

Univerzita Karlova

Filozofická fakulta

Fonetický ústav

Diplomová práce

Bc. Alžběta Růžičková

Sofistikované strategie maskování hlasu a jejich fonetická podoba

Sophisticated strategies of voice disguise and their phonetic character

Praha 2018

Vedoucí práce: Ing. Tomáš Bořil, Ph.D.

Děkuji vedoucímu práce Ing. Tomáši Bořilovi, Ph.D. za odborné vedení, cenné rady a vstřícnost. Mé poděkování patří také Michailu Monstrovovi, který se podílel na manuálním zpracování řečového materiálu využitého v této práci.

Prohlašuji, že jsem diplomovou 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 9. srpna
2018

.....
Jméno a příjmení

ABSTRAKT

V řečovém projevu lze nalézt některé prvky, které jsou charakteristické pro daného mluvčího, tzv. idiosynkratické rysy. Tato práce se zaměřuje na podobu těchto rysů při záměrném maskování hlasu – zda jsou mluvčí schopni je zásadním způsobem měnit, či zda spíše zůstávají stabilní i navzdory záměrným modifikacím řečového projevu. Bylo také sledováno, zda u zkoumaných rysů při maskování hlasu existují obecné tendence k podobným změnám napříč mluvčími. Analyzovány byly statistické ukazatele základní frekvence, kontury f_0 , vokální formanty, dlouhodobé formantové distribuce, spektrální vlastnosti sibilant, intenzita řečového projevu, intenzitní kontury, mluvní a artikulační tempo, %V a kontury lokálního artikulačního tempa. U mediánu a směrodatné odchylky f_0 , vokálních formantů, dlouhodobých formantových distribucí, intenzity, artikulačního tempa i %V byly při maskování hlasu obecně zaznamenány znatelné posuny; u většiny z těchto parametrů se tyto posuny lišily mezi mluvčími. Bylo však zjištěno, že hodnota %V má obecnou tendenci při maskování hlasu stoupat. Také intenzita ve většině případů maskování hlasu vykazovala nárůst. U kontur f_0 byly pozorovány podobné vzorce napříč mluvčími při nemaskovaném projevu; při projevu maskovaném se však v užití různých kontur f_0 napříč jednotlivci objevily větší rozdíly, mluvčí mají tedy při maskování hlasu zřejmě častěji tendenci se uchýlovat k nestandardním vzorcům průběhů základní frekvence. Podobné výsledky byly zaznamenány také u kontur temporálních. Užití intenzitních kontur se mezi maskovanými a nemaskovanými projevy příliš nelišilo, byly však pozorovány znatelné odlišnosti u některých mluvčích, které zůstávaly relativně stabilní i při maskování hlasu.

Klíčová slova:

Forenzní fonetika, maskování hlasu, idiosynkratické rysy

ABSTRACT

Speech contains certain attributes characteristic for a speaker, so-called idiosyncratic features. This thesis focuses on the form of these features in intentional voice disguise – whether speakers are able to change them in a substantial way, or if they tend to remain stable in spite of intentional speech modifications. It was also investigated whether any general tendencies to similar changes of such features under voice disguise exist among the speakers. The observed features were statistical f0 indicators, f0 contours, vowel formants, long-term formant distributions, spectral characteristics of sibilants, intensity, intensity contours, speech and articulation rate, %V and local articulation rate contours. In f0 median and standard deviation, vowel formants, LTFDs, intensity, articulation rate, and %V, prominent shifts under voice disguise were observed in general; in the majority of these parameters, the shifts differed among speakers. However it was found that the value of %V generally tends to rise under voice disguise. Also, intensity showed an increase in majority of cases. In f0 contours, similar patterns were observed among speakers in normal speech, however, in disguised speech, greater differences appeared among speakers; speakers tend to employ nonstandard dynamic f0 patterns more often while attempting voice disguise. Similar results were also obtained for temporal contours. Intensity contours usage did not differ much between normal and disguised speech, however, significant differences were observed among some speakers that remained relatively stable even under voice disguise.

Key words:

Forensic phonetics, voice disguise, idiosyncratic features

OBSAH

| | |
|-----------------------------------------------------------------|----|
| 1. Úvod..... | 8 |
| 1.1. Forenzní fonetika | 8 |
| 1.2. Profilování mluvčího..... | 10 |
| 1.3. Srovnávání mluvčích..... | 12 |
| 1.4. Idiosynkratické rysy | 13 |
| 1.4.1. Základní frekvence..... | 13 |
| 1.4.2. Vokální formanty..... | 15 |
| 1.4.3. Spektrální charakteristiky..... | 16 |
| 1.4.4. Intenzita..... | 18 |
| 1.4.5. Globální temporální ukazatele | 19 |
| 1.4.6. Lokální temporální ukazatele..... | 21 |
| 1.4.7. Disfluence | 21 |
| 1.5. Proměnlivost hlasu v závislosti na momentální situaci..... | 22 |
| 1.6. Maskování hlasu | 24 |
| 1.6.1. Změny základní frekvence | 25 |
| 1.6.2. Imitace dialektu..... | 28 |
| 1.6.3. Imitace cizineckého přízvuku | 28 |
| 1.6.4. Umístění cizího předmětu do úst..... | 29 |
| 1.7. Cíle práce | 30 |
| 2. Metoda | 31 |
| 2.1. Materiál..... | 31 |
| 2.2. Poslechová analýza | 32 |
| 2.3. Akustická analýza | 33 |
| 2.3.1. Základní frekvence..... | 34 |
| 2.3.2. Vokální formanty..... | 35 |
| 2.3.3. Spektrální těžiště | 36 |
| 2.3.4. Intenzita..... | 36 |
| 2.3.5. Globální temporální charakteristiky..... | 37 |
| 2.3.6. Lokální temporální ukazatele..... | 38 |
| 3. Výsledky | 39 |
| 3.1. Poslechová analýza | 39 |
| 3.1.1. Nemaskovaný projev..... | 39 |
| 3.1.2. Maskovaný projev | 39 |
| 3.2. Akustická analýza | 40 |
| 3.2.1. Základní frekvence..... | 40 |
| 3.2.2. Vokální formanty..... | 51 |
| 3.2.3. Spektrální těžiště | 62 |
| 3.2.4. Intenzita..... | 64 |

| | | |
|--------|------------------------------------------|----|
| 3.2.5. | Globální temporální charakteristiky..... | 71 |
| 3.2.6. | Lokální temporální ukazatele..... | 75 |
| 4. | Diskuze..... | 77 |
| 4.1. | Statická analýza..... | 78 |
| 4.2. | Dynamická analýza..... | 80 |
| 5. | Závěr..... | 82 |
| | Seznam referencí..... | 84 |
| | Příloha č. 1..... | 90 |
| | Příloha č. 2..... | 91 |
| | Příloha č. 3..... | 92 |

1. ÚVOD

1.1. FORENZNÍ FONETIKA

Mluvený projev každého jedince je specifický. Stejně jako lze vypožorovat charakteristické vzorce v chůzi, psaní a jiných automatizovaných aktivitách, existují takové charakteristické znaky také v řeči. Tyto jevy podávají informace o daném mluvčím a mohou odlišit jeho řečový projev od ostatních jedinců. Díky této skutečnosti je možné provádět fonetickou identifikaci mluvčího ve forenzně fonetické praxi.

Podoba mluveného projevu může o daném jedinci poskytnout tři kategorie informací: (1) informace fyziologické, tj. jaká je anatomie vokálního traktu mluvčího, případně jak fungují jeho orgány činné při řeči; tyto charakteristiky ovlivňuje např. věk, zdravotní stav či užívání omamných látek; (2) informace sociálního charakteru; jedná se o jevy osvojené mluvčím, které vyplývají z jeho příslušnosti k určité sociální skupině. Mezi tyto informace patří např. regionální původ či socioekonomický status; (3) krátkodobé jevy – do této kategorie spadají rysy krátkodobého charakteru, jakými jsou např. afektivní stavy či aktuální komunikační nastavení, tedy mluvní styl, případně akomodace mluvčího, tedy přizpůsobení se momentálnímu komunikačnímu partnerovi (Giles et al., 1991; Nolan 1999; Skarnitzl, 2014).

K zachycování charakteristik řečového projevu mluvčího existují dva přístupy: přístup poslechový a přístup akustický (Nolan, 1999).

Pomocí poslechového přístupu lze lépe zachytit některé řečové jevy, např. dialekt či řečové vady, ale také posoudit celkový dojem, jakým daný hlas působí. Poslechem lze také často odhalit maskování hlasu. Nevýhodou tohoto přístupu je však jeho subjektivita a obtížná vyčíslitelnost jednotlivých parametrů, kvůli čemuž mohou být výsledky takovéto analýzy shledány jako nepříliš spolehlivé. Některé detaily navíc tímto způsobem nemusejí být vůbec zachytitelné, např. drobné výkyvy hlasitosti či melodie hlasu. Tato skutečnost je způsobena jednak omezeními lidského sluchu, ale především podobou lidské percepce, která je kategorická – pro co nejvyšší efektivitu řečové komunikace jsou při percepci jednotlivé řečové jevy zařazovány do kategorií, což umožňuje rychlé

zpracovávání řečové informace, zároveň ale nejsou zaznamenány některé drobné, pro účely komunikace nerelevantní nuance.

Naproti tomu akustický přístup je objektivní a výsledky akustické analýzy jsou kvantifikovatelné. Pomocí akustické analýzy lze zachytit i takové jevy, které nejsou zaznamenatelné poslechem. Různé akustické jevy také mohou udávat podobu jednoho komplexního percepčního korelátu, např. na vnímané výšce hlasu se podílí nejen základní frekvence, která je dána fyziologií hlasivek, ale také vokálníké formanty, jež udává tvar a velikost vokálního traktu (Skarnitzl & Hývlová, 2014). Působení těchto dvou ukazatelů bývá obtížné oddělit sluchem, je však možné je pozorovat každý zvlášť při akustické analýze řečového signálu.

Vzhledem ke specifčnosti řečového projevu pro daného jedince vznikla myšlenka, že by bylo možné nalézt jakousi hlasovou alternativu otisku prstu, pomocí které by bylo možné jednoznačně identifikovat jedince ve forenzní praxi, podobně jako je tomu u daktyloskopie. Tento koncept, tzv. voiceprint neboli otisk hlasu, byl poprvé uveden ve 40. letech 20. století. Při tomto postupu bylo na základě srovnávání vizuální podoby spektrogramů jednotlivých slov pronášených pachatelem a podezřelým usuzováno, zda se jedná o téhož jedince. Tento přístup se však ukázal jako nevhodný (Hollien, 2002, pp. 133).

V současné době se ve forenzní praxi využívá kombinace jak akustického, tak i poslechového přístupu. Při akustické analýze jsou sledovány jednotlivé dílčí parametry, které mohou být pro mluvčího charakteristické – tzv. idiosynkratické rysy. Na základě jejich kombinace je pak prováděna identifikace mluvčího.

Idiosynkratický rys činí ideálním pro identifikaci mluvčího následující vlastnosti: (1) velká variabilita mezi mluvčími, tj. možnost na základě rysu jednoznačně odlišit dva jedince, a zároveň (2) stabilita v rámci projevu jednoho mluvčího; (3) častý výskyt v řeči – ve forenzních podmínkách bývá často k dispozici jen malé množství materiálu, a je tudíž výhodou, pokud lze předpokládat, že se daný rys vyskytne i v krátké nahrávce, a to pokud možno nejen okrajově; (4) snadná měřitelnost, (5) odolnost vůči zkreslení vlivem podmínek při nahrávání a (6) odolnost vůči maskování hlasu, tedy aby mluvčí nebyl schopen daný parametr vědomě ovlivnit (Nolan, 1999).

Ideální idiosynkratický rys, který by splňoval všechna uvedená kritéria, však dosud nebyl objeven. Při akustické analýze projevu mluvčího je tudíž třeba

sledovat větší množství parametrů, v různé míře tato kritéria splňující, a teprve na základě jejich kombinace provádět identifikaci mluvčího.

Ve forenzních podmínkách lze analýzu idiosynkratických rysů jedince uplatňovat tehdy, pokud je k dispozici nahrávka hlasu neznámého mluvčího, jejíž je třeba charakterizovat. V takových případech mohou nastat dvě situace: (1) je k dispozici pouze nahrávka neznámého mluvčího (obvykle pachatele), tj. sporná nahrávka, nebo (2) existuje také nahrávka projevu známého mluvčího (obvykle podezřelého), tj. nahrávka srovnávací.

Pokud je k dispozici srovnávací nahrávka, je možné provést její srovnání s nahrávkou spornou a s určitou měrou jistoty potvrdit či vyloučit totožnost mluvčího, tj. zda na obou nahrávkách promlouvá tentýž člověk, nebo nikoliv. Pokud srovnávací nahrávka k dispozici není, lze pomocí analýzy nahrávky sporné provést profilování mluvčího, tj. co nejpřesněji jej charakterizovat, aby bylo možné zúžit skupinu potenciálních podezřelých.

1.2. PROFILOVÁNÍ MLUVČÍHO

Je-li k dispozici nahrávka hlasu pachatele, ale neexistuje podezřelý, je potřeba na základě sporné nahrávky pachatele co nejpřesněji popsat, a tudíž co nejlépe identifikovat cílovou populaci, v níž by se pachatel mohl nacházet. Musí být tedy charakterizovány parametry sociolingvistické povahy, jakými jsou pohlaví, věk, regionální původ či socioekonomický status, ale také neobvyklé osobní charakteristiky, jako např. řečové vady nebo výskyt neobvyklých lexikálních jednotek (Skarnitzl, 2014: 15). Jessen (2007a) označuje za charakteristiky, které je možné nejsnáze identifikovat u neznámého mluvčího, pohlaví, věk, sociolekt, cizinecký přízvuk, regionální dialekt mateřského jazyka a zdravotní stav.

Identifikace pohlaví mluvčího nebývá ve většině případů problematická vzhledem k odlišné morfologii vokálního traktu mužů a žen, které se odrážejí zejména v základní frekvenci a vokálních formantech.

V prepubertálním věku existuje lineární vztah mezi fyzickým vzrůstem jedince a délkou vokálního traktu, který určuje hodnoty vokálních formantů, a to nezávisle na pohlaví jedince. U žen tento vzorec růstu vokálního traktu pokračuje také během puberty až do dospělosti. Naproti tomu u mužů během puberty nastává neproporční prodlužování vokálního traktu způsobené zejména

sestoupením hrtanu. Tato skutečnost vede k signifikantním rozdílům jak v celkové délce, tak i ve tvaru vokálního traktu mezi muži a ženami. Akustickým dopadem těchto rozdílů jsou nižší frekvence vokálních formantů u mužů, jejichž vokální trakt je delší, než u žen (Fitch & Giedd, 1999).

Nejvíce prominentním korelátem pohlaví je ovšem průměrná základní frekvence, která bývá u žen výrazně vyšší než u mužů, což je zapříčiněno rozdílem v délce hlasivek (Jessen, 2007a); „blaná část glotis je u žen přibližně o 60 % kratší“ (Titze, 1989; in: Skarnitzl, 2014).

Jak již bylo zmíněno, frekvence vokálních formantů přímo korelují s rozměry vokálního traktu, který do jisté míry souvisí s fyzickým vzrůstem jedince. Z této skutečnosti by tedy bylo možné odvodit, že z hodnot vokálních formantů lze získat informace o tělesné velikosti mluvčího. Gonzáles (2004) však upozorňuje, že existuje pouze slabý vztah mezi frekvencemi vokálních formantů (a dalších souvisejících hodnot, jako např. šířkou formantových pásem) a výškou či hmotností mluvčího.

Určení věku je ve srovnání s identifikací pohlaví mluvčího obtížnější, neboť neexistuje konkrétní fonetický parametr, jehož pomocí by bylo možné věk mluvčího stanovit. Klasifikace věku ve forenzně fonetické praxi bývá většinou prováděna na základě poslechové analýzy a vyvozována z obecného dojmu, jakým řečový projev mluvčího působí. Odhady věku nebývají příliš přesné, z toho důvodu se předpokládaný věk mluvčího uvádí v intervalech 10 až 20 let (Jessen, 2007a).

Podle klasifikovaného dialektu je určován region, v němž daný mluvčí strávil většinu času v dětství a během dospívání, na základě předpokladu, že v dospělosti se dialekt mluvčího příliš nemění, a to ani v případě, že dlouhodobě žije v jiném regionu (Jessen, 2007a).

Odhalit prvky cizineckého přízvuku v řečovém projevu mluvčího, resp. odlišit je od prvků dialektálních, obvykle nebývá problematické, nicméně rozpoznat mateřský jazyk mluvčího může být obtížné; často je možné formulovat pouze obecnou charakteristiku cizineckého přízvuku, např. jazykovou rodinu (Jessen, 2007a).

1.3. SROVNÁVÁNÍ MLUVČÍCH

Pokud je k dispozici jak sporná nahrávka, tak i nahrávka srovnávací, tj. vzorek řečového projevu podezřelého, je možné provést srovnávání mluvčích a s určitou měrou jistoty stanovit, zda se jedná o tutéž osobu, či o dva různé jedince.

Porovnávání takových dvou nahrávek bývá komplikováno odlišnými podmínkami při jejich pořizování.

Kvalita nahrávky hlasu pachatele, která je často odposlechem nebo záznamem telefonního hovoru, může být degradována, a signál tudíž zkreslen; např. při telefonním přenosu je přenášeno pásmo o rozsahu cca 300-3400 Hz v případě pevné telefonní sítě a cca 100-3600 Hz v případě sítě mobilní, jejíž frekvenční odezva je však proměnlivá. Pokud se tedy jedná o záznam telefonního hovoru, dochází ke ztrátě značné části informací; tímto způsobem může být zapříčiněno např. zkreslení základní frekvence (Vaňková & Bořil, 2014) či znesnadnění extrakce vokalických formantů, které mohou být vlivem telefonního (konkrétně mobilního) přenosu zvýšeny či sníženy, přičemž tyto efekty se liší napříč jednotlivými vokály i různými mluvčími (Byrne & Foulkes, 2004, Kaiser & Bořil, 2018), a vyšší formanty (F4 a výše) mohou v signálu zcela chybět (Skarnitzl et al., 2014). Problematičnost sporné nahrávky také často spočívá v jejím krátkém trvání.

Srovnávací nahrávka bývá většinou získávána při výslechu, a to nikoli ve studiových podmínkách. Kvalita sporné a srovnávací nahrávky může být tedy značně odlišná, což znesnadňuje jejich porovnávání.

Při srovnávání mluvčích je třeba zaměřit se na jejich podobnost a typičnost. Tj. nakolik jsou si řečové projevy zkoumaných mluvčích podobné, a pokud je u nich nápadná podobnost nalezena, nakolik jsou jejich projevy typické pro danou populaci. Pakliže jsou si dva mluvčí podobní, ale jejich hlasy lze považovat za typické, tedy nijak nápadně výjimečné v rámci populace, není možné takovou podobnost považovat za stěžejní důkaz, že se jedná o téhož mluvčího. Pokud se ale hlasy zachycené na obou nahrávkách zřetelně podobají a navíc se výrazně odchyľují od populačního standardu, je mnohem pravděpodobnější, že se skutečně jedná o nahrávky řečových projevů jediného mluvčího. Typičnost projevů mluvčích je stanovována na základě populačních statistik, pokud jsou tyto k dispozici. Populační statistiky hodnot základní frekvence existují např. pro

angličtinu, němčinu, švédštinu či polštinu (Skarnitzl, 2014). V českém prostředí byla vytvořena databáze obecné češtiny pro forenzní účely, která obsahuje nahrávky několika typů projevů sta rodilých mluvčích obecné češtiny, a na jejím základě Skarnitzl & Vaňková (2017) představili také populační statistiky f_0 pro české mluvčí. Zmíněná databáze byla také využita v praktické části této práce.

1.4. IDIOSYNKRATICKÉ RYSY

V řečovém signálu lze nalézt několik idiosynkratických rysů, které mohou napomoci identifikaci mluvčího v různé míře. Jak již bylo uvedeno, ideální idiosynkratický rys, který by k identifikaci postačil jako jediný, nebyl nalezen, ve forenzně fonetické praxi se tudíž využívá kombinace analýz více jevů; takovýto komplexní přístup může poskytnout poměrně spolehlivou oporu pro určení identity jedince.

1.4.1. ZÁKLADNÍ FREKVENCE

Základní frekvence, která je akustickým korelátem rychlosti kmitání hlasivek, je ve forenzní praxi často využívaným parametrem. Její značnou výhodou je dostupnost i v krátkých řečových vzorcích, vzhledem k tomu, že je přítomna ve všech (foneticky) znělých segmentech. Mimoto je poměrně robustní vůči kvalitativnímu zkreslení a její extrakce ze signálu není problematická. Nadto je poměrně variabilní napříč mluvčími. Nevýhodou tohoto parametru je však poměrně značná variabilita intraindividuální, tj. v rámci řečového projevu jednoho mluvčího (Skarnitzl & Hývlová, 2014: 51).

Průběh základní frekvence je možné sledovat několika způsoby. Nejjednodušším z nich je zachycení hodnot f_0 pomocí statistických ukazatelů, které popisují charakteristické znaky celkového průběhu základní frekvence v signálu pomocí konkrétních numerických hodnot. Takovými ukazateli jsou vyjádření střední hlasové frekvence, tj. jakési jedné frekvenční hladiny, která je pro řečový projev mluvčího charakteristická, a variabilita f_0 v průběhu promluvy. Střední hodnotu f_0 lze vyjádřit pomocí aritmetického průměru, mediánu či modu; proměnlivost f_0

během řečového projevu lze zachytit pomocí směrodatné odchylky, variačního koeficientu či percentilového rozpětí.

Pokud jde o zachycení střední hodnoty základní frekvence, doporučuje Lindh (2006) vyčíslení mediánu jakožto spolehlivějšího parametru, spíše než aritmetického průměru f_0 , neboť medián není náchylný ke zkreslení vlivem extrémních hodnot, které se při extrakci f_0 mohou objevit v podobě tzv. oktávových skoků, které jsou následkem chybné detekce f_0 využívaným softwarem.

Jinou možností zachycení jakési neutrální hodnoty f_0 charakteristické pro daného mluvčího je tzv. základní hladina. Jedná se o určitou výchozí polohu základní frekvence, ze které vycházejí jednotlivé melodické úkony, a ke které se mluvčí vždy opět vrací. Hodnota základní hladiny by měla odpovídat běžné (tedy nikoli třepené) intonaci na konci výpovědi (Skarnitzl & Vaňková, 2017), měla by tedy pro daného mluvčího zůstat stabilní, a to i navzdory změnám f_0 vyplývajícím např. z emočního zabarvení promluvy (Skarnitzl & Hývlová, 2014). Tento ukazatel tedy postihuje neutrální nastavení vokálního traktu jedince lépe než střední hodnota f_0 , a je tudíž považován za robustnější faktor při identifikaci mluvčího (Lindh & Eriksson, 2007).

Jiným způsobem analýzy základní frekvence je zkoumání jejího dynamického průběhu v řeči, tj. změn v čase. Na rozdíl od statistických ukazatelů f_0 lze tímto způsobem postihnout tvary kontur f_0 , jež mluvčí ve svém řečovém projevu využívá. Ukazuje se, že k melodické realizaci promluv s týmž lexikálním obsahem mluvčí přistupují různě, a kontury, které využívají v promluvách s podobným obsahem, jsou u jednotlivých mluvčích nenáhodné. Analýza průběhu f_0 v čase by tudíž mohla napomoci identifikaci mluvčího (Volín & Bořil, 2014). Využití tohoto postupu ve forenzní praxi však může být problematické, vyžaduje totiž nalezení takových promluv, které by byly obsahově podobné, a tudíž porovnatelné. To může v reálných podmínkách představovat problém, neboť takový postup je časově náročný, a nadto se vhodné srovnatelné promluvy v analyzovaném řečovém materiálu nemusí vůbec vyskytnout.

1.4.2. VOKALICKÉ FORMANTY

Za velmi přínosný ukazatel pro účely identifikace mluvího bývají považovány vokalické formanty, neboť odrážejí fyziologii vokálního traktu daného mluvího, ale i určitá habituální artikulační nastavení. Ačkoliv nejsou hodnoty vokalických formantů invariantní v rámci mluvího, jeví se jako méně náchylné k proměnám zapříčiněným kontextem než parametry odvíjející se od činnosti hlasivek (např. f_0 nebo spektrální sklon), a jsou tudíž shledávány přínosným ukazatelem při identifikaci mluvího (Nolan & Grigoras, 2005). Zatímco hodnoty F1, F2 a částečně také F3 jsou akustickými koreláty kvality vokálu, tj. závisí na momentálním artikulačním nastavení mluvího, vyšší formanty (F4 a F5) jsou spíše závislé na fyziologii mluvího a během promluvy zůstávají více méně stabilní bez ohledu na artikulaci (Reetz & Jongman, 2009 in: Skarnitzl et al., 2014). Díky tomu se F4 a F5 jeví jako užitečný parametr pro identifikaci mluvího, překážku v jejich užití však může představovat jejich zhoršená čitelnost, případně nepřítomnost v řečovém materiálu, který je k dispozici v reálných forenzních podmínkách.

Podobně jako je tomu u základní frekvence, i u vokalických formantů lze extrahovat buď jednu konkrétní hodnotu v měřeném úseku (tj. v případě formantů v daném vokálu), nebo sledovat jejich dynamický průběh během artikulace samohlásek.

Při zkoumání statických hodnot vokalických formantů dané samohlásky se zpravidla provádí analýza formantových frekvencí v artikulačním cíli vokálu, tedy v místě, kde formanty dosáhly stabilního průběhu, anebo v bodě nacházejícím se uprostřed samohlásky. Takto získané hodnoty lze mezi mluvími porovnávat pro jednotlivé vokalické kvality. Druhou možností je analýza dlouhodobých formantových distribucí (*long-term formant distribution, LTF*), při které jsou porovnávány jednotlivé formanty bez ohledu na to, ze kterého vokálu byly extrahovány. Tento ukazatel odráží celkovou fyziologii vokálního traktu mluvího a také jeho obecné artikulační návyky, např. labializaci či horizontální i vertikální posuny artikulace vokálů (Gold et al., 2013). Podle Skarnitzla et al. (2014) se jako nejlépe odlišující různé mluví a zároveň nejrobustnější vůči zkreslení jeví LTF3.

Pro zachycení pohybu formantů v čase je možné extrahovat několik hodnot v průběhu vokálu a na jejich základě vytvořit tzv. formantové trajektorie. Ty na

rozdíl od hodnot vokalických formantů extrahovaných pouze v jednom bodě v rámci jednotlivých vokálů poskytují informace o dynamickém průběhu formantů během artikulace samohlásek, a tedy o pohybech artikulačních orgánů mluvího během jejich realizace. Podle McDougall (2006) jsou tyto pohyby charakteristické pro daného jedince, podobně jako jiné automatizované motorické aktivity, např. chůze. Zatímco jednotlivé artikulační cíle (např. v rámci vokálu) jsou v rámci jazyka více méně pevně určeny, způsob, jakým se od jednoho artikulačního cíle artikulační orgány mluvího přenastaví na cíl další, poskytuje prostor pro variabilitu, a právě individuální artikulační návyky mluvích, kteří se v tomto ohledu od sebe navzájem mohou značně lišit, mohou poskytovat idiosynkratické informace využitelné při identifikaci mluvího (Skarnitzl et al., 2014).

1.4.3. SPEKTRÁLNÍ CHARAKTERISTIKY

Co se týče analýzy frekvenční domény řeči, pro identifikaci mluvího lze využít také tzv. spektrální sklon, tedy postupný pokles energie ve vyšších frekvencích. Ten je určen řadou faktorů, především však fonačním nastavením. Obecně lze říci, že celkový spektrální sklon řečového projevu mluvího je akustickým korelátem percepčního dojmu barvy hlasu mluvího. Hlasy s nízkým spektrálním sklonem, tj. s menším poklesem energie ve vyšších frekvencích, znějí ostřeji, zatímco hlasy, u nichž je možné pozorovat strmější spektrální sklon, znějí měkčeji. Spektrální sklon bývá obvykle vyrovnáván v důsledku vyššího mluvního úsilí, např. při tlačené fonaci, a naopak např. při fonaci dyšné je spektrální sklon řečového signálu strmější.

Při identifikaci mluvího je možné využít dlouhodobé ukazatele spektrálního sklonu, které vycházejí z dlouhodobého průměrného spektra (long-term average spectrum; LTAS); to zachycuje rozložení energie v jednotlivých frekvencích během delšího úseku mluveného projevu, tj. bez ohledu na identitu jednotlivých řečových segmentů. Existuje několik ukazatelů, jež lze na základě LTAS vyčíslit, a které by mohly být využívány při identifikaci mluvího; obvykle se jedná o porovnávání spektrálních vrcholů či pozorování rozdílů množství energie v určitých frekvenčních pásmech. Podle dosavadních výzkumů lze na základě

LTAS poměrně spolehlivě rozlišit pohlaví mluvčího či rozdíly ve věku mluvčích (Weingartová et al., 2014).

Mimo dlouhodobých ukazatelů spektrálního sklonu lze využít také jeho ukazatele krátkodobé, tj. získané na základě jednotlivých hlásek. Spektrální charakteristiky některých řečových segmentů se ukazují jako přínosnější pro identifikaci mluvčího, než je tomu v případě segmentů jiných; např. realizace fonému /s/ se u mluvčích projevuje jako poměrně stabilní v rámci projevu jednoho mluvčího a má také potenciál od sebe navzájem odlišovat různé mluvčí (French, 1994).

Za přínosný parametr pro identifikaci mluvčího je považován také spektrální sklon vokálů. Při identifikaci mluvčího se obvykle provádí porovnávání jednotlivých harmonických složek; nejčastěji se jedná o první harmonickou složku a nějakou další, především druhou harmonickou (H1-H2), nejsilnější harmonickou složku v oblasti prvního formantu (H1-A1) či třetího formantu (H1-A3) (Weingartová et al., 2014). Kitamura & Akagi (2007) považují za specifické pro daného mluvčího především rozložení energie ve vyšších frekvencích, cca mezi 1700 a 2500 Hz.

Spektrální sklon může být ovšem náchylný ke zkreslení způsobenému akustickými podmínkami při nahrávání; impulzní odezva prostoru, v němž byla nahrávka pořízena, může výrazně ovlivnit tvar výsledného spektra analyzovaného signálu, a zkreslit spektrální sklon může také šum (Weingartová et al., 2014). Z tohoto důvodu je vhodné porovnávat pouze takové nahrávky, jejichž kvalita je srovnatelná, a pokud nejsou k dispozici, pak jednu z nich dodatečně kvalitativně degradovat tak, aby byla druhé nahrávce co nejpodobnější (French, 1994).

Spektrum jednotlivých segmentů lze kromě spektrálního sklonu analyzovat ještě pomocí dalších faktorů, jakými jsou spektrální těžiště (*centre of gravity*) a jeho směrodatná odchylka (*standard deviation; SD*), sešikmení (*skewness*) a špičatost (*kurtosis*) spektra. Spektrální těžiště charakterizuje rozložení energie ve spektru, jedná se o hodnotu frekvence, resp. frekvenčního pásma, v jehož oblasti se nachází pomyslné centrum této energie. Směrodatná odchylka pak udává šířku tohoto frekvenčního pásma.

Kavanagh (2012) sledovala spektrální charakteristiky hlásek /m/, /n/, /ŋ/, /l/ a /s/ v angličtině; v případě sonor využila spektrální těžiště, SD, frekvenční pásmo v oblasti vrcholu amplitudy a také naopak v oblasti jejího největšího propadu. Pro

analýzu frikativy pak využila spektrální těžiště, SD, sešikmení a špičatost. Ve své studii dospěla k závěru, že nejpřínosnějšími faktory pro identifikaci mluvího je pro všechny analyzované segmenty spektrální těžiště, pro nazály pak také SD. Jako nejvýhodnější pro účely identifikace se ukázala analýza realizací fonémů /m/ a /s/, kde se u všech faktorů ukázal jako signifikantní faktor mluvího. Naopak /l/ považuje Kavanagh za nejméně přínosné. Příčinu specifčnosti nazál pro daného mluvího shledává Kavanagh v relativní stabilitě podoby nosních dutin, které se zároveň značně liší mezi jednotlivci (Amino et al., 2006; citováno v: Kavanagh, 2012); SD a těžiště nazálních konsonantů se tudíž ukazují jako charakteristické pro daného mluvího.

Schindler & Draxler (2013) analyzovali spektrální momenty anglických nazál a frikativ. Spektrální moment je hodnotou, která popisuje celkový tvar spektra segmentu, a to pomocí součtu amplitud jednotlivých frekvencí, spektrálního těžiště, rozložení energie ve spektru, sešikmení a špičatosti spektra. Největší potenciál pro účely identifikace mluvího autoři studie našli u fonémů /m/, /n/, /f/ a /s/. U těchto segmentů byl vždy mluví vyhodnocen jako signifikantní faktor, a to dokonce ve větší míře než v případě vokálních formantů.

1.4.4. INTENZITA

Z řečového signálu je možné získat hodnoty intenzity řečového projevu, která je akustickým korelátem vnímané hlasitosti. Střední hodnoty intenzity nemohou být považovány za vhodné pro využití při identifikaci mluvího, neboť spíše než na jeho hlase závisejí na charakteru samotné nahrávky; odvíjejí se od prostředí, v němž nahrávání probíhalo, vzdálenosti mluvího od mikrofону, vlastností nahrávacího zařízení atd. Aby bylo možné porovnat střední hodnoty intenzity mezi dvěma nahrávkami, muselo by při získávání obou z nich být docíleno totožných podmínek, což ve forenzní praxi nelze pokládat za možné. Analýzu střední hodnoty intenzity tudíž není možné považovat za přínosnou pro identifikaci mluvího.

Je nicméně možné z řečového signálu získat dynamické ukazatele intenzity, tj. změny intenzity v čase. He & Dellwo (2014) sledovali, nakolik jsou pro mluvího charakteristické změny intenzity mezi jednotlivými slabikami, a to jednak holisticky pomocí směrodatné odchylky změn intenzity mezi slabikami, a také

lokálně, tedy párovým porovnáváním intenzitních vrcholů u dvojic sousedících slabik. V obou typech měření byl faktor mluvčího vyhodnocen jako statisticky signifikantní, což naznačuje, že tyto metody mohou mít potenciál aplikace při identifikaci mluvčího.

Jiným způsobem sledování průběhu intenzity v řeči je pozorování temporální organizace pozitivní a negativní dynamiky, tj. úseků, v nichž intenzita narůstá směrem od propadů amplitudy k jejím vrcholům, či naopak klesá. I v tomto případě se faktor mluvčího ukázal jako statisticky signifikantní, přičemž asi 70 % variability napříč mluvčími bylo identifikováno pomocí měření negativní dynamiky, tj. úseků, v nichž docházelo k poklesu amplitudy. Tento výsledek je zřejmě dán skutečností, že zatímco pozitivní dynamika, která odráží fázi otevírání vokálního traktu, nese zásadnější lingvistickou informaci, a je tudíž mluvčím více korigována, negativní dynamika odpovídá naopak fázi zavírání vokálního traktu, která již zásadní lingvistickou informaci nenesení, tudíž poskytuje větší prostor pro variabilitu (He & Dellwo, 2017). Podobně jako u formantových trajektorií lze tedy i zde pozorovat variabilitu mezi mluvčími, avšak určitou stabilitu v rámci projevu mluvčího, neboť v případě zavírání vokálního traktu mluvčí taktéž provádí automatizované pohyby.

1.4.5. GLOBÁLNÍ TEMPORÁLNÍ UKAZATELE

V řečovém projevu mluvčího je možné změřit ukazatele, které souvisejí s jeho časovým průběhem. Výhoda těchto jevů spočívá v jejich odolnosti vůči kvalitativní degradaci signálu, neboť pro jejich analýzu nejsou zapotřebí podrobné informace z frekvenční oblasti, jako je tomu u základní frekvence, vokálních formantů nebo spektrálních charakteristik. Výzkumy také nasvědčují, že temporální charakteristiky jsou odolné vůči maskování hlasu, neboť si jich mluvčí často nejsou vědomi, a tudíž nezařazují změny temporálních jevů do svých maskovacích strategií (Eriksson & Wretling, 1997; Wretling & Eriksson, 1998; Dellwo, Ramyeed & Dankovičová, 2009; citováno ve Weingartová & Volín, 2014).

Jedním z přístupů k analýze temporálních charakteristik projevu mluvčího je měření počtu lingvistických jednotek za jednotku časovou – obvykle je sledován počet slabik za sekundu. Tímto způsobem lze vyčíslit dva ukazatele: (1) mluvní

tempo, tedy průměrný počet vyslovených slabik za jednu sekundu v průběhu celé promluvy včetně pauz a hezitačních prvků, a (2) tempo artikulační, které kvantifikuje rychlost artikulace mluvího; jedná se taktéž o průměrný počet slabik za sekundu, z celkového trvání promluvy jsou však v tomto případě vyloučeny pauzy a hezitace. Jessen (2007b) považuje artikulační tempo za potenciálně přínosné pro identifikaci mluvího, pokud by byly vytvořeny populační statistiky, s nimiž by se hodnoty artikulačního tempa analyzovaných nahrávek mohly porovnávat, a sledovat, nakolik se artikulačního tempa daného mluvího odchyluje od populačního standardu. Upozorňuje však, že artikulační tempo se významně liší při spontánním a čteném projevu, a je zásadní mít na paměti, že ve forenzních podmínkách je často mluvní styl zachycený na sporné a srovnávací nahrávce rozdílný, a hodnoty artikulačního tempa získané na základě těchto nahrávek tudíž nemusejí být srovnatelné; na značnou variabilitu artikulačního tempa napříč různými mluvními styly u jednoho mluvího upozorňují také Cao & Wang (2011).

Jiným přístupem k analýze temporálního aspektu řečového projevu je měření načasování určitých úkonů v řeči – tj. globálních temporálních ukazatelů, nepřesně často také nazývaných ukazateli rytmickými (Volín, 2017). Jedná se o vyjádření temporálních pravidelností, jež lze vysledovat v řečovém kontinuu, obvykle v podobě podílu či variability trvání vokalických a konsonantických intervalů (Weingartová & Volín, 2014). Hodnoty těchto ukazatelů jsou do značné míry závislé na fonotaktice daného jazyka, nicméně poskytují prostor i pro variabilitu mezi mluvími. Individuální vzorce globálních temporálních ukazatelů mohou záviset na osvojených artikulačních návycích či na fyziologii vokálního traktu, případně je mohou ovlivňovat oba tyto faktory. Ukazují se jako charakteristické pro daného mluvího a poměrně stabilní navzdory prozodické a lingvistické variabilitě (Dellwo et al., 2015). Pro identifikaci mluvího v češtině se jako nejvhodnější globální temporální ukazatel jeví %V, tedy podíl vokalických intervalů v řeči, resp. podíl trvání vokalických intervalů a součtu trvání intervalů vokalických a konsonantických (Weingartová & Volín, 2014; Weingartová, 2015).

1.4.6. LOKÁLNÍ TEMPORÁLNÍ UKAZATELE

Kromě pozorování temporálních jevů v průběhu celé promluvy se lze zaměřit také na temporální jevy lokální, které zohledňují skutečnost, že tempo řeči vyjádřené jakýmkoliv způsobem nebývá u daného mluvčího konstantní v průběhu promluvy; jeho hodnota variuje v závislosti na různých faktorech, např. délce prozodické fráze či závěrovém zpomalování (Weingartová & Volín, 2014). Pro tyto účely byl zaveden ukazatel lokálního artikulačního tempa (*local articulation rate*, *LAR*; Pfitzinger, 1996; Volín, 2009b, citováno v: Weingartová & Volín, 2014), které je převrácenou hodnotou časové vzdálenosti mezi dvěma sousedními slabičnými vrcholy. Pomocí vyhlazených kontur lokálního artikulačního tempa v průběhu promluvy, tj. několika hodnot popisujících jeden promluvový úsek, lze zachytit rozdíly mezi mluvčími lépe než pomocí globálních temporálních ukazatelů, které poskytují pouze průměrné hodnoty pro celé úseky řečového signálu (Weingartová, 2013).

1.4.7. DISFLUENCE

V řečovém projevu jedince, především pak v projevu nepřipraveném, se mohou vyskytovat tzv. disfluenční jevy, tj. narušení plynulosti promluvy. Tyto prvky jsou mluvčími využívány zcela běžně, jestliže potřebují získat čas pro formulaci své výpovědi během hovoru. U některých mluvčích se pak mohou objevit také prvky dysfluenční, které jsou již projevy patologické neplynulosti řečového projevu.

Mezi disfluenční jevy patří vyplněné nebo tiché pauzy, opakování celých slov či jejich částí, opakování celých frází, tzv. falešné začátky (tedy přerušení započaté promluvy a započetí nové) a dloužení hlásek.

Braun a Resin (2015) našly u jednotlivých mluvčích charakteristické vzorce využívání různých disfluenčních prvků; většina mluvčích se převážně uchyluje k jednomu či dvěma typům disfluencí, nicméně žádný z pozorovaných typů nedominuje u všech mluvčích zároveň. Ačkoliv je disfluenční chování v rámci projevu jednoho mluvčího do jisté míry variabilní, stále zůstává pro daného jedince charakteristické, a tudíž je jeho analýza potenciálně přínosná při identifikaci mluvčího. Autorky studie ovšem upozorňují, že v reálných forenzních případech je často k dispozici malé množství řečového materiálu, a o disfluenčním

chování mluvčího lze tedy získat jen omezené informace. Analýzu disfluencí tudíž nelze při identifikaci mluvčího využít samostatně, může ovšem představovat oporou při identifikaci v kombinaci s dalšími metodami.

1.5. PROMĚNLIVOST HLASU V ZÁVISLOSTI NA MOMENTÁLNÍ SITUACI

Jednotlivé aspekty řečového projevu jsou, jak již bylo naznačeno, do jisté míry variabilní a závisejí na mnoha faktorech.

Podoba řečového signálu se může u jednoho mluvčího značně lišit v závislosti na mluvním stylu, tj. zda se jedná o projev spontánní, připravený či čtený, a také podle toho, k jakému publiku jedinec promlouvá, tedy zda je jeho projev formální či neformální.

Hlas se může zásadně měnit také v závislosti na prostředí, ve kterém je řečový projev pronášen. Pokud mluvčí hovoří v hlučném prostředí, je podoba řečového signálu odlišná než tehdy, když je stejný hluk do nahrávky daného projevu přidán teprve dodatečně. Tuto skutečnost způsobuje tzv. Lombardův efekt, tj. tendence mluvčích v hlučném prostředí měnit určité parametry svého řečového projevu. Tyto změny zahrnují např. zvýšení základní frekvence, zvýšení intenzity, narovnání spektrálního sklonu, posuny vokalických formantů či prodlužování vokálů (Junqua, J.-C., 1993; Van Summers et al., 1988).

Charakteristické změny v řečovém projevu lze předpokládat také v dalších specifických situacích. Podobně jako je možné u lidí, kteří vědí, že jsou fotografováni, pozorovat změny výrazu tváře, dá se očekávat, že je-li si mluvčí vědom, že je nahráván, do jisté míry pozmění svůj hlas. I při telefonátu může být charakter tzv. telefonního hlasu adaptován dané situaci díky představě, že dotyčný hovoří s někým na delší vzdálenost a srozumitelnost přenášeného hlasu bude díky přenosové cestě a ruchům snížena. Takový volající se pak pravděpodobně může pokoušet o zjednodušení svých formulací a možná i částečně přeladí své formanty podobně, jako tomu dochází u Lombardova jevu v hlučném prostředí; tedy intuitivně přeladí svůj hlas za účelem zvýšení šance porozumění na straně posluchače.

Podobu promluvy jedince ovlivňuje také jeho komunikační partner; jeho projevu má mluvčí v některých ohledech tendenci svou řeč připodobňovat, v jiných se od ní naopak vzdalovat (Bilous & Krauss, 1988; citováno v: Giles & Ogay, 2007). Giles et al. (1991: 48) upozorňují, že „akomodační procesy, byť v téže části hovoru, mohou vykazovat odlišné, dokonce i protichůdné trendy.“

Na řečový projev mají dopad také afektivní stavy, v nichž se mluvčí při jeho realizaci nachází. Při některých afektivních stavech, např. ve vzteku, může mít mluvčí tendenci zvyšovat své mluvní úsilí, což má za následek narovnávání spektrálního sklonu. Ukazuje se, že pokud má mluvčí radost či vztek, dojde ke zvýšení f_0 i její variability; naopak je tomu např. při smutku (Scherer, 2003; citováno v: Skarnitzl & Hývlová, 2014: 52).

Afektivním stavem, který se často vyskytuje v nahrávkách využívaných při reálných forenzních případech, je stres. Jeho dopady na řečový projev se mohou značně lišit napříč mluvčími (Skarnitzl & Hývlová, 2014: 52). Giddens et al. (2013: 390.e27) popisují obecný trend zvyšování f_0 při stresu, které může být následkem větší tenze svalů, zejména svalů štítnoprstencového. Podotýkají však, že u jednotlivých mluvčích se dopady stresu různí.

Aspektem, jehož vliv na podobu řečového signálu je ve forenzních případech zásadní, je také lež. Komerční systémy pro její detekci, které jsou na trhu k dispozici, se zaměřují zároveň na lež a stres, a to na základě předpokladu, že mluvčí při lži vykazují více stresu než ti, kteří říkají pravdu. Ukazuje se však, že takový jednoznačný vztah mezi těmito dvěma faktory neexistuje, alespoň ne univerzálně u všech mluvčích. Efektivita těchto automatických systémů je tudíž diskutabilní (Harnsberger et al., 2009). Kirchhübel & Howard (2011) sledovali rozdíly v průměru a směrodatné odchylce f_0 , intenzitě a vokálních formantech při stresu a lži. Podle výsledků jejich studie nelze na základě těchto parametrů rozlišit mluvčí, kteří lžou, od těch, kteří říkají pravdu; jejich měření ukázala „absenci zásadních rozdílů, a pokud změny nastaly, byly obvykle totožné pro pravdu i lež. Značná variabilita v rámci projevu mluvčího i napříč jednotlivými mluvčími zdůrazňuje skutečnost, že projevy lži v řeči jsou individuální, mnohotvárné a nejsou v žádném případě jednoznačně identifikovatelné“ (Kirchhübel & Howard, 2011: 1096).

Anolli & Ciceri (1997) rozlišují dvojí typ lži: připravenou a nepřipravenou. Připravená lež je předem promyšlena a je směřována k posluchači, který mluvčího

ze lži nepodezírání, a ten je tudíž vystaven menšímu stresu, neboť není nucen přímo během rozhovoru formulovat nová lživá tvrzení. V takovém případě je podoba řečového projevu velmi blízká pravdivým promluvám. Naproti tomu druhý typ lži, kteří autoři studie popisují, je lež nepřipravená; ta je předkládána aktivnímu posluchači, který mluvčího přímo podezírání ze lži a klade mluvčímu otázky, čímž jej nutí okamžitě v průběhu dialogu lež doplňovat. Mluvčí je tudíž ve větším stresu a v jeho řečovém projevu se mohou projevat změny s ním související, např. zvýšení f_0 a její variability či častější výskyt disfluenčních prvků (Anolli & Ciceri, 1997).

1.6. MASKOVÁNÍ HLASU

Kromě změn řečového projevu, jež ovlivňují okolnosti, a kterých si mluvčí často ani není vědom, může jedinec také záměrně modifikovat svůj hlas s tím cílem, aby znemožnil svou identifikaci na základě nahrávky své promluvy. S takovým jednáním se ve forenzní praxi lze setkat tehdy, pokud si je pachatel při činu (např. při výhrůžném telefonátu) vědom, že jeho projev může být nahráván.

Případy maskování hlasu nejsou obecně ve forenzní praxi příliš časté; Masthoff (1996) uvádí, že ze všech případů trestné činnosti zkoumaných na univerzitě v Trevíru mezi lety 1988 a 1995 se pouze v 5 % vyskytlo maskování hlasu. U některých konkrétních typů trestné činnosti se nicméně tento jev vyskytuje poměrně hojně, a to zejména v případech únosů, vydírání, sexuálního obtěžování a falešných telefonických oznámení (Künzel, 2000). V případech, kdy pachatel při činu promlouval a mohl předpokládat, že jeho projev bude nahráván, se maskování hlasu objevilo dokonce v 52 % případů, jimiž se zabýval německý Spolkový kriminální úřad (*Bundeskriminalamt; BKA*) v letech 1988-1995. „Tento podíl zahrnuje i případy vydírání, u nějž poměr maskování hlasu dosahuje 69 %“ (Masthoff, 1996: 161).

Za nejefektivnější způsob maskování hlasu lze považovat maskování elektronické, tj. např. využití elektronické syntézy řeči, které může i zcela znemožnit identifikaci mluvčího (Clark & Foulkes, 2007). Tyto případy však podle dostupných statistik nejsou příliš frekventované; Masthoff (1996: 161) uvádí, že mezi výskyty maskování hlasu se objevilo pouze 1 % případů maskování elektronického, Gfroerer (1994, citováno v: Eriksson, 2010) jeho

výskyt vyčísluje na 10 % případů maskování hlasu. Künzel (2000) nicméně upozorňuje, že ačkoliv je elektronické maskování hlasu relativně vzácné, jeho četnost v posledních letech narůstá; jedná se především o předem připravené promluvy vytvořené manipulací nahrávek řeči jiných mluvčích.

Ve většině případů maskování hlasu však mluvčí volí méně sofistikované strategie; nejčastějšími způsoby záměrné modifikace řečového projevu jsou falzet, třepená fonace, šepot, imitace cizineckého přízvuku či ucpání nosu. Mezi další využívané maskovací strategie patří např. umístění cizího předmětu do úst, využití rezonátoru (např. plechovky přiložené ke rtům), kapesník přidržovaný před ústy či napodobování regionálního dialektu (Künzel, 2000; Masthoff, 1996).

Některé změny projevu při maskování mají tendenci k častějšímu souvyskytu, např. zvýšení f_0 a umělá monotónnost (tj. snížení směrodatné odchylky f_0) či snížení mluvního tempa a umělá monotónnost (Künzel, 2000). Podle Masthoffovy (1996) studie uplatňuje většina mluvčích při maskování hlasu změnu pouze jednoho, případně dvou parametrů, ne však více; ke změně jediného parametru přistoupilo 55 % mluvčích. Tuto relativní jednoduchost maskovacích strategií Masthoff připisuje skutečnosti, že mluvčí při současné formulaci svých promluv a udržování maskovací strategie musejí vynakládat poměrně značné kognitivní úsilí.

Ať už se daný jedinec rozhodne pro jakoukoliv maskovací strategii, v průběhu promluvy má modifikace hlasu tendenci k postupné deklinaci, zejména tehdy, pokud mluvčí uskutečňuje více maskovaných promluv (např. telefonátů) v průběhu času. V některých případech může nakonec dojít i k úplnému vymizení maskování hlasu (Künzel, 2000).

1.6.1. ZMĚNY ZÁKLADNÍ FREKVENCE

Nejčastěji modifikovaným parametrem při maskování hlasu je, jak již bylo zmíněno, základní frekvence. Její změny jsou pro mluvčí poměrně snadno proveditelné a zároveň velmi výrazné, tudíž se jeví jako efektivní pro skrytí identity.

Při zvýšení základní frekvence je možné dosáhnout jiného způsobu fonace – falzetu. Naopak při výrazném snížení f_0 lze docílit třepené fonace, tedy nepravidelných hlasivkových pulzů. Tento způsob fonace se běžně objevuje

v řeči, obvykle na konci prozodických frází, je nicméně možné jej využívat během celé promluvy jakožto strategii maskování hlasu.

Künzel (2000) pozoroval 100 mluvčích (50 mužů a 50 žen), kteří v rámci studie záměrně zvýšili či naopak snížili svůj hlas. Při zvyšování f_0 někteří mluvčí zůstávali u své modální fonace, u některých došlo ke změně hlasového rejstříku a přechodu do falzetu (tento jev byl častěji pozorován u mužů). Celkově byly u mužů při zvyšování základní frekvence zaznamenány znatelnější posuny. Naproti tomu snižování základní frekvence se objevovalo častěji u žen, a u nich také docházelo k výraznějším změnám (v průměru o 3,5 půltónu, naproti tomu muži svou základní frekvenci snižovali v průměru o 2,5 půltónu). Příčinu tohoto výsledku vidí Künzel v rozdílech mezi přirozenou f_0 mužů a žen; zatímco základní frekvence mužských mluvčích je přirozeně nižší, a mají tudíž více prostoru hlasem stoupat, u žen je tomu právě naopak. Dále Künzel nicméně uvádí, že mluvčí, jejichž přirozená f_0 byla relativně nízká, měli tendenci hlasem ještě více klesat, zatímco ti s poměrně vysokou přirozenou f_0 hlasem častěji stoupali; tento trend byl nápadnější u žen než u mužů.

Také u českých mluvčích, kteří měli libovolně maskovat svůj hlas, byly nejčastěji pozorovány modifikace základní frekvence. Z analýzy dat, jež obsahuje databáze obecné češtiny pro forenzní účely (Skarnitzl & Vaňková, 2017), vyplynulo, že ke změnám f_0 dochází přibližně u 70 % mluvčích; z toho 43 % mluvčích má tendenci svou základní frekvenci zvyšovat a naopak u 27 % lze zaznamenat její snížení (Růžičková, 2016).

Schopnost posluchačů rozpoznat mluvčího, který záměrně maskuje svou základní frekvenci, pomocí poslechu je oproti rozpoznání mluvčího, který svůj hlas nemaskuje, značně zhoršena; podle studie McGehee (1937; citováno v: Eriksson, 2010) je správnost identifikace snížena z 80 % na 63 %. Pakliže mluvčí hovoří falzetem, správnost jeho identifikace posluchači klesá až na 4 % (Wagner & Köster, 1999; citováno v: Eriksson, 2010).

Co se týče možností získání informací o přirozené základní frekvenci i při jejím maskování, lze přirozenou f_0 do jisté míry predikovat, pokud je maskováním snížena. Pokud však dojde k jejímu zvýšení, jsou možnosti získání informací o přirozené f_0 mluvčího značně omezené, a to bez ohledu na to, zda dojde ke změně hlasového rejstříku, či nikoliv (Künzel, 2000).

Při výrazné změně základní frekvence dochází zároveň také k posunu vokálních formantů, konkrétně F1-3, přestože obvykle jsou na sobě tyto parametry více méně nezávislé. To je zřejmě zapříčiněno pohybem hrtanu – při extrémním snižování hlasu se hrtan snižuje, a dochází tudíž k celkovému prodloužení vokálního traktu, a tedy také ke snižování vokálních formantů. Naopak při výrazném zvýšení základní frekvence dojde obdobným způsobem ke zkrácení vokálního traktu, a tedy i ke zvýšení formantů (Künzel, 2000).

K úplné ztrátě ukazatelů spojených se základní frekvencí, jakými jsou střední hodnota f_0 nebo její průběh v řeči, dochází při šepotu, který se také objevuje mezi maskovacími strategiemi ve forenzní praxi. Bartle & Dellwo (2015) sledovali schopnost posluchačů rozlišit mluvčí pomocí diskriminačního percepčního testu; posluchačům byly předkládány dvojice položek, a to obou realizovaných buď s běžnou fonací, nebo šeptem, a respondenti měli určit, zda jsou obě položky vzorky řečového projevu téhož mluvčího. Celkem se studie účastnilo 33 posluchačů, z toho 11 forezních fonetiků a 22 laiků. Výsledky ukázaly, že při rozpoznávání položek realizovaných s modální fonací se úspěšnost odborníků a laiků příliš nelišila; průměrná úspěšnost činila 92 % u forezních fonetiků a 89 % u laických posluchačů. Výraznější rozdíly se objevily při identifikaci šeptaných hlasů; úspěšnost forezních fonetiků klesla na 79 %, zatímco u laiků byla snížena na 72 %. Obecně je tedy vlivem šepotu snížena schopnost posluchačů identifikovat mluvčího na základě hlasu, nicméně školení posluchači dosahují lepších výsledků.

Orchard & Yarmey (1995) taktéž pozorovali lepší výsledky posluchačů při identifikaci hlasu, pokud byly porovnávány dva vzorky realizované s modální fonací než dva vzorky řeči pronášené šeptem, a to u distinktivních i nedistinktivních hlasů. Doporučují opatrnost při sestavování hlasové rekognice svědkem s využitím dvou šeptaných nahrávek, neboť v takové situaci častěji dochází k mylným tvrzením, že se jedná o téhož mluvčího, než při porovnávání dvou nahrávek s běžnou fonací; hlasovou rekognici, při které je využita jedna nahrávka běžného hlasu a jedna nahrávka šepotu, však považují za ještě méně spolehlivou.

1.6.2. IMITACE DIALEKTU

Jedním ze způsobů maskování hlasu, které se ve forenzní praxi objevují, je napodobování jiného regionálního dialektu, než kterým daný mluvčí běžně hovoří, ovšem stále v rámci jeho mateřského jazyka.

Dialekt lze považovat za velmi zásadní součást identity mluvčího a při jeho identifikaci pomocí poslechu může v některých případech do značné míry zastřít další aspekty řečového projevu jedince – pro posluchače může být dialekt mluvčího natolik silným vjemem, že další jevy v jeho řeči již nezaznamená (Eriksson, 2010).

Napodobení dialektu zpravidla působí přirozenějším dojmem než např. maskování hlasu pomocí objektu umístěného v ústech či kapesníku přidrženého před ústy, proto může v takovém případě být obtížné identifikovat, že se jedná o maskovaný projev (Markham, 1999).

Markham (1999) testoval, nakolik jsou švédští mluvčí schopni napodobit regionální dialekt odlišný od svého vlastního tak, aby toto napodobení znělo přesvědčivě dalším rodilým mluvčím švédštiny; sledoval zejména schopnost imitace alofonů některých švédských fonémů charakteristických pro dané nářeční oblasti. Podle výsledků jeho studie se úspěšnost mluvčích při imitaci dialektů různila, nicméně u žádného z nich se posluchačům nepodařilo identifikovat jejich dialekt původní. Pokud tedy jako maskovací strategii mluvčí zvolí změnu realizace hlásek charakteristických pro svůj dialekt, jsou zřejmě schopni před posluchači úspěšně skrýt svůj skutečný regionální původ, a to i tehdy, pokud se jim zvolený dialekt nepodaří zcela věrně napodobit.

1.6.3. IMITACE CIZINECKÉHO PŘÍZVUKU

Při záměrné modifikaci řečového projevu lze uplatnit také imitaci cizineckého přízvuku. I tento způsob maskování hlasu se vyskytuje ve forenzní praxi (Künzel, 2000).

Neuhauser (2008) pozorovala, nakolik jsou rodilí mluvčí němčiny schopni imitovat francouzský přízvuk, a to z produkční perspektivy; analyzovala některé fonetické jevy charakteristické pro francouzský přízvuk, např. elize /h/, vypouštění iniciálního prevokalického rázu, potlačení aspirace, nazalizace vokálů

či posun artikulace vokálů dopředu. U všech mluvčích bylo do určité míry zaznamenáno potlačení aspirace, u některých dokonce zcela; u většiny mluvčích se také objevily změny artikulace vokálů. Celkově se úspěšnost napodobování přízvuku lišila napříč mluvčími, většina z nich byla také při jeho imitaci nekonzistentní a jednotlivé segmenty nerealizovala způsobem charakteristickým pro francouzský přízvuk po celou dobu promluvy.

Neuhauser & Simpson (2007) sledovali úspěšnost imitace francouzského přízvuku v němčině také z odlišné perspektivy; zjišťovali, nakolik jsou rodilí mluvčí němčiny úspěšní při rozpoznávání autentického a imitovaného francouzského přízvuku. Jak již bylo zmíněno, úspěšnost imitace francouzského přízvuku rodilými mluvčími němčiny se různila. Lišila se však také úspěšnost, s jakou respondenti skutečný a imitovaný francouzský přízvuk rozeznávali; bylo zaznamenáno několik případů, ve kterých zněl imitovaný přízvuk posluchačům autentičtěji než ten skutečný. Tyto výsledky naznačují, že někteří mluvčí jsou při maskování hlasu schopni napodobit cizinecký přízvuk natolik věrně, že rodilým mluvčím zní jako přízvuk autentický, a takováto maskovací strategie tedy může zkomplikovat profilování mluvčího či jeho identifikaci.

1.6.4. UMÍSTĚNÍ CIZÍHO PŘEDMĚTU DO ÚST

Jak již bylo uvedeno, mimo modifikací nastavení vokálního traktu mluvčí při maskování hlasu v některých případech využívají také různé objekty umístěné do úst nebo před ústa; jedná se např. o využití plechovky přidržované před ústy coby rezonátoru, přiložení kapesníku před ústa či fixaci spodní čelisti vložení tužky mezi přední zuby.

Dopadem posledního zmíněného typu maskování na řečový projev mluvčího se zabývali Figueiredo & Britto (1996), kteří se právě s tímto typem maskování setkali ve své forenzně fonetické praxi, a to v souvislosti s vyšetřováním několika případů únosů. Upozorňují na skutečnost, že zatímco způsoby maskování, které více souvisí s fonací (např. třepená fonace, šepot nebo falzet) mají stabilnější akustické dopady, než přidržování tužky mezi zuby, neboť všechny řečové segmenty jsou ovlivněny více méně stejně; na druhou stranu dopady předmětu vloženého mezi přední zuby se u různých segmentů značně liší. Z tohoto důvodu může dojít např. k zakrytí dialektu mluvčího, a tudíž ke značné komplikaci jeho

identifikace. Tento způsob maskování vyžaduje konstantní čelistní úhel během celé promluvy, rty jsou obvykle zaostřené a předmět mezi zuby znemožňuje jejich zaokrouhlování, také pohyby jazyka jsou do značné míry omezeny. Podle výsledků studie dochází k posunům vokalických formantů v důsledku omezení vokalického prostoru; směr těchto posunů se odvíjí od identity daného vokálu. Změny vokalických formantů se objevují u všech mluvčích, jejich podoba se však napříč mluvčími ukazuje jako značně variabilní.

1.7. CÍLE PRÁCE

Tato práce předkládá analýzu změn, ke kterým dochází v řečovém projevu, snaží-li se mluvčí záměrně měnit svůj hlas takovým způsobem, aby nebylo možné provést jejich identifikaci na základě pořízené nahrávky. Bude provedena analýza maskovaných a nemaskovaných projevů mužských mluvčích, kteří využívali komplexní maskovací strategie, tj. využili změn více řečových parametrů zároveň.

Cílem práce je předložit informace o změnách různých řečových parametrů při maskování hlasu a vyhodnotit, zda lze u jednotlivých parametrů pozorovat stabilitu i navzdory záměrným modifikacím řečového projevu jedince, či jsou mluvčí schopni dané parametry ve svém řečovém projevu významně měnit v případě, že maskují svůj hlas. V rámci této práce bude zároveň také hledána odpověď na otázku, zda u zkoumaných parametrů existují obecné tendence ke konkrétní podobě posunů při maskování hlasu, tj. zda lze pozorovat podobné vzorce změn jednotlivých parametrů napříč mluvčími.

2. METODA

2.1. MATERIÁL

Při získávání materiálu pro tuto práci byla využita databáze obecné češtiny vytvořené pro forenzní účely (Skarnitzl & Vaňková, 2017), která obsahuje nahrávky projevů 100 mužů ve věku mezi 18 a 40 lety. Nahrání mluvčí pocházejí ze středočeské nářeční oblasti a lze tedy předpokládat, že jsou rodilými mluvčími obecné češtiny, tudíž je možné zanedbat případný vliv dialektu z oblasti Moravy či českého pohraničí na jejich mluvený projev.

Je třeba upozornit na skutečnost, že statistiky, které byly provedeny na tomto vzorku a jsou považovány za reprezentativní pro středočeskou nářeční oblast, nemusejí být totožné pro nářeční oblasti jiné, a pro forenzní účely v těchto oblastech by bylo vhodné vytvořit databáze vlastní.

Využitá databáze obsahuje od každého mluvčího 5 typů projevu: spontánní projev, krátké izolované čtené věty, popis obrázku, krátký čtený text a krátký text předčítaný maskovaným hlasem. Nahrávky byly pořizovány v tichém prostředí nahrávacím zařízením Edirol R09 nekomprimovaně s vzorkovací frekvencí 48 kHz a bitovou hloubkou 16 bit.

Při nahrávání maskovaného textu byli mluvčí požádáni, aby libovolným způsobem změnili svůj hlas tak, aby podle něj nemohli být identifikováni případným svědkem. V této úloze mluvčí volili různé strategie, ve většině případů se však uchylovali k výrazné změně jediného řečového parametru; povětšinou se jednalo o zvýšení základní frekvence, méně pak o její snížení (Růžičková, 2016). Maskovací strategie většiny mluvčích byly tedy poměrně jednoduché a zároveň také snadno odhalitelné – jednalo se o natolik výrazné změny, že by případnému svědku bylo na první poslech patrné, že se nejedná o přirozený mluvený projev.

Někteří mluvčí však pro maskování svého hlasu volili složitější strategie, při kterých přistoupili ke změnám více řečových parametrů zároveň. V mnohých případech působil výsledný projev relativně přirozeným dojmem, tj. případný posluchač, který by nepředpokládal, že se jedná o maskovaný projev, by pravděpodobně ihned neidentifikoval daný hlas jako záměrně modifikovaný.

Nahrávky, jež obsahuje databáze obecné češtiny pro forenzní účely, nebyly pro účely této práce shledány zcela vhodnými. Zprv, jedná se o poměrně krátké texty, jejichž přednes trvá asi jednu minutu, a poskytují tudíž relativně málo materiálu pro analýzu. Zásadnější překážkou pro srovnávání je pak skutečnost, že se texty pro tyto dva typy projevu navzájem liší: text pro nemaskovaný projev je dialogem dvou postav, Milánka a maminky, a text pro projev maskovaný je jednou stranou telefonního hovoru dvou „mafianů“. Oba texty tedy mohou mluvčího vést k případné stylizaci, nadto v každém z textů odlišné. V obou textech lze předpokládat výskyt pauz, který by se při předčítání souvislého textu bez střídání replik postav lišil; v případě prvního textu jde o střídání replik dvou postav, v případě textu druhého pak o pauzy na místech, na kterých by hovořil druhý účastník telefonního hovoru.

Mluvčí využívající komplexní maskovací strategie tedy byli požádáni o opětovné nahrávání. 12 z oslovených mluvčích souhlasilo a byly pořízeny nové nahrávky pro lepší srovnatelnost maskovaného a nemaskovaného projevu. Jednalo se o populárně naučný text, jehož přednes trval asi 2,5 minuty; tentýž text byl mluvčími předčítán jak běžným, tak maskovaným hlasem. Znění tohoto textu je předloženo v Příloze č. 1.

Nové nahrávky byly opět pořizovány pomocí nahrávacího zařízení Edirol R09 s vzorkovací frekvencí 48 kHz. Přes snahu o nahrávání v tichém prostředí bylo nutné pořízené nahrávky dvou mluvčích z analýzy vyřadit pro ruch na pozadí. V této práci tedy byly využity nahrávky maskovaného a nemaskovaného projevu deseti mluvčích.

2.2. POSLECHOVÁ ANALÝZA

Samotné akustické analýze jednotlivých idiosynkratických jevů předcházel pečlivý poslech všech nahrávek, při kterém byly zaznamenány percepční charakteristiky jednotlivých projevů.

Byly pozorovány charakteristické znaky přirozených, tedy nemaskovaných projevů mluvčích, kam patří celková vnímaná výška hlasu, dynamika intonace,

artikulační vlastnosti (např. otevřenost vokálů, laryngalizace, labializace, palatalizace, ostrost sykavek apod.) či vnímané tempo, ale také např. řečové vady.

Následně byly u těchto charakteristik pozorovány posuny při maskovaném projevu.

Poslechová analýza byla provedena pouze autorkou práce a data nebyla podrobena percepčním testům s jinými posluchači, výsledky analýzy jsou tedy subjektivní. Přesto je možné výsledky poslechové analýzy pokládat za spolehlivé, vzhledem k tomu, že se jedná o hrubé charakteristiky projevu mluvčích, nikoliv o obtížně postřehnutelné detaily.

V rámci akustické analýzy pak byly sledovány akustické koreláty těchto jevů a na závěr byly pozorovány souvislosti mezi výsledky poslechové analýzy a analýzy akustické.

2.3. AKUSTICKÁ ANALÝZA

Pro účely akustické analýzy byla v programu Praat (Boersma & Weenink, 2017) provedena anotace nahrávek v několika vrstvách: hlásky, slabiky, melodémy a nádechové úseky.

Nejprve byly nahrávky rozděleny na jednotlivé hlásky, a to pomocí automatického segmentačního nástroje Prague Labeller (Pollák et al., 2007). Hranice hlásek byly následně manuálně korigovány.

Manuálně byly také označeny hranice melodémů, které byly podle typu a pozice rozřazeny do pěti kategorií: melodém neukončující v iniciální, mediální a finální pozici v rámci nádechového úseku, melodém ukončující klesavý a melodém ukončující klesavý příznakový, tj. sahající přes více než jeden mluvnický takt.

Na základě segmentace na jednotlivé hlásky byla provedena automatická segmentace na slabiky, která byla také manuálně korigována. Slabiky byly děleny v souladu s principem maximální přetury.

V rámci akustické analýzy byla sledována základní frekvence, vokalické formanty, intenzita a temporální jevy.

U analyzovaných jevů bylo testováno, zda maskování hlasu představuje signifikantní faktor pro změny hodnot daných parametrů. Při testování rozdílnosti mediánů hodnot byl využit lineární smíšený model; pro účely analýzy byly vytvořeny dva modely, jeden z nich se zahrnutím fixního faktoru maskování, druhý pak bez něj. Mluvčí byl v obou modelech zohledněn jakožto náhodný faktor. Pomocí testu ANOVA pak bylo na hladině významnosti 5 % vyhodnoceno, zda se tyto dva získané modely signifikantně liší, či nikoliv.

Druhým způsobem statistické analýzy bylo využití Pearsonova chí kvadrátu neboli testu nezávislosti. Tento postup byl využíván v případě analýz dynamických průběhů základní frekvence, intenzity a lokálního artikulačního tempa, v jejichž případě bylo zkoumáno typické využívání kontur. Pomocí této statistické metody bylo testováno, zda je maskování hlasu signifikantním faktorem pro změny rozdělení, tj. četnosti využívání jednotlivých kontur danými mluvčími. Statistická významnost faktoru maskování byla u těchto jevů testována pro každého mluvčího zvlášť, a z toho důvodu byla provedena Bonferroniho korekce pro vícenásobné testy – při tomto postupu je nulová hypotéza zamítnuta tehdy, pokud je její p-hodnota menší nebo rovna podílu zvolené hladiny významnosti (5 %) a počtu provedených testů (v tomto případě 10, neboť byly pozorovány projevy celkem deseti mluvčích); u jednotlivých mluvčích byl tedy faktor maskování vyhodnocen jako statisticky signifikantní tehdy, pokud byla p-hodnota získaná při daném testování menší nebo rovna 0,005.

Veškerá data využitá v této práci, stejně jako skripty vytvořené pro jejich zpracování, jsou umístěny v archivu Fonetického ústavu Filozofické fakulty Univerzity Karlovy.

2.3.1. ZÁKLADNÍ FREKVENCE

Hodnoty základní frekvence byly automaticky extrahovány pomocí programu Praat s časovým krokem 0,01 s; frekvenční rozsah byl stanoven na 60 až 300 Hz. Získané hodnoty nebyly pro velký objem dat manuálně korigovány, byly však eliminovány detekované hodnoty v neznělých segmentech, ve kterých nebyla základní frekvence přítomna, případně byl přítomen pouze její dozvuk způsobený hláskovým okolím; tyto hodnoty byly tudíž extrahovány chybně a mohly by způsobit zkreslení výsledků.

U jednotlivých nahrávek byly srovnány statistické ukazatele základní frekvence, tedy medián a směrodatná odchylka, a také variabilita těchto ukazatelů mezi jednotlivými melodémy. Medián hodnot základní frekvence byl upřednostněn před průměrnou f_0 , neboť je u něj menší riziko zkreslení vlivem extrémních hodnot.

Dále byly pozorovány průběhy kontur f_0 v jednotlivých melodémech. Hodnoty f_0 byly z Hz převedeny na pultónovou stupnici a následně normalizovány, a to následujícím způsobem: pro každou celou nahrávku byla změřena průměrná hodnota základní frekvence, která byla následně odečtena od průběhu f_0 v nahrávce. Tímto způsobem bylo dosaženo vystředění hodnot f_0 každé nahrávky kolem nuly, zůstala však uchována informace o relativní střední hodnotě f_0 v rámci jednotlivých melodémů, tj. zda je kontura f_0 v určitém melodému položena výše či níže než je celková průměrná hodnota f_0 dané nahrávky.

Pro kontury f_0 v jednotlivých melodémech byly pomocí jazyka R za využití knihovny rPraat (Bořil & Skarnitzl, 2016) vyčísleny koeficienty Legendreových polynomů $c_1 - c_4$, které charakterizují hrubý tvar kontury: c_1 určuje střední hodnotu, c_2 sklon křivky, c_3 vypovídá o jejím konvexním či konkávním pronutí a c_4 charakterizuje pronutí esovitě.

Pomocí shlukové analýzy k-means (Hartigan & Wong, 1979) bylo nalezeno 6 clusterů, tedy typických kontur navzájem se lišících tvarem či relativní f_0 . Tento celkový počet clusterů byl shledán nejvhodnějším; při nižším počtu nebylo dosaženo dostatečné specifičnosti jednotlivých průběhů, naopak při vyšším počtu se již v analýze začaly projevovat outliery, tedy hodnoty, které nebyly mezi daty běžné nebo byly výsledkem chybné extrakce.

Jednotlivé clustery byly přiřazeny k melodémům a bylo srovnáváno, které typické průběhy využívali mluvčí v obou typech projevu při realizaci jednotlivých typů melodémů.

2.3.2. VOKALICKÉ FORMANTY

V analyzovaných nahrávkách byly sledovány hodnoty vokalických formantů F_1-4 mezi jednotlivými vokalickými kvalitami a také dlouhodobé formantové

distribuce LTF1-4. Hodnoty formantů byly automaticky detekovány v programu Praat; byly detekovány formanty F1-F5 ve frekvenčním rozsahu do 5,5 kHz, a to pomocí metody Burg.

Z analyzovaných dat byly vyřazeny vokály v okolí nazál a likvid, v nichž mohou být formanty ovlivněny koartikulací a zkreslovat tudíž výsledky.

Výsledky analýzy vokálních formantů byly následně porovnávány u maskovaných a nemaskovaných projevů jednotlivých mluvčích.

2.3.3. SPEKTRÁLNÍ TĚŽIŠTĚ

Pro neznělé sibilanty, tedy [s] a [ʃ], bylo v Praatu vyčísleno spektrální těžiště, tj. průměrná hodnota celé oblasti ve spektru segmentu, a to uprostřed trvání daného segmentu. Získané hodnoty pak byly porovnány mezi jednotlivými mluvčími a jejich maskovanými a nemaskovanými projevy.

2.3.4. INTENZITA

Pomocí programu Praat byly extrahovány intenzitní kontury jednotlivých melodémů, na něž byly rozděleny analyzované nahrávky, a to s automatickým časovým krokem a určením minimální základní frekvence jako 50 Hz. Výsledky extrakce hodnot intenzity v Praatu jsou udávány v dB, nejedná se však o dB FS, ale o hodnoty dB SPL normované k arbitrárně stanovené hodnotě v programu Praat. Je tedy třeba mít na zřeteli fakt, že tato analýza neodpovídá skutečným hodnotám intenzity zvuku při nahrávání.

Na základě získaných intenzitních kontur byly změřeny průměrné hodnoty intenzity v analyzovaných nahrávkách a byla provedena jejich párová srovnání v rámci jednotlivých mluvčích. Je třeba upozornit, že rozdíly v průměrné naměřené intenzitě mezi mluvčími nemusejí být směrodatné vzhledem k možným odlišným podmínkám při nahrávání. Z tohoto důvodu byla pro účely srovnávání intenzity maskovaných a nemaskovaných projevů provedena normalizace, a to následujícím způsobem: byly vypočítány průměrné hodnoty intenzity nemaskovaných projevů jednotlivých mluvčích, a ty byly následně odečteny od

hodnot intenzity v jejich projevech obou typů. Takto bylo dosaženo vystředění intenzity nemaskovaných projevů všech mluvčích kolem nuly, zatímco informace o rozdílu v intenzitě u jejich maskovaných projevů oproti projevům nemaskovaným zůstala zachována.

Podobně jako v případě kontur základní frekvence byly také pro intenzitní kontury získané z jednotlivých melodémů vyčísleny koeficienty Legendreových polynomů. Jedná se o hrubé proložení křivkou, ačkoliv ve skutečnosti intenzitní kontury poměrně výrazně kolísají. Je nicméně vhodné pozorovat, zda existují typické průběhy intenzitních kontur, ke kterým se mluvčí uchylují, a pokud ano, zda se četnost užití různých průběhů liší mezi mluvčími a zda má tendenci podléhat změnám při maskování hlasu.

Při hledání odpovědí na tyto otázky byla, opět podobně jako v případě základní frekvence, využita shluková analýza, konkrétně metoda k-means. Bylo nalezeno šest typických hrubých kontur. Stejně jako u základní frekvence byla provedena normalizace podle mluvčího, a to odečtením hodnot jednotlivých vzorků od průměrné naměřené intenzity dané nahrávky.

Pro clustery nalezené metodou k-means bylo porovnáno užití jednotlivými mluvčími při maskovaném a nemaskovaném projevu v jednotlivých typech melodémů.

2.3.5. GLOBÁLNÍ TEMPORÁLNÍ CHARAKTERISTIKY

U každé nahrávky byly analyzovány tři globální temporální ukazatele: mluvní tempo, artikulační tempo a %V, tj. podíl vokalických intervalů v průběhu artikulace (tedy podíl délky vokalických intervalů a součtu délek vokalických a konsonantických intervalů).

Hodnoty mluvního tempa byly získány z celých nahrávek, a to v časovém úseku od začátku artikulace po její konec. Mluvní tempo bylo vyčísleno v průměrném počtu slabik za sekundu.

Také průměrné hodnoty artikulačního tempa byly vypočítány pro celé nahrávky. Tento ukazatel byl však také vyčíslen zvláště pro každý melodém, aby bylo možné napříč nahrávkami porovnat i jeho variabilitu. Stejně jako v případě

mluvního tempa byla zvolena jednotka syll/s. Při výpočtech artikulačního tempa byly vyloučeny pauzy trvající 100 ms a více.

Podíl vokálních intervalů %V byl vypočten v celých nahrávkách i v jednotlivých melodémech, stejně jako artikulační tempo.

Získané hodnoty byly porovnávány u maskovaných a nemaskovaných projevů jednotlivých mluvčích.

2.3.6. LOKÁLNÍ TEMPORÁLNÍ UKAZATELE

V nahrávkách byly sledovány změny trvání slabik v průběhu nádechových úseků. Pro jednotlivé slabiky bylo vypočteno trvání, a tyto hodnoty byly seřazeny podle pořadí slabik. Z analýzy bylo vyloučeno trvání posledních dvou slabik v nádechovém úseku, které ve většině případů podléhají závěrovému zpomalování, a pravděpodobně by tudíž nebyly pro analýzu přínosem.

Pro takto získané průběhy byly nalezeny Legendreovy polynomy, jejichž pomocí lze charakterizovat trendy ve změnách trvání po sobě jdoucích slabik během nádechových úseků.

Metodou k-means byly nalezeny čtyři typické kontury, a ty byly přiřazeny jednotlivým nádechovým úsekům. Následně bylo užití těchto kontur porovnáno u jednotlivých mluvčích a jejich maskovaných a nemaskovaných projevů.

3. VÝSLEDKY

3.1. POSLECHOVÁ ANALÝZA

3.1.1. NEMASKOVANÝ PROJEV

Charakteristiky maskovaných i nemaskovaných projevů mluvčích obsahuje Příloha č. 2.

Co se týče nemaskovaných projevů, čtyři mluvčí hovořili spíše vyšším hlasem, čtyři spíše hlubším hlasem a výška hlasu dvou zbývajících mluvčích zněla průměrně. Na percipované výšce hlasu se podílejí dva akustické koreláty, a to základní frekvence (která vypovídá o velikosti hlasivek) a vokální formanty (ty souvisejí s velikostí vokálního traktu, a tedy také s celkovým fyzickým vzrůstem jedince).

Melodie řeči se u pěti mluvčích jevila jako výrazně variabilní, což pravděpodobně souviselo se skutečností, že se jednalo o čtený text, při jehož přednesu mohli mluvčí mít tendenci k určité stylizaci; je otázkou, nakolik by byla melodie variabilní, pokud by byl nahráván spontánní projev.

U většiny mluvčích nebylo zaznamenáno nestandardní tempo, pouze dva mluvčí hovořili relativně rychle a tempo jednoho mluvčího se jevila jako značně nízké.

U tří mluvčích se objevila mírně nestandardní nastavení vokálního traktu: u jednoho mluvčího byl zaznamenán posun artikulace mírně dopředu, jeden mluvčí vykazoval mírnou nazalizaci a projev jednoho mluvčího působil dojmem artikulace bez pohybu spodní čelisti.

3.1.2. MASKOVANÝ PROJEV

Většina mluvčích přistoupila k modifikaci výšky hlasu: pět mluvčích se rozhodlo položit svůj hlas výše, dva mluvčí naopak níže. U ostatních tří mluvčích zůstal dojem výšky hlasu bez výrazných změn. Pro zvýšení hlasu se rozhodli dva ze čtyř mluvčích, kteří i při nemaskovaném projevu hovořili poměrně vysokým hlasem, naproti tomu pouze jeden ze čtyř mluvčích, jejichž přirozený hlas byl spíše hlubší. Snížení hlasu pak zvolil jeden mluvčí s vyšším přirozeným hlasem a

naopak jeden s hlubším. Oba mluvčí, u kterých byla výška nemaskovaného hlasu více méně průměrná, přistoupili k jejímu zvýšení.

U většiny mluvčích nebyly pozorovány změny ve variabilitě melodie řeči; pouze u dvou mluvčích byla znatelně větší než při maskovaném projevu, přičemž v jednom případě se jednalo o mluvčího, jehož melodie při nemaskovaném projevu byla hodnocena jako relativně variabilní, a ve druhém případě šlo naopak o mluvčího, jehož přirozená melodie byla vyhodnocena jako relativně plochá. Naproti tomu melodie maskovaného projevu dvou mluvčích byla výrazně plošší než u jejich projevu nemaskovaného, i v tomto případě se k této strategii uchýlil jeden mluvčí s relativně plochou a jeden s relativně variabilní melodií při nemaskovaném hovoru.

U většiny mluvčích bylo zachováno také tempo bez znatelných změn. Ke zvýšení tempa se uchýlili dva mluvčí, jejichž tempo při nemaskovaném projevu bylo vnímáno jako průměrné, naopak ke zpomalení přistoupil jeden mluvčí, jehož nemaskované tempo bylo již vnímáno jako relativně nízké, a jeden mluvčí, jehož nemaskované tempo bylo vyhodnoceno jako relativně vysoké – v jeho případě se při maskování jednalo zejména o výrazné pauzy, spíše než o zpomalení artikulace.

U sedmi mluvčích došlo k nápadným změnám melodických průběhů.

Čtyři mluvčí přistoupili ke změně způsobu fonace, a to na fonaci tlačenu.

Byly zaznamenány také artikulační modifikace: snížení hrtanu (u 2 mluvčích), zvýšení hrtanu (2), posun kořene jazyka nahoru (2), posun kořene jazyka dozadu (3), nazalizace (3) a hovor se sevřenými čelistmi (1).

3.2. AKUSTICKÁ ANALÝZA

3.2.1. ZÁKLADNÍ FREKVENCE

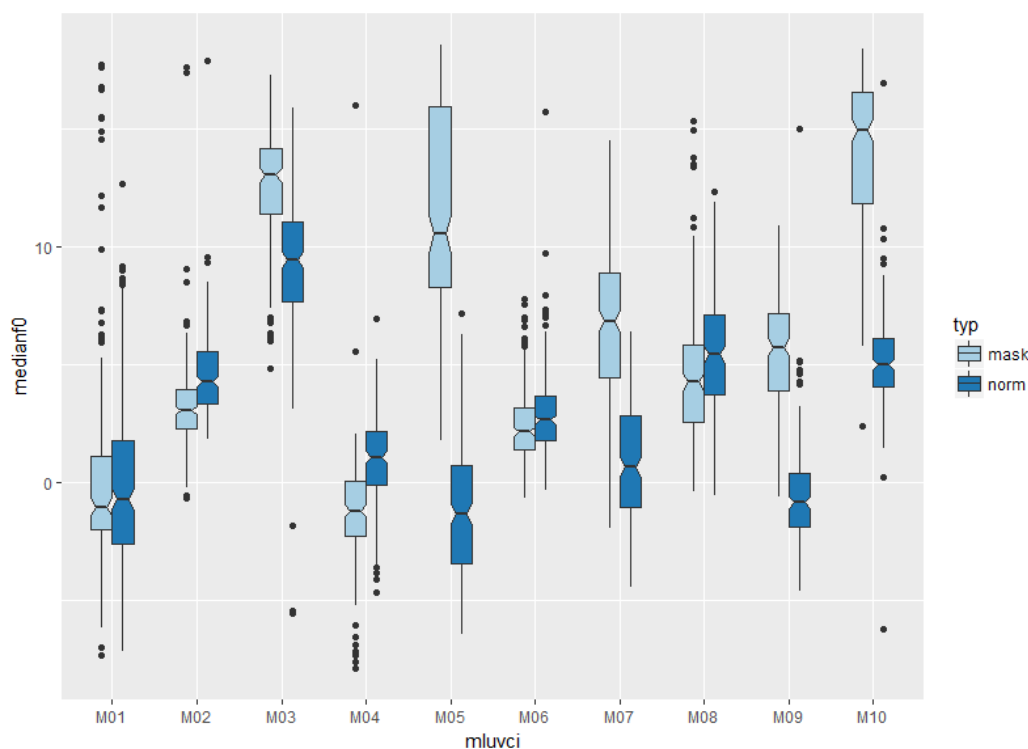
Medián

U šesti mluvčích došlo k signifikantní změně střední hodnoty základní frekvence, jak znázorňuje obrázek 3.2.1-1. Jak již bylo zmíněno, jedná se o parametr, který při maskování hlasu podléhá změnám velmi často, neboť taková modifikace je poměrně snadná a zároveň percepčně výrazná.

Nadprůměrné nemaskované hodnoty f_0 korelují s percepčním dojmem vyššího hlasu mluvčích (M02, M03, M08 a M10).

U hodnot pohybujících se kolem průměru a nižších není tento vztah vždy platný; např. hlas mluvčího M06 byl vnímán jako hlubší, zatímco výška hlasu mluvčího M05 byla vyhodnocena jako spíše průměrná, ačkoliv hodnota f_0 je u mluvčího M05 signifikantně nižší, než je tomu v případě mluvčího M06. Tento percepční dojem mohl být způsoben hodnotami vokalických formantů.

U mluvčích, u nichž byl zaznamenán percepční dojem zvýšení hlasu při maskování, došlo k signifikantnímu nárůstu f_0 (jednalo se o mluvčí M03, M05, M07, M09 a M10). Také mluvčí, kteří při maskování hovořili hlouběji znějícím hlasem (M02 a M06), vykázali snížení naměřené f_0 .



Obrázek 3.2.1-1: Medián f_0 [ST]; mask – maskovaný projev (světle modrá pole), norm – normální, tj. nemaskovaný projev (tmavě modrá pole).

V případě dvou ze tří mluvčích, u nichž při poslechu nebyla zaznamenána výrazná změna výšky hlasu (M04 a M08), však bylo zjištěno snížení f_0 ; výsledky akustické analýzy tudíž nekorelují se závěry analýzy poslechové. Tato skutečnost je zřejmě zapříčiněna vlivem změny fonace u těchto mluvčích: oba přistoupili k fonaci tlačené, při které se zároveň objevovalo neperiodické kmitání hlasivek.

Hrtan tedy nebyl snížen, což by vyvolalo dojem hlubšího hlasu, a došlo pouze ke zpomalení, resp. občasnému narušení periodicity kmitání hlasivek, což se promítlo do výsledků akustické analýzy.

Významnost vlivu maskování f_0 na projev jedince byla testována za využití lineárního smíšeného modelu, přičemž typ projevu (tj. maskovaný či nemaskovaný) figuroval jako fixní faktor, mluvčí byl stanoven jako faktor náhodný. Pro medián základní frekvence nebyl faktor mluvčího na hladině významnosti 5 % vyhodnocen jako statisticky signifikantní (p-hodnota byla vyčíslena na 0,052), a to navzdory skutečnosti, že u některých mluvčích byly zaznamenány změny mediánu f_0 v řádu celých půltónů; toto ilustruje tabulka 3.2.1-1, která byla získána v rámci využití lineárního smíšeného modelu.

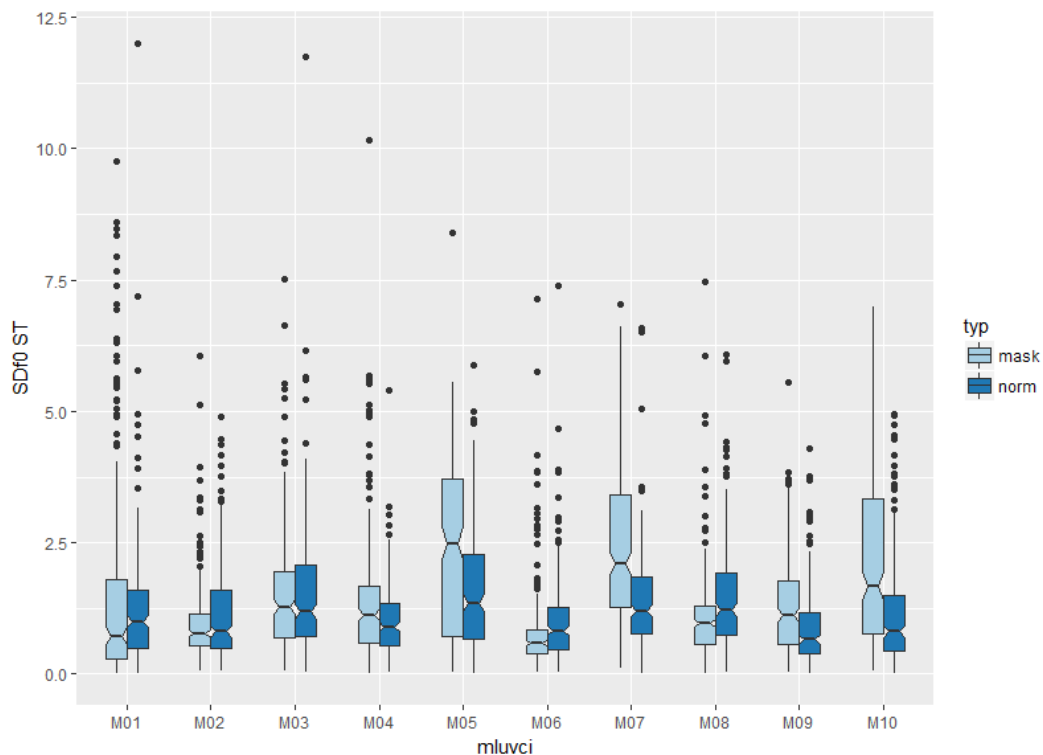
Na základě analyzovaných dat tedy nelze tvrdit, že mluvčí jsou při maskování schopni signifikantním způsobem měnit střední hodnotu své základní frekvence; tento výsledek je pravděpodobně zapříčiněn malým vzorkem mluvčích.

| | medián f_0 při nemaskovaném projevu | posun mediánu f_0 při maskovaném projevu |
|-----|---------------------------------------|--------------------------------------------|
| M01 | -0,18 | 0,45 |
| M02 | 4,59 | -1,30 |
| M03 | 9,25 | 3,25 |
| M04 | 0,90 | -2,12 |
| M05 | -1,14 | 13,03 |
| M06 | 2,93 | -0,47 |
| M07 | 0,97 | 5,41 |
| M08 | 5,39 | -0,78 |
| M09 | -0,55 | 6,04 |
| M10 | 5,10 | 8,82 |

Tabulka 3.2.1-1: Rozdíly mediánu f_0 mezi maskovaným a nemaskovaným projevem; v prvním sloupci posun oproti průměrné střední hodnotě f_0 všech mluvčích, v pravém sloupci posun mediánu f_0 oproti nemaskovanému projevu daného mluvčího.

Směrodatná odchylka

U poloviny mluvčích přesáhl medián směrodatné odchylky základní frekvence (SD f_0) nemaskovaného projevu 1 ST; u těchto mluvčích byla také melodie řeči vnímána jako variabilní, zatímco u ostatních, u kterých se směrodatná odchylka naopak pohybovala pod jedním půltónem, byla melodie nemaskovaného projevu hodnocena jako plošší. Hodnocení variability melodie tedy koreluje s naměřenou směrodatnou odchylkou základní frekvence, kterou znázorňuje obrázek 3.2.1-2.



Obrázek 3.2.1-2: Směrodatná odchylka f_0 [ST]; mask – maskovaný projev (světle modrá pole), norm – normální, tj. nemaskovaný projev (tmavě modrá pole).

Naproti tomu percepční hodnocení změn variability melodie se s naměřenými hodnotami směrodatné odchylky u maskovaných projevů rozcházejí. Jako plošší oproti nemaskovanému projevu byla melodie hodnocena u dvou mluvčích (M04 a M05), ovšem v obou těchto případech došlo naopak ke statisticky signifikantnímu nárůstu SD f_0 ; percepční dojem menší variability byl pravděpodobně vyvolán skutečností, že oba mluvčí sice na začátku a na konci nádechových úseků hlasem stoupali či klesali, nicméně v prostředních částech nádechových úseků melodie

jejich hlasu poměrně dlouho setrvala na jedné úrovni. Toto platí zejména u mluvčího M05, u jehož maskovaného projevu byla směrodatná odchylka f_0 vyčíslena jako výrazně vyšší než u jeho projevu nemaskovaného, avšak percepční dojem ploché melodie uvnitř nádechových úseků byl poměrně výrazný.

Jako variabilnější než v případě nahrávek nemaskovaného hlasu byla vnímána melodie maskovaných projevů mluvčích M07 a M10. V obou případech byl zaznamenán také statisticky signifikantní nárůst směrodatné odchylky f_0 .

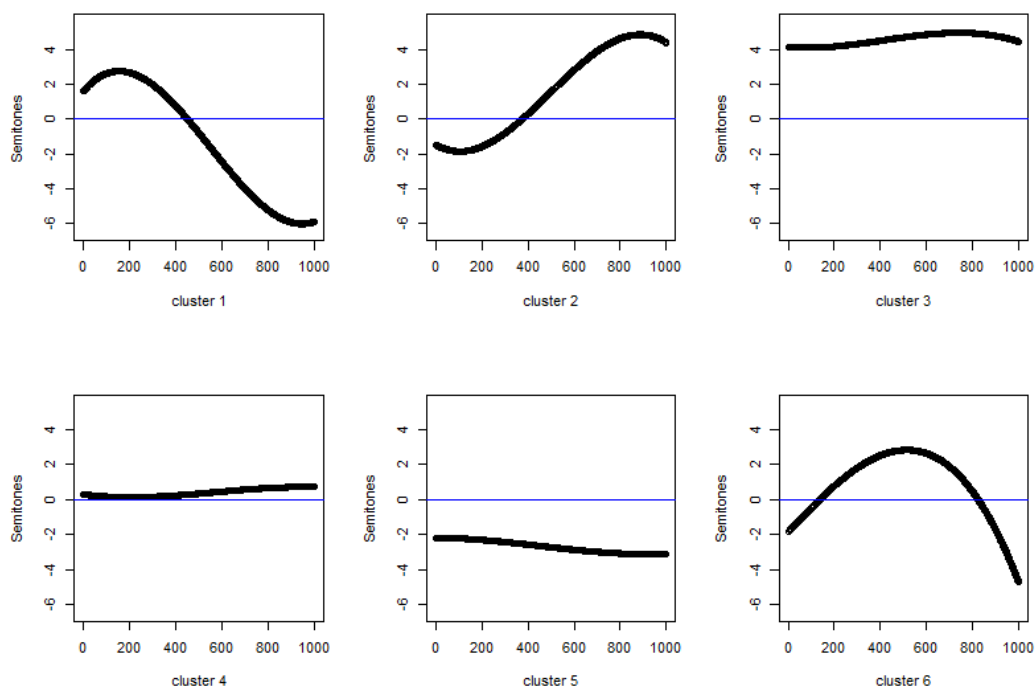
U ostatních šesti mluvčích nebyla výrazná změna variability melodie při poslechové analýze zachycena, avšak ke statisticky významné změně směrodatné odchylky f_0 skutečně nedošlo pouze u dvou z nich (M02 a M03). Tři zbývající mluvčí vykázali snížení a jeden naopak zvýšení SD f_0 .

Také SD f_0 u maskovaných projevů mluvčích zachycuje obrázek 3.2.1-2.

Je třeba říci, že při testování statistické významnosti změn směrodatné odchylky při maskování hlasu nebyl, podobně jako v případě mediánu f_0 , faktor maskování vyhodnocen jako statisticky signifikantní na hladině významnosti 5 % ($p=0,058$). I tento výsledek může být zapříčiněn nedostatečným množstvím dat; na základě této analýzy nelze usuzovat, že směrodatná odchylka či medián f_0 zůstávají při maskování hlasu napříč mluvčími stabilní. S využitím dostupného vzorku pouze nelze potvrdit hypotézu, že maskování hlasu je pro změny střední hodnoty f_0 a její směrodatné odchylky statisticky významným faktorem.

Kontury f0

Pomocí shlukové analýzy k-means bylo nalezeno šest typických průběhů kontur základní frekvence: klesavá (cluster 1; C1), stoupavá (C2), stoupavě-klesavá (C6), plochá pohybující se kolem střední hodnoty f0 (C4), plochá na vyšší úrovni (C3) a plochá na nižší úrovni (C5). Jednotlivé typické kontury předkládá obrázek 3.2.1-3.



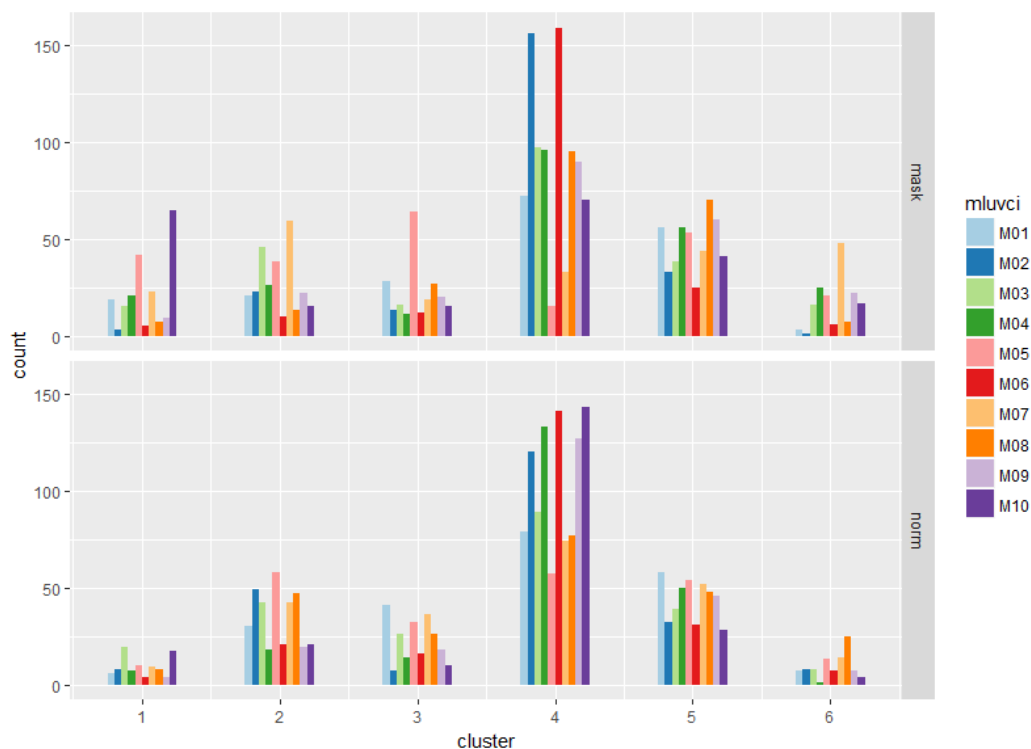
Obrázek 3.2.1-3: Typické intonační kontury nalezené ve všech projevech mluvčích po provedení normalizace f0, tj. vystředění melodických průběhů nahrávek kolem 0.

U jednotlivých mluvčích bylo srovnáno využití těchto typických kontur, a to jak celkově, tak v rámci různých melodémů.

Souhrn užití kontur napříč všemi typy melodémů představuje obrázek 3.2.1-4. V maskovaných i nemaskovaných projevech se nejčastěji vyskytuje C4, tedy plochá kontura pohybující se na úrovni střední hodnoty základní frekvence.

Obecně lze tvrdit, že četnost užití jednotlivých clusterů je mezi mluvčími podobná v případě nemaskovaných projevů; naproti tomu u projevů maskovaných četnosti užití jednotlivých kontur mezi mluvčími více kolísají. Toto lze pozorovat zejména u C1 (klesající kontura), jejíž četnost užití u dvou mluvčích výrazně

vzrostla, nebo u C4 – užívání této kontury bylo naopak čtyřmi mluvčími znatelně redukováno.



Obrázek 3.2.1-4: Četnost typických intonačních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích.

Pro testování statistické významnosti rozdílů v užití typických kontur při maskovaném a nemaskovaném projevu jednotlivých mluvčích byl využit Pearsonův chí kvadrát, neboli test nezávislosti, a byla provedena Bonferroniho korekce hladiny významnosti, tj. její vydělení počtem mluvčích – hladina významnosti byla tudíž stanovena na 0,5 %. Jako statisticky signifikantní se ukázaly změny ve využití jednotlivých clusterů u sedmi mluvčích (M02 $p=0,0003$, M04 $p=6 \cdot 10^{-7}$, M05 $p=8,5 \cdot 10^{-12}$, M07 $p=2 \cdot 10^{-9}$, M08 $p=1,4 \cdot 10^{-6}$, M09 $p=0,003$ a M10 $p=8,4 \cdot 10^{-13}$). Někteří mluvčí jsou tedy zřejmě schopni při maskování hlasu významně měnit vzorce užívání melodických kontur.

Zaměříme se nyní na výskyt jednotlivých clusterů v různých melodémech.

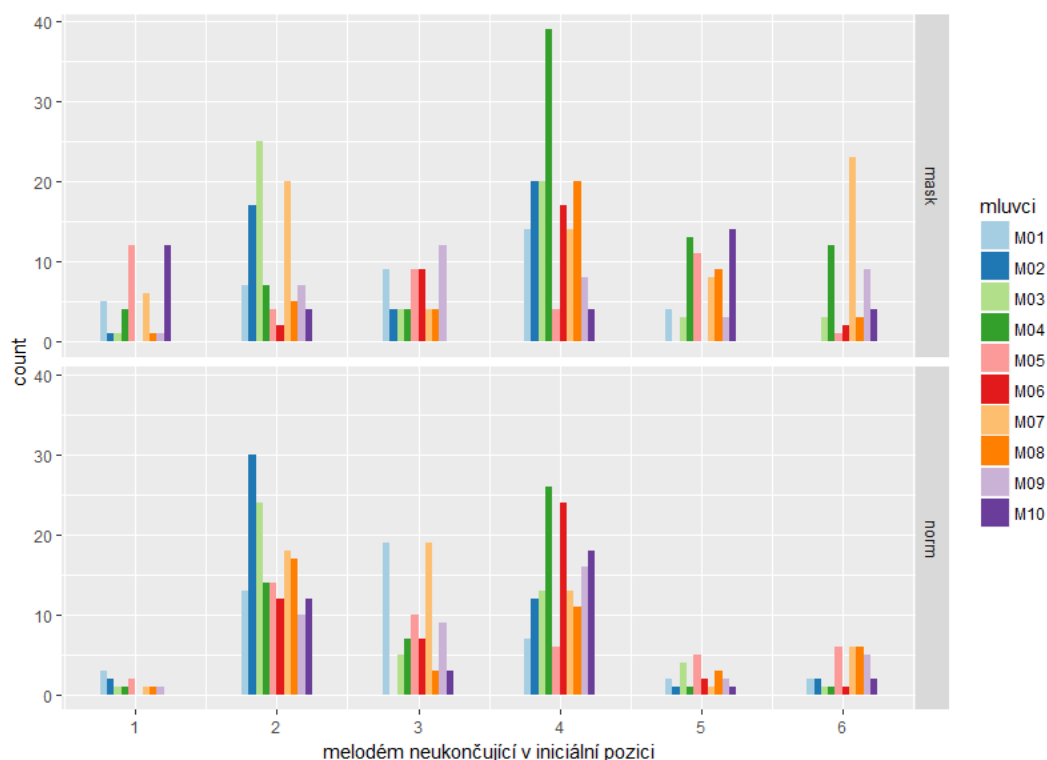
V případě melodému neukončujícího v iniciální pozici byly v nemaskované řeči nejčastěji využívány kontury C2 a C4, tedy kontury stoupající a ploché pohybující

se kolem střední hodnoty f_0 . Naproti tomu kontury C1 (klesavá), C5 (plochá pod úrovní střední hodnoty f_0) a C6 (stoupavě-klesavá) byly využívány spíše zřídka.

Při maskovaných projevech byl však u pěti mluvčích (M04, M05, M07, M08 a M10) zaznamenán nárůst četnosti užití právě těchto tří, pro melodém neukončující v iniciální pozici poněkud neočekávaných, melodických průběhů. Tato skutečnost nasvědčuje výše zmíněné hypotéze, že užití kontur odchylojící se od standardu může nasvědčovat záměrné modifikaci hlasu.

Užití jednotlivých clusterů u melodému neukončujícího v iniciální pozici zobrazuje obrázek 3.2.1-5.

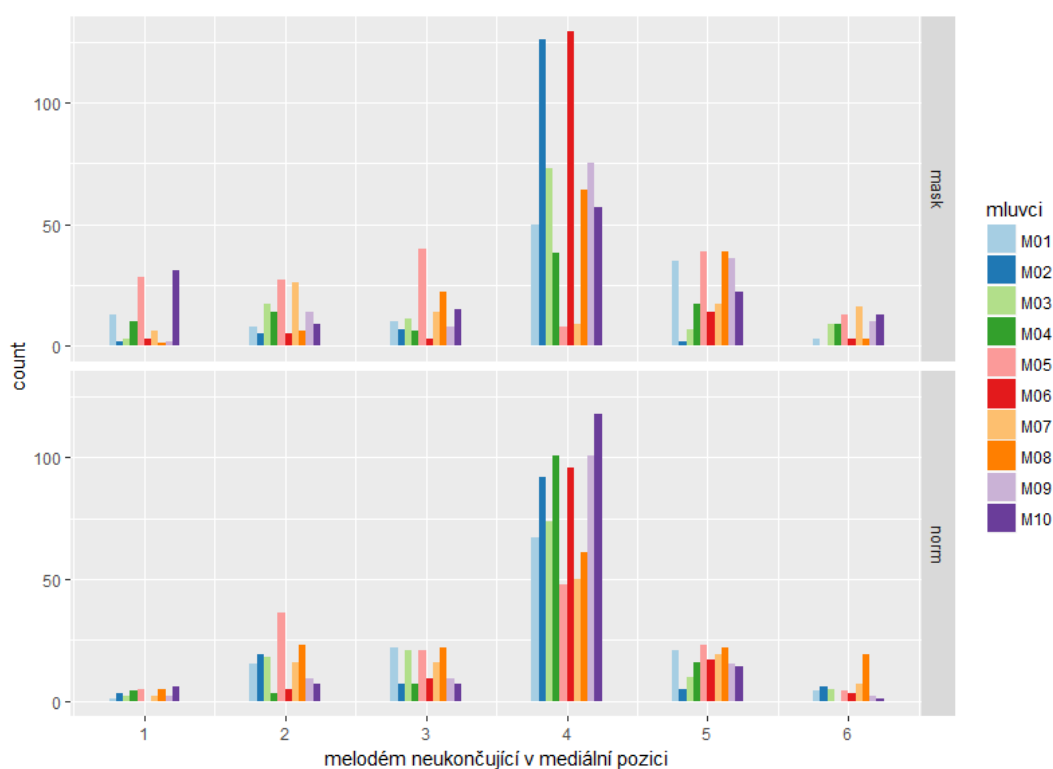
Co se týče statistické významnosti změn využití jednotlivých clusterů u mluvčích, v rámci realizací melodému neukončujícího v iniciální pozici byly změny na hladině významnosti 0,5 % vyhodnoceny jako signifikantní u čtyř mluvčích (M04 $p=0,0006$, M05 $p=0,002$, M07 $p=5,3 \cdot 10^{-5}$ a M10 $p=1,6 \cdot 10^{-7}$).



Obrázek 3.2.1-5: Četnost typických intonačních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích při realizaci melodému neukončujícího v iniciální pozici.

U melodému neukončujícího v mediální pozici byl při nemaskovaném projevu všemi mluvčími využíván cluster C4 (plochá kontura pohybující se kolem střední hodnoty f_0), jak je patrné z obrázku 3.2.1-6. V jeho případě lze u maskovaných projevů pozorovat větší rozkolísanost než u projevů nemaskovaných; u většiny mluvčích četnost jeho užití výrazně klesla. Naproti tomu u tří mluvčích byl zaznamenán nárůst užití clusteru C1 (klesavá kontura), který byl v nemaskovaných projevech využíván pouze okrajově. Větší rozkolísanost četnosti užití jednotlivými mluvčími při maskovaném projevu lze pozorovat také u clusteru C5 (plochá kontura pod úrovní střední hodnoty f_0).

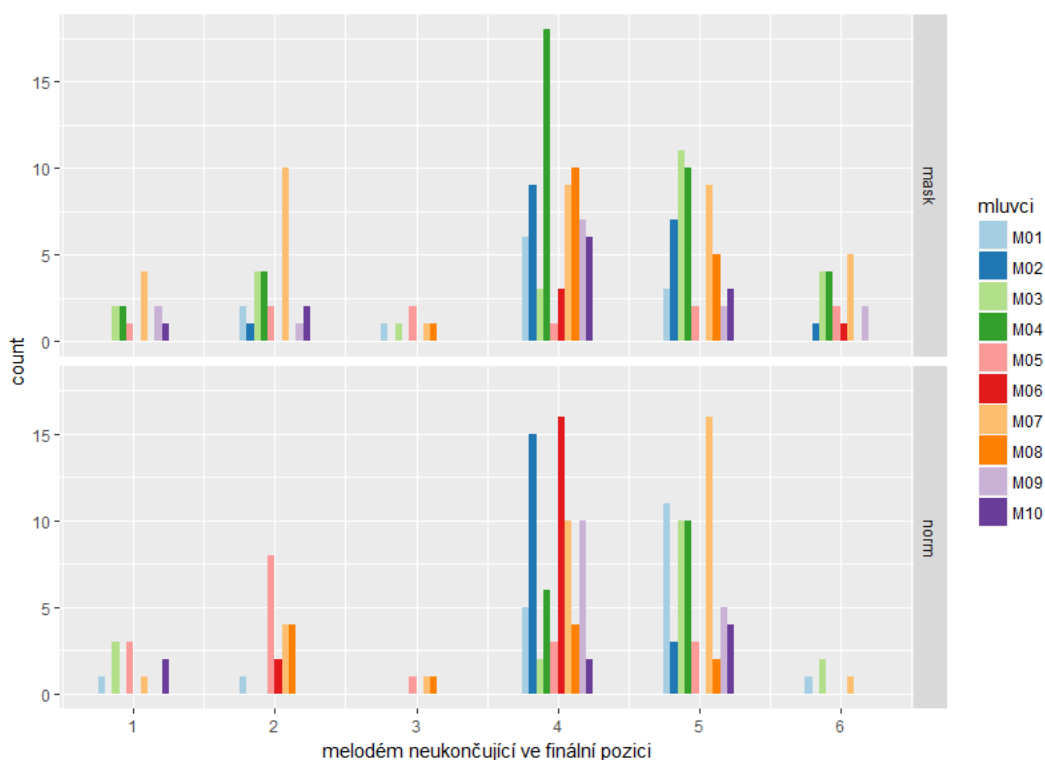
U tohoto typu melodému došlo ke statisticky signifikantním změnám četnosti využití pozorovaných clusterů dokonce u osmi mluvčích (M01 $p=0,0004$, M02 $p=0,001$, M04 $p=4,9 \cdot 10^{-8}$, M05 $p=1,3 \cdot 10^{-11}$, M07 $p=1,8 \cdot 10^{-6}$, M08 $p=3,4 \cdot 10^{-5}$, M09 $p=0,002$ a M10 $p=3 \cdot 10^{-10}$).



Obrázek 3.2.1-6: Četnost typických intonačních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích při realizaci melodému neukončujícího v mediální pozici.

Při realizaci melodému neukončujícího ve finální pozici se nejčastěji uplatňovaly clustery C4 a C5, a to jak v nemaskovaných, tak v maskovaných projevech. Užití jednotlivých melodémů se však napříč mluvčími výrazně liší v obou typech projevů. Nadto byl výskyt tohoto typu melodému v nahrávkách poměrně zřídka, a není tedy možné na základě jeho realizací formulovat jednoznačné závěry. Při využití testu nezávislosti nebyly změny užívání jednotlivých kontur vyhodnoceny jako statisticky významné u žádného mluvčího.

Četnost využívání jednotlivých clusterů při realizaci melodému neukončujícího ve finální pozici znázorňuje obrázek 3.2.1-7.



Obrázek 3.2.1-7: Četnost typických intonačních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích při realizaci melodému neukončujícího ve finální pozici.

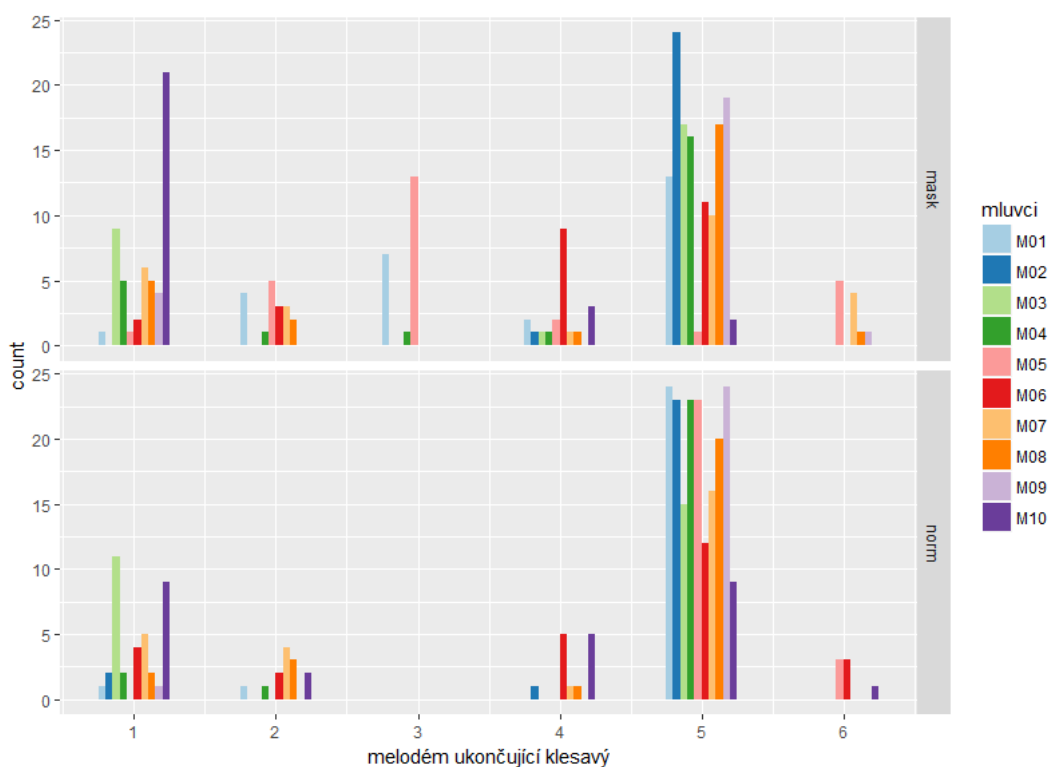
Při realizaci melodému ukončujícího klesavého se u většiny mluvčích v maskovaných i nemaskovaných projevech nejčastěji objevoval cluster C5, tedy plochá melodie pod úrovní střední hodnoty základní frekvence.

U nemaskovaných projevů se naproti tomu nevyskytoval cluster C3 (plochá melodie nad úrovní střední hodnoty základní frekvence), k jehož využívání se však uchýlili dva mluvčí při maskování hlasu. Naopak využívání clusteru C5 při maskování hlasu u pěti mluvčích výrazně pokleslo.

Test nezávislosti odhalil statistickou významnost změn užívání jednotlivých clusterů při realizaci melodému ukončujícího klesavého pouze u jediného mluvčího (M05 $p=6,9 \cdot 10^{-8}$); je možné, že je tento výsledek zapříčiněn malým množstvím analyzovaných dat.

Realizace melodému ukončujícího klesavého znázorňuje obrázek 3.2.1-8.

Obecně lze říci, že existují typické kontury, jež mluvčí využívají v jednotlivých typech melodémů při nemaskovaném projevu, a také to, že při maskování hlasu mají mluvčí tendenci se častěji uchýlovat k nestandardním konturám f0.



Obrázek 3.2.1-8: Četnost typických intonačních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích při realizaci melodému ukončujícího klesavého.

3.2.2. VOKALICKÉ FORMANTY

Hodnoty vokalických formantů byly vyčísleny nejprve pro jednotlivé vokály. Výsledky těchto měření zobrazují grafy obsažené v Příloze č. 3. Byly vytvořeny houslové grafy a také krabicové grafy s vykrojením znázorňujícím 95% konfidenční interval pozice mediánu; pokud se vykrojení dvou grafu nepřekrývají, lze tvrdit, že tyto dva mediány jsou signifikantně odlišné, a to na hladině významnosti 5 % (Chambers et al., 1983, p. 62, in: Wickham, 2016). U jednotlivých formantů těch vokálů, u nichž nebyl k dispozici dostatek dat k tomu, aby bylo možné vytvořit reprezentativní krabicové grafy, které by adekvátně zobrazily rozložení hodnot, byly předloženy pouze houslové grafy.

Vokály byly rozřazeny do skupin podle místa artikulace; hlásky [a] a [a:] byly analyzovány dohromady, neboť jejich místo artikulace se neliší (Skarnitzl et al., 2016), stejně je tomu také u [e] a [e:], které byly taktéž zařazeny do jedné kategorie. Vokál [o] byl analyzován samostatně bez svého fonologicky dlouhého protějšku, neboť [o:] se v analyzovaném textu neobjevilo. Naproti tomu zavřené vokály, tj. [ɪ], [i:], [u] a [u:], byly analyzovány samostatně, neboť jejich místa artikulace – je-li jejich podoba ortoepická – se v češtině signifikantně liší (Skarnitzl et al., 2016).

Jako popsané hlásky jsou pro účely analýzy považovány všechny reprezentace daných fonémů, ačkoliv se jejich fonetická podoba v jednotlivých případech může odchylovat od podoby ortoepické (např. v důsledku fonetických redukcí).

F1

Směr posunů prvního formantu v daných vokálech u jednotlivých mluvčích znázorňuje tabulka 3.2.2-1.

U sedmi mluvčích došlo u vokálů [a] a [a:] ke zvýšení prvního formantu při maskovaném projevu oproti projevu nemaskovanému, u ostatních tří naopak k jeho snížení. Vyšší F1 je akustickým korelátem větší otevřenosti vokálu, výsledek analýzy těchto hlásek lze tudíž interpretovat dvojím způsobem: (1) obecně větší otevřenost vokálů či (2) pečlivější artikulace, tedy decentralizace vokálů.

| vokál | zvýšení F1 | bez signifikantní změny | snížení F1 |
|--------|------------|-------------------------|------------|
| [i:] | M01 | M04 | M02 |
| | M03 | M06 | |
| | M05 | M07 | |
| | M08 | | |
| | M09 | | |
| | M10 | | |
| [ɪ] | M01 | M04 | M02 |
| | M03 | | M06 |
| | M05 | | M08 |
| | M07 | | |
| | M09 | | |
| | M10 | | |
| [e e:] | M01 | | M02 |
| | M03 | | M06 |
| | M04 | | M08 |
| | M05 | | |
| | M07 | | |
| | M09 | | |
| [a a:] | M01 | | M02 |
| | M03 | | M06 |
| | M04 | | M08 |
| | M05 | | |
| | M07 | | |
| | M09 | | |
| [o] | M01 | M06 | M02 |
| | M03 | M07 | M08 |
| | M04 | | |
| | M05 | | |
| | M09 | | |
| | M10 | | |

Tabulka 3.2.2-1: posuny F1 při maskovaném projevu oproti projevu nemaskovanému u jednotlivých mluvčích

U polootevřených vokálů, tedy [e], [e:] a [o] (hláska [o:] byla vyřazena z analýzy pro nedostatek dat), docházelo většinou opět k nárůstu F1, přičemž se jednalo o tytéž mluvčí jako v případě [a a:], s výjimkou jednoho mluvčího, u kterého nebyl zaznamenán signifikantní posun F1 u vokálu [o].

Také u předních zavřených vokálů [ɪ] a [i:] byly zaznamenány obdobné posuny.

V případě vokálů [u] a [u:] došlo ke zvýšení F1 u dvou mluvčích, v ostatních případech pro nedostatek dat nelze s určitostí říci, zda F1 prodělal zaznamatelný posun.

Tyto výsledky naznačují, že většina mluvčích se při maskování hlasu uchýlila k větší otevřenosti vokálů, nikoliv k jejich pečlivější artikulaci, a že jsou mluvčí obecně schopni signifikantně měnit hodnotu prvního formantu.

Při provedení statistického modelu pomocí lineárního smíšeného modelu za aplikace Bonferroniho korekce, kdy byla hladina významnosti 5 % vydělena počtem analyzovaných vokalických kvalit, tj. sedmi, a tudíž stanovena na 0,7 %, nebyl faktor maskování hlasu při změnách F1 vyhodnocen jako statisticky signifikantní u žádného vokálu. Pro získání statisticky významnějších výsledků by zřejmě byl zapotřebí větší objem dat, resp. větší vzorek mluvčích.

F2

V případě druhého formantu také povětšinou docházelo ke změnám při projevech maskovaných ve srovnání s těmi nemaskovanými, avšak nikoli v tolika případech jako v případě formantu prvního. Byly zaznamenány posuny oběma směry, tedy zvýšení i snížení (na rozdíl od formantu prvního, u něž byl v naprosté většině zaznamenán posun k vyšším hodnotám). Posuny mediánu F2 jednotlivých vokálů u mluvčích předkládá tabulka 3.2.2-2.

U vokálu, jehož hodnota F2 bývá v češtině nejvyšší, tedy u [i:], bylo u tří mluvčích zaznamenáno zvýšení F2, u tří naopak jeho snížení, tj. horizontální centralizace místa artikulace.

U vokálu [ɪ], který bývá v češtině obecně zadnější než jeho fonologicky dlouhý protějšek [i:], došlo ke zvýšení F2 u dvou mluvčích, k jeho snížení pak u tří mluvčích. K poklesu F2 došlo u tohoto vokálu u stejných mluvčích jako u vokálu předchozího. Zvýšení F2 bylo pozorováno u jediného mluvčího, u něž bylo zaznamenáno také při artikulaci vokálu [i:]; z této skutečnosti však nemusí vyplývat, že posuny artikulace [i:] a [ɪ] se u ostatních mluvčích navzájem liší; je možné, že pro vyhodnocení hodnot F2 těchto vokálů a jejich posouzení jako statisticky významně odlišných mezi maskovanými a nemaskovanými projevy pouze není k dispozici dostatek dat.

U vokálů [e e:] došlo u pěti mluvčích ke zvýšení F2, tedy k posunu artikulace dopředu. Naproti tomu u tří mluvčích hodnoty druhého formantu klesly. Ke zvýšení i snížení hodnot F2 došlo mimo jiné u týchž mluvčích jako v případě vokálu [ɪ].

Artikulace vokálů [a a:] se u šesti mluvčích posunula dopředu, tj. došlo ke zvýšení F2; jednalo se o tytéž mluvčí, kteří vykazali zvýšení F2 také u předcházejících vokálu, a jednoho, u něž v předchozích vokálech nedošlo k signifikantní změně F2. U jediného mluvčího byl F2 naopak snížen, a jednalo se o jednoho z mluvčích, u nichž bylo snížení F2 pozorováno také u všech předchozích vokálů.

U vokálu [o] u sedmi mluvčích nedošlo k signifikantnímu posunu F2, u ostatních tří byl zaznamenán jeho posun nahoru.

Signifikantní posuny F2 u vokálů [u] a [u:] nebylo možné pozorovat kvůli nedostatku dat.

| vokál | zvýšení F2 | bez signifikantní změny | snížení F2 |
|---------------|------------|-------------------------|------------|
| [i:] | M01 | M04 | M02 |
| | M07 | M05 | M03 |
| | M08 | M09 | M06 |
| | | M10 | |
| [ɪ] | M04 | M01 | M02 |
| | M07 | M05 | M03 |
| | | M08 | M06 |
| | | M09 | |
| | | M10 | |
| [e e:] | M01 | M08 | M02 |
| | M04 | M10 | M03 |
| | M05 | | M06 |
| | M07 | | |
| | M09 | | |
| [a a:] | M01 | M03 | M02 |
| | M04 | M06 | |
| | M05 | M08 | |
| | M07 | | |
| | M09 | | |
| | M10 | | |
| [o] | M01 | M02 | |
| | M04 | M03 | |
| | M09 | M05 | |
| | | M06 | |
| | | M07 | |
| | | M08 | |
| | | M10 | |

Tabulka 3.2.2-2: Posuny F2 při maskovaném projevu oproti projevu nemaskovanému u jednotlivých mluvčích.

Obecně je možné říci, že co se týče posunů F2, tedy změn horizontální polohy jazyka, volili mluvčí různé strategie. U tří mluvčích nebyly u žádného vokálu pozorovány signifikantní změny F2, nedošlo tedy k horizontálnímu posunu jejich artikulace. U dvou mluvčích bylo pozorováno zvýšení F2 u všech vokálů, tedy obecně artikulace více vpředu. Naopak snížení F2 u všech vokálů došlo u jednoho mluvčího.

U čtyř mluvčích pak byla pozorována konvergence F2 směrem k zadním či předním vokálům; u dvou mluvčích nebyl pozorován posun F2 u [i:], bylo však pozorováno jeho zvýšení u ostatních vokálů, lze tedy předpokládat, že tito mluvčí neartikulovali při maskování více vpředu ten vokál, který tvořili nejvíce vpředu i při běžném hovoru, zatímco artikulaci ostatních vokálů posunuli jeho směrem. Naopak u jiných dvou mluvčích nebyl pozorován posun F2 u vokálů [a a:] a [o], zatímco u [ɪ] a [i:], tedy předních vokálů, došlo k jeho snížení, a tedy posunu artikulace více dozadu.

Stejně jako v případě prvního formantu, ani u F2 ale nebyl za využití lineárního smíšeného modelu faktor maskování shledán statisticky významným u žádné z analyzovaných vokalických kvalit.

Na základě analyzovaných se dat lze domnívat, že jsou mluvčí při maskování hlasu schopni výrazně měnit tento parametr, neexistuje však obecná tendence posouvat F2 konkrétním směrem, neboť pro posuny horizontální polohy jazyka při artikulaci vokálů volí odlišné strategie. Statistický test nicméně nepotvrdil, že by maskování hlasu představovalo signifikantní faktor při posunech F2; je možné, že příčinou je i v tomto případě nedostatek dat.

F3

Na rozdíl od prvních dvou formantů je F3 závislý na kvalitě vokálu jen částečně, lze tedy předpokládat, že jej mluvčí budou schopni při maskování hlasu modifikovat méně než F1 a F2. Výsledky analýzy maskovaných a nemaskovaných projevů tomuto předpokladu však příliš nenasvědčují, jak dokládá tabulka 3.2.2-3.

Ve většině případů nedošlo ke statisticky signifikantním posunům F3, nicméně ani počet případů, kdy k takovým posunům došlo, nebyl zanedbatelný. U některých mluvčích nedocházelo ke konzistentním posunům F3 napříč vokály; např. u mluvčího M03 se objevilo zvýšení F3 u vokálů [e e:] a [ɪ], ale naopak jeho snížení u vokálů [o] a [u].

| vokál | zvýšení F3 | bez signifikantní změny | snížení F3 |
|--------|---------------------------------|----------------------------------------|---------------------------------|
| [i:] | M04 M09 | M01 M03 M07 | M02 M05 M06 M08 M10 |
| [ɪ] | M03 M04 M07 | M01 M05 M06 M08 M09 M10 | M02 |
| [e e:] | M01 M03 M04 M07 M10 | M05 M06 M08 M09 | M02 |
| [a a:] | M01 M07 M08 | M02 M03 M04 M06 M10 | M05 M09 |
| [o] | M01 M04 M08 | M06 M07 M10 | M02 M03 M05 M09 |
| [u] | | M01 M04-09 | M02 M03 M10 |
| [u:] | | M01-10 | |

Tabulka 3.2.2-3: Posuny F3 při maskovaném projevu oproti projevu nemaskovanému u jednotlivých mluvčích.

Tento parametr se tedy neukázal jako stabilní navzdory maskování hlasu, a nevykázal ani tendenci se při maskování hlasu posouvat konkrétním směrem napříč mluvčími, resp. různými maskovacími strategiemi.

Při testování pomocí lineárního smíšeného modelu nebyl faktor maskování hlasu vyhodnocen jako statisticky signifikantní pro hodnoty F3 u žádné ze zkoumaných vokalických kvalit.

F4

Čtvrtý formant je považován za zcela závislý na mluvčím, nikoliv ale na nastavení jeho vokálního traktu, tudíž je možné předpokládat, že bude v rámci mluvčího stabilní a že nebude ovlivněn maskováním hlasu. Akustická analýza však svědčí o tom, že někteří mluvčí jsou schopni F4 ve svých vokálech signifikantně modifikovat.

Směry posunů F4 u jednotlivých vokálů znázorňuje tabulka 3.2.2-4.

U pěti mluvčích nedošlo ve většině vokálů ke statisticky významným změnám mediánu F4; dva mluvčí nevykázali žádné signifikantní posuny F4, u dalších tří mluvčích se takové změny objevily jen v menšině vokálů. U jednoho mluvčího (M09) však u šesti ze sedmi analyzovaných vokalických kvalit došlo k signifikantnímu zvýšení F4, u jednoho mluvčího (M02) byl naopak F4 ve všech vokalických kvalitách snížen; popsané posuny byly poměrně výrazné, nejednalo se tedy o zanedbatelné rozdíly.

U ostatních mluvčích nebyla podoba F4 u maskovaného projevu ve srovnání s maskovaným konzistentní. U mluvčího M05 se dokonce objevilo jak zvýšení F4 (u předních vokálů), tak i jeho snížení (u vokálu [u:]).

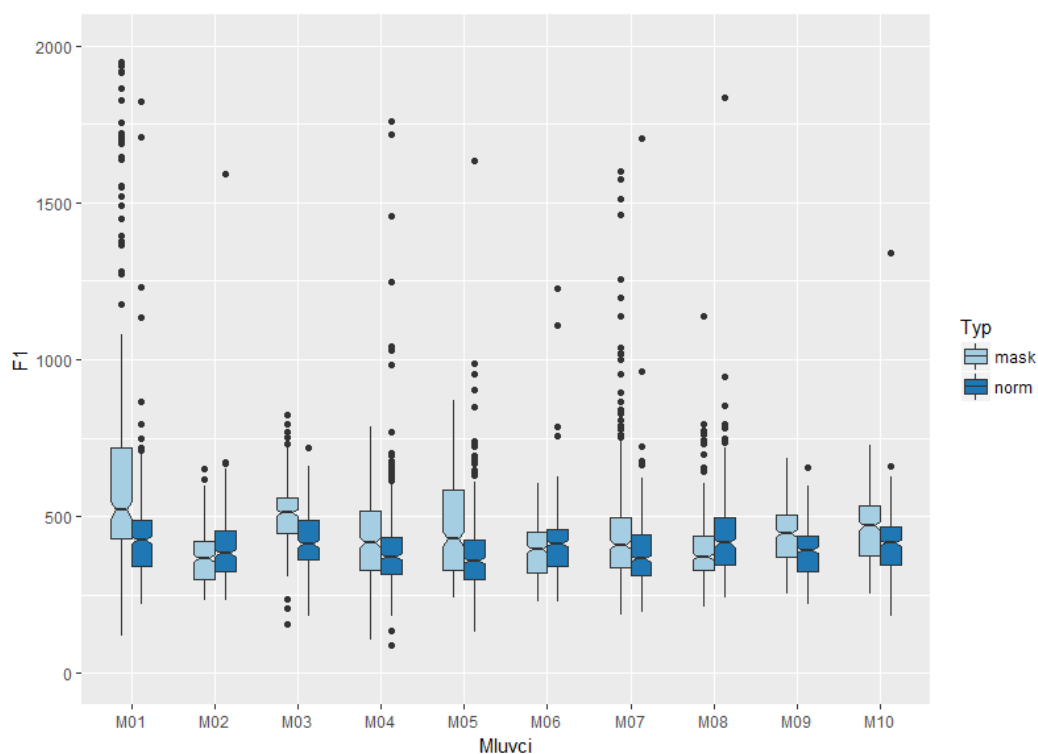
Lineární smíšený model nevyhodnotil faktor maskování hlasu jako statisticky signifikantní u žádného z analyzovaných vokálů.

| vokál | zvýšení F4 | bez signifikantní změny | snížení F4 |
|--------|--------------------------|---------------------------------|-------------------|
| [i:] | M01 M05 M09 | M03 M04 M07 M08 M10 | M02 M06 |
| [ɪ] | M01 M05 M07 M09 | M03 M04 M08 M10 | M02 M06 |
| [e e:] | M05 M07 M09 M10 | M01 M03 M04 M08 | M02 M06 |
| [a a:] | M07 M09 | M01 M04 M05 M08 M10 | M02 M03 M06 |
| [o] | M01 M09 | M04 M05-08 M10 | M02 M03 |
| [u] | M09 | M01 M04-08 M10 | M02 M03 |
| [u:] | | M01 M03 M04 M06-M10 | M02 M05 |

Tabulka 3.2.2-4: Posuny F4 při maskovaném projevu oproti projevu nemaskovanému u jednotlivých mluvčích.

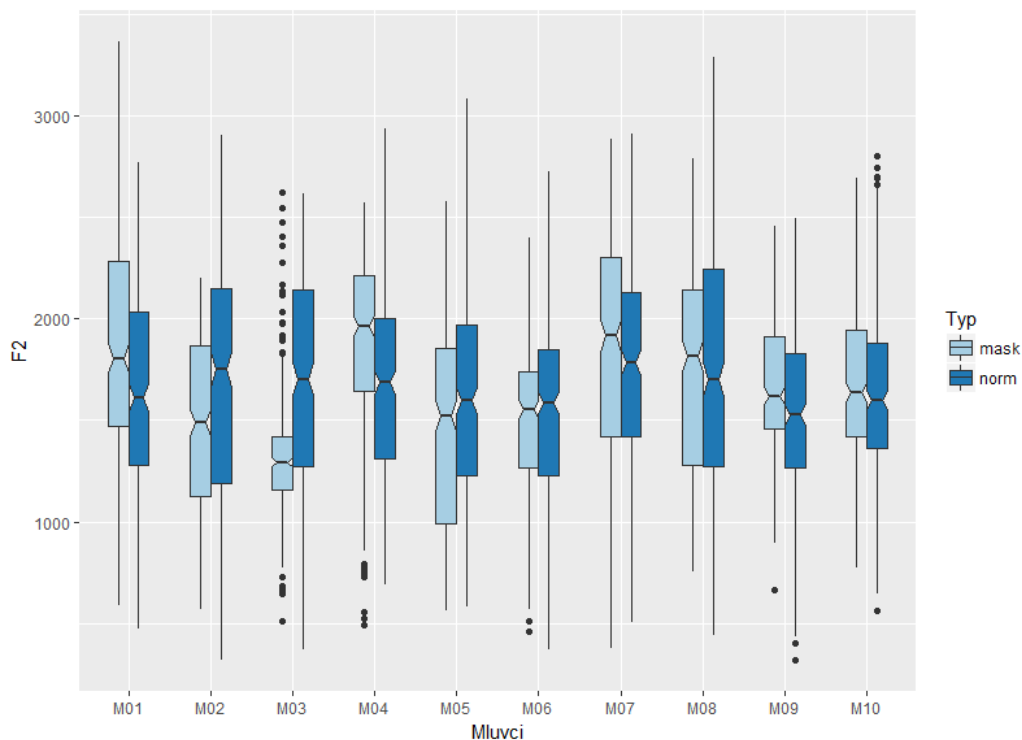
Dlouhodobé formantové distribuce

Jak již vyplývá z výsledků akustické analýzy prvního formantu provedené v jednotlivých vokálech, u většiny mluvčích byly zaznamenány znatelné změny F1, a to většinou směrem k vyšším hodnotám, což odpovídá otevřenější artikulaci při maskování hlasu než při běžném projevu. Také mediány LTF1 byly u osmi z deseti analyzovaných mluvčích signifikantně odlišné při porovnání jejich běžných a maskovaných projevů, z toho v sedmi případech došlo při maskování k jeho nárůstu. Tento výsledek demonstruje obrázek 3.2.2-1.



Obrázek 3.2.2-1: LTF1 [Hz] u jednotlivých mluvčích při maskovaném (mask) a nemaskovaném (norm) projevu.

V případě LTF2 bylo zaznamenáno pouze pět signifikantních posunů, jednalo se tedy o polovinu pozorovaných mluvčích. U tří mluvčích došlo ke zvýšení hodnot F2, což odpovídá přednější artikulaci vokálů, naopak u dvou mluvčích došlo k jejich snížení, jak je patrné na obrázku 3.2.2-2. Nižší počet statisticky signifikantních posunů mediánu F2 bez ohledu na kvalitu jednotlivých vokálů patrně vyplývá z toho, že horizontální posuny jazyka nebyly u všech mluvčích konzistentní napříč vokály.



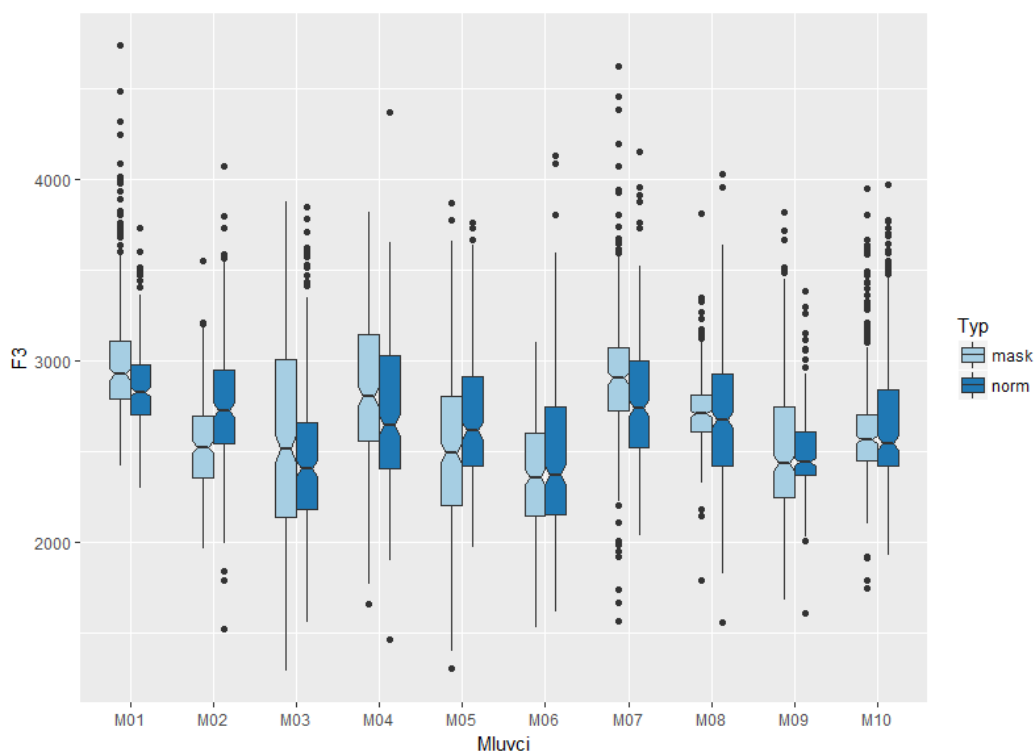
Obrázek 3.2.2-2: LTF2 [Hz] u jednotlivých mluvčích při maskovaném (mask) a nemaskovaném (norm) projevu.

Z propojení výsledných hodnot LTF1 a LTF2 vyplývá, že se mezi mluvčími při maskování hlasu objevovaly následující artikulační posuny vokálů:

- otevřenější + přednější: M01; M04; M09
- otevřenější + zadnější: M03
- pouze otevřenější: M05; M07; M10
- pouze zavřenější: M08
- pouze zadnější: M02
- bez signifikantního posunu: M06

Na základě těchto výsledků je tedy možné říci, že neexistuje výrazně převažující směr posunu jazyka při artikulaci vokálů, který by se uplatňoval při maskovacích strategiích.

U pěti, tj. poloviny mluvčích došlo k signifikantnímu posunu LTF3, z toho u tří mluvčích byl zaznamenán nárůst hodnoty mediánu LTF3, u dvou pak její pokles, jak ukazuje obrázek 3.2.2-3.

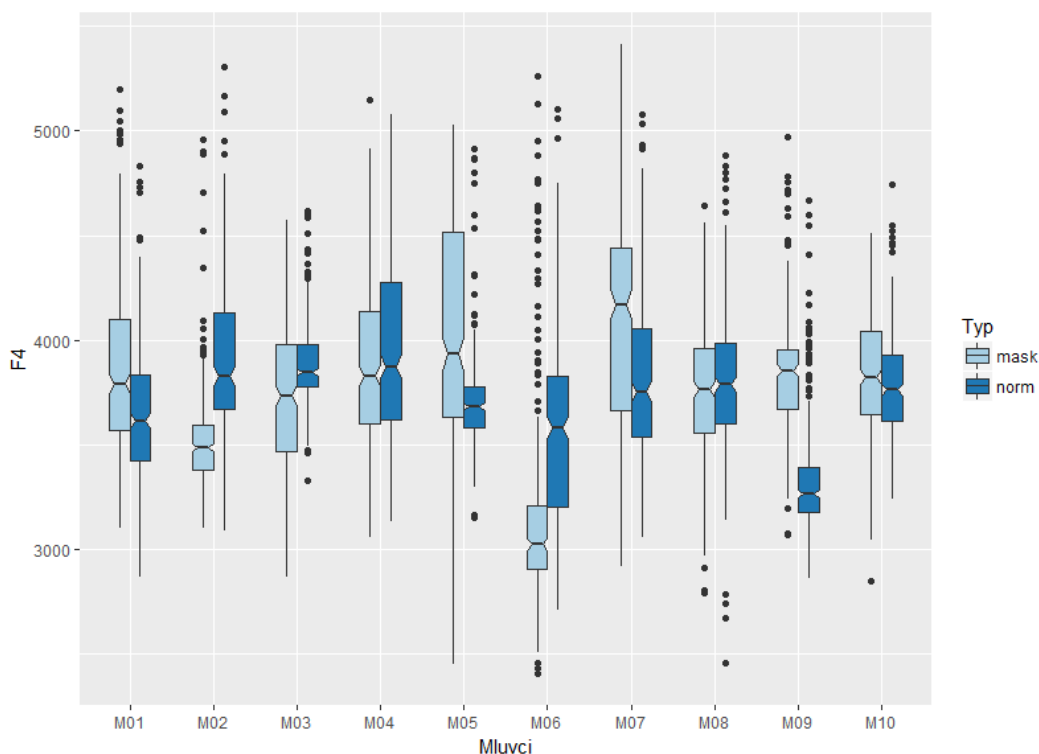


Obrázek 3.2.2-3: LTF3 [Hz] u jednotlivých mluvčích při maskovaném (mask) a nemaskovaném (norm) projevu.

Statisticky signifikantní posun LTF4 byl zaznamenán u vyššího počtu mluvčích, než tomu bylo v případě LTF3, a to u sedmi z deseti, přičemž ve čtyřech případech se jednalo o zvýšení, ve třech pak o snížení mediánu. Tento výsledek naznačuje obrázek 3.2.2-4.

Skutečnost, že se signifikantní změna LTF4 odehrála u více mluvčích než v případě LTF3, lze do jisté míry označit za překvapivou, neboť se čtvrtý formant, na rozdíl od třetího, považuje za zcela nezávislý na momentálním nastavení vokálního traktu mluvčího (Reetz & Jongman, 2009; citováno v: Skarnitzl et al., 2014).

Je nicméně třeba uvést, že při testování pomocí lineárního smíšeného modelu nebyl faktor maskování hlasu u posunů hodnot dlouhodobých formantových distribucí shledán statisticky signifikantním na hladině významnosti 5 %, a to u žádného z pozorovaných formantů. Je ovšem možné, že tento výsledek je způsoben omezeným množstvím analyzovaných dat, resp. mluvčích.



Obrázek 3.2.2-4: LTF4 [Hz] u jednotlivých mluvčích při maskovaném (mask) a nemaskovaném (norm) projevu.

3.2.3. SPEKTRÁLNÍ TĚŽIŠTĚ

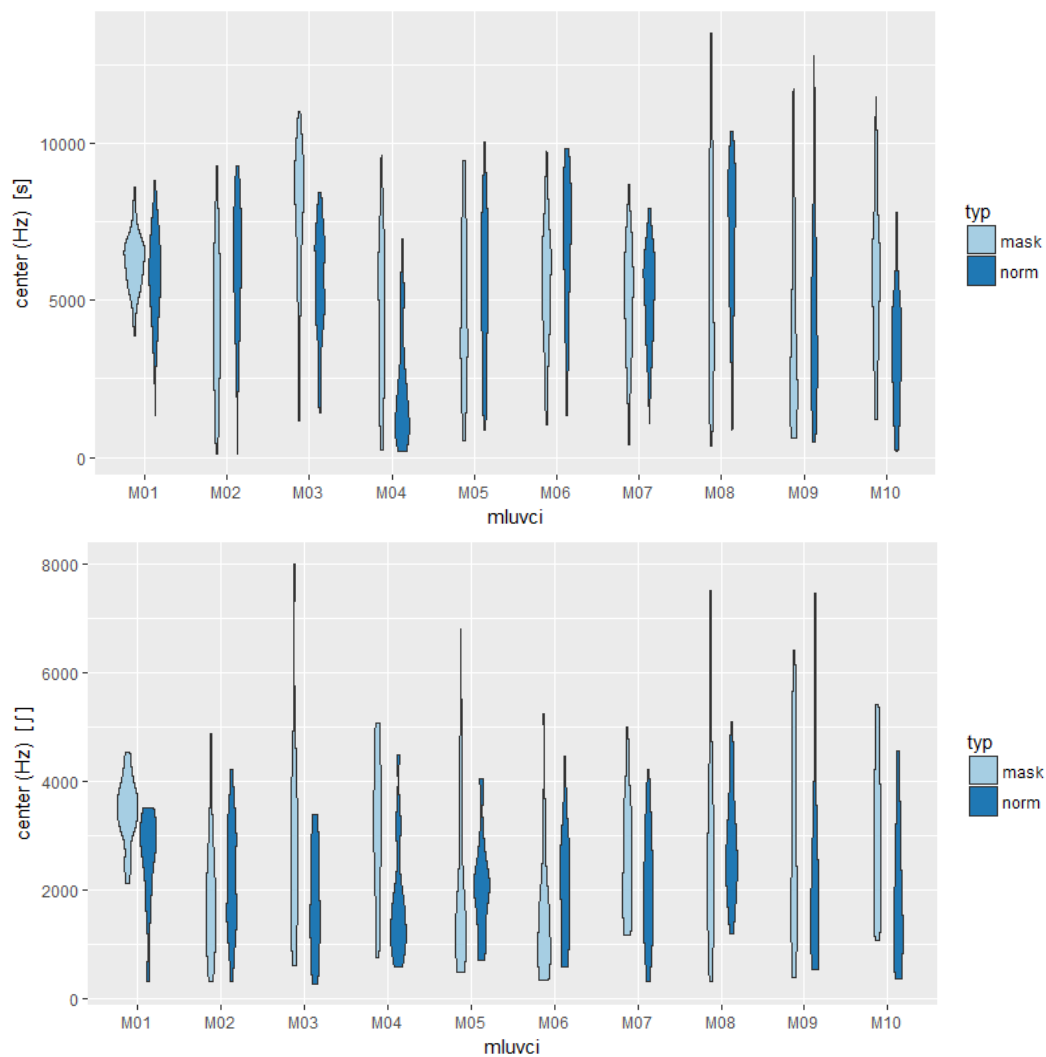
Pro neznělé sibilanty, tedy [s] a [ʃ], bylo identifikováno spektrální těžiště.

V případě hlásky [s] u většiny mluvčích nebyly pozorovány signifikantní změny spektrálního těžiště, jak předkládá obrázek 3.2.3-1. U některých mluvčích je možné sledovat určitý posun, pro velký rozptyl hodnot však není možné s určitostí říci, že skutečně došlo k signifikantní změně spektrálního těžiště, či nikoliv.

Obdobné výsledky lze sledovat také u hlásky [ʃ], viz obrázek 3.2.3-2.

U obou frikativ je možné pozorovat nárůst hodnoty spektrálního těžiště u mluvčího M04 při maskování hlasu, u ostatních mluvčích však není zřejmý žádný takový posun. Jednotlivé mluvčí navíc od sebe navzájem není možné pomocí tohoto ukazatele spolehlivě odlišit. Je možné, že tento výsledek je zapříčiněn nedostatkem dat.

Při využití lineárního smíšeného modelu nebyl faktor maskování na hladině významnosti 5 % vyhodnocen jako signifikantní, a to pro spektrální těžiště nebo směrodatnou odchylku žádné ze zkoumaných sibilant. Není však možné říci, zda je tento výsledek zapříčiněn stabilitou těchto parametrů, či spíše nedostatkem analyzovaných dat.

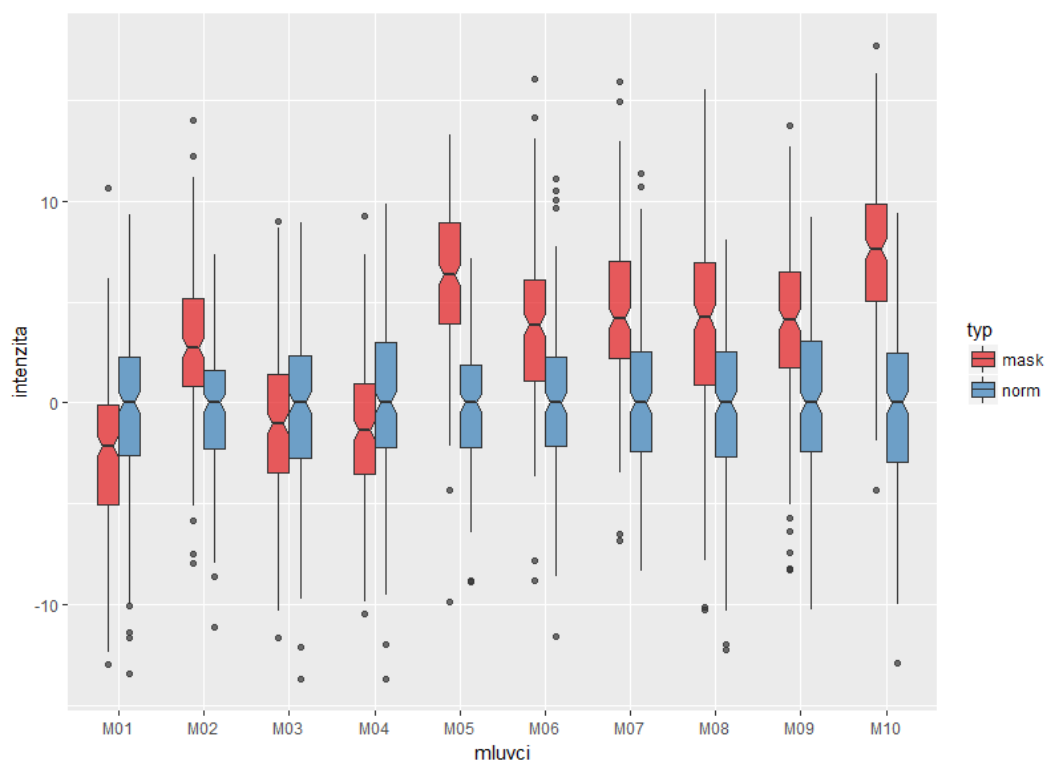


Obrázek 3.2.3-1 (nahore) a 3.2.3-2 (dole): Spektrální těžiště [s] a [j] [Hz] u jednotlivých mluvčích při maskovaném (mask) a nemaskovaném (norm) projevu.

3.2.4. INTENZITA

Medián

U jednotlivých mluvčích byly porovnány mediány naměřené intenzity maskovaných a nemaskovaných projevů. Vzhledem k tomu, že nahrávání mluvčích probíhalo za různých podmínek (mezi nahrávkami jednotlivých mluvčích se mohla lišit odezva daného prostředí, okolní šum či vzdálenost mluvčího od nahrávacího zařízení), nemohly být rozdíly v naměřené intenzitě mezi mluvčími považovány za směrodatné. Intenzita jednotlivých nemaskovaných projevů byla tudíž normalizována tak, aby hodnota mediánu u každého z nich byla rovna nule – od všech vzorků obsažených v intenzitní kontuře dané nahrávky byla odečtena hodnota jejího mediánu. Tatáž hodnota (tj. medián intenzity nemaskovaného projevu) pak byla odečtena od vzorků, z nichž se skládala intenzitní kontura maskovaného projevu téhož mluvčího. Tímto způsobem je možné sledovat změny intenzity mezi maskovanými a nemaskovanými projevy jednotlivých mluvčích. U všech mluvčích došlo při maskování hlasu ke statisticky významnému posunu mediánu intenzity; u sedmi z nich bylo pozorováno jeho zvýšení, u tří naopak snížení – toto pozorování předkládá obrázek 3.2.4-1.



Obrázek 3.2.4-1: Posuny mediánu intenzity při maskovaném projevu oproti projevu nemaskovanému.

| | medián intenzity při nemaskovaném projevu (dB) | posun mediánu intenzity při maskovaném projevu (dB) |
|-----|------------------------------------------------|-----------------------------------------------------|
| M01 | 0 | -2,5 |
| M02 | 0 | 3,1 |
| M03 | 0 | -0,8 |
| M04 | 0 | -1,3 |
| M05 | 0 | 6,2 |
| M06 | 0 | 3,9 |
| M07 | 0 | 4,5 |
| M08 | 0 | 3,9 |
| M09 | 0 | 4,1 |
| M10 | 0 | 7,6 |

Tabulka 3.2.4-1: Posuny F4 při maskovaném projevu oproti projevu nemaskovanému u jednotlivých mluvčích.

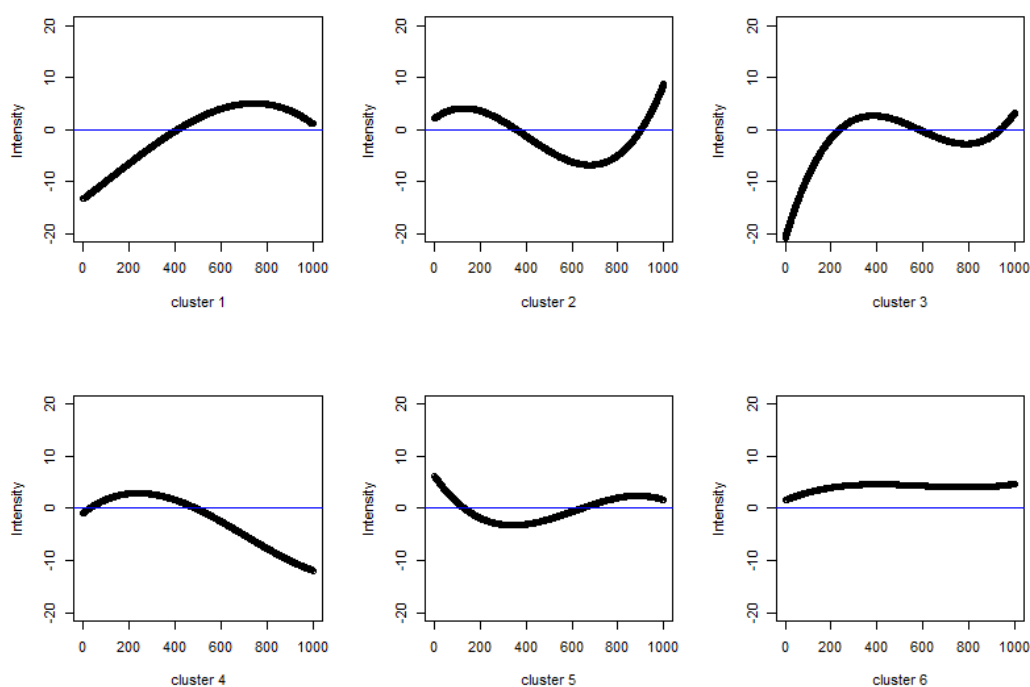
Při testování změn hodnot intenzity pomocí lineárního smíšeného modelu s typem promluvy jakožto fixním faktorem byla p-hodnota vyčíslena na 0,01 – na hladině významnosti 5 % se tudíž faktor maskování hlasu projevil jako statisticky signifikantní ($p=0,01$). Celkové posuny intenzity při maskování hlasu u jednotlivých mluvčích předkládá tabulka 3.2.4-1.

Tento výsledek je možné interpretovat dvěma způsoby: (1) mluvčí mají většinou tendenci mluvit více nahlas, ale je také možné, že (2) většina mluvčích se při maskování hlasu více přiblížila k nahrávacímu zařízení než při předčítání textu nemaskovaným hlasem.

Intenzitní kontury

Podobně jako u kontur f_0 byly pomocí shlukové analýzy k-means nalezeny typické průběhy intenzitních kontur v jednotlivých melodémech. V tomto případě byly kontury všech nahrávek – tj. maskovaných i nemaskovaných projevů – normalizovány, tedy vystředěny během nuly.

Bylo identifikováno celkem šest typických průběhů; při vyšším počtu se objevovaly velmi podobné, a tudíž redundantní vzorce. Tvary získaných kontur ukazuje obrázek 3.2.4-2.



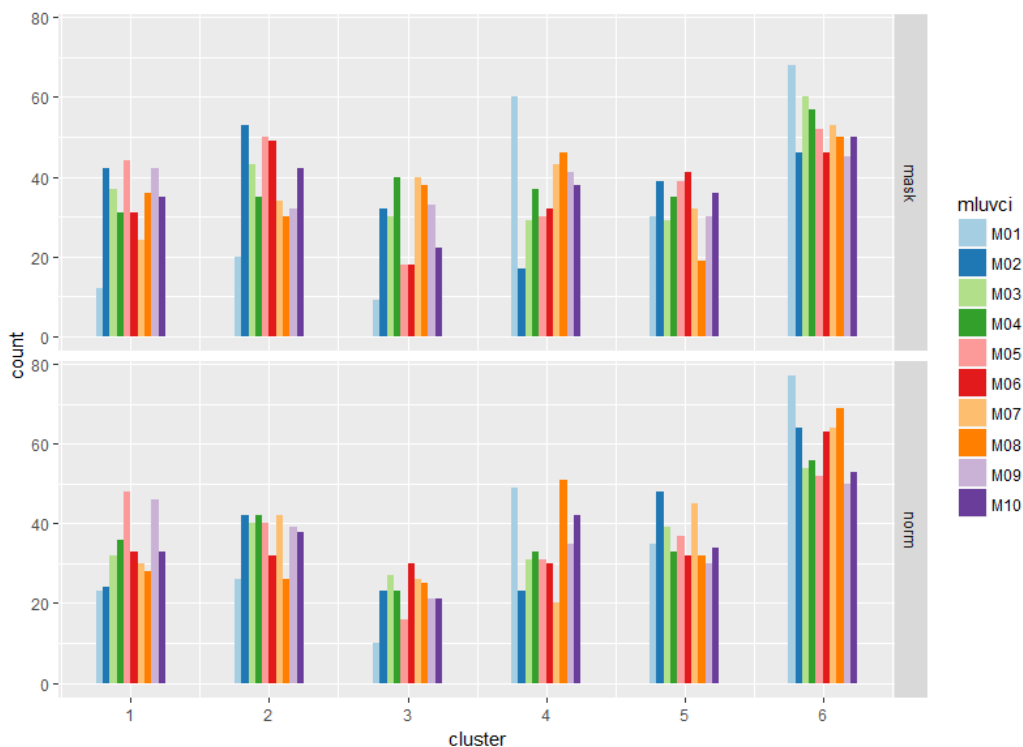
Obrázek 3.2.4-2: Typické intenzitní kontury.

Výskyt těchto typických průběhů byl pozorován u jednotlivých mluvčích, a to jak celkově, tak v rámci jednotlivých typů melodémů.

Celkový souhrn užití sledovaných kontur zachycuje obrázek 3.2.4-3. Jak je z tohoto obrázku patrné, obecně byl v obou typech projevu nejčastěji využíván cluster 6 (tj. téměř plochý průběh, mírně nad průměrnou hodnotou intenzity celé nahrávky), nicméně jeho prevalence nebyla příliš výrazná (jako tomu bylo u převažujícího clusteru v případě kontur f0, viz oddíl 3.2.1).

Lze říci, že celkové užití jednotlivých clusterů v průběhu celého projevu se mezi maskovanými a nemaskovanými projevy zásadním způsobem neliší.

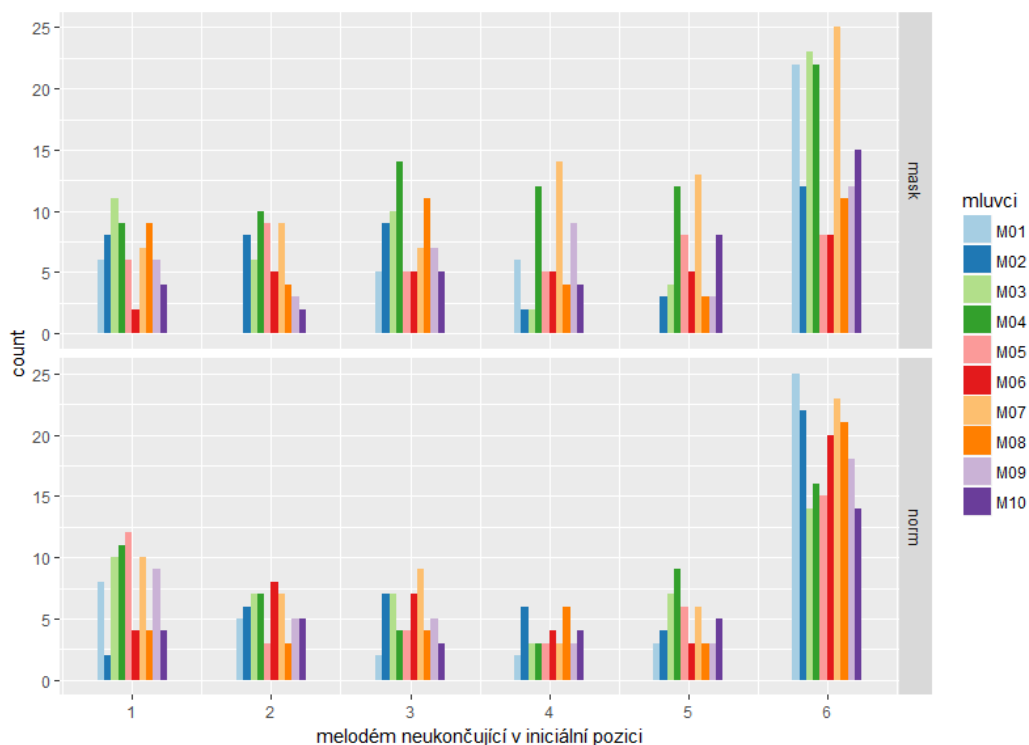
U některých mluvčích se však zřejmě mohou objevit nestandardní četnosti užití clusterů, které zůstávají stabilní i při maskovaném projevu; např. mluvčí M01 vykazoval frekventovanější užití clusterů č. 4 a 6, a naopak méně časté užití clusterů č. 1, 2 a 3. V tomto směru se lišil např. od mluvčího M02. Zdá se tedy, že užití analýzy intenzitních kontur by mohlo mít potenciál od sebe odlišit dva mluvčí.



Obrázek 3.2.4-3: Četnost typických intenzitních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích.

Užití typických intenzitních kontur bylo sledováno také v jednotlivých typech melodémů.

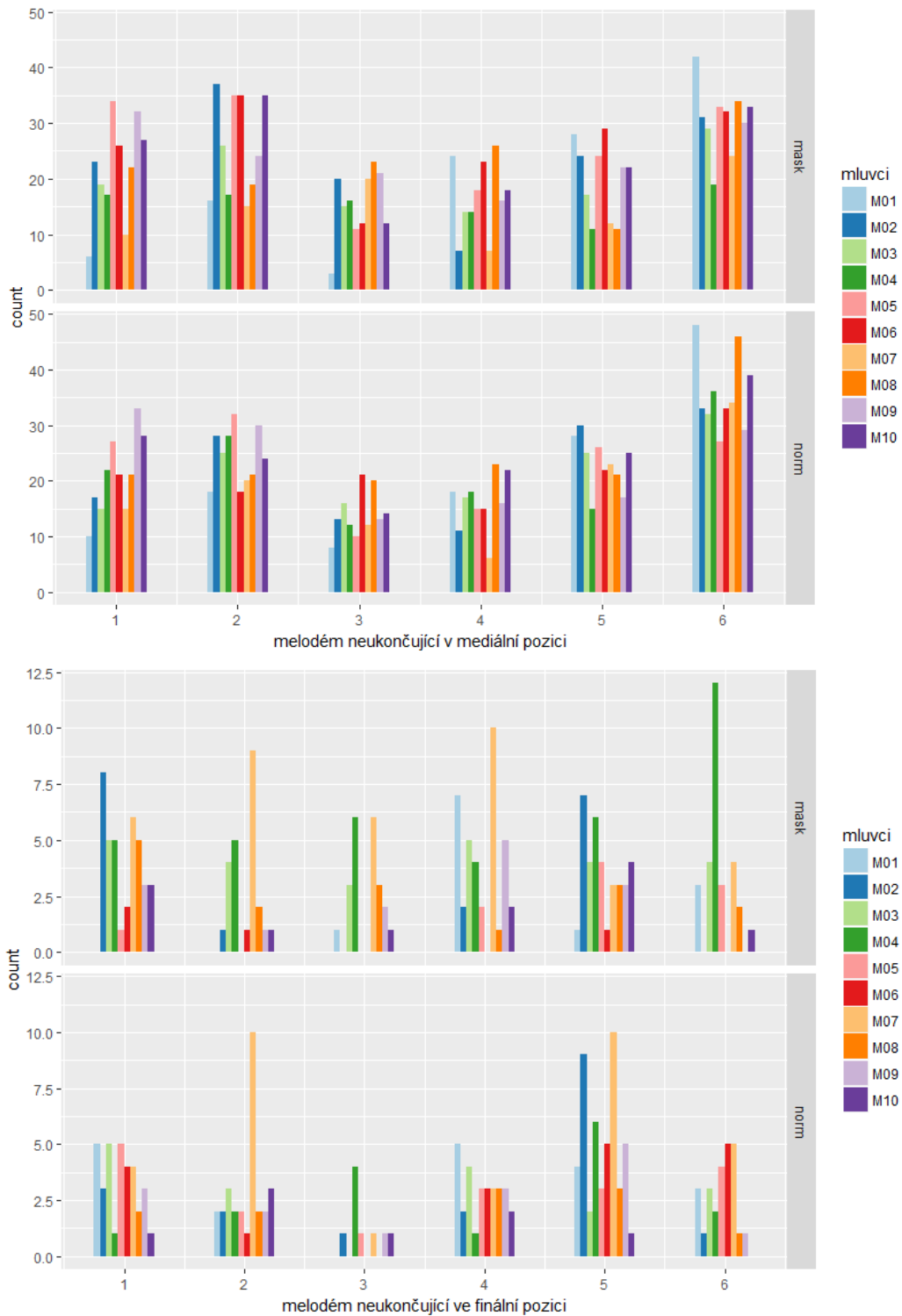
Rozložení užívání jednotlivých clusterů při realizaci melodému neukončujícího v iniciální pozici ukazuje obrázek 3.2.4-4. Nejčastěji zde byl využíván cluster č. 6, a to především v nemaskovaných projevech. U clusterů č. 3, 4, 5 a 6 lze při maskovaných projevech sledovat větší rozkolísanost napříč mluvčími, zatímco při nemaskovaných projevech je četnost užití těchto kontur mezi mluvčími velmi podobná.



Obrázek 3.2.4-4: Četnost typických intenzitních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích při realizaci melodému neukončujícího v iničiální pozici.

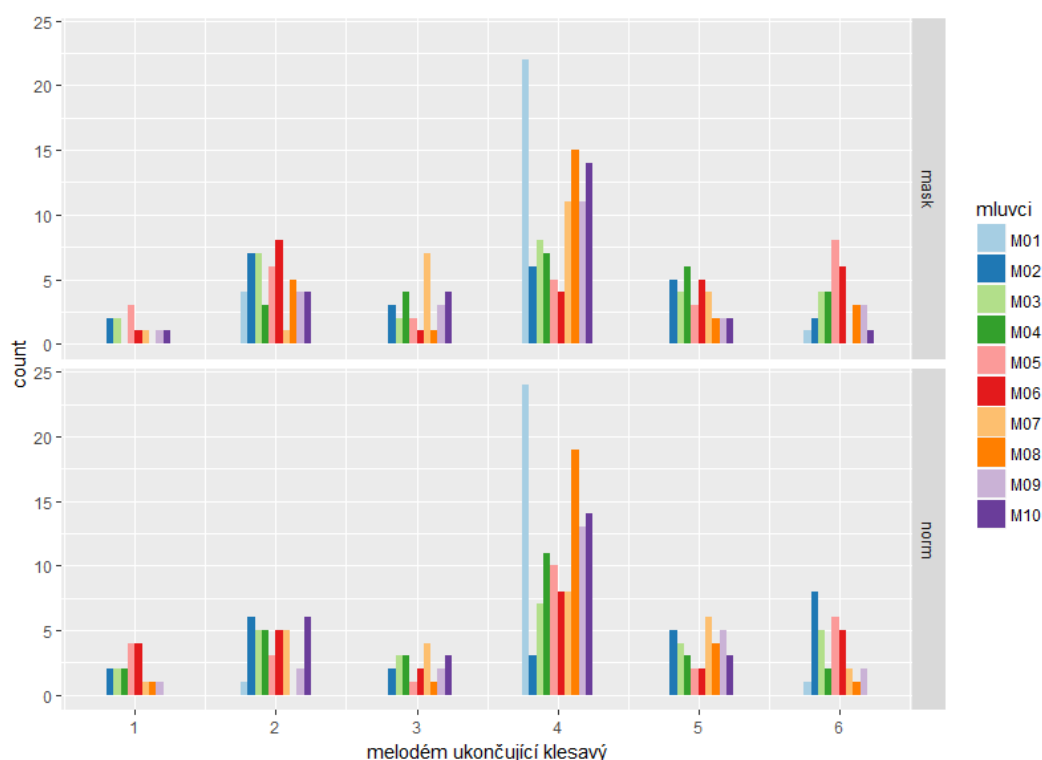
Výskyt typických kontur při realizaci melodému neukončujícího v mediální pozici předkládá obrázek 3.2.4-5. V tomto případě se četnosti užití clusterů u jednotlivých mluvčích při maskovaném projevu velmi blíží projevu nemaskovanému. U některých dvojic mluvčích lze pozorovat obdobné rozdíly v užití jednotlivých clusterů při maskovaných a nemaskovaných projevech; např. mluvčí M01 a M02 nebo M07 a M08, analýza kontur f_0 by tedy mohla být považována za přínosnou při odlišování dvou jednotlivců i navzdory maskování hlasu. U jiných dvojic mluvčích však poměry četností užití clusterů tuto stabilitu při maskovaném i nemaskovaném projevu nevykazují – např. u mluvčích M06 a M08.

Výskyt jednotlivých typických kontur při realizaci melodému neukončujícího ve finální pozici zachycuje obrázek 3.2.4-6. V tomto případě je však interpretace výsledků obtížná, neboť četnost takových realizací se pohybovala pouze v řádech jednotek.



Obrázky 3.2.4-5 (nahore) a 3.2.4-6 (dole): Četnost typických intenzitních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích při realizaci melodému neukončujícího v iniciální a finální pozici.

Zastoupení pozorovaných clusterů při realizaci melodému ukončujícího klesavého ukazuje obrázek 3.2.4-7. Je patrné, že u většiny mluvčích byl nejčastěji využíván cluster č. 4, tj. počáteční mírný nárůst a následně postupný pokles intenzity. U tohoto melodému lze pozorovat změny četnosti užití jednotlivých clusterů při maskovaném a nemaskovaném projevu (např. pokles užití clusteru č. 6 mluvčím M02), jedná se však opět pouze o jednotky případů, a tyto rozdíly tudíž nemohou být považovány za signifikantní.



Obrázek 3.2.4-7: Četnost typických intenzitních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích při realizaci melodému ukončujícího klesavého.

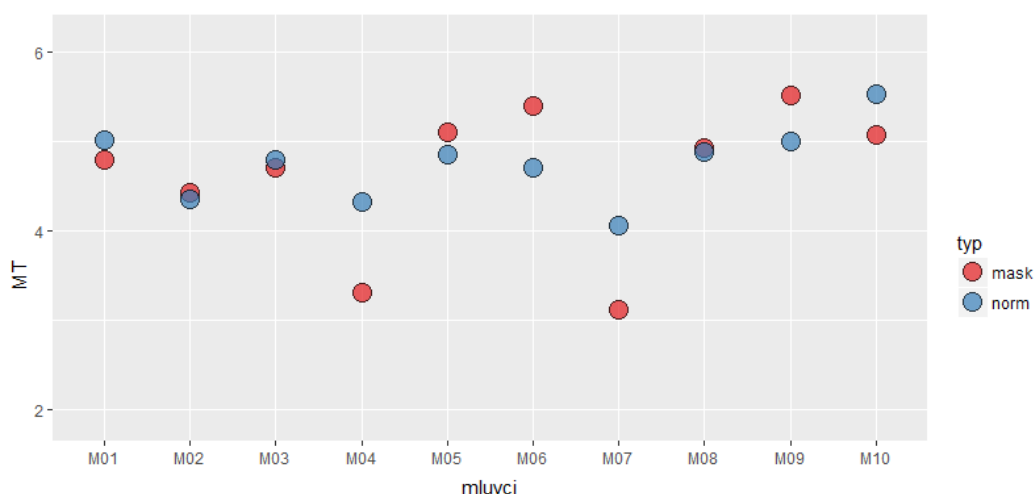
Pro testování významnosti vlivu maskování na změny ve využívání vzorců intenzitních kontur jednotlivými mluvčími byl využit test nezávislosti, tj. Pearsonův chí kvadrát, a byla uskutečněna Bonferroniho korekce. Na hladině významnosti 0,5 % nebyl faktor maskování vyhodnocen jako statisticky signifikantní u žádného mluvčího, a to u žádného ze zkoumaných melodémů; je možné, že tento výsledek je způsoben malým množstvím dat.

3.2.5. GLOBÁLNÍ TEMPORÁLNÍ CHARAKTERISTIKY

Mluvní tempo

U všech nahrávek bylo vyčísleno celkové mluvní tempo; výsledky zobrazuje obrázek 3.2.5-1. Jak je patrné, u čtyř mluvčích došlo ke zdatelnému poklesu mluvního tempa při maskovaném projevu, naopak u tří mluvčích byl pozorován jeho nárůst. U ostatních tří mluvčích se mluvní tempo při maskovaném projevu oproti projevu nemaskovanému výrazně nelišilo.

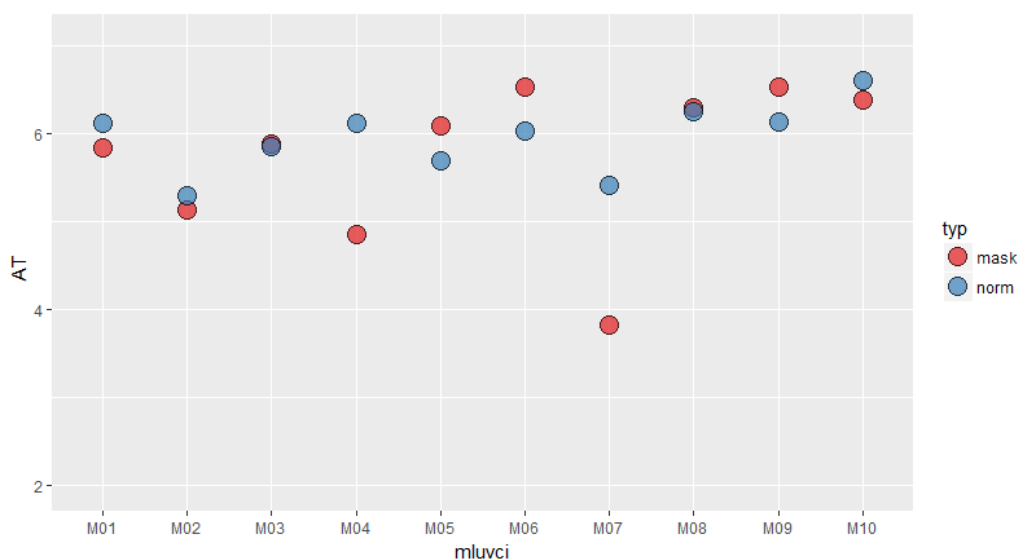
Tyto výsledky jsou spíše orientační, neboť z každé nahrávky byla získána pouze jedna průměrná hodnota mluvního tempa.



Obrázek 3.2.5-1: Celkové hodnoty mluvního tempa [syll/s] v maskovaném (mask) a nemaskovaném (norm) projevu u jednotlivých mluvčích.

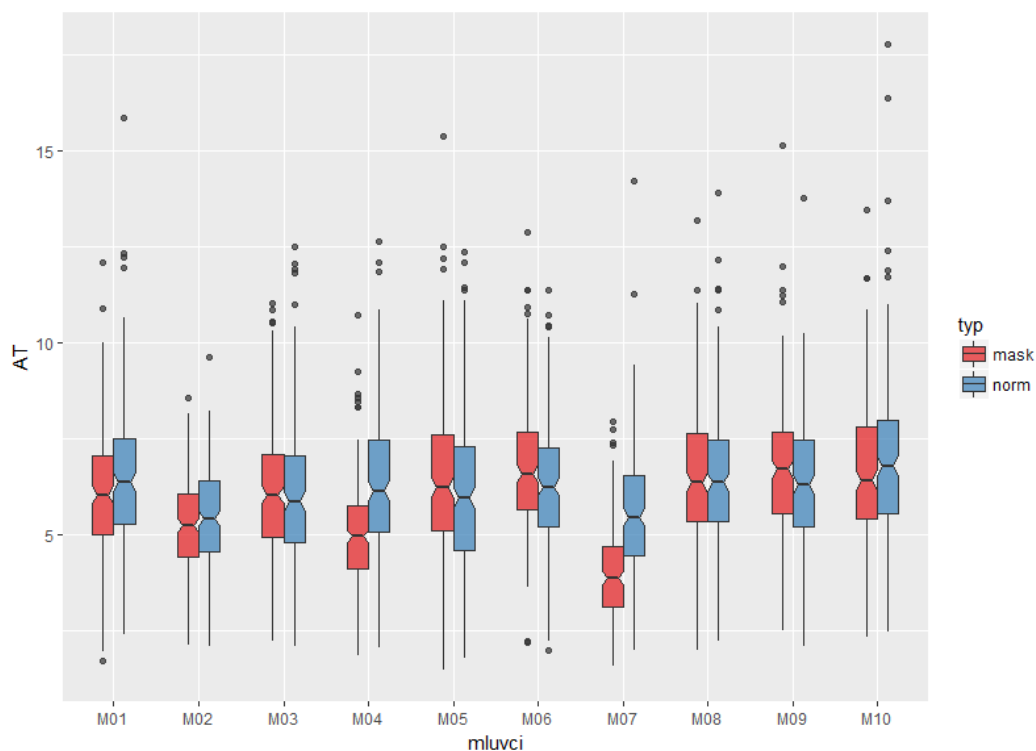
Artikulační tempo

Také pro artikulační tempo byly získány průměrné hodnoty jednotlivých nahrávek pro srovnání s výsledky analýzy mluvního tempa. Výsledné hodnoty zachycuje obrázek 3.2.5-2. Při maskování hlasu prodělaly průměrné hodnoty mluvního tempa u všech mluvčích posuny stejným směrem, jako tomu bylo v případě tempa mluvního; neobjevila se tedy taková maskovací strategie, při které by došlo ke zvýšení AT v důsledku rychlejší artikulace a zároveň ke snížení MT vlivem častějších či delších pauz.



Obrázek 3.2.5-2: Celkové hodnoty artikulačního tempa [syll/s] v maskovaném (mask) a nemaskovaném (norm) projevu u jednotlivých mluvčích.

Artikulační tempo bylo také vyčísleno v rámci jednotlivých melodémů (resp. mluvních taktů), aby bylo možné zachytit jeho variabilitu v průběhu promluv. Výsledky této analýzy zobrazuje obrázek 3.2.5-3.



Obrázek 3.2.5-3: Artikulační tempo [syll/s] v maskovaném (mask) a nemaskovaném (norm) projevu u jednotlivých mluvčích.

Výsledky, jež podává tato analýza, podávají upřesňující informace o posunech artikulačního tempa mezi maskovanými a nemaskovanými projevy. Jak ukazuje obrázek 3.2.5-3, pouze u dvou mluvčích (M04 a M07) došlo ke statisticky signifikantní změně artikulačního tempa, a to směrem dolů. U většiny mluvčích tedy zůstává artikulační tempo bez výrazné změny nezávisle na maskování hlasu.

Při nemaskovaných projevech se většina mluvčích pohybuje mezi pěti a sedmi slabikami za sekundu. Jediné dvě hodnoty mediánu artikulačního tempa, které se pohybovaly pod 5 syll/s, byly získány z maskovaných projevů.

Je tedy možné říci, že při maskovaném projevu mají mluvčí častěji tendenci vzdalovat se od standardních hodnot AT, nicméně nejedná se o výrazný podíl jednotlivců.

U některých se medián AT nemaskovaného projevu signifikantně liší (např. mluvčí M06 a M07), ovšem většina mluvčích se v tomto ohledu navzájem podobá. Tento parametr tudíž zřejmě není možné považovat za příliš přínosný při identifikaci mluvčího, pokud není artikulační tempo daného jedince výrazně nestandardní.

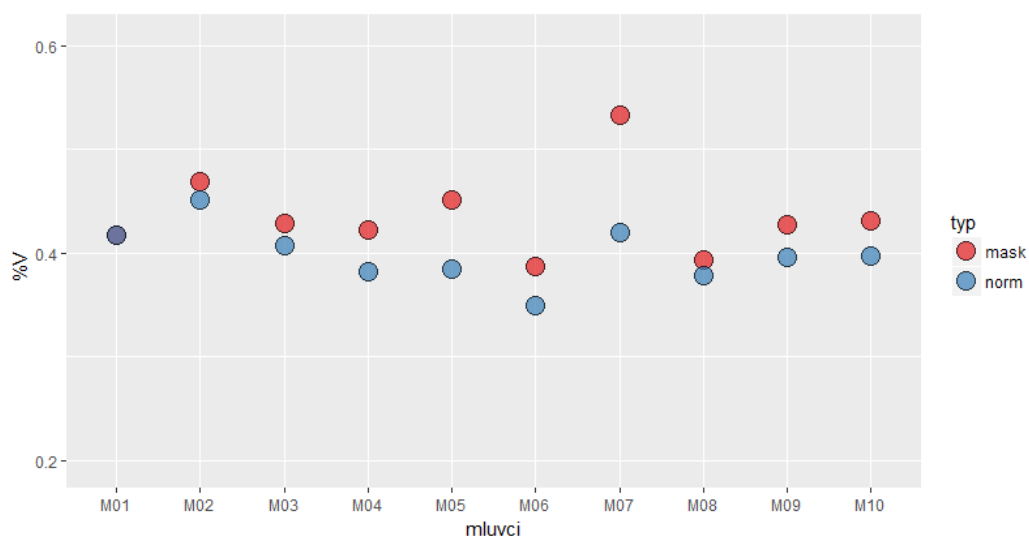
Pokud jde obecně o vliv faktoru maskování na změny artikulačního tempa, za využití lineárního smíšeného modelu nebyl tento faktor vyhodnocen jako statisticky signifikantní na hladině významnosti 5 %. Je možné, že je tato skutečnost způsobena nedostatečným množstvím analyzovaných dat.

%V

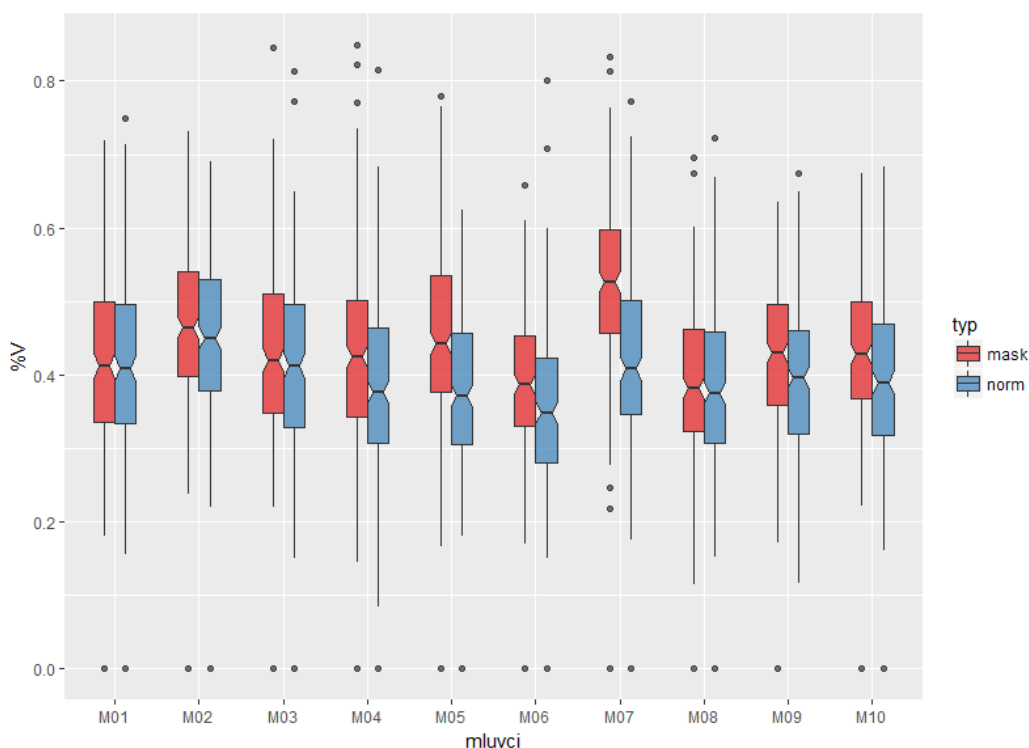
Ve zkoumaných nahrávkách byl vypočítán podíl vokalických intervalů v řeči (resp. v součtu vokalických a konsonantických intervalů, tj. s vyloučením pauz a disfluenčních prvků). Průměrné hodnoty předkládá obrázek 3.2.5-4.

Tyto výsledky ukazují, že obecně má %V tendenci v maskovaných projevech narůstat. U žádného mluvčího nebyl pozorován pokles hodnoty tohoto ukazatele.

Variabilitu %V napříč mluvnickými taktami lze sledovat na obrázku 3.2.5-5. U šesti mluvčích došlo ke statisticky signifikantnímu nárůstu mediánu %V; jeho mírné zvýšení je patrné i u ostatních mluvčích, nejedná se však o statisticky signifikantní změny.



Obrázek 3.2.5-4: Celkové hodnoty %V v maskovaném (mask) a nemaskovaném (norm) projevu u jednotlivých mluvčích.



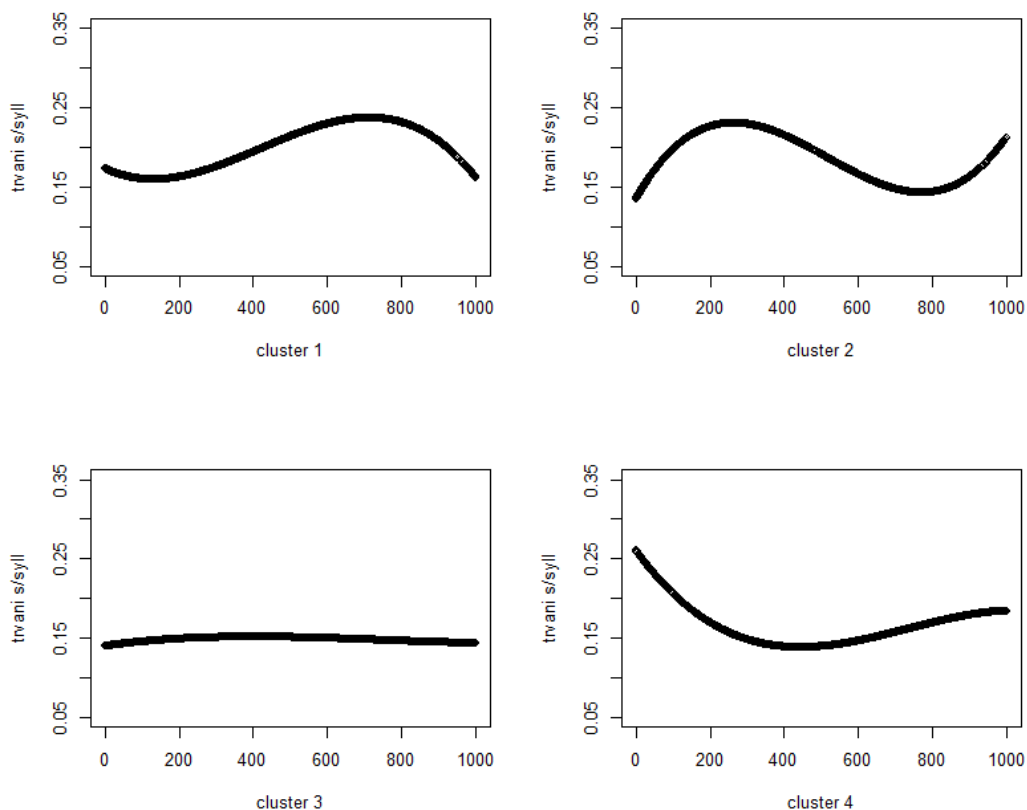
Obrázek 3.2.5-5: %V v maskovaném (mask) a nemaskovaném (norm) projevu u jednotlivých mluvčích.

Tento ukazatel má tedy při maskování hlasu obecně tendenci ubírat se konkrétním směrem, a to k vyšším hodnotám.

Faktor maskování hlasu byl v tomto případě na hladině významnosti 5 % vyhodnocen jako statisticky signifikantní, a to s p-hodnotou 0,002.

3.2.6. LOKÁLNÍ TEMPORÁLNÍ UKAZATELE

Pro analýzu lokálního tempa bylo změřeno trvání jednotlivých slabik v mluvních taktech, a pro ty byly nalezeny čtyři typické kontury. Ty ukazuje obrázek 3.2.6-1.



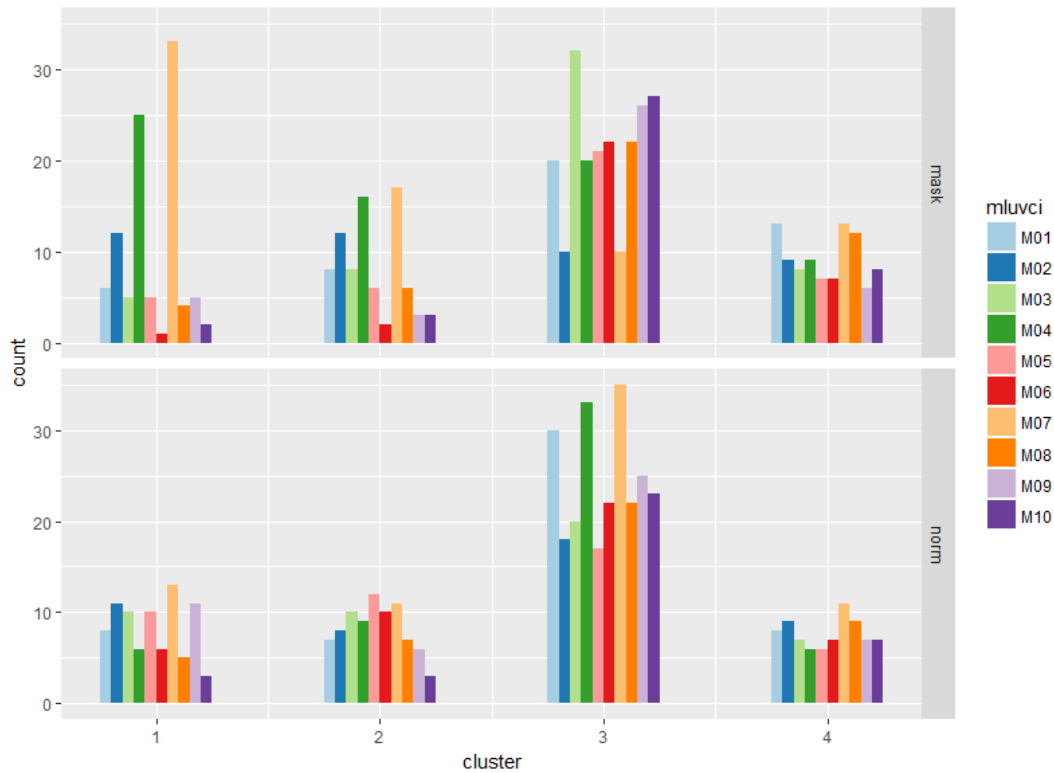
Obrázek 3.2.6-1: Typické temporální kontury.

Četnost výskytu těchto kontur v mluvních taktech zachycuje obrázek 3.2.6-2.

Při nemaskovaných projevech se využití jednotlivých kontur napříč mluvčími výrazně neliší. Naproti tomu při projevech maskovaných výskyt clusterů mezi mluvčími místy kolísá, což lze pozorovat zejména u clusterů č. 1 a 2.

Na základě získaných dat nelze předpokládat, že by bylo možné mluvčí navzájem odlišit pomocí těchto temporálních kontur, ani že četnost jejich užití zůstává stabilní i při maskování hlasu. Je však možné říci, že při maskování hlasu mají mluvčí tendenci uchýlovat se k méně standardním vzorcům uplatňování jednotlivých typických kontur.

Při využití lineárního smíšeného modelu byl nicméně faktor maskování u dvou mluvčích vyhodnocen jako statisticky významný, a to na hladině významnosti 0,5 % (po aplikaci Bonferroniho korekce; jednalo se o mluvčí M04 $p=0,001$ a M07 $p=2,5 \cdot 10^{-5}$). Někteří mluvčí tedy při maskování hlasu zřejmě jsou schopni signifikantně měnit využívané vzorce lokálního artikulačního tempa.



Obrázek 3.2.6-2: Četnost typických temporálních kontur v maskovaných (mask) a nemaskovaných (norm) projevech jednotlivých mluvčích.

4. DISKUZE

V rámci této práce byly podrobeny analýze maskované a nemaskované projevy deseti mluvčích, kteří pro maskování hlasu volili sofistikované strategie, tj. uplatňovali změny více než jednoho řečového parametru. Ačkoliv obecně většina mluvčích při maskování hlasu přistupuje k jednodušším strategiím a modifikuje pouze jediný řečový parametr, přičemž ve většině případů se jedná o modifikaci základní frekvence (Masthoff, 1996; Künzel, 2000), lze se setkat i se strategiemi komplexnějšími; právě některé takové případy byly v této práci zkoumány. Při pořizování maskovaných nahrávek byli mluvčí instruováni, aby libovolným způsobem změnili svůj hlas tak, aby na základě jeho nahrávky nebylo možné provést jejich identifikaci. Strategie, které mluvčí při této úloze uplatňovali, se tedy různí.

Nejprve byla provedena poslechová analýza a bylo popsáno, jak znějí přirozené, tj. nemaskované projevy mluvčích, a k jakým sluchem zaznamenaným změnám v jejich hlasech dochází při využití maskovacích strategií.

Tato poslechová analýza byla provedena autorkou práce a může tudíž být do jisté míry subjektivní. Charakteristiky jednotlivých projevů, které byly popsány, jsou však natolik prominentní, že se výsledky poslechové analýzy dají považovat za spolehlivé. Pro naprostou objektivitu by nicméně bylo zapotřebí provést percepční testy, aby bylo zjištěno, jak jednotlivé projevy znějí různým posluchačům. Poslechová analýza v této práci však představuje spíše doplněk k analýze akustické, a získání objektivních a detailních charakteristik pozorovaných řečových projevů pomocí percepčních testů tudíž nebylo předmětem této práce.

V rámci akustické analýzy byly obecně uplatňovány dva přístupy: (1) statický přístup, tedy takový, který popisuje průběhy daných ukazatelů pomocí jedné hodnoty, v tomto případě mediánu či směrodatné odchylky, a (2) přístup dynamický, který bere v úvahu průběh ukazatelů v čase.

4.1. STATICKÁ ANALÝZA

Pomocí statického přístupu byly analyzovány statistické hodnoty základní frekvence, vokalické formanty, dlouhodobé formantové distribuce, spektrální těžiště sibilant, intenzita a globální temporální charakteristiky.

Ve zkoumaném vzorku se většina mluvčích uchýlila ke změně střední hodnoty své základní frekvence, což bylo pozorováno jak při poslechové, tak i při akustické analýze. Tento výsledek je v souladu se závěry Künzela (2000) a Masthoffa (1996), že většina mluvčích se při maskování hlasu uchyluje právě ke změně tohoto parametru, a to zřejmě proto, že takové změny jsou snadno uskutečnitelné a poslechově velmi výrazné, a nadto nezpůsobují nesrozumitelnost řečového projevu.

U některých mluvčích byly zaznamenány také změny variability základní frekvence, a to směrem k vyšším i nižším hodnotám. Vyhodnocené posuny se však liší mezi poslechovou a akustickou analýzou. Dojem variability základní frekvence samozřejmě může být považován za subjektivní, pro získání lépe vypovídajících výsledků poslechové analýzy by bylo vhodné provést percepční testy.

Analýza prvního formantu odhalila, že u většiny mluvčích docházelo k jeho zvyšování, které je akustickým korelátem větší otevřenosti vokálů. Naproti tomu v případě druhého formantu došlo mezi mluvčími k posunům jak k vyšším, tak i k nižším hodnotám, zdá se tedy, že směr vertikálního posunu jazyka se u mluvčích při maskování hlasu různí. Ačkoliv F3 je již na kvalitě vokálu závislý jen částečně, i u něj docházelo při maskovaném projevu některých mluvčích ke znatelným posunům, ani tento parametr se tudíž neprojevil jako stabilní navzdory maskování hlasu. Ani čtvrtý formant, který je považován za zcela nezávislý na kvalitě vokálu, nezůstával mezi maskovanými a nemaskovanými projevy mluvčích vždy stabilní; ve většině případů k signifikantním posunům nedošlo, nicméně za zcela stabilní parametr jej na základě analýzy dostupných dat považovat nelze.

U dlouhodobé distribuce prvního formantu byly u většiny mluvčích pozorovány posuny směrem nahoru (jejich artikulace vokálů tedy byla otevřenější). V případě LTF2 byly zaznamenány signifikantní posuny u poloviny mluvčích, jednalo se jak

o jeho zvyšování, tak i o snižování, což zřejmě vyplývá z nekonzistence horizontálních posunů jazyka napříč vokalickými kvalitami u některých mluvčích. Pokud vezmeme v úvahu pozorované změny LTF1 a LTF2, je možné říci, že někteří mluvčí se při maskování hlasu uchylují ke změnám artikulace vokálů, a to různými směry; v maskovacích strategiích tudíž výrazně nepřevládá konkrétní směr posunu.

U pěti mluvčích byl pozorován signifikantní posun LTF3 a v případě LTF4 došlo k významným změnám hodnot dokonce u sedmi mluvčích. Ani tyto parametry tudíž nelze na základě analýzy dostupných dat považovat za odolné vůči záměrným modifikacím hlasu.

Spektrální těžiště sibilant bylo vyhodnoceno jako spíše stabilní i navzdory maskování hlasu, neboť u většiny mluvčích nebyly pozorovány výrazné posuny jeho hodnoty, stejně jako tomu bylo také v případě směrodatné odchylky spektrálního těžiště. U těchto parametrů se ovšem objevily velké intervaly spolehlivosti, zřejmě zapříčiněné omezeným množstvím dat. Nebyly tedy pozorovány signifikantní posuny, nicméně na základě těchto hodnot zároveň nebylo až na výjimky možné od sebe mluvčí spolehlivě odlišit.

Při pozorování artikulačního tempa nebyly u většiny mluvčích nalezeny signifikantní změny při maskovaných projevech. Nadto byly nalezené hodnoty artikulačního tempa do značné míry podobné napříč mluvčími, tento parametr se tudíž neukazuje jako příliš prospěšný pro identifikaci mluvčího, pokud se jeho hodnoty u daného jedince výrazně neodchylují od standardu.

Je třeba uvést, že při analýze pomocí lineárního smíšeného modelu nebyl faktor maskování u žádného z uvedených parametrů vyhodnocen jako statisticky signifikantní; na základě dostupných dat tedy není možné tvrdit, že jsou mluvčí obecně schopni jejich hodnoty měnit při maskování hlasu. Tento výsledek však může být zapříčiněn skutečností, že byl využit materiál od deseti mluvčích, což je poměrně malý vzorek, a informace, které lze na jeho základě získat o celé populaci, jsou tudíž značně omezené.

U všech mluvčích došlo při maskovaném projevu ke statisticky signifikantním změnám intenzity při maskování hlasu, většinou se jednalo o její zvýšení. Také při testování pomocí lineárního smíšeného modelu byl faktor maskování vyhodnocen

jako statisticky významný, je však třeba upozornit na to, že projevy všech mluvčích byly normalizovány – medián intenzity nemaskovaného projevu byl stanoven na nulovou hodnotu a maskovaný projev byl pak k tomu nemaskovanému vztahován. Jak bylo uvedeno v oddílu 2.2.2.4, intenzitu projevů nebylo možné porovnávat mezi jednotlivými mluvčími, neboť nebylo možné zajistit naprosto totožné podmínky při nahrávání.

Jako zajímavý parametr se ukázalo %V, tedy podíl vokalických intervalů v řeči. U většiny mluvčích byl při maskování hlasu zaznamenán signifikantní nárůst jeho mediánu a u žádného mluvčího naopak nebyl pozorován jeho pokles. Také užití lineárního smíšeného modelu ukázalo faktor maskování u tohoto ukazatele jako statisticky významný. %V má tedy zřejmě při maskování hlasu obecně tendenci stoupat. Pozoruhodné je, že nárůst %V je možné pozorovat i při jiných podmínkách, než jakou je maskování hlasu – tento proces byl mimo jiné zaznamenán např. u pacientů s Parkinsonovou chorobou (Pettorino et al., 2018). Z jakého důvodu k nárůstu %V za různých podmínek dochází, zůstává otázkou pro další výzkum.

4.2. DYNAMICKÁ ANALÝZA

K identifikaci mluvčího mohou přispět informace o charakteristických pohybech jeho vokálního traktu, u nichž lze očekávat rozdíly mezi mluvčími dané individuální anatomii i artikulačními návyky, a zároveň relativní stabilitu u daného mluvčího, podobně jako je tomu u dalších automatizovaných činností, např. chůze (McDougall, 2006; He & Dellwo, 2017). Tyto charakteristické jevy lze zachytit pomocí dynamických ukazatelů.

V rámci akustické analýzy byly tímto způsobem zkoumány základní frekvence, intenzita a lokální tempo. Pro každý z těchto ukazatelů byly nalezeny typické kontury, které se objevovaly při realizaci mluvních taktů, a bylo pozorováno, s jakou četností se jednotlivé kontury vyskytují u mluvčích při maskovaném a nemaskovaném projevu. Signifikance faktoru maskování pro změny distribucí kontur byla vyhodnocována pomocí testu nezávislosti.

Distribuce jednotlivých typických kontur základní frekvence se při nemaskovaných projevech značně podobaly napříč mluvčími. Při maskovaném projevu však lze pozorovat větší rozkolísanost mezi mluvčími – při maskování hlasu se tedy zřejmě u mluvčích častěji vyskytují vzorce, které lze považovat za nestandardní. U většiny mluvčích byl faktor maskování vyhodnocen jako statisticky signifikantní.

Podobné tendence je do jisté míry možné pozorovat také u temporálních kontur; i zde se distribuce kontur při nemaskovaných projevech podobaly napříč mluvčími, zatímco při maskování hlasu se někteří mluvčí měli tendenci uchýlovat k méně standardním vzorcům. Faktor maskování se ukázal jako statisticky signifikantní u dvou mluvčích.

Naproti tomu v případě intenzitních kontur se distribuce mezi maskovanými a nemaskovanými projevy obecně výrazně nelišila. Při aplikaci testu nezávislosti nebyl faktor maskování vyhodnocen jako statisticky signifikantní u žádného mluvčího. Mezi některými mluvčími však byly pozorovány výrazně odlišné distribuce užití typických kontur, které při maskovaném i nemaskovaném projevu zůstávaly více méně stabilní; zdá se tedy, že využití intenzitních kontur by mohlo napomoci odlišení dvou mluvčích, pakliže se četnost užití jednotlivých typických kontur mezi nimi výrazně liší, a to i navzdory maskování hlasu. Je však třeba, aby byl k dispozici dostatek řečového materiálu, což může být v reálných forenzních podmínkách problematické.

5. ZÁVĚR

Tato práce předkládá přehled dosavadního výzkumu zaměřeného na aplikaci fonetických poznatků ve forenzní praxi. Věnuje se problematice profilování a identifikace mluvčího na základě idiosynkratických rysů, tedy charakteristických jevů, které je možné nalézt v řečovém projevu jedince. Představuje také dosavadní výzkumy zaměřené na různé způsoby maskování hlasu a jejich dopady na idiosynkratické rysy.

V rámci této práce byly zpracovány nahrávky maskovaných a nemaskovaných projevů deseti mluvčích a byla provedena jejich poslechová i akustická analýza. Pro účely akustické analýzy byla provedena automatická segmentace na jednotlivé hlásky a slabiky, která byla následně manuálně korigována. Následně byla provedena detailní akustická analýza vybraných řečových jevů a byly kvantifikovány rozdíly mezi přirozenou a maskovanou řečí. U výsledků akustické analýzy pak byly také sledovány souvislosti s výsledky analýzy poslechové.

Akustické analýze byly podrobeny následující řečové parametry: (1) statistické hodnoty základní frekvence, (2) kontury f_0 při realizaci melodémů, (3) vokalické formanty pro jednotlivé vokály a dlouhodobé formantové distribuce, (4) spektrální těžiště sibilant, (5) posuny mediánu intenzity a využití intenzitních kontur při realizaci melodémů, (6) globální temporální charakteristiky a (7) lokální temporální ukazatele.

U jednotlivých parametrů bylo sledováno, zda zůstávají spíše stabilní navzdory maskování hlasu, či zda jsou mluvčí schopni je při záměrné modifikaci svého řečového projevu výrazně měnit, a také zda u daných parametrů existuje při maskování obecná tendence k podobným posunům napříč mluvčími.

Zajímavé výsledky byly pozorovány zejména u dynamických ukazatelů, tj. kontur f_0 a intenzitních a temporálních kontur. Zatímco vzorce využívání intenzitních kontur se u mluvčích ukázaly jako více méně stabilní i navzdory maskování hlasu, u kontur f_0 a temporálních kontur měli mluvčí při maskování hlasu častěji tendence uchylovat se k nestandardním vzorcům.

Jako pozoruhodný faktor se projevilo %V, u nějž byla pozorována obecná tendence při maskování hlasu stoupat.

Pro další výzkum by bylo vhodné shromáždit data od více mluvčích, aby bylo dosaženo větší reprezentativnosti vzorku. Již existující databáze obecné češtiny pro forenzní účely (Skarnitzl & Vaňková, 2017) sice obsahuje maskované a nemaskované projevy sta mluvčích, využití tohoto materiálu je však problematické, neboť se od sebe texty pro běžný a maskovaný projev výrazně liší. Bylo by tudíž potřeba pořídit nové nahrávky, aby byl maskovaný a nemaskovaný materiál srovnatelný.

SEZNAM REFERENCÍ

- Anolli, L. & Ciceri, R. (1997). The Voice of Deception: Vocal Strategies of Naive and Able Liars. *Journal of Nonverbal Behavior*, 21, 259-284.
- Bartle, A. & Dellwo, V. (2015). Auditory speaker discrimination by forensic phoneticians and naive listeners in voiced and whispered speech. *Speech, Language and the Law*, 22.2, 229-248.
- 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.
- Byrne, C. & Foulkes, P. (2004). The ‚Mobile Phone Effect‘ on Vowel Formants. *Speech, Language and the Law*, 11, 83-102.
- Braun, A. & Rosin, A. (2015). On the Speaker-Specificity of Hesitation Markers. *Proceedings of 18th ICPhS*.
- Cao, H. & Wang, Y. (2011). A Forensic Aspect of Articulation Rate Variation in Chinese. *Proceedings of the 17th International Congress of Phonetic Sciences 2011*, 396-399.
- Clark, J. & Foulkes, P. (2007). Identification of voices in electronically disguised speech. *Speech, Language and the Law*, 14, 195-221.
- 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.
- Eriksson, A. (2010). The Disguised Voice: Imitating Accents or Speech Styles and Impersonating Individuals. In: Llamas, C. & Watt, D. (Eds.), *Language and Identities*, 86-96. Edinburgh: Edinburgh University Press.
- Figueiredo, R. M. & Britto, H. S. (1996). A report on the acoustic effects of one type of disguise. *Forensic Linguistics*, 3, 168-175.

- Fitch, W. T. & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, 106, 1511-1522.
- French, P. (1994). An overview of forensic phonetics with particular reference to speaker identification. *Forensic Linguistics*, 1, 169-181.
- Giddens, C. L.; Barron, K. W.; Byrd-Craven, J.; Clark, K. F. & Winter, A. S. (2013). Vocal Indices of Stress: A Review. *Journal of Voice*, 27, 390.e21-390.e29.
- Giles. H.; Coupland, N. & Coupland, J. (1991). Accomodation Theory: Communication, Context, and Consequence. In H. Giles, J. Coupland & N. Coupland (eds.), *Contexts of accomodation: Developments in applied sociolinguistics*, 1-68. New York: Cambridge University Press.
- Giles, H. & Ogay, T. (2007). Communication Accomodation Theory. In B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars*, 293-310. Mahwah: Lawrence Erlbaum.
- 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*.
- González, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics*, 32, 277-287.
- Harnsberger. J. D., Hollien, H., Martin, C. A. & Hollien, K. A. (2009). Stress and Deception in Speech: Evaluating Layered Voice Analysis. *Journal of Forensic Sciences*, 54, 642-650.
- Hartigan, J. A. & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, 28, 100-108.
- He, L. & Dellwo, V. (2014). Speaker Idiosyncratic Variability of Intensity across Syllables. *Proceedings of Interspeech 2014*, 233-237.

- He, L. & Dellwo, V. (2017). Between-speaker variability in temporal organizations of intensity contours. *The Journal of the Acoustical Society of America*, 141, 488-494.
- Hollien, H. (2002). *Forensic Voice Identification*. London: Academic Press.
- Jessen, M. (2007a). Speaker Classification in Forensic Phonetics and Acoustics. In C. Müller (Ed.), *Speaker Classification I. Lecture Notes in Computer Science*, vol 4343, 180-204. Berlin: Springer, Berlin, Heidelberg.
- Jessen, M. (2007b). Forensic reference data on articulation rate in German. *Science and Justice*, 47, 50-67.
- Junqua, J.-C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America*, 93, 510-424.
- Kaiser, J. & Bořil, T. (2018). Impact of the GSM AMR Codec on Automatic Vowel Formant Measurement in Praat and VoiceSauce. In *2018 41th International Conference on Telecