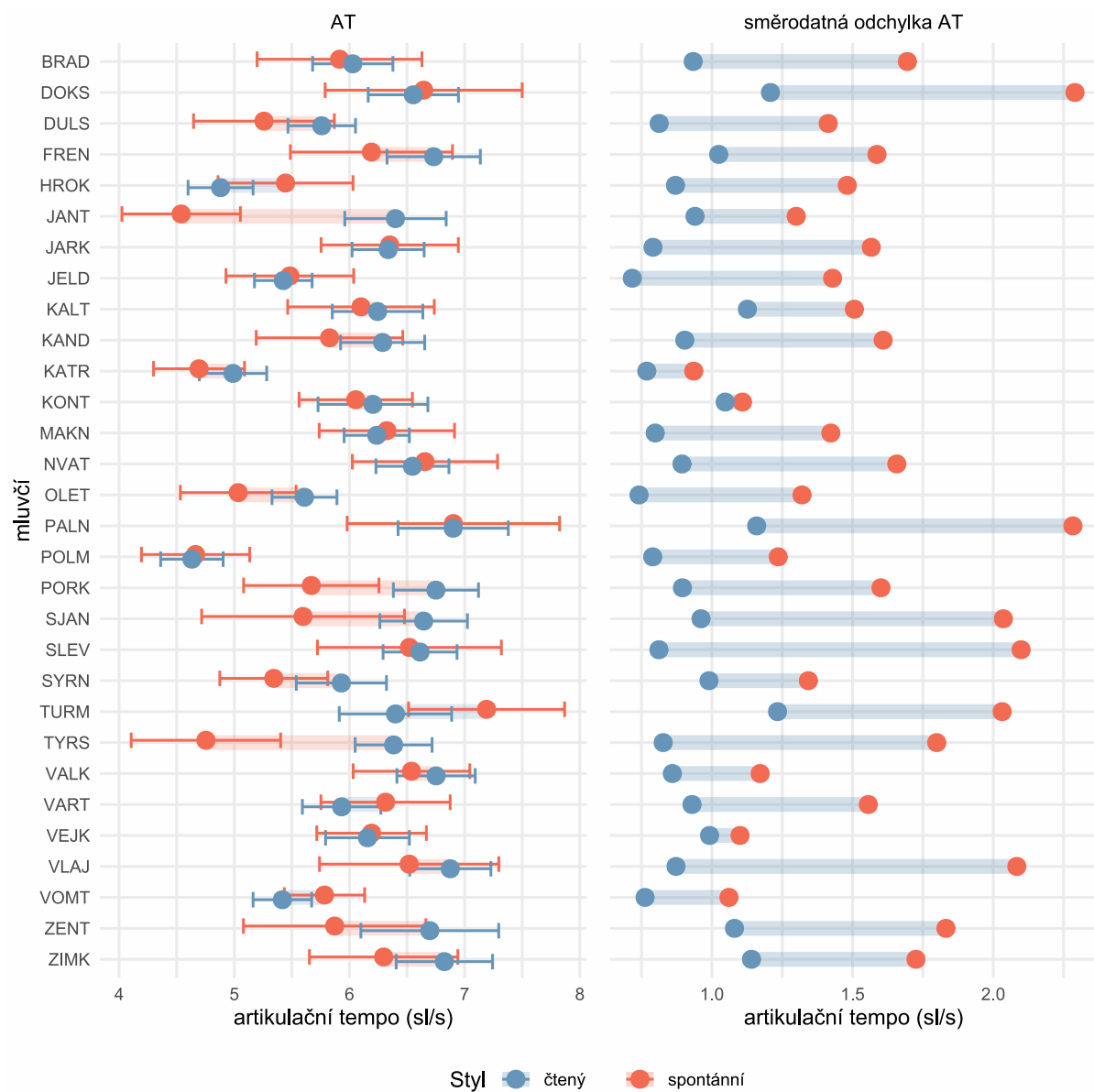
Významnost rozdílů mezi čteným a spontánním projevem byla u všech parametrů vyhodnocena pomocí Wilcoxonova testu, přičemž při určení hladiny významnosti byla aplikována Bonferroniho korekce, v rámci které byl zohledněn počet sledovaných parametrů: $\alpha = 0,05/8 = 0,0063$. Byla také sledována vzájemná korelace jednotlivých parametrů.

6.4.2 Výsledky

Jak naznačuje obrázek 6.2, u většiny mluvčích nebyly pozorovány podstatné rozdíly mezi průměrnými hodnotami artikulačního tempa ve čteném a spontánním projevu, zároveň ovšem nejsou přítomny natolik výrazné interindividuální rozdíly, aby bylo možné na základě artikulačního tempa spolehlivě odlišit mluvčí. V případě jednotlivců, jejichž artikulační tempo se výrazně vymyká celkovému průměru, by však patrně mohlo přispět k jejich identifikaci (např. hodnoty AT u mluvčího POLM jsou relativně nízké, zároveň patrně zůstávají stabilní nezávisle na mluvním stylu, a artikulační tempo by patrně napomohlo jeho odlišení např. od mluvčího VALK, u kterého také nedochází k výrazným posunům AT mezi čteným a spontánním projevem a celkově se pohybuje spíše ve vyšších hodnotách). Sledovaný materiál nicméně zahrnuje i mluvčí, jejichž artikulační tempo nabývá značně rozdílných hodnot ve spontánním a čteném projevu, tento parametr tudíž nelze považovat za obecně nezávislý na mluvním stylu.

Ačkoliv na obrázku 6.2 není u většiny mluvčích patrný výrazný posun artikulačního tempa v závislosti na mluvním stylu (viz výše), při jeho celkovém srovnání mezi čteným a spontánním projevem, které předkládá obrázek 6.3, je patrná určitá tendence k vyšším hodnotám AT ve čteném projevu; rozdíl mezi hodnotami AT ve sledovaných mluvních stylech nicméně nebyl vyhodnocen jako statisticky

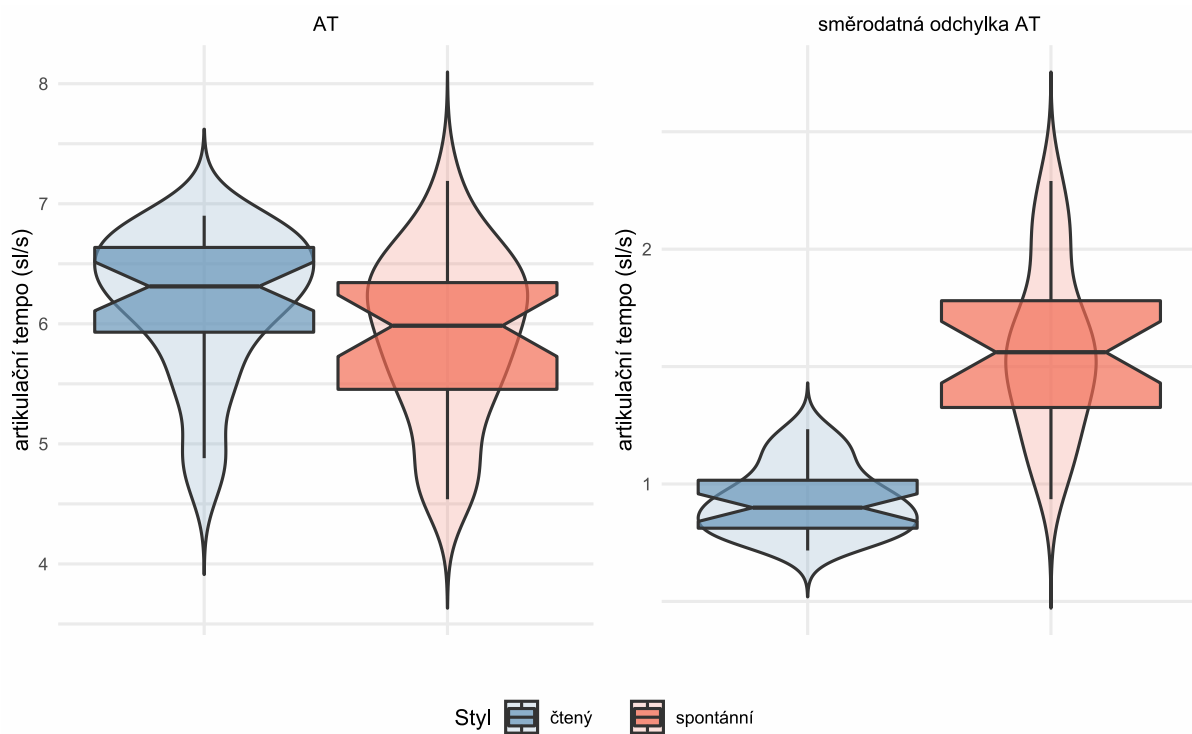
signifikantní ($p = 0,0164$). (Pro konkrétní hodnoty všech temporálních parametrů naměřené v obou mluvních stylech viz přílohu B.)



Obrázek 6.2. Artikulační tempo a směrodatná odchylka artikulačního tempa jednotlivých mluvčích ve čteném a spontánním projevu. Body u AT reprezentují průměr hodnot z jednotlivých promluvo- vých úseků, linky znázorňují 95% konfidenční intervaly.

Jakkoliv o markantních posunech průměrného artikulačního tempa mezi mluvčími styly u většiny mluvčích hovořit nelze, je možné si povšimnout obecného trendu, který se týká šířky konfidenčních intervalů, jež reflektují proměnlivost

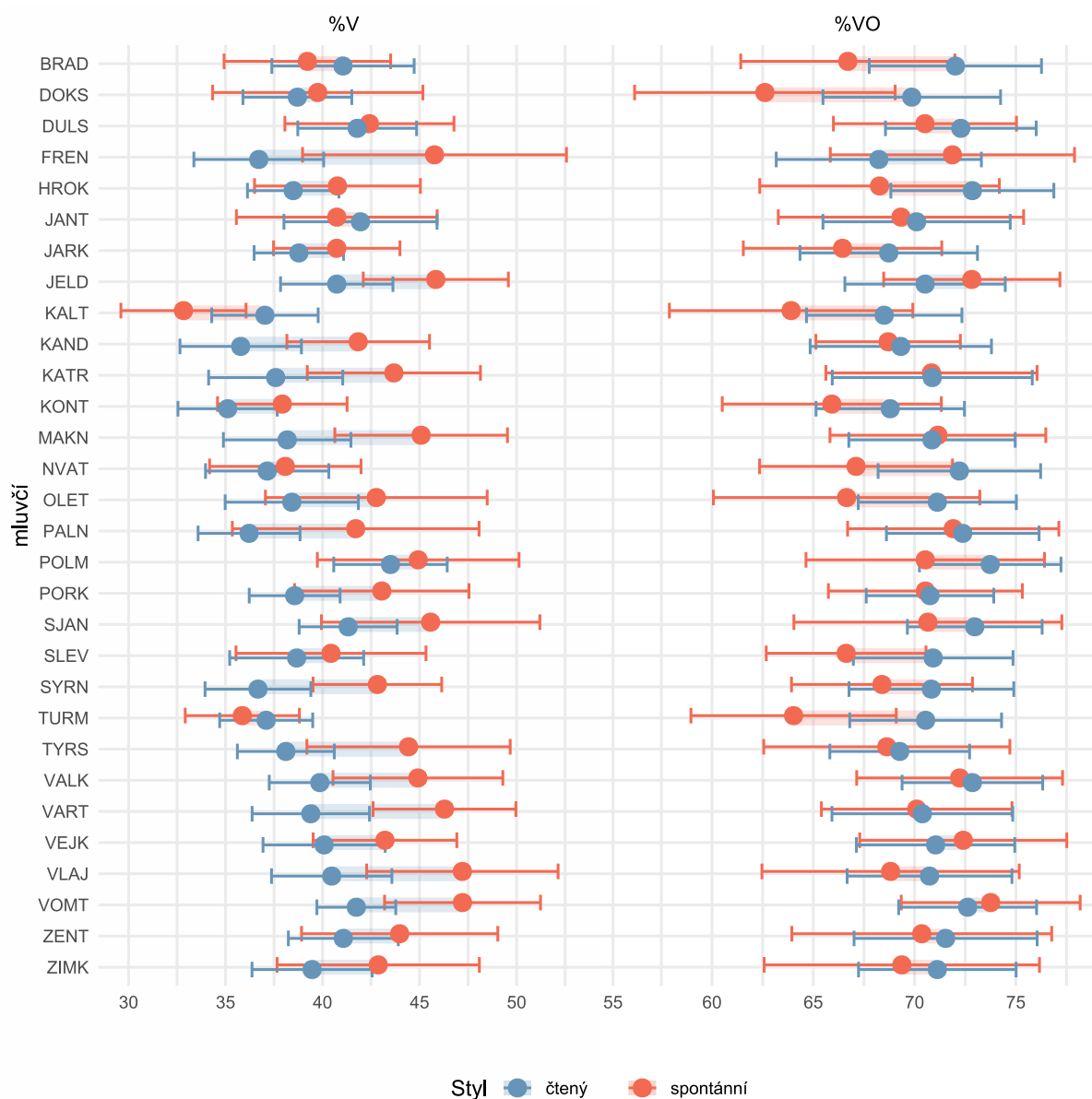
parametru uvnitř řečových vzorků (ačkoli mohou být také ovlivněny počtem hodnot extrahovaných v rámci daného vzorku). V pravé části obrázku 6.2 jsou vynešeny hodnoty směrodatné odchylky artikulačního tempa v jednotlivých nahrávkách, které charakterizují právě variabilitu AT. V tomto zobrazení je viditelná zcela konzistentní tendence artikulačního tempa k dosahování výraznější variability ve spontánním projevu ve srovnání s projevem čteným. Tyto rozdíly napříč mluvčími ilustruje pravá část obrázku 6.3, na níž je viditelný značný rozdíl mezi směrodatnou odchylkou AT ve sledovaných mluvních stylech, který se také ukázal jako statisticky signifikantní ($p < 0,0001$). Ze stejného obrázku je také patrné, že ve čteném projevu jsou hodnoty SD AT napříč mluvčími jednotnější, než je tomu v případě projevu spontánního.



Obrázek 6.3. Artikulační tempo a jeho směrodatná odchylka napříč mluvčími ve čteném a spontánním projevu.

Obrázek 6.4 znázorňuje hodnoty %V a %VO naměřené v jednotlivých řečových vzorcích. U obou parametrů je zřejmý výrazný překryv napříč mluvčími, pro

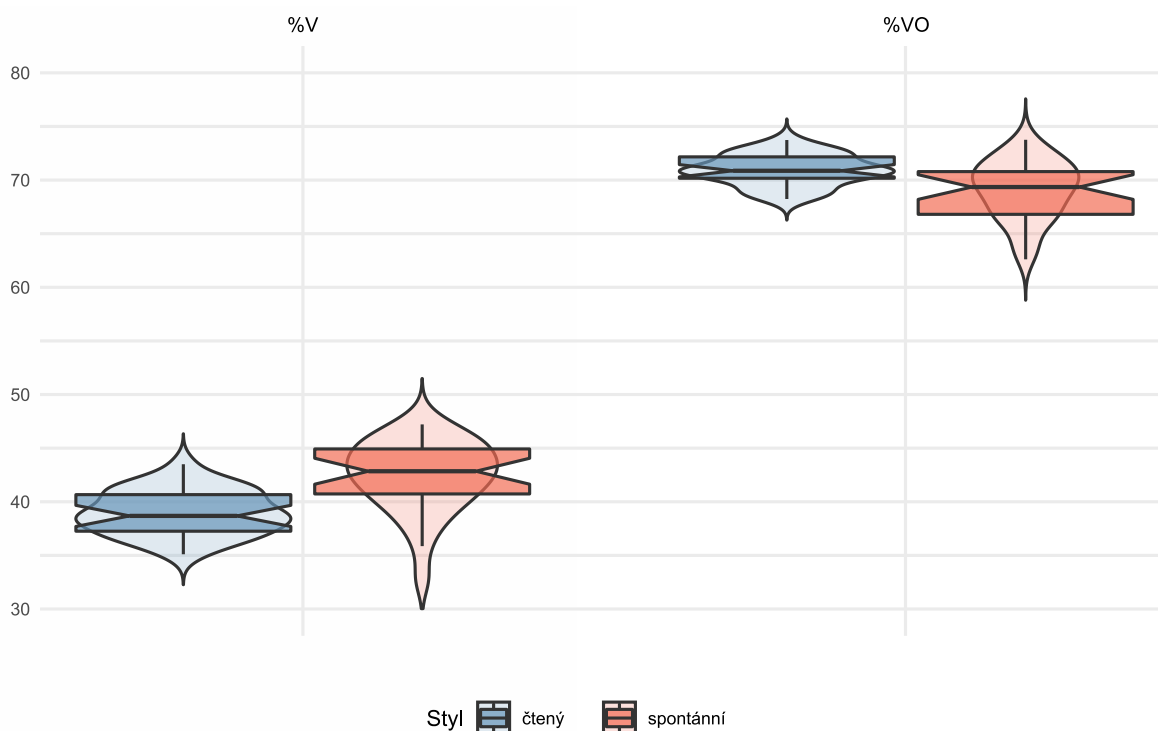
vzájemné rozlišování jednotlivců by tedy tento parametr přílišný přínos nepředstavoval. Výrazné rozdíly nejsou přítomny ani mezi mluvními styly, pozorovatelný je nicméně téměř jednotný, byť mírný trend napříč mluvčími, jímž je nižší %V a naopak vyšší %VO ve čteném projevu. Tuto obecnou tendenci demonstruje obrázek 6.5, který zachycuje %V a %VO napříč mluvčími, a také podle výsledků Wilcoxonova testu se jedná o významné rozdíly ($p < 0,001$ a $p = 0,0002$). Vedle toho je také na zmíněném obrázku patrný větší rozptyl hodnot %VO napříč mluvčími ve spontánním projevu; v případě parametru %V je obdobný trend taktéž pozorovatelný, ovšem v nižší míře.



Obrázek 6.4. Podíl vokalických a znělých intervalů v řeči u jednotlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot z jednotlivých promluvovacích úseků, linky znázorňují 95% konfidenční intervaly.

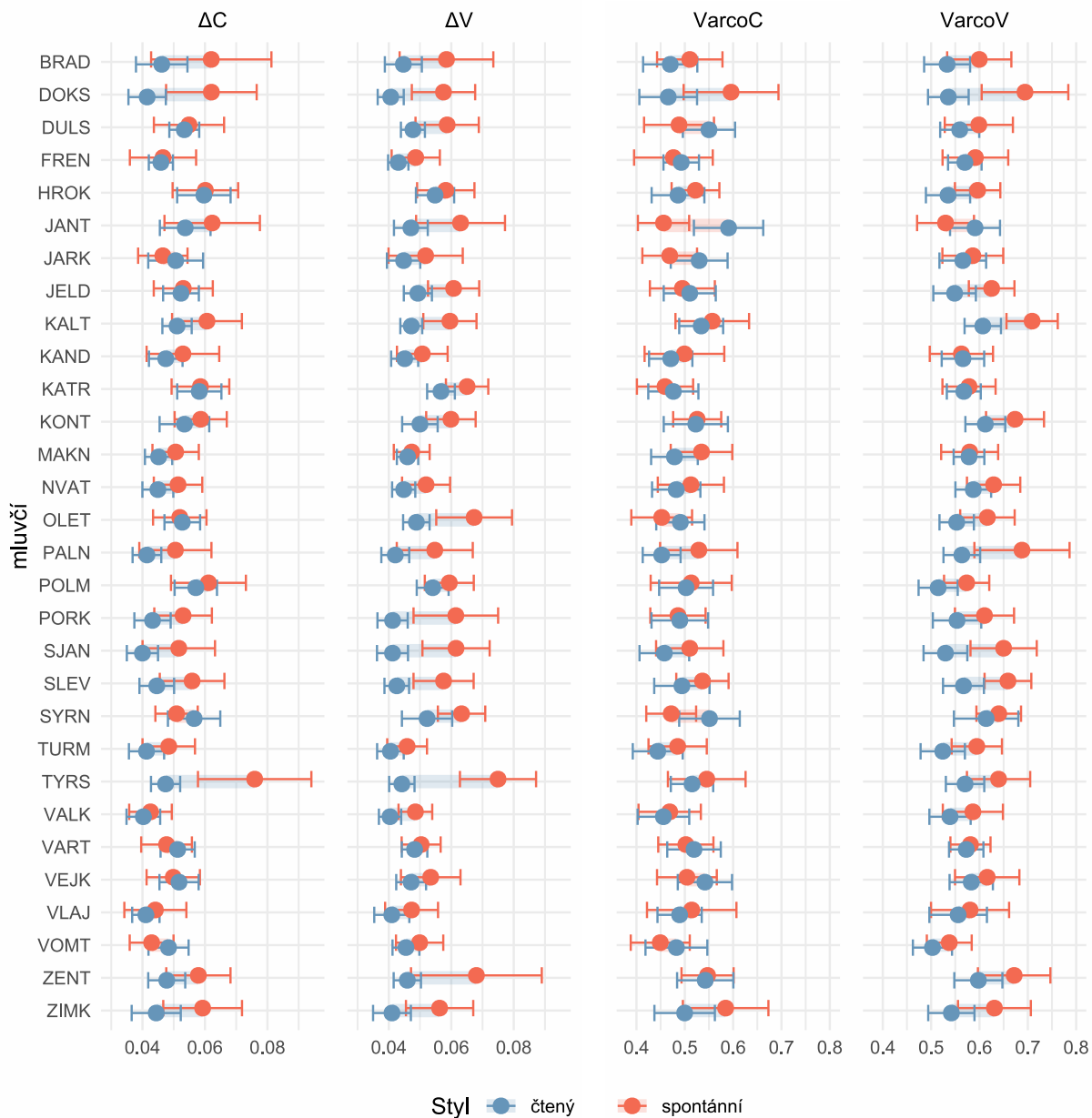
Je ovšem potřeba upozornit, že získané hodnoty nelze považovat za zobecnitelné pro čtený projev celkově – jak již bylo zmíněno, do textu, jež mluvčí předčítali, byly cíleně ve větším množství zařazeny segmenty, které jsou v běžném jazyce méně frekventované. Rozdíly v průměrném zastoupení jednotlivých tříd hlásek ve využitých nahrávkách ilustruje obrázek 6.1; zatímco vokalických segmentů se ve spontánním projevu objevovalo průměrně 42 %, v případě čteného textu se

jednalo o 40 %. Naproti tomu znělé hlásky (tedy vokály, sonory a znělé obstruenty) tvořily průměrně 75 % všech segmentů ve spontánním projevu a 77 % v projevu čteném. Ačkoliv spolu popsané ukazatele, tedy podíl vokalických a znělých intervalů v řeči, souvisejí, nebyla u nich pozorována silná korelace; jejich korelační koeficient je roven pouze 0,36 (viz obr. 6.8).



Obrázek 6.5. Podíl vokalických a znělých intervalů v řeči napříč mluvčími ve čteném a spontánním projevu.

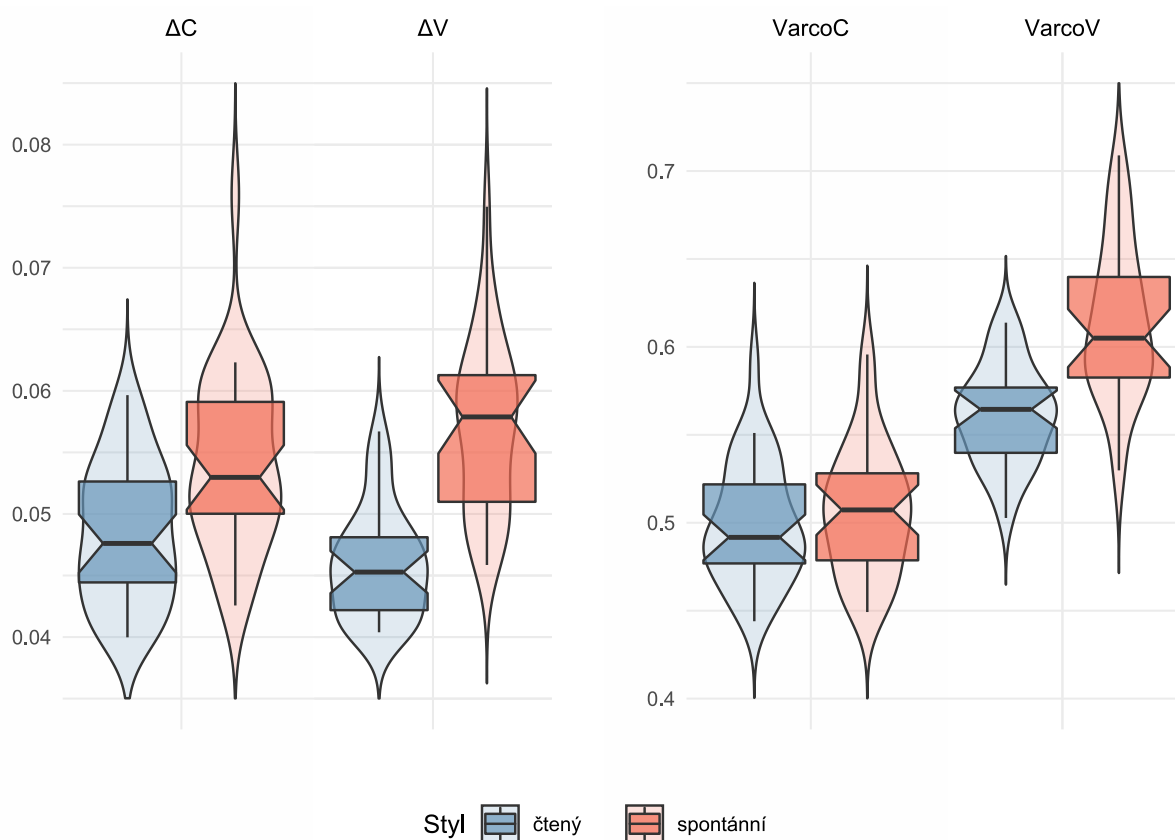
Na základě srovnání sledovaných ukazatelů variability trvání vokalických a konsonantických intervalů, tedy ΔV , ΔC , VarcoV a VarcoC , u jednotlivých mluvčích, které znázorňuje obrázek 6.6, lze konstatovat, že mezi spontánním a čteným projevem jednotlivce obecně nebývá přítomen výrazný rozdíl, ovšem ani rozdíly mezi jednotlivci nejsou natolik markantní, aby indikovaly individualitu mluvčího a bylo na jejich základě možné jej odlišit od ostatních.



Obrázek 6.6. Směrodatná odchylka trvání konsonantických a vokálních intervalů (vlevo) a vari-
ační koeficient směrodatné odchylky trvání konsonantických a vokálních intervalů (vpravo) u jed-
notlivých mluvčích ve čteném a spontánním projevu. Body reprezentují průměr hodnot
z jednotlivých promluвовých úseků, linky znázorňují 95% konfidenční intervaly.

Jakkoliv je ovšem rozdíl mezi spontánním a čteným projevem drobný a v rámci jednotlivých vzorků by patrně nebyl statisticky významný, je i zde, obdobně jako v případě %V a %VO, znatelný jakýsi trend jednotný téměř pro všechny mluvčí, kterým jsou mírně vyšší hodnoty (tj. větší variabilita trvání vokálních a

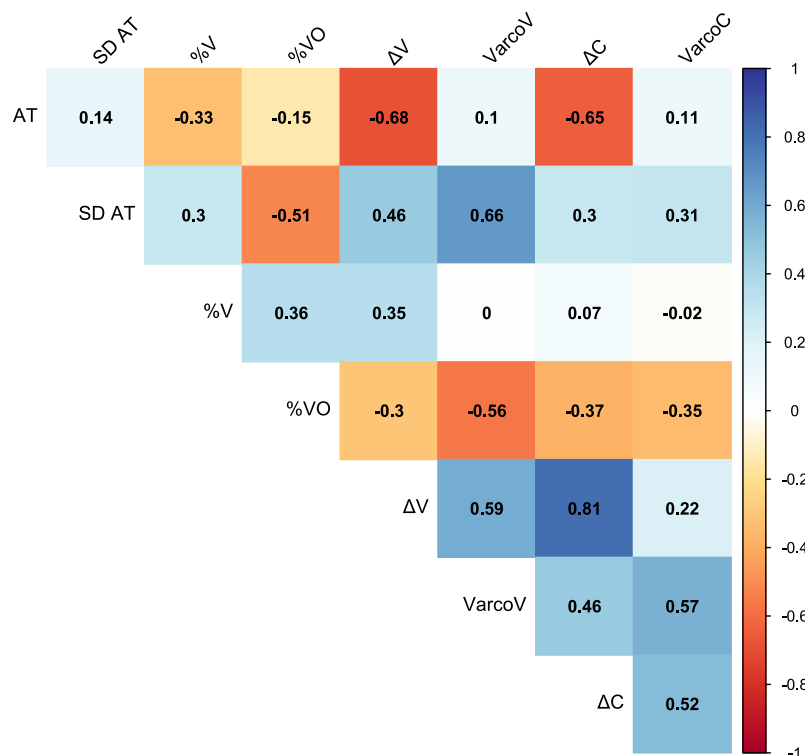
konsonantických intervalů) ve spontánním projevu. Přítomnost tohoto trendu ilustruje obrázek 6.7, který obsahuje srovnání průměrných hodnot sledovaných parametrů ze všech řečových vzorků. Výraznější rozdíly ukazatelů variability lze pozorovat u vokálních intervalů, přičemž podle výsledků Wilcoxonova testu se jedná o rozdíly signifikantní (v obou případech je $p < 0,0001$). Poněkud menší, ovšem taktéž signifikantní rozdíl ve sledovaných mluvních stylech je přítomen u ΔC ($p = 0,0003$). Naproti tomu v případě parametru VarcoC se rozdíl mezi čteným a spontánním projevem jako statisticky významný neprojevil ($p = 0,371$) a také na obrázku 6.7 si lze povšimnout, že na rozdíl od ostatních tří parametrů zde mezi oběma mluvními styly není přítomen nápadný rozestup.



Obrázek 6.7. směrodatná odchylka trvání konsonantických a vokálních intervalů (vlevo) a variační koeficient směrodatné odchylky trvání konsonantických a vokálních intervalů (vpravo) napříč mluvními ve čteném a spontánním projevu.

Směrodatné odchylky trvání vokalických a konsonantických intervalů, tedy ΔV a ΔC , spolu velmi silně korelují (viz obr. 6.8) – jejich korelační koeficient je roven 0,81. Při normalizaci průměrným trváním intervalů spolu však výsledné parametry, tedy VarcoV a VarcoC, korelují pouze středně silně ($r = 0,57$). Středně silnou vzájemnou korelaci pak vykazují i ΔV a VarcoV ($r = 0,59$) a ΔC a VarcoC ($r = 0,52$).

Jak ukazuje obrázek 6.8, mimo ΔV a VarcoV není u žádné další dvojice sledovaných temporálních ukazatelů přítomna velmi silná korelace. Silnou korelaci však vykazují ΔV a ΔC s artikulačním tempem, přičemž v obou případech se jedná o korelaci negativní ($r = -0,68$ a $r = -0,65$), a také VarcoV (nikoli však VarcoC) se směrodatnou odchylkou artikulačního tempa ($r = 0,66$).



Obrázek 6.8. Korelační matice sledovaných temporálních parametrů.

6.4.3 Diskuse a závěr

V rámci experimentu 6 byly analyzovány vybrané temporální parametry řečového signálu: artikulační tempo (AT), podíl vokalických intervalů v řeči (%V), podíl znělých intervalů v řeči (%VO), směrodatná odchylka trvání vokalických a konsonantických intervalů v řeči (ΔV a ΔC) a variační koeficient ΔV a ΔC (VarcoV a VarcoC). Byla sledována jejich variabilita napříč mluvčími a také u jednotlivců ve čteném a spontánním projevu.

Artikulační tempo vyjádřené coby počet realizovaných slabik za sekundu se ukázalo jako parametr poměrně málo variabilní napříč mluvčími, obecně tedy nelze předpokládat, že by artikulační tempo mohlo hrát zásadní roli při jejich vzájemném rozlišování. Ve vzorku se ovšem objevili i jednotlivci, jejichž artikulační tempo se znatelně odchylovalo od průměrného chování dané populace – v takových případech by tento parametr o individualitě mluvčích vypovídat mohl. Jedná se tedy o obdobné závěry, k jakým ve své studii dospěla Gold (2018). Je však zapotřebí neopomíjet aspekt mluvního stylu; jakkoli u většiny mluvčích nebyl zaznamenán výrazný rozdíl mezi hodnotami AT ve čteném a spontánním projevu, sledovaný vzorek obsahoval i nahrávky projevu jednotlivců, u nichž výrazné rozdíly přítomny byly.

Napříč mluvčími byla pozorována určitá tendence k o něco vyššímu artikulačnímu tempu ve čteném projevu, která však nebyla vyhodnocena jako statisticky signifikantní. Její přítomnosti ovšem nepřímo nasvědčují hodnoty indexu kumulativní strmosti základní frekvence analyzovaného v kapitole 3 (oddíl 3.4), který využívá normalizaci trváním či počtem slabik: zatímco při normalizaci slabikami u uvedeného parametru nebyl pozorován významný rozdíl mezi sledovanými mluvními styly, při normalizaci trváním byly ve čteném projevu pozorovány vyšší hodnoty CSI – tato diskrepance souvisí právě s rozdíly v artikulačním tempu. Tento výsledek koresponduje se závěry, k nimž dospěl Jessen (2007), výsledky

jehož experimentu prokázaly tendenci k vyššímu artikulačnímu tempu ve čteném projevu, která se ukázala jako statisticky signifikantní.

Ve srovnání s celkovým artikulačním tempem doznala výraznějších posunů mezi mluvními styly jeho směrodatná odchylka, tedy metrika odrážející variabilitu tohoto parametru uvnitř jednotlivých řečových vzorků. Jak se ukázalo, mluvčí mají tendenci s artikulačním tempem výrazněji kolísat ve spontánním projevu, zatímco ve čteném projevu jsou hodnoty tohoto ukazatele obecně stabilnější – i tento závěr je v souladu s poznatky Jessena (2007), který v rámci své studie pozoroval signifikantně vyšší směrodatnou odchylku artikulačního tempa ve spontánním projevu než v projevu čteném. Tato skutečnost přitom není způsobena případným výskytem disfluenčních prvků či delších pauz, které byly při výpočtu artikulačního tempa vyřazeny. Sledovaný materiál obsahoval nahrávky mluvčích, jejichž AT ve spontánním projevu dosahovalo jak nižších, tak vyšších hodnot než v projevu čteném, nelze tedy obecně říci, že by se ve spontánním projevu vyskytovaly pouze úseky s obdobným či nižším AT, k čemuž by vedla např. větší míra dloužení hlásek. Na základě získaných výsledků je možné konstatovat, že ve spontánním projevu existuje obecná tendence výskytu úseků jak s vyšším, tak s nižším artikulačním tempem ve srovnání s projevem čteným. I napříč mluvčími byla směrodatná odchylka artikulačního tempa jednotnější v případě čteného projevu, míra proměnlivosti artikulačního tempa ve spontánním projevu se napříč mluvčími různila zřetelně více.

U parametrů postihujících relativní zastoupení vokalických a znělých intervalů v řeči, tedy %V a %VO, byl zaznamenán značný překryv hodnot napříč mluvčími, nejedná se tedy zřejmě o ukazatel s příliš vysokým idiosynkratickým potenciálem. Byla pozorována všeobecná tendence k nižším hodnotám %V a naopak vyšším hodnotám %VO ve čteném projevu – byť spolu tedy tyto ukazatele souvisí, projevil se u nich opačný trend. Zde lze pozorovat určitý rozpor se závěry Leemana et al. (2014), kteří tyto parametry naopak hodnotí coby stabilní napříč

mluvními styly. Na tomto místě je ovšem potřeba uvést, že zatímco obsah spontánního projevu závisel na komunikačních strategiích samotných mluvčích, text pro čtený projev byl jednotný a záměrně formulovaný tak, aby ve větší míře obsahoval segmenty méně frekventované v běžné řeči. V důsledku této skutečnosti se ve využitých vzorcích čteného projevu objevovalo relativně méně vokálů a naopak více všech znělých segmentů než v nahrávkách projevu spontánního. Roli také mohlo hrát zastoupení fonologicky dlouhých vokálů, jichž se ve čteném projevu vyskytovalo relativně více než v projevu spontánním, jak je zmiňováno v kapitole 4 v rámci experimentu 2: Parametry vokalických formantů (viz obrázek 4.4). Zůstává tedy otázkou, nakolik by se hodnoty %V a %VO lišily v závislosti na mluvším stylu, pakliže by byly jednotlivé segmenty v předčítaném textu reprezentovány s četností odpovídající běžné řeči.

Co se týče parametrů postihujících variabilitu trvání vokalických a konsonantických intervalů, tj. ΔV , ΔC , VarcoV a VarcoC, ani zde nelze hovořit o výrazné interindividuální variabilitě. Zejména v případě vokalických intervalů byla napříč mluvčími zjištěna určitá tendence k relativně vyšší proměnlivosti trvání ve spontánním projevu. U parametrů ΔV a ΔC byla podle očekávání nalezena korelace s artikulačním tempem (vzhledem k níž Volín, 2007 doporučuje při srovnávání nahrávek s odlišným artikulačním tempem využívat normalizované ukazatele VarcoV a VarcoC, viz oddíl 6.2).

Zásadní výhoda temporálních parametrů oproti parametrům odvozeným od frekvenčních charakteristik signálu spočívá v jejich odolnosti vůči kvalitativní degradaci signálu, jeví se tudíž jako vhodné i pro vzájemné srovnávání řečových vzorků s podstatně rozdílnou kvalitou. Nicméně temporální parametry sledované v rámci popsaného experimentu zřejmě nepředstavují příliš nosné ukazatele pro vzájemné odlišování mluvčích, ačkoliv např. artikulační tempo by patrně mohlo přispívat analýze využívající jiné parametry v případě jednotlivců, u kterých se hodnoty tohoto ukazatele vymykají populačnímu standardu. Obecně byla

u temporálních parametrů pozorována tendence k většímu kolísání v rámci řečových vzorků i k nejednotnosti napříč mluvčími ve spontánním projevu ve srovnání s projevem čteným.

7 Intenzita

Řečovým rysem, kterým se od sebe mluvčí mohou vzájemně odlišovat, je hlasitost, s jakou hovoří. Jejím akustickým korelátem je intenzita zvukového signálu. Srovnávání intenzity napříč řečovými vzorky je však problematické, neboť vedle vlastní řečové produkce závisí na mnoha dalších faktorech, jakými jsou např. vlastnosti nahrávacího zařízení, vzdálenost mluvčího od mikrofonu či akustické dispozice prostorů, v nichž proběhlo nahrávání. Pro vzájemnou srovnatelnost celkové intenzity dvou řečových vzorků by tedy bylo zapotřebí zajistit shodné podmínky při jejich pořizování, jejichž dosažení je však prakticky nemožné. Nabízí se ovšem možnost sledovat kolísání intenzity signálu v čase, které může vykazovat vzorce charakteristické pro mluvčího. Parametry založené na intenzitě, které jsou využívány k vzájemnému srovnávání různých nahrávek, tudíž vždy zahrnují temporální aspekt.

7.1 Proměnlivost intenzity

Jak již bylo uvedeno v kapitole 3 (oddíl 3.3), kolísání intenzity v řečovém signálu kvantifikovali např. Adami et al. (2003), kteří jej sledovali uvnitř jednotlivých znělých segmentů společně s pohyby základní frekvence. Jestliže v průběhu daného segmentu intenzita vykazala nárůst, byl označen symbolem „+“, v případě poklesu naopak „-“ (dělení segmentů tedy bylo pouze binární, o rovných konturách se autoři nezmiňují); obdobně byl vyhodnocován také průběh f_0 a pozornost byla věnována také trvání jednotlivých segmentů. V tomto způsobu parametrizace průběhu intenzity společně se základní frekvencí autoři spatřují značný potenciál pro vzájemné rozlišování mluvčích.

He a Dellwo (2017) představili ukazatel nazvaný intenzitní dynamika, který postihuje trvání přechodů amplitudy mezi jejími vrcholy a propady; jako pozitivní dynamika je označována rychlost nárůstu amplitudy od jejího propadu po následující vrchol, zatímco negativní dynamikou autoři nazývají rychlost poklesu amplitudy od vrcholu k následujícímu propadu v amplitudové obálce. Amplitudová dynamika je spojována s pohyby artikulačních orgánů, které odrážejí individuální fyziologii i kinematiku vokálního traktu daného mluvčího (podobný předpoklad uvádí např. McDougall, 2006; viz kapitolu 1); jejich studie je založena na premise, že tvar intenzitní kontury může být úzce provázán s artikulačními pohyby, konkrétně otevíráním a zavíráním vokálního traktu, přičemž v okamžiku největšího otevření vokálního traktu nastává amplitudový vrchol, při jeho zavření naopak amplitudový propad³⁵. Autoři tedy usuzují, že jestliže je rychlost artikulačních pohybů charakteristická pro daného mluvčího, bude tuto individualitu odrážet také tvar intenzitní kontury.

Ve zmíněné studii sledovali He a Dellwo (2017) podobu intenzitní dynamiky v nahrávkách čteného projevu ve švýcarské němčině. V řečovém materiálu byly automaticky umístěny slabičné hranice a následně byly označeny vrcholy a propady intenzitní kontury – vrcholy ležely v místech intenzitních maxim mezi slabičnými hranicemi a za body propadů byla považována amplitudová minima nacházející se mezi dvěma vrcholy. Intenzitní kontury byly normalizovány tak, aby ve všech analyzovaných vzorcích dosahovaly stejné maximální hodnoty. Následně byla měřena rychlost a strmost nárůstů a propadů intenzity mezi jejími vrcholy a propady: ve dvourozměrném prostoru definovaném časem (vyčísleným v sekundách) a intenzitou (vyjádřenou prostřednictvím decibelů) byly vyznačeny body vrcholů a propadů intenzitní kontury; pozitivní dynamika byla stanovena jakožto velikost vektoru směřujícího od propadu k následujícímu vrcholu, negativní dynamika pak od vrcholu k následujícímu propadu. Na základě získaných hodnot bylo

³⁵ Podobný princip autoři sledovali také v rámci studie zaměřené na dynamiku prvního formantu (He et al., 2019), která byla popsána v oddílu 4.3 kapitoly 4.

vyčísleno několik parametrů: průměr, směrodatná odchylka a PVI (index párové variability) pozitivní i negativní dynamiky. Výsledky popsaného experimentu prokázaly značnou variabilitu sledovaných parametrů napříč mluvčími, přičemž výrazně více se mezi jednotlivci lišila dynamika negativní. Menší interindividuální variabilitu parametrů pozitivní dynamiky autoři přisuzují důslednějšímu ovládnutí artikulačních gest při otevírání vokálního traktu, neboť mohou obsahovat více lingvisticky relevantní informace, zatímco artikulační gesta při zavírání vokálního traktu mohou poskytovat větší prostor pro individuální variaci.

Na uvedenou studii navázala Lins Machado (2021) při zkoumání spontánního projevu rodilých mluvčích holandštiny, a to v jejich mateřském jazyce a v angličtině (která představovala jejich druhý jazyk). I v případě jejího experimentu vykazovala větší potenciál pro rozlišení mluvčích dynamika negativní, nicméně na rozdíl od He a Dellwa (2017) nepozorovala tento rozdíl v tak vysoké míře, což dává do spojitosti s povahou využitého materiálu – spontánní projev vykazuje vyšší různorodost artikulačních a temporálních vzorců než projev čtený (De Nil, 1991; Illa, 2020; obojí citováno v: Lins Machado, 2021: 169).

Volín et al. (2017) využili pro parametrizaci proměnlivosti intenzity v signálu parametr nazvaný index kumulativní strmosti (cumulative slope index; CSI), který byl již popsán v kapitole 3 (oddíl 3.2) a který zachycuje rozsah a rychlost změn daného parametru v čase. Ve své studii, zaměřené na změny prozodických charakteristik u českých mluvčích v závislosti na věku, extrahovali autoři hodnotu intenzity v 10ms intervalech a následně vyčíslili CSI pro jednotlivé nahrávky. Výsledky jejich analýzy poukazují na větší variabilitu intenzity u starších mluvčích; předmětem studie nebylo vyhodnocení, nakolik by tento parametr mohl být přínosný při vzájemném rozlišování jednotlivců, lze nicméně očekávat, že by v něm, podobně jako v případě intenzitní dynamiky, mohly být pozorovány určité individuální vzorce.

Vzhledem ke značné náchylnosti intenzity k vlivům okolností, jakými jsou vzdálenost mluvčího od mikrofonu či šum na pozadí, a s ohledem na vlastnosti materiálu, který je v této práci využíván (každý mluvčí byl nahráván na odlišném místě, akustické charakteristiky prostředí tudíž nebyly totožné, a není možné zaručit, že po celou dobu nahrávání mluvčí udržovali stejnou vzdálenost od nahrávacího zařízení; pro podrobnější informace o vlastnostech tohoto materiálu viz kapitolu 2), není vlastní analýza parametrů intenzity řečového signálu součástí této práce.

8 Experiment 7: Souhrnná analýza akustických parametrů řečového signálu

8.1 Metoda

Na základě výsledků všech předchozích kapitol byl sestaven soubor akustických parametrů odvozených od jednotlivých proměnných (základní frekvence, vokálních formantů F_1 , F_2 a F_3 , frekvenčního spektra vokálních a konsonantických segmentů, dlouhodobého průměrného spektra a temporální roviny řeči), přičemž byly vyřazeny vzájemně velmi silně korelované parametry (tj. takové, jejichž korelační koeficient dosáhl 0,8). Výsledný soubor tvoří 33 metrik:

parametry základní frekvence	základní hladina f_0 (ZH f_0 ; mezi dalšími silně korelovanými parametry, kterými byly průměr a medián f_0 , byla zvolena právě základní hladina vzhledem k její stabilitě v rámci projevu jednoho mluvčího, viz kapitolu 3, oddíl 3.4)
	směrodatná odchylka f_0 (SD f_0)
	CSI normalizovaný počtem slabik získaný z vyhlazené kontury f_0 (dále pouze „CSI f_0 “; ze čtyř sledovaných parametrů CSI, které vzájemně silně korelovaly, byl zvolen vzhledem k tomu, že u parametrů získaných z vyhlazených kontur f_0 byla pozorována větší stabilita v rámci projevu mluvčího, a normalizace počtem slabik byla vybrána vzhledem k provázanosti melodie řeči, jejímž je f_0 akustickým korelátem, spíše s lingvistickým obsahem promluvy než s jejím trváním)

parametry vokalických formantů	Velikost vokalického prostoru (VSA)
	LTF ₁ , LTF ₂ , LTF ₃
	F ₁ [i:], F ₁ [u]
	F ₂ [i:], F ₂ [u] (vzhledem k tomu, že velmi vysoká korelace nebyla zaznamenána ani u F ₂ [o], by stanovenému postupu odpovídalo také jeho zařazení, nicméně v zájmu konzistentnosti byl tento parametr vyčleněn)
	F ₃ [i:], F ₃ [u]
parametry frekvenčního spektra vokálů	H1*-H2*, H2*-H4*
	H1*-A1*, H1*-A2*
	H4*-H2kHz*
parametry frekvenčního spektra konsonantů	Spektrální těžiště [s], směrodatná odchylka [s] (SD [s]), špičatost [s]
	Spektrální těžiště [m], spektrální těžiště [l]
dlouhodobé parametry spektrálního sklonu	BgNoFo, index α
	Hammarberg index
temporální parametry	artikulační tempo (AT), směrodatná odchylka artikulačního tempa (SD AT)
	%V, %VO
	ΔV , VarcoV, VarcoC

U vybraných parametrů byla sledována míra korelace napříč proměnnými pro případné vyřazení silně korelovaných parametrů odvozených od různých proměnných.

Na základě průměrných hodnot jmenovaných akustických parametrů v jednotlivých nahrávkách byla v programu R provedena analýza hlavních komponent³⁶ (*principal component analysis*, PCA) a byla vypočítána vlastní čísla (*eigenvalues*) jednotlivých komponent, která reprezentují jimi postihovanou míru variability dat. Komponenty s vysokými vlastními čísly vysvětlují velkou část variability dat, zatímco ty s nízkými vlastními čísly představují spíše náhodný šum v datech (podrobněji viz např. Volín, 2007: 299–307). Ze 33 výsledných komponent byly podle Kaiserova kritéria (Volín, 2007: 302) vybrány ty, jejichž vlastní číslo bylo větší než 1.

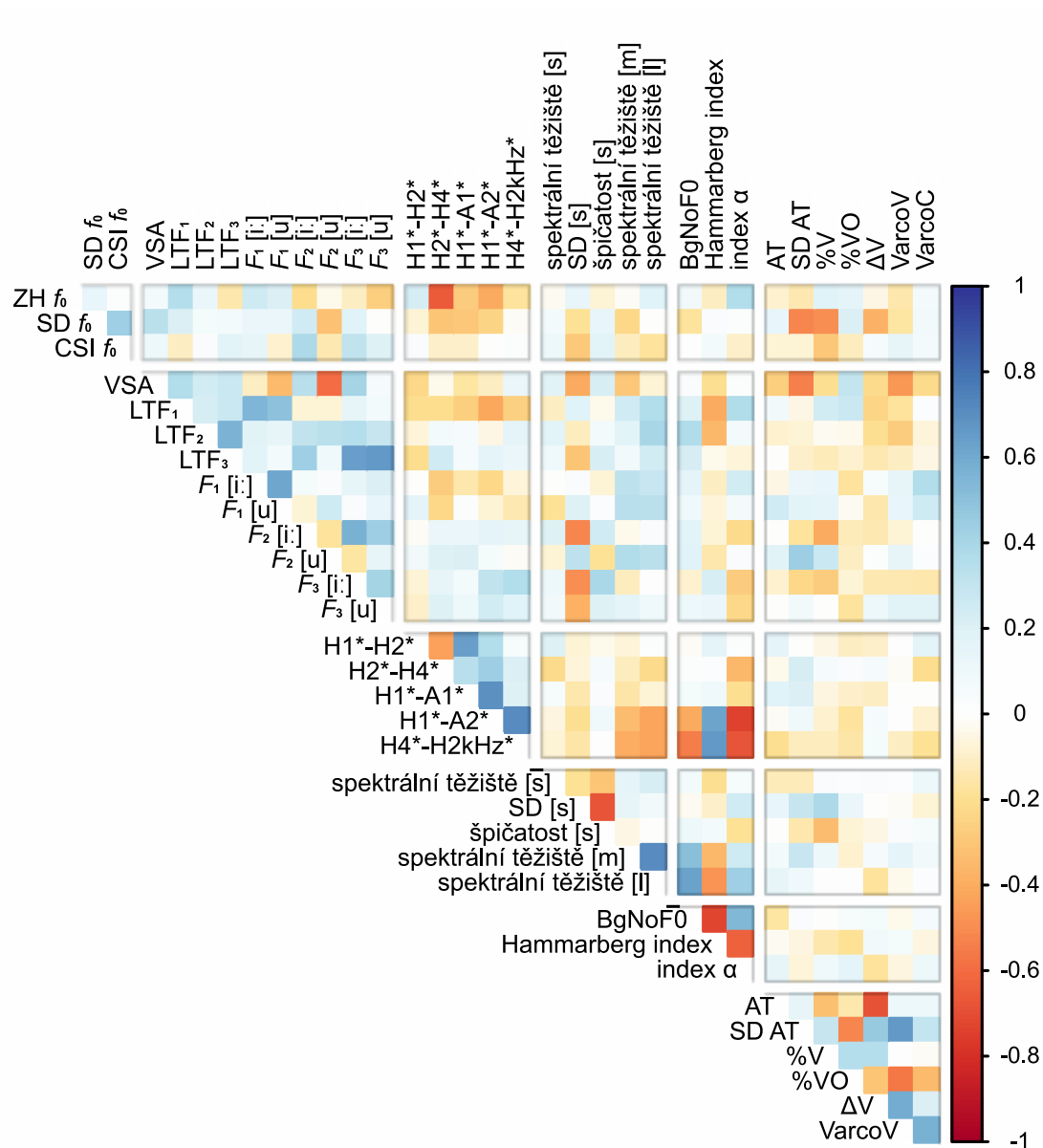
Pro každou z takto získaných hlavních komponent bylo sledováno, jak velký podíl celkového rozptylu dat postihuje, a dále její korelace s původními akustickými parametry, aby bylo možné posoudit, zda mají některé z měřených parametrů (či proměnných) zásadnější podíl na akustické variabilitě řečového signálu ve srovnání s ostatními. Hodnoty jednotlivých komponent byly také porovnány u jednotlivých mluvčích a ve čteném a spontánním projevu.

8.2 Výsledky

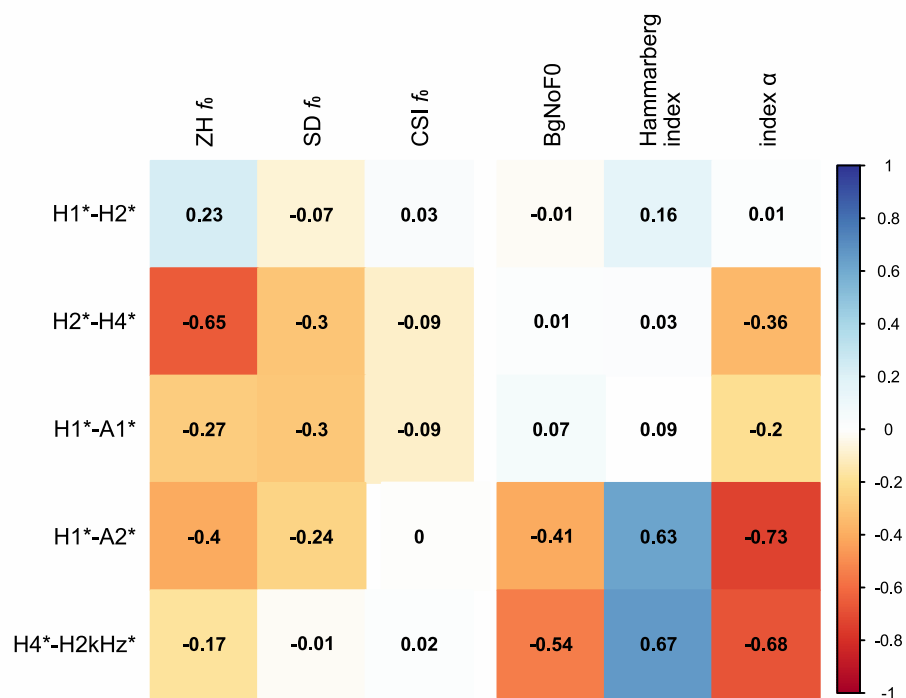
Míru korelace napříč všemi parametry ukazuje obrázek 8.1. Jak je z něj patrné, mezi parametry odvozenými od různých proměnných ve většině případů neexistuje silná korelace; výjimku tvoří základní hladina f_0 s $H2^*$ - $H4^*$, tj. parametry odvozené od základní frekvence a frekvenčního spektra vokálů ($r = -0,65$) a dále

³⁶ Volín (2007: 299) se kloní k označení „analýza hlavních komponentů“ v souladu s doporučením Slovníku spisovné češtiny, nicméně vzhledem k častějšímu zastoupení spojení „analýza hlavních komponent“ v dostupné literatuře napříč obory bude i v rámci této práce užívána tato varianta.

H1*-A2* a H4*-H2kHz* s Hammarberg indexem a indexem α , tj. parametry odvozené od frekvenčního spektra vokálů, které odrážejí krátkodobý spektrální sklon, a dlouhodobého frekvenčního spektra, které odrážejí spektrální sklon dlouhodobý ($|r| \in \langle 0,63; 0,73 \rangle$); podrobnější informace o vzájemné korelaci parametrů založených na těchto proměnných poskytuje obrázek 8.2. U žádné dvojice parametrů napříč proměnnými však nebyla zjištěna velmi silná korelace (tj. taková, jejíž korelační koeficient by dosáhl hodnoty 0,8).

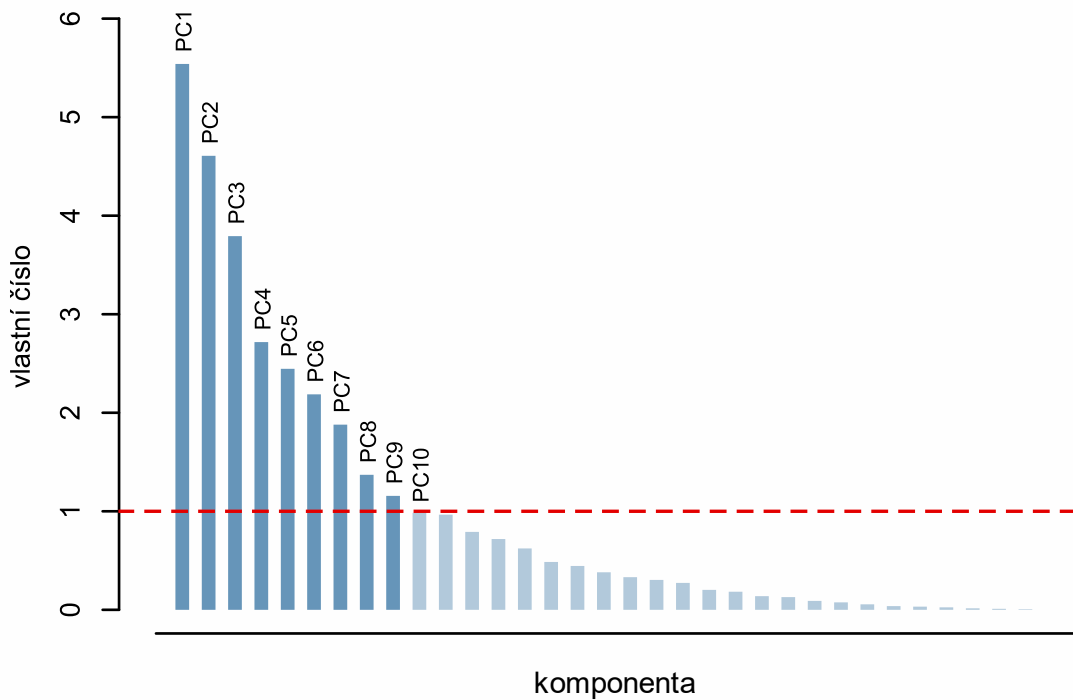


Obrázek 8.1. Korelační matice vybraných akustických parametrů řečového signálu. Parametry jsou rozděleny do skupin podle proměnné, od níž jsou odvozeny.

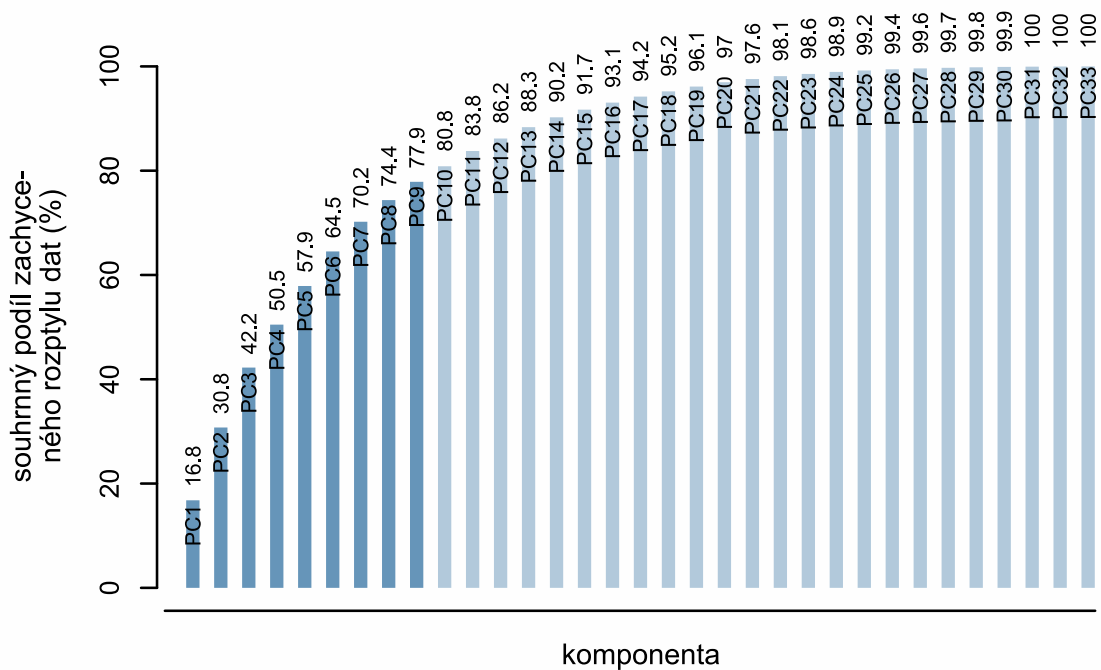


Obrázek 8.2. Korelační matice parametrů založených na základní frekvenci a dlouhodobém průměrném spektru (vodorovně) a frekvenčním spektru vokálních formantů (svisle).

Pro další analýzu bylo vybráno celkem 9 komponent, jejichž vlastní číslo bylo vyšší než 1; výši vlastních čísel jednotlivých komponent ukazuje obrázek 8.3. Vlastní čísla prvních devíti komponent se pohybují mezi 1,15 (PC9) a 5,54 (PC1). Dohromady dané komponenty vysvětlují 77,9 % variability dat; celkový podíl variability dat zachycovaný jednotlivými komponentami předkládá obrázek 8.4.



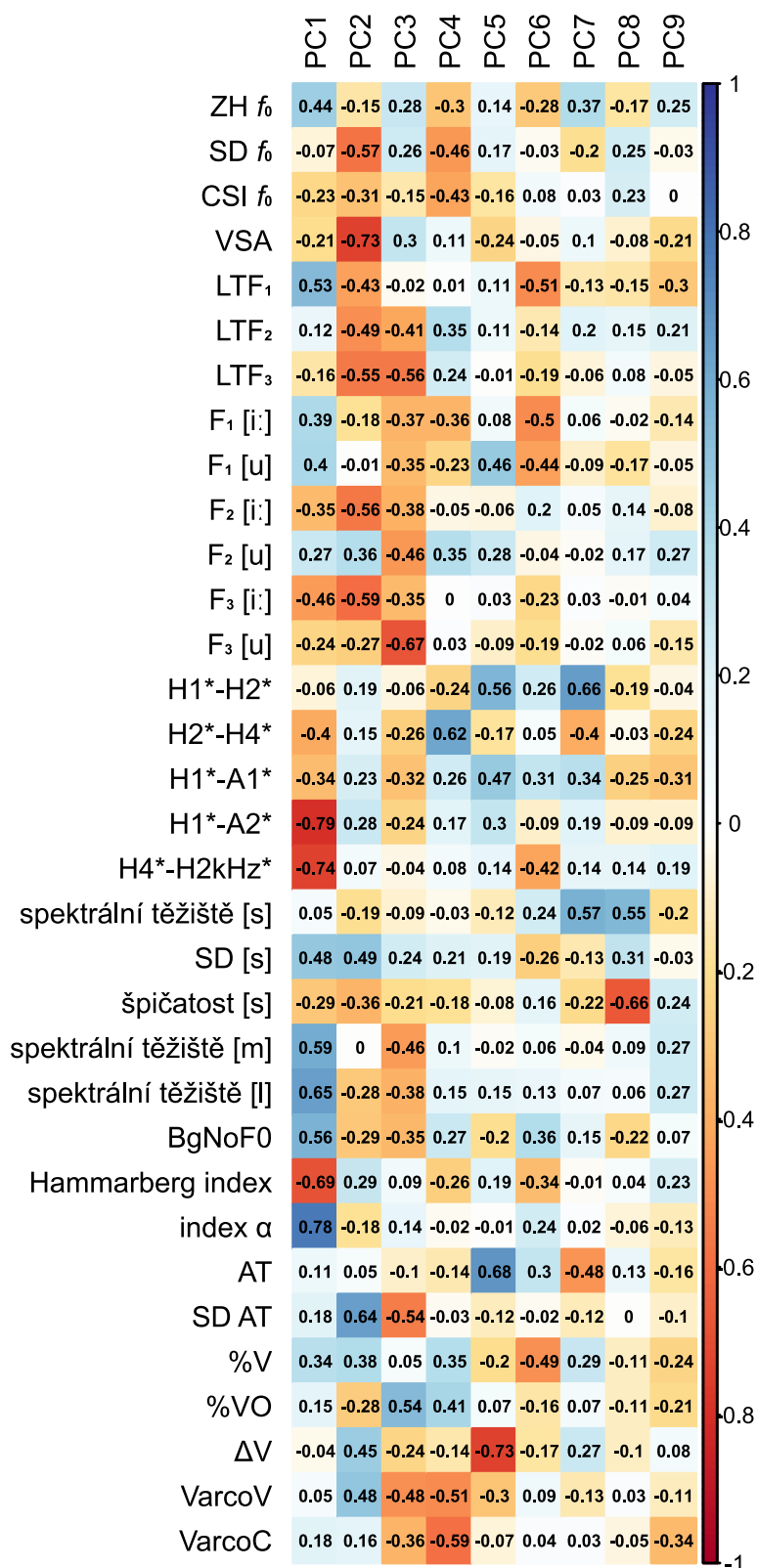
Obrázek 8.3. Vlastní čísla jednotlivých komponent; vodorovná přerušovaná čára znázorňuje hranici pro určení, které komponenty jsou zásadní a které naopak představují spíše šum v datech (hranice byla stanovena v souladu s Kaiserovým kritériem; viz výše).



Obrázek 8.4. Celkový podíl rozptylu v datech zachycený danou komponentou společně s komponentami předchozími.

Obrázek 8.5 prezentuje míru korelace hlavních komponent PC1–PC9 s hodnotami sledovaných akustických parametrů. Na základě získaných výsledků nelze vyvozovat přímočarou korelaci jednotlivých komponent s určitými parametry v tom smyslu, že by se u nich ukázala silná či velmi silná korelace s jediným parametrem, zatímco s ostatními nikoli (tento výsledek ovšem není překvapivý). U PC1 je nicméně přítomna silná korelace ($|r| \in \langle 0,65; 0,79 \rangle$) s parametry postihujícími krátkodobý spektrální sklon ve vokálech ($H1^*-A2^*$, $H4^*-H2kHz^*$) i spektrální sklon dlouhodobého průměrného spektra (Hammarberg index, index α), a dále také u spektrálního těžiště konsonantu [l]. PC2 vykázala korelaci zejména s parametry odvozenými od vokálních formantů; silná korelace byla pozorována s velikostí vokálního prostoru ($r = -0,73$), u některých dalších parametrů pak byla pozorována středně silná korelace. Parametrem z odlišné kategorie, který rovněž silně koreloval s PC2, byla směrodatná odchylka artikulačního tempa ($r = 0,64$). Třetí komponenta také vykázala nejnápadnější korelaci s některými parametry vokálních formantů, její silná korelace byla pozorována s třetím formantem vokálu [u]. (Pro podrobné informace o míře korelace dalších komponent s akustickými parametry viz obrázek 8.5.)

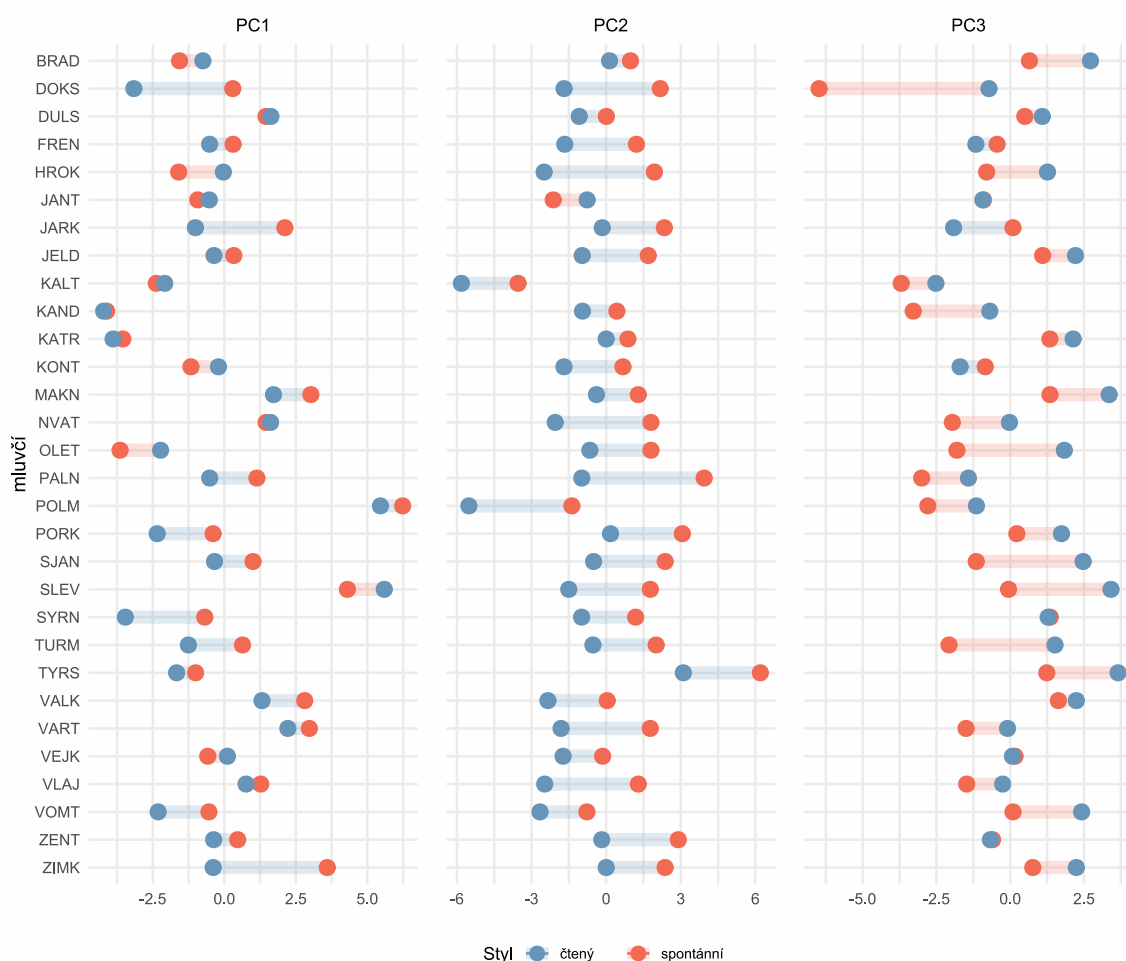
Na základě těchto výsledků lze konstatovat, že nejvýraznější část variability v řečovém signálu je určována vokálními formanty a (krátkodobým i dlouhodobým) spektrálním sklonem, zatímco role ostatních akustických parametrů je nižší; např. spektrální momenty konsonantu [s] se na variabilitě dat podílejí spíše okrajově, podobně jako globální temporální ukazatele.



Obrázek 8.5. Korelační matice prvních devíti komponent a jednotlivých akustických parametrů.

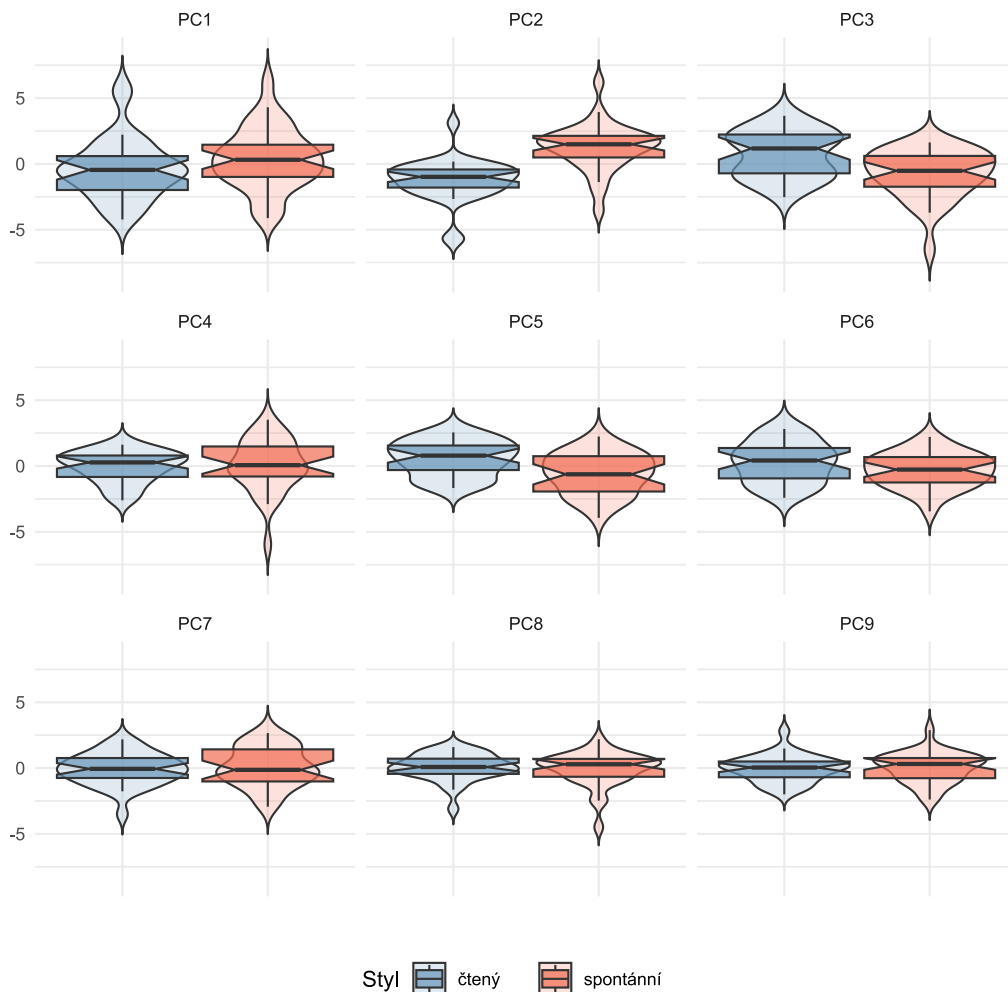
Hodnoty prvních tří výše zmíněných komponent u jednotlivých mluvčích ve sledovaných mluvních stylech jsou pro ilustraci znázorněny na obrázku 8.6. Zejména PC1 by patrně mohla vzájemně odlišit některé mluvčí, neboť interindividuální variabilita v případě velké části mluvčích přesahuje variabilitu intraindividuální, zatímco zejména PC2 se u většiny mluvčích příliš neliší, avšak nápadný je u ní téměř konzistentní posun k vyšším hodnotám ve spontánním projevu; u PC3 se naopak téměř u všech mluvčích objevují vyšší hodnoty v projevu čteném.

Pro analýzu hlavních komponent byly ovšem využity pouze průměrné hodnoty sledovaných parametrů z jednotlivých nahrávek, a nejsou tudíž k dispozici konfidenční intervaly, spolehlivost získaných hodnot pro jednotlivé řečové vzorky tak není možné vyhodnotit.



Obrázek 8.6. Hodnoty prvních tří hlavních komponent u jednotlivých mluvčích ve čteném a spontánním projevu.

Ze srovnání komponent mezi čteným a spontánním projevem, které znázorňuje obrázek 8.7, vyplývá, že nejvýrazněji se v těchto mluvních stylech liší PC2 a dále také PC3 a PC5; u těchto komponent byl rozdíl mezi mluvními styly za využití Wilcoxonova testu vyhodnocen jako statisticky signifikantní (PC2 a PC3: $p < 0,0001$; PC5: $p = 0,0002$).



Obrázek 8.7. Hodnoty jednotlivých hlavních komponent ve čteném a spontánním projevu.

8.3 Diskuse a závěr

V rámci experimentu 7 byla provedena analýza hlavních komponent na základě parametrů analyzovaných v předchozích experimentech, u kterých nebyla

pozorována velmi silná vzájemná korelace. Jednalo se o parametry odvozené od základní frekvence řečového signálu, vokálních formantů, frekvenčního spektra vybraných vokálních a konsonantických segmentů, dlouhodobého průměrného spektra a temporální roviny řeči. U hlavních komponent získaných za využití hodnot daných parametrů byla pozorována míra jejich korelace s jednotlivými parametry, z níž byla odvozována míra dopadu těchto parametrů na variabilitu řečového signálu, a také jejich rozdíly napříč mluvčími a mezi čteným a spontánním mluvním stylem.

Před provedením vlastní analýzy hlavních komponent byla vyhodnocena korelace parametrů napříč sledovanými proměnnými; v předchozích experimentech byla korelace sledována pouze u parametrů příslušejících k téže proměnné. Jak se ukázalo, většina parametrů napříč různými proměnnými nevykázala silnou korelaci, výjimku ovšem představovaly parametry odvozené od frekvenčního spektra vokálů, které odrážejí krátkodobý spektrální sklon, a od dlouhodobého frekvenčního spektra, které odrážejí spektrální sklon dlouhodobý – jednalo se o parametry $H1^*-A2^*$ a $H4^*-H2kHz^*$ z první jmenované kategorie a Hammarberg index a index α z kategorie druhé. Tuto skutečnost bylo možné předpokládat vzhledem k tomu, že všechny jmenované ukazatele parametrizují tentýž jev, tedy spektrální sklon, byť se liší typem využívaného frekvenčního spektra (krátkodobé \times dlouhodobé) a analyzovanými úseky řečového signálu (vokální segmenty \times znělé řečové proudy). Dále se korelace objevila také mezi základní hladinou f_0 a $H2^*-H4^*$, v tomto případě však důvod pro takovou korelaci není zcela zřejmý. U žádné dvojice parametrů odvozených od odlišných proměnných ovšem nebyla zaznamenána velmi silná korelace.

V rámci analýzy hlavních komponent bylo vybráno prvních devět komponent s vlastním číslem větším než 1, které dohromady postihovaly 77,9 % variability dat. Z pozorování míry korelace mezi jednotlivými komponentami a sledovanými akustickými parametry vyplynulo, že první komponenta (vysvětlující 16,8 %

variability dat) koreluje zejména s ukazateli spektrálního sklonu, a to jak krátkodobého, tak dlouhodobého; druhá a třetí komponenta (postihující společně s komponentou první 30,8 %, resp. 42,2 % variability dat) pak korelovaly převážně s parametry odvozenými od vokálních formantů. Nejvýraznější dopad na celkovou variabilitu dat se tedy ukázal právě u spektrálního sklonu a vokálních formantů, zatímco např. role temporálních ukazatelů či spektrálních momentů konsonantických segmentů se projevila jako méně významná a taktéž hodnoty ukazatelů základní frekvence nevykázaly silnou korelaci s žádnou z prvních devíti hlavních komponent (nabízí se ovšem otázka, nakolik by výsledky analýzy hlavních komponent pozměnilo zahrnutí průměru či mediánu základní frekvence namísto její základní hladiny, která – ačkoli s oběma velmi silně korelovala – se ve srovnání s nimi projevila jako stabilní pro mluvčího nezávisle na mluvním stylu, viz kapitolu 3, oddíl 3.4).

Porovnání hlavních komponent napříč mluvčími a mluvními styly ukázalo (jak ilustruje obrázek 8.6), že určité komponenty se vyznačují výraznější interindividuální variabilitou ve srovnání s variabilitou intraindividuální (např. PC1), zatímco u jiných byly nápadné zejména rozdíly mezi čteným a spontánním projevem (nejvýrazněji v případě PC2). Vzhledem k využití průměrných hodnot jednotlivých parametrů z každého řečového vzorku byly pro každého mluvčího získány pouze dvě hodnoty od každé hlavní komponenty (jedné odrážející čtený a druhé odrážející spontánní projev), vynášení závěrů ohledně stability komponent v rámci projevu mluvčího je tudíž problematické, nicméně výsledky experimentu naznačují, že právě určité hlavní komponenty odvozené od hodnot akustických parametrů řeči by mohly odrážet individualitu mluvčího, zatímco jiné by mohly poukazovat na daný mluvní styl či další faktory ovlivňující podobu řečového signálu. Do budoucna by tedy mohlo být přínosné provést analýzu hlavních komponent na základě akustických parametrů odvozených od různých proměnných s využitím materiálu obsahujícího více typů projevu (lišících se mluvními styly,

ale i dalšími faktory, např. podmínkami při nahrávání) od jednotlivých mluvčích a pozorovat možné způsoby jejího využití v této oblasti.

9 Celková diskuse a závěr

Předchozí kapitoly představily způsoby parametrizace řečového signálu pomocí široké škály akustických parametrů odvozených od různých akustických proměnných; shrnuly dosavadní metodologické postupy aplikované v rámci dostupných studií při zkoumání variability řečového projevu a popsaly také analýzu intraindividuální a interindividuální variability vybraných parametrů v rámci vlastních experimentů. Tato práce tedy představuje ucelený popis metodologie využívané při analýze řady akustických parametrů, pomocí kterých je možné charakterizovat podobu řečového signálu, včetně demonstrace vybraných postupů na jednotném řečovém materiálu. Tato práce obsahuje (alespoň v českém prostředí) jedinečnou souhrnnou analýzu takto obsáhlého souboru akustických parametrů na jednotném řečovém materiálu.

Jednotlivé dílčí experimenty sledovaly, nakolik se liší či naopak podobají hodnoty akustických parametrů příslušejících k jednotlivým proměnným (základní frekvenci, vokálním formantům, frekvenčnímu spektru a temporální rovině řeči; celkem se jednalo o 61 metrik) mezi různými mluvčími a také mezi čteným a spontánním projevem. Výstupem těchto experimentů je mimo jiné zmapování distribuce hodnot sledovaných parametrů v obou mluvních stylech u dané populace, kterou představují rodilí mluvčí obecné češtiny mužského pohlaví mezi 19 a 50 lety (viz přílohu B; analýza byla provedena na vzorku čítajícím 30 jedinců). Při vzájemném rozlišování mluvčích, resp. určení, zda na dvou různých nahrávkách hovoří tentýž jedinec, na základě akustických parametrů hraje roli nejen podobnost získaných hodnot, ale také jejich typičnost v rámci dané populace, a právě populační statistiky mohou sloužit jako opora k jejímu stanovení.

Mezi sledované vlastnosti jednotlivých parametrů patřil jejich potenciál pro vzájemné rozpoznávání mluvčích, pro nějž je zásadní jednak malá proměnlivost v projevu jednoho mluvčího – a to nejen v rámci jednoho řečového vzorku, ale

také mezi více různými vzorky (v této práci byly využity nahrávky lišící se mluvčím stylem) – a jednak rozdílnost mezi různými mluvčími. Z tohoto hlediska se jako nejslibnější parametry projeví základní hladina f_0 (obdobný závěr uvádějí např. Lindh a Eriksson, 2007 či Skarnitzl a Hývlová, 2014), dlouhodobý průměr F_3 (tento formant bývá obecně považován za přínosný ukazatel při identifikaci mluvčího, viz např. Skarnitzl et al., 2014) a spektrální momenty frikativy [s], zejména pak spektrální těžiště (a s ním velmi silně korelující sešikmení; tyto parametry považuje za přínosné při rozlišování mluvčích také např. Kavanagh, 2012). Žádný z uvedených parametrů ovšem neposkytoval dostatečné informace pro rozlišení všech mluvčích; při analýze s tímto cílem lze tudíž na základě výsledků provedených experimentů doporučit kombinaci více parametrů. Tento závěr koresponduje s výsledky dosavadních výzkumů; autoři dostupných studií zaměřených na idiosynkratický potenciál akustických parametrů zpravidla rovněž konstatují, že dílčí akustické parametry nemohou samy o sobě spolehlivě identifikovat jednotlivce, a doporučují aplikovat analýzu více parametrů zároveň (viz např. Hollien, 1990, citováno v: Nolan, 1999: 13; Adami et al., 2003 či Skarnitzl & Hývlová, 2014).

Při analýze akustických parametrů u jednotlivých mluvčích byla mimo jiné věnována pozornost jedincům, u nichž byly zjištěny hodnoty zřetelně se odlišující od hodnot naměřených u ostatních mluvčích, a bylo zjišťováno, zda lze ve využitém řečovém materiálu nalézt jednotlivce, u kterých by takto odchylné hodnoty byly přítomny v parametrech odvozených od více proměnných. Mezi sledovanými mluvčími se vyskytli tací, kteří vykazovali nestandardní hodnoty v parametrech příslušejících ke dvěma různým proměnným:

- u mluvčího JANT byla jako u jediného zjištěna vyšší hodnota vokalického prostoru (viz kapitolu 4, oddíl 4.5) a poměrně nízké artikulační tempo (viz kapitolu 6, oddíl 6.4) ve spontánním projevu ve srovnání s projevem čteným;

- mluvčí KALT jako jediný vykázal větší rozsah f_0 ve spontánním projevu (a to poměrně výrazně; viz kapitolu 3, oddíl 3.4), zároveň u něj byl změřen největší vokální prostor ve čteném projevu (viz kapitolu 4, oddíl 4.5);
- u mluvčího POLM byly pozorovány znatelně vyšší hodnoty parametrů střední hlasové frekvence ve čteném projevu oproti projevu spontánnímu, a to včetně základní hladiny, která u ostatních mluvčích zůstávala spíše stabilní nezávisle na mluvním stylu (viz kapitolu 3, oddíl 3.4), a tento mluvčí se také vyznačoval relativně nízkým artikulačním tempem v obou mluvních stylech (viz kapitolu 6, oddíl 6.4).

Na základě těchto pozorování tedy nelze říci, že by se mluvčí, kteří se odchylojí od pomyslného standardu, takto odchylovali ve shodných parametrech, ani že by při nestandardní hodnotě určitého parametru odvozeného od jedné proměnné u daného mluvčího bylo možné předpokládat také nestandardní hodnoty některého parametru v oblasti další proměnné. Mluvčí, u kterého by se výrazně vymykaly hodnoty parametrů odvozených od více než dvou proměnných, ve sledovaném materiálu nalezen nebyl.

Vedle odlišností hodnot jednotlivých parametrů mezi mluvčími byla věnována pozornost také jejich rozdílům mezi čteným a spontánním projevem. Jako výrazně závislé na mluvním stylu se ukázaly průměr a medián základní frekvence a také ukazatele odrážející její rozsah, tedy směrodatná odchylka a percentilová rozpětí, u kterých byly zjištěny vyšší hodnoty ve čteném projevu. Jak již ovšem bylo zmíněno, základní hladina, která se blíží spodní hranici rozsahu základní frekvence v daném řečovém vzorku, se naopak projevila jakožto ukazatel stabilní pro daného mluvčího. Tyto výsledky naznačují, že mluvčí mají tendenci využívat větší rozpětí f_0 (které je akustickým korelátem intonačního rozsahu) ve čteném projevu ve srovnání s projevem spontánním, přičemž dochází k posouvání horní hranice výše, zatímco spodní hranice zůstává spíše neměnná (toto pozorování je v souladu se závěry, ke kterým dospěli Lindh a Eriksson, 2007). Na nejen větší

intonační variabilitu, ale také větší variabilitu artikulační ve čteném projevu poukazují některé ukazatele F_1 a F_2 (které korespondují s vertikální a horizontální dimenzí artikulace vokálů), jmenovitě F_1 vokálu [a], jehož hodnoty nasvědčují otevřenější artikulaci, a F_2 vokálů [i:], [ɪ] a [ɛ], které poukazují na jejich přednější artikulaci ve čteném projevu; o tendenci k perifernější artikulaci ve čteném projevu vypovídají také vyšší hodnoty velikosti vokalického prostoru. Na základě ukazatelů základní frekvence (odrážející melodii řeči) i vokalických formantů (reflektujících artikulační nastavení) lze tedy usuzovat, že při předčítání textu mluvčí využívají většího intonačního rozsahu a pečlivější artikulace, zatímco při spontánním hovoru se uchylují spíše k plošší intonaci a k větší centralizaci vokálů.

Mezi parametry lišící se v závislosti na mluvním stylu patřily také některé temporální ukazatele, a to směrodatná odchylka artikulačního tempa a parametry odrážející variabilitu trvání vokalických a konsonantických intervalů v řeči: ΔV , ΔC a VarcoV (nikoli ovšem VarcoC, tj. při využití normalizace, jejímž cílem je eliminace vlivu artikulačního tempa, se již závislost trvání konsonantických intervalů na mluvním stylu neprojevila). Z tohoto výsledku lze odvozovat, že ve čteném projevu mluvčí udržují relativně stabilnější temporální průběh řeči než v projevu spontánním.

V obou mluvních stylech bylo mimo výše samotných hodnot akustických parametrů možné pozorovat také jejich distribuce, tedy to, zda jsou průměrné hodnoty parametrů spíše unifikované napříč mluvčími, či zda se v nich mluvčí naopak rozcházejí (při tomto srovnávání tedy není zohledňován rozptyl hodnot parametrů uvnitř jednotlivých řečových vzorků). U parametrů, v jejichž případě byly pozorovatelné odlišnosti mezi rozděleními průměrných hodnot ve sledovaných mluvních stylech, byl obecně nalezen větší rozptyl ve spontánním projevu; jednalo se o ukazatele střední hodnoty základní frekvence, artikulační tempo a směrodatnou odchylku artikulačního tempa. Je tedy možné konstatovat, že ve čteném projevu

jsou průměrné hodnoty těchto parametrů jednotnější napříč mluvčími, zatímco v projevu spontánním se jednotlivci navzájem více rozcházejí.

U výsledků srovnávání uvedených mluvních stylů je ovšem třeba brát v úvahu možný vliv obsahu promluv; zatímco spontánní projev se odvíjel od individuálních komunikačních strategií, jež volili jednotliví mluvčí, text pro čtený projev byl jednotný a jeho podoba byla specifická: jednalo se o popis interakce dvou postav obsahující přímou řeč, který byl záměrně formulován tak, aby obsahoval v češtině méně obvyklé hlásky a hláskové kombinace ve vyšší míře než běžná řeč. (Tento text je uveden v příloze A.)

V rámci závěrečného experimentu byla uskutečněna souhrnná analýza všech zkoumaných parametrů, která se zaměřila na jejich vzájemné souvislosti a také jejich vztah k celkové podobě řečového signálu. Byla pozorována silná korelace určitých parametrů krátkodobého a dlouhodobého spektrálního sklonu; tento výsledek není zcela překvapivý, neboť tyto metriky odrážejí tentýž jev, ač první jmenované využívají krátkodobé spektrum získané z jednotlivých segmentů, zatímco druhé vychází z dlouhodobého průměrného spektra. Dlouhodobé parametry spektrálního sklonu jsou relativně náročné na materiál, neboť pro stabilizaci podoby LTAS je zapotřebí řečový signál o trvání v řádu desítek sekund (viz např. Sergeant & Welch, 2007: 660). Vzhledem k silné korelaci s některými dlouhodobými parametry spektrálního sklonu ovšem určité parametry krátkodobé patrně odrážejí obdobné charakteristiky řečového projevu, a pro parametrizaci těchto charakteristik by tudíž patrně mohly posloužit i příslušné krátkodobé parametry spektrálního sklonu; z jejich menší náročnosti na materiál také vyplývá možnost výběru kvalitativně lépe vyhovujících vzorků (např. vokálů bez obsahu vnějších ruchů), která je v případě LTAS výrazně problematičtější. Dále se silná korelace projevila také mezi základní hladinou f_0 a jedním z parametrů krátkodobého spektrálního sklonu (konkrétně $H2^*-H4^*$), v tomto případě však vztah mezi jmenovanými

ukazateli není natolik zřejmý. U dalších parametrů napříč proměnnými se již silná korelace neukázala.

Pro zhodnocení, nakolik jednotlivé parametry postihují celkovou variabilitu řečového signálu, byla v závěrečném experimentu využita analýza hlavních komponent s využitím 33 metrik (jejich soubor byl sestaven tak, aby neobsahoval velmi silně korelované parametry). Výsledkem analýzy bylo celkem devět komponent s vlastním číslem vyšším než 1, jež dohromady vysvětlily 77,9 % variability využitých dat. S první komponentou, která sama postihovala 16,8 % variability dat, korelovaly zejména parametry odrážející (krátkodobý i dlouhodobý) spektrální sklon, s druhou a třetí komponentou, jejichž souhrnný podíl zachyceného rozptylu dat společně s první komponentou tvořil 42,2 %, vykázaly nejvýraznější korelaci určité parametry vokálních formantů. Je tedy možné říci, že na variabilitu řečového signálu nejvýrazněji dopadají právě parametry spektrálního sklonu a vokálních formantů, naproti tomu temporální parametry, ukazatele základní frekvence či spektrální momenty konsonantů se na této variabilitě podílejí spíše v nižší míře.

Na základě uvedených skutečností lze uzavřít, že tato práce předkládá pozorování intraindividuální a interindividuální variability rozsáhlého souboru akustických parametrů řečového signálu, údaje o jejich potenciálu pro vzájemnou diferenciaci jednotlivců a mluvních stylů a také zhodnocení jejich celkového podílu na akustické variabilitě řeči v češtině. Obsahuje také podrobný souhrn metodologických postupů, jež lze využít při akustické analýze řečového signálu, popsanych v rámci dosavadního výzkumu.

Seznam použité literatury

- Adami, A. G., Mihaescu, R., Reynolds, D. A., & Godfrey, J. J. (2003). Modeling prosodic dynamics for speaker recognition. *Proceedings of ICASSP'03*, IV-788.
- Arantes, P., & Eriksson, A. (2014). Temporal stability of long-term measures of fundamental frequency. In *Speech Prosody*, 2014, 1149-1152.
- Asadi, H., Nourbakhsh, M., He, L., Pellegrino, E., & Dellwo, V. (2018). Between-speaker rhythmic variability is not dependent on language rhythm, as evidence from Persian reveals. *International Journal of Speech, Language and the Law*, 25(2), 151-174.
- Bahmanbiglu, S. A., Mojiri, F., & Abnavi, F. (2017). The impact of language on voice: an LTAS study. *Journal of Voice*, 31(2), 249.e9-249.e12.
- Balkó, I. (2005). K výzkumu tempa řeči a tempa artikulace v různých řečových úlohách. *Bohemistika*, 3, 185-198.
- Bartoň, T., Cvrček, V., Čermák, F., Jelínek, T. & Petkevič, V. (2009). *Statistiky češtiny*. Nakladatelství Lidové noviny.
- Beck, J. M. (2010). Organic variation of the vocal apparatus. In W.J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences*, 256-297. Blackwell Publishers.
- Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer (Version 6.3.09). Staženo z <http://www.praat.org/>
- Bořil, T., & Skarnitzl, R. (2016). Tools rPraat and mPraat. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech, and Dialogue*, 367-374. Springer International Publishing.
- Braun, A. (1995). Fundamental frequency – How speaker-specific is it? *BEIPHOL 64, Studies in Forensic Phonetics*, 9-23.
- Cambier-Langeveld, T. (2007). Current methods in forensic speaker identification: Results of a collaborative exercise. *International Journal of Speech, Language & the Law*, 14(2), 223-243.
- Cardoso, A., Foulkes, P., French, J. P., Gully, A. J., Harrison, P. T., & Hughes, V. (2019). Forensic voice comparison using long-term acoustic measures of voice quality. *Proceedings of the 19th ICPHS*, 1455-1459.
- Cavalcanti, J. C., Eriksson, A., & Barbosa, P. A. (2021). Acoustic analysis of vowel formant frequencies in genetically related and non-genetically related speakers with implications for forensic speaker comparison. *Plos one*, 16(2), e0246645.

- Corrette, R. (2023). Praat Vocal Toolkit. Dostupné na <https://www.praatvocal-toolkit.com>.
- Czaplicki, B., Żygis, M., Pape, D., & Jesus, L. M. (2016). Acoustic evidence of new sibilants in the pronunciation of young Polish women. *Poznan Studies in Contemporary Linguistics*, 52(1), 1-42.
- Da Silva, R. R., da Costa, J. P. C., Miranda, R. K., & Del Galdo, G. (2016). Applying base value of fundamental frequency via the multivariate kernel-density in forensic speaker comparison. In *2016 10th International Conference on Signal Processing and Communication Systems*, 1-8.
- Dankovičová, J. (1997). The domain of articulation rate variation in Czech. *Journal of Phonetics*, 25(3), 287-312.
- Dankovičová, J. (1999). Articulation rate variation within the intonation phrase in Czech and English. *Proceedings of the 14th ICPHS*, 269-272.
- de Looze, C. (2010). *Analyse et Interprétation de l'Empan Temporel des Variations Prosodiques en Français et en Anglais*. Nepublikovaná disertační práce. Laboratoire Parole et Lang, Université de Provence - Aix-Marseille
- Dellwo, V., Leemann, A., & Kolly, M-J. (2015). Rhythmic variability between speakers: Articulatory, prosodic and linguistic factors. *Journal of the Acoustical Society of America*, 137, 1513-1528.
- Eichhorn, J. T., Kent, R. D., Austin, D., & Vorperian, H. K. (2018). Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice*, 32(5), 644.e1-644.e9.
- Eriksson, A., Thunberg, G. C., & Traunmller, H. (2001). Syllable prominence: a matter of vocal effort, phonetic distinctness and top-down processing. *Proceedings of Interspeech 2001*, 399-402.
- Esposito, C. M., Sleeper, M., & Schäfer, K. (2021). Examining the relationship between vowel quality and voice quality. *Journal of the International Phonetic Association*, 51(3), 361-392.
- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co.
- Fant, G. (1960). *Acoustic theory of speech production*. Mouton.
- Farrús, M., & Hernando, J. (2009). Using jitter and shimmer in speaker verification. *IET Signal Processing*, 3(4), 247-257.

- Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., & Liss, J. M. (2015). The relationship between speech segment duration and vowel centralization in a group of older speakers. *The Journal of the Acoustical Society of America*, 138(4), 2132–2139.
- Fuchs, S. (2017). Changes and challenges in explaining speech variation: A brief review. Dostupné na: https://www.researchgate.net/publication/320991961_Changes_and_challenges_in_explaining_speech_variation_A_brief_review.
- Garellek, M., Samlan, R. A., Kreiman, J., & Gerratt, B. R. (2013). Perceptual sensitivity to a model of the source spectrum. *Proceedings of Meetings on Acoustics*, Vol. 19, 060157, 1–5.
- Gold, E. (2018). Articulation rate as a speaker discriminant in British English. *Proceedings of Interspeech 2018*, 1828–1832.
- Gold, E., French, P. & Harrison, P. (2013). Examining long-term formant distributions as a discriminant in forensic speaker comparisons under a likelihood ratio framework. *Proceedings of Meetings on Acoustics 2013*, vol 19, 060041.
- Gottsmann, F. & Harwardt, C. (2011). Investigating robustness of spectral moments on normal- and high-effort speech. *Proceedings of Interspeech 2011*, 2937–2940.
- Grabe, E., Kochanski, G., & Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech*, 50(3), 281–310.
- Hammarberg, B., Fritzell, B., Gaufin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngologica*, 90(1–6), 441–451.
- Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America*, 101(1), 466–481.
- He, L., & Dellwo, V. (2017). Between-speaker variability in temporal organizations of intensity contours. *The Journal of the Acoustical Society of America*, 141(5), EL488–EL494.
- He, L., Zhang, Y., & Dellwo, V. (2019). Between-speaker variability and temporal organization of the first formant. *The Journal of the Acoustical Society of America*, 145(3), EL209–EL214.
- Heeren, W. F. L., & He, L. (2021). Between-speaker variability in segmental F1 dynamics in spontaneous speech. *Proceedings of IAFPA 2021*.
- Hollien, H. (2002). *Forensic Voice Identification*. Academic Press.

- Holmes, E. (2022). Recognising socio-phonetically comparable speakers using phonetic approaches to automatic speaker recognition. *Proceedings of IAFPA 2022*, 5–6.
- Houzar, A., Nechanský, T., & Skarnitzl, R. (2022). *Impact of vocal tract resonance modifications on LTF and f₀*. Prezentováno na konferenci IAFPA 2022, Praha, Česká republika.
- Houzar A. & Skarnitzl, R. (2022). Intra- and inter-speaker variability of vowel space using three different formant extraction methods. *AUC Philologica*, 2022(1), 83–95.
- Houzar, A., Nechanský, T. & Skarnitzl, R. (v tisku). The effect of targeted voice manipulations on long-term acoustic characteristics. *Proceedings of the 20th ICPHS*.
- Hruška, R. (2016). *Vlastnosti kontur základní frekvence s ohledem na segmentální složení promluv*. Nepublikovaná bakalářská práce. Fonetický ústav, Filozofická fakulta Univerzity Karlovy.
- Hruška, R., & Bořil, T. (2017). Temporal variability of fundamental frequency contours. *AUC Philologica*, 2017(3), 35–44.
- Hudson, T., De Jong, G., McDougall, K., Harrison, P., & Nolan, F. (2007). FO statistics for 100 young male speakers of Standard Southern British English. *Proceedings of the 16th ICPHS*, 1809–1812.
- Iseli, M., Shue, Y. L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America*, 121(4), 2283–2295.
- Jacewicz, E., & Fox, R. A. (2013). Cross-dialectal differences in dynamic formant patterns in American English vowels. In G. S. Morrison & P. F. Assmann (Eds.), *Vowel inherent spectral change*, 177–198. Springer Berlin.
- Jannedy, S., & Weirich, M. (2017). Spectral moments vs discrete cosine transformation coefficients: Evaluation of acoustic measures distinguishing two merging German fricatives. *The Journal of the Acoustical Society of America*, 142(1), 395–405.
- Jessen, M. (2007). Forensic reference data on articulation rate in German. *Science & Justice*, 47(2), 50–67.
- Jessen, M. (2010). The forensic phonetician: Forensic speaker identification by experts. In M. Coulthard & A. Johnson (Eds.), *The Routledge handbook of forensic linguistics*, 378–394. New York: Routledge.
- Jessen, M., Koster, O., & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *The International Journal of Speech, Language and the Law*, 12(2), 174–213.

- Jesus, L. M., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, 30(3), 437–464.
- Jin, S. H., & Liu, C. (2013). The vowel inherent spectral change of English vowels spoken by native and non-native speakers. *The Journal of the Acoustical Society of America*, 133(5), EL363–EL369.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3), 1252–1263.
- Kavanagh, C. (2012). *New consonantal acoustic parameters for forensic speaker comparison*. Nepublikovaná disertační práce. York: Department of Language and Linguistic Science, University of York.
- Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, 118(2), 1038–1054.
- Kreiman J., & Garellek M. (2011). Perceptual importance of the voice source spectrum from H2 to 2 kHz. *The Journal of the Acoustical Society of America*, 130, 2570.
- Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons.
- Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In: W.J.Hardcastle & N.Hewlett (Eds.), *Coarticulation: Theory, data and techniques*, 7–30. Cambridge University Press.
- Künzel, H. J. (1997). Some general phonetic and forensic aspects of speaking tempo. *International Journal of Speech, Language and the Law*, 4(1), 48–83.
- Leemann, A.; Kolly, M.-J. & Dellwo, V. (2014). Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison. *Forensic Science International*, 238, 59–67.
- Lechta, V. (2013). Diagnostika poruch fluence. In Z. Cséfalvay, V. Lechta & M. Wdowczynová (Eds.), *Diagnostika narušené komunikační schopnosti u dospělých*, 23–52. Portál.
- Lindh, J. (2006). Preliminary Descriptive Fo-statistics for Young Male Speakers. *Lund University Working Papers*, 52, 89–92.
- Lindh, J. & Eriksson, A. (2007). Robustness of Long Time Measures of Fundamental Frequency. *Proceedings of Interspeech 2007*, 2025–2028.
- Lins Machado, C. (2021). A cross-linguistic study of between-speaker variability in intensity dynamics in L1 and L2 spontaneous speech. In C. Bernardasci, D. Dipino,

- D. Garassino, S. Negrinelli, E. Pellegrino, & S. Schmid (Eds.), *L'individualità del parlante nelle Scienze Fonetiche: Applicazioni tecnologiche e forensi*, 157–174. Officina-ventuno.
- Machač, P., & Skarnitzl, R. (2009). *Fonetická segmentace hlásek*. Epocha.
- McDougall, K. (2006). Dynamic features of speech and the characterization of speakers: towards a new approach using formant frequencies. *International Journal of Speech, Language and the Law*, 13, 89–126.
- Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, 131(3), 2249–2260.
- Mooshammer, C. (2010). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *The Journal of the Acoustical Society of America*, 127(2), 1047–1058.
- Morrison, G. S. (2009). L1-Spanish Speakers' Acquisition of the English/i/–/I/Contrast II: Perception of Vowel Inherent Spectral Change. *Language and Speech*, 52(4), 437–462.
- Morrison, G. S. (2013). Vowel inherent spectral change in forensic voice comparison. In G. S. Morrison & P. F. Assmann (Eds.), *Vowel inherent spectral change*, 263–282. Springer Berlin.
- Nara, K. (2017). Acoustic and Electroglottographic Study of Breathy and Modal Vowels as Produced by Heritage and Native Gujarati Speakers. *Proceedings of Interspeech 2017*, 1054–1058.
- Nearey, T. M. (2013). Vowel inherent spectral change in the vowels of North American English. In G. S. Morrison & P. F. Assmann (Eds.), *Vowel inherent spectral change*, 49–85. Springer Berlin.
- Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America*, 80(5), 1297–1308.
- Nechanský, T., Houzar, A., Bořil, T. & Skarnitzl, R. (v recenzním řízení). Controlled voice quality modifications: acoustic, perceptual and ASR analysis. *International Journal of Speech, Language and the Law*.
- Ng, M. L., Chen, Y., & Chan, E. Y. (2012). Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese-English bilingual speakers—a long-term average spectral analysis. *Journal of Voice*, 26(4), e171–e176.

- Nolan, F. (1999). Speaker Recognition and Forensic Phonetics. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences*, 744–767. Blackwell Publishers.
- Nolan, F., & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law*, 12(2), 143–173.
- Nordenberg, M., & Sundberg, J. (2004). Effect on LTAS of vocal loudness variation. *Logopedics Phoniatrics Vocology*, 29(4), 183–191.
- Nunes, A., Coimbra, R. L., & Teixeira, A. (2010). Voice quality of european portuguese emotional speech. *Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language*, 142–151.
- Pandit, M., & Kittler, J. (1998). Feature selection for a DTW-based speaker verification system. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98* (Vol. 2), 769–772.
- Pollák, P., Volín, J., & Skarnitzl, R. (2007). HMM-Based Phonetic Segmentation in Praat Environment. *Proceedings of SPECOM 2007*, 537–541. MSLU.
- Pritschet, L., Powell, D., & Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: Changing attitudes over four decades. *Psychological Science*, 27(7), 1036–1042.
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vídeň, Rakousko. Staženo z <https://www.R-project.org/>.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265–292.
- Rose, P. (2015). Forensic voice comparison with monophthongal formant trajectories—a likelihood ratio-based discrimination of “schwa” vowel acoustics in a close social group of young Australian females. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4819–4823.
- Schenker, N. & Gengleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, 55(3), 182–186.
- Sergeant, D. & Welch, G. F. (2007). Age-related changes in long-term average spectra of children’s voices. *Journal of voice*, 22(6), 658–670.
- Seyfarth, S., & Garellek, M. (2015). Coda glottalization in American English. *Proceedings of the 18th ICPHS*, paper 807.

- Shadle, C. H. & Scully, C. (1995). An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics*, 23(1-2), 53-66.
- Shue, Y. (2013). VoiceSauce: A program for voice analysis (Version 1.14). K dispozici na <http://www.phonetics.ucla.edu/voicesauce/>.
- Shue, Y.-L., Keating, P., Vicenik, C. & Yu, K. (2011). VoiceSauce: A rprogram for voice analysis. *Proceedings of the 17th ICPhS*, 1846-1849.
- Skarnitzl, R. (2014). Forezní fonetika. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvího*, 11-20. Univerzita Karlova, Filozofická fakulta.
- Skarnitzl, R. & Hývlová, D. (2014). Statistický popis hodnot základní frekvence. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvího*, 49-64. Univerzita Karlova, Filozofická fakulta.
- Skarnitzl, R., Lazárková, D., Nechanský, T. & Šturm, P. (2014). Vokalické formanty. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvího*, 21-48. Univerzita Karlova, Filozofická fakulta.
- Skarnitzl, R., Šturm, P. & Volín, J. (2016). *Zvuková báze řečové komunikace: Fonetický a fonologický popis řeči*. Karolinum.
- Skarnitzl, R. & Vaňková, J. (2017). Fundamental frequency statistics for male speakers of Common Czech. *AUC Philologica*, 2017(3), 7-17.
- Skarnitzl, R.; Vaňková, J. & Bořil, T. (2015). Optimizing formant extraction of vowel formants. In O. Niebuhr & R. Skarnitzl (Eds.), *Tackling the complexity in speech*, 165-182. Univerzita Karlova, Filozofická fakulta.
- Skarnitzl, R., & Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické listy*, 18(1), 7-11.
- Smirnova, N., Starshinov, A., Oparin, I., & Goloshchapova, T. (2007). Speaker identification using selective comparison of pitch contour parameters. *Proceedings of the 16th ICPhS*, 1801-1804.
- Story, B. H., & Bunton, K. (2017). Vowel space density as an indicator of speech performance. *The Journal of the Acoustical Society of America*, 141(5), EL458-EL464.
- Svatošová, M., & Volín, J. (2022). Description of FO contours with Legendre polynomials. *AUC Philologica*, 2022(1), 97-113.
- Tamarit, L., Goudbeek, M., & Scherer, K. (2008). Spectral slope measurements in emotionally expressive speech. *Proceedings of Speech Analysis and Processing for Knowledge Discovery*, 007.

- Tavi, L. (2019). Classifying females' stressed and neutral voices using acoustic-phonetic analysis of vowels: an exploratory investigation with emergency calls. *International Journal of Speech Technology*, 22(3), 511–520.
- Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis-jitter, shimmer and hnr parameters. *Procedia Technology*, 9, 1112–1122.
- Themistocleous, C., Fyndanis, V., & Tsapkini, K. (2022). Sonorant spectra and coarticulation distinguish speakers with different dialects. *Speech Communication*, 142, 1–14.
- Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., Howard, D. M., Hunter, E. J., Kaelin, D., Kent, R. D., Kreiman, J., Kob, M., Löfqvist, A., McCoy, S., Miller, D. G., Noé, H., Scherer, R. C., Smith, J. R., Story, B. H., Švec, J. G., Ternström, S., & Wolfe, J. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America*, 137(5), 3005–3007.
- Traunmüller, H. & Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. Staženo z <https://www.researchgate.net/publication/240312210> The frequency range of the voice fundamental in the speech of male and female adults
- Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering*, 59(5), 1264–1271.
- Tylečková, L., & Skarnitzl, R. (2019). The mapping of voice parameters in connected speech of healthy Common Czech male speakers. *Akustické listy*, 25(1–2), 10–18.
- Van der Harst, S., Van de Velde, H., & Schouten, B. (2007). Acoustic characteristics of standard Dutch /ɣ/. *Proceedings of the 16th ICPHS*, 1469–1472.
- Van der Hulst, H., & Smith, N. (1982). An overview of autosegmental and metrical phonology. *Linguistic Models* 2, 1–46.
- Vaňková, J., & Skarnitzl, R. (2014). Within-and between-speaker variability of parameters expressing short-term voice quality. *Proceedings of Speech Prosody* 7, 1081–1085.
- Volín, J. (2007). *Statistické metody ve fonetickém výzkumu*. Epocha.
- Volín, J. (2017). Appeal and disrepute of the so-called global rhythm metrics. *AUC Philologica*, 2017(3), 79–94.
- Volín, J. & Bořil, T. (2014). Základní frekvence v konturách a průběžích. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvího*, 65–76. Praha: Univerzita Karlova, Filozofická fakulta.

- Volín, J., & Skarnitzl, R. (2007). Temporal downtrends in Czech read speech. *Proceedings of Interspeech 2007*, 442–445.
- Volín, J. & Skarnitzl, R. (2018). *Segmentální plán češtiny*. Univerzita Karlova, Filozofická fakulta.
- Volín, J., Tykalová, T., & Bořil, T. (2017). Stability of Prosodic Characteristics Across Age and Gender Groups. *Proceedings of Interspeech 2017*, 3902–3906.
- Volín, J., & Weingartová, L. (2012). Idiosyncrasies in local articulation rate trajectories in Czech. *Proceedings of Perspectives on Rhythm and Timing*, 67.
- Wagner, A., & Braun, A. (2003). Is voice quality language-dependent? Acoustic analyses based on speakers of three different languages. *Proceedings of the 15th ICPhS*, 651–654.
- Wei, T. & Simko, V. (2021). R package 'corrplot': Visualization of a Correlation Matrix (Version 0.92). Dostupné na <https://github.com/taiyun/corrplot>.
- Weingartová, L. (2014). Rhythm metrics for speaker identification in Czech. *AUC Philologica*, 2014(1), 33–42.
- Weingartová, L.; Bořil, T. & Vaňková, J. (2014). Spektrální sklon. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvčího*, 77–94. Univerzita Karlova, Filozofická fakulta.
- Weingartová, L. & Volín, J. (2014a). Short-term spectral slope measures and their sensitivity to speaker, vowel identity and prominence. *Akustické listy*, 20, 5–12.
- Weingartová, L. & Volín, J. (2014b). Temprální charakteristiky. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvčího*, 95–103. Univerzita Karlova, Filozofická fakulta.
- Weirich, M. (2015). Organic Sources of Inter-Speaker Variability in Articulation: Insights from Twin Studies and Male and Female Speech. In S. Fuchs, D. Pape, C. Petrone & P. Perrier (Eds.), *Individual Differences in Speech Production and Perception*, 189–222. Peter Lang Edition.
- Weirich, M., & Simpson, A. (2013). Investigating the relationship between average speaker fundamental frequency and acoustic vowel space size. *The Journal of the Acoustical Society of America*, 134(4), 2965–2974.
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35(4), 501–522.
- Whitfield, J. A., & Goberman, A. M. (2014). Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *Journal of communication disorders*, 51, 19–28.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Dostupné na <https://ggplot2.tidyverse.org/>.

Williams, D., van Leussen, J. W., & Escudero, P. (2015). Beyond North American English: modelling vowel inherent spectral change in British English and Dutch. *Proceedings of the 18th ICPHS*, paper 596.

Příloha A: Text čteného projevu

Maminka se zeptala Milana: „Milánku, už máš hotový úkol? Kdy ho budeš psát?“

Milan chvílku přemýšlel a pak odpověděl: „Já musím napsat pár souvětí na Říhovou, kde budou nějaké gramatické figle. Například vztažné věty, čárky před „a“ a podobně. Vlastně jsem tě chtěl poprosit, jestli mi s tím nepomůžeš.“

„Můžeme se na to mrknout klidně hned,“ řekla maminka, „jen bych dala vařit vodu na čaj a podívám se, jestli máme citrony. Jak dlouho nám ten úkol zabere? Bude to těžké?“

„No, mají tam být i různé příklady na zastaralou a knižní slovní zásobu. Skoro půlku jsem už ve škole udělal, ale moc dobře mi to nešlo. Chtěl bych začít co nejdřív. Až budu hotov, došel bych ti do lékárny pro ten Pneumocyt. A potom půjdu hrát fotbal, včera jsem dal čtyři góly. Nebýt Láďových faulů, mohlo jich být víc. Jen míč budu muset přifouknout. Neboj se, dám pozor na auta.“

Příloha B: Hodnoty akustických parametrů

Následující tabulka obsahuje hodnoty akustických parametrů měřených v rámci této práce. Pro analýzu byly využity nahrávky čteného (čt.) a spontánního (sp.) projevu třiceti rodilých mluvčích češtiny mužského pohlaví ve věku od 19 do 50 let. Každý ze zkoumaných řečových vzorků trval asi jednu minutu.

proměnná	parametr	mluvní styl	průměr	medián	5. percentil	10. percentil	90. percentil	95. percentil	směrodatná odchylka
základní frekvence	medián (Hz)	čt.	125,9	125,8	103,2	105,4	140,5	159,8	16,72
		sp.	114,5	113,6	95,2	95,5	133,2	142,0	15,49
	průměr (Hz)	čt.	127,7	127,8	104,5	106,9	145,7	159,5	16,53
		sp.	115,8	114,5	95,6	97,2	136,6	146,0	16,07
	základní hladina (Hz)	čt.	104,4	103,4	84,8	89,0	122,5	133,4	14,82
		sp.	103,6	103,5	85,1	86,6	118,4	128,3	13,96
	SD (Hz)	čt.	18,37	18,50	12,00	12,37	23,98	25,86	4,18
		sp.	10,64	8,93	5,63	5,86	14,38	18,34	6,07
	10.-90. percentil (Hz)	čt.	45,92	46,11	29,32	31,37	63,56	63,80	10,32
		sp.	23,18	19,87	13,06	14,31	29,64	35,10	12,96
	5.-95. percentil (Hz)	čt.	56,63	57,54	38,33	39,84	73,92	77,25	12,60
		sp.	31,88	26,23	17,39	18,31	43,30	53,77	17,31
	CSI/sl vyhlazený (ST)	čt.	2,01	1,80	1,31	1,42	2,63	3,19	0,63
		sp.	1,86	1,57	0,87	0,92	2,48	3,40	1,21
	CSI/t vyhlazený (ST)	čt.	12,25	11,36	7,99	9,23	16,64	19,03	3,99
		sp.	9,03	7,90	4,62	5,13	11,91	15,00	4,79
	CSI/sl původní (ST)	čt.	2,94	2,71	1,94	2,14	3,92	4,45	0,83
		sp.	3,22	2,66	1,38	1,55	5,23	8,24	2,08
CSI/t původní (ST)	čt.	17,84	17,46	11,92	13,03	22,09	26,88	5,11	
	sp.	15,60	13,24	8,25	8,68	27,75	35,97	8,58	

proměnná	parametr	mluvní styl	průměr	medián	5. percentil	10. percentil	90. percentil	95. percentil	směrodatná odchylka
vokálníké formanty	VSA (Hz)	čt.	107 592	106 830	47 062	58 112	145 266	169 535	40 839
		sp.	70 456	68 055	25 664	30 491	110 716	116 670	30 382
	LTF1 (Hz)	čt.	459,4	457,9	408,6	428,0	501,2	520,5	33,52
		sp.	456,6	457,0	412,0	419,0	506,5	507,6	32,71
	LTF2 (Hz)	čt.	1 444,2	1 443,8	1 315,1	1 328,1	1 520,8	1 538,8	69,92
		sp.	1 432,2	1 424,2	1 290,3	1 355,0	1 534,5	1 571,0	83,73
	LTF3 (Hz)	čt.	2 508,8	2 508,6	2 377,9	2 386,2	2 642,8	2 662,2	109,55
		sp.	2 493,6	2 471,2	2 321,2	2 362,7	2 655,1	2 668,9	117,60
	F1 [i:] (Hz)	čt.	321,7	314,0	288,5	289,8	354,0	374,4	29,51
		sp.	330,2	319,6	286,6	294,6	390,8	405,9	37,92
	F1 [i] (Hz)	čt.	389,5	389,4	334,7	346,2	426,3	434,7	29,85
		sp.	378,7	377,1	357,9	361,5	408,2	412,1	25,12
	F1 [ɛ] (Hz)	čt.	457,8	452,4	404,0	411,2	512,4	521,1	36,58
		sp.	446,3	443,7	409,8	411,9	486,9	493,8	33,01
	F1 [a] (Hz)	čt.	563,2	556,0	499,9	502,5	633,5	652,1	47,44
		sp.	532,7	527,5	470,5	478,4	594,2	597,4	45,30
	F1 [o] (Hz)	čt.	437,4	437,9	368,4	387,2	481,7	484,6	34,92
		sp.	427,6	430,8	367,1	391,0	468,0	473,1	31,47
	F1 [u] (Hz)	čt.	360,3	355,6	316,6	330,1	396,4	402,9	27,65
		sp.	364,3	359,2	318,8	329,5	403,3	427,3	33,61
	F2 [i:] (Hz)	čt.	2 009,8	2 015,2	1 776,7	1 816,1	2 195,9	2 235,9	134,03
		sp.	1 922,7	1 901,5	1 745,1	1 755,6	2 114,5	2 123,2	142,87
	F2 [i] (Hz)	čt.	1 799,2	1 791,8	1 627,7	1 639,9	1 978,9	2 033,7	131,01
		sp.	1 729,6	1 729,5	1 565,4	1 610,6	1 905,0	1 954,4	127,52
	F2 [ɛ] (Hz)	čt.	1 648,2	1 632,0	1 482,4	1 523,5	1 772,1	1 823,4	105,50
		sp.	1 580,7	1 580,9	1 392,3	1 452,8	1 677,8	1 741,0	101,19
	F2 [a] (Hz)	čt.	1 280,6	1 291,0	1 144,3	1 183,8	1 381,6	1 412,7	92,06
		sp.	1 305,0	1 296,1	1 151,6	1 186,5	1 427,1	1 451,1	94,99
	F2 [o] (Hz)	čt.	1 105,9	1 110,1	958,0	991,2	1 201,3	1 218,5	86,44
		sp.	1 152,2	1 146,1	1 025,7	1 054,0	1 250,9	1 315,5	91,28
	F2 [u] (Hz)	čt.	1 099,6	1 103,2	968,1	1 000,2	1 217,8	1 260,4	104,12
		sp.	1 178,3	1 189,2	952,6	1 018,6	1 345,9	1 354,6	146,74

proměnná	parametr	mluvní styl	průměr	medián	5. percentil	10. percentil	90. percentil	95. percentil	směrodatná odchylka	
vokálníké formanty	F3 [i:] (Hz)	čt.	2 876,3	2 873,2	2 665,9	2 715,9	3 045,2	3 152,1	164,30	
		sp.	2 812,0	2 816,1	2 520,5	2 590,0	3 017,4	3 067,8	175,87	
	F3 [i] (Hz)	čt.	2 558,4	2 559,0	2 396,4	2 414,5	2 718,7	2 752,6	135,60	
		sp.	2 543,2	2 512,7	2 334,1	2 350,7	2 682,2	2 784,4	169,77	
	F3 [ɛ] (Hz)	čt.	2 555,9	2 547,5	2 383,8	2 417,2	2 710,8	2 742,3	133,71	
		sp.	2 504,5	2 477,6	2 325,7	2 363,0	2 652,63	2 689,0	125,65	
	F3 [a] (Hz)	čt.	2 412,0	2 440,6	2 179,7	2 199,7	2 574,8	2 609,2	149,16	
		sp.	2 420,1	2 395,7	2 200,7	2 247,9	2 602,5	2 667,5	144,56	
	F3 [o] (Hz)	čt.	2 429,7	2 436,1	2 193,6	2 263,3	2 577,9	2 604,5	121,83	
		sp.	2 428,9	2 413,5	2 217,1	2 288,9	2 618,2	2 654,4	133,25	
	F3 [u] (Hz)	čt.	2 418,1	2 415,3	2 233,6	2 251,0	2 586,7	2 599,7	122,29	
		sp.	2 436,1	2 410,5	2 310,5	2 319,4	2 580,5	2 598,5	101,38	
	frekvenční spektrum vokálů	H1*-H2* (dB)	čt.	3,40	3,25	0,44	0,73	5,84	7,18	2,48
			sp.	2,97	3,00	-1,34	-0,80	7,41	7,81	3,13
H2*-H4* (dB)		čt.	5,20	5,18	2,19	2,78	7,56	7,85	1,84	
		sp.	5,90	5,04	3,04	3,76	9,59	10,42	2,42	
H1*-A1* (dB)		čt.	16,85	16,54	12,61	13,29	20,36	21,36	2,89	
		sp.	17,02	16,94	11,91	12,27	21,34	22,08	3,45	
H1*-A2* (dB)		čt.	23,17	24,26	15,30	17,65	27,46	28,42	4,47	
		sp.	23,49	23,72	17,04	18,61	29,00	29,37	4,14	
H1*-A3* (dB)		čt.	17,41	17,52	9,98	12,61	23,09	23,18	4,66	
		sp.	17,42	16,90	10,02	11,28	22,87	23,82	4,25	
H4*-H2kHz* (dB)	čt.	6,41	5,83	2,42	3,44	9,76	10,88	2,95		
	sp.	6,23	5,84	2,36	2,72	8,87	10,38	2,83		
H2kHz*-H5kHz* (dB)	čt.	22,48	22,87	17,06	17,96	25,06	27,68	3,22		
	sp.	22,12	21,98	17,66	19,51	25,02	25,78	2,65		
frekvenční spektrum konsonantů	spektrální těžiště [s] (Hz)	čt.	6 673,8	6 717,6	4 532,7	4 792,2	7 979,0	8 088,9	1 131,7	
		sp.	6 526,7	6 743,5	4 482,2	5 282,7	7 468,5	7 756,8	1 020,6	
	SD [s] (Hz)	čt.	4 007,6	4 087,2	3 470,3	3 674,7	4 350,0	4 452,6	334,4	
		sp.	4 140,6	4 204,0	3 584,4	3 806,9	4 409,2	4 497,7	262,4	
	sešikmení [s] (Hz)	čt.	0,08	0,04	-0,30	-0,26	0,57	0,75	0,36	
		sp.	0,15	0,01	-0,14	-0,13	0,53	0,77	0,32	

proměnná	parametr	mluvní styl	průměr	medián	5. percentil	10. percentil	90. percentil	95. percentil	směrodatná odchylka
frekvenční spektrum konsonantů	špičatost [s] (Hz)	čt.	-0,58	-0,63	-0,99	-0,90	-0,10	0,08	0,34
		sp.	-0,67	-0,73	-1,03	-0,94	-0,35	-0,13	0,31
	spektrální těžiště [m] (Hz)	čt.	935,1	892,5	700,5	745,5	1 234,1	1 335,3	187,5
		sp.	1 031,7	1 023,0	782,2	812,0	1 383,6	1 416,5	193,9
	SD [m] (Hz)	čt.	2 105,0	2 033,4	1 767,1	1 813,9	2 451,7	2 667,0	285,9
		sp.	2 339,4	2 251,4	2 014,7	2 028,1	2 818,7	2 880,5	289,2
	spektrální těžiště [n] (Hz)	čt.	1 052,5	1 017,9	773,8	806,0	1 375,2	1 461,8	234,4
		sp.	1 124,2	1 089,9	818,3	840,9	1 405,6	1 548,5	216,6
	SD [n] (Hz)	čt.	2 245,8	2 187,4	1 834,3	1 911,8	2 663,8	2 835,6	303,4
		sp.	2 418,9	2 338,7	2 050,0	2 072,9	2 883,9	2 936,9	301,2
	spektrální těžiště [l] (Hz)	čt.	1 093,0	1 091,2	804,4	850,3	1 333,5	1 375,3	191,6
		sp.	1 091,1	1 111,7	824,5	886,6	1 310,7	1 362,3	183,2
	SD [l] (Hz)	čt.	2 080,5	2 088,8	1 725,4	1 776,9	2 363,0	2 434,1	238,9
		sp.	2 174,5	2 170,7	1 736,6	1 847,3	2 516,4	2 625,9	273,1
dlouhodobé průměrné spektrum	Bg	čt.	0,0039	0,0020	0,0005	0,0007	0,0090	0,0107	0,0039
		sp.	0,0037	0,0020	0,0006	0,0010	0,0081	0,0096	0,0037
	BgNoFo	čt.	0,0072	0,0045	0,0010	0,0010	0,0163	0,0201	0,0072
		sp.	0,0075	0,0050	0,0015	0,0020	0,0149	0,0230	0,0075
	index α	čt.	0,0320	0,0200	0,0048	0,0070	0,0717	0,0830	0,0320
		sp.	0,0285	0,0165	0,0035	0,0067	0,0628	0,0946	0,0285
Hammarberg index (dB)	čt.	28,46	27,88	21,00	21,31	34,87	35,91	4,77	
	sp.	28,33	28,62	20,31	23,34	33,87	34,27	4,51	
temporální rovina řeči	AT	čt.	6,17	6,31	4,93	5,37	6,76	6,85	0,61
		sp.	5,89	5,98	4,68	4,75	6,65	6,79	0,70
	SD AT	čt.	0,93	0,90	0,75	0,77	1,14	1,19	0,14
		sp.	1,58	1,56	1,08	1,11	2,09	2,20	0,36
	%V	čt.	38,99	38,68	35,97	36,62	41,74	41,88	2,08
		sp.	42,40	42,85	36,79	38,06	45,88	46,79	3,36
	%VO	čt.	70,93	70,87	68,60	68,79	72,85	72,92	1,44
		sp.	69,04	69,35	63,96	65,73	72,24	72,64	2,79
ΔV	čt.	0,046	0,045	0,041	0,041	0,052	0,054	0,046	
	sp.	0,057	0,058	0,047	0,048	0,065	0,068	0,057	

proměnná	parametr	mluvní styl	průměr	medián	5. percentil	10. percentil	90. percentil	95. percentil	směrodatná odchylka
temporální rovina řeči	VarcoV	čt.	0,561	0,564	0,519	0,529	0,599	0,610	0,561
		sp.	0,614	0,605	0,549	0,572	0,675	0,691	0,614
	ΔC	čt.	0,048	0,048	0,041	0,041	0,057	0,058	0,048
		sp.	0,054	0,053	0,043	0,046	0,062	0,062	0,054
	VarcoC	čt.	0,500	0,492	0,454	0,458	0,543	0,551	0,500
		sp.	0,507	0,507	0,454	0,459	0,548	0,572	0,507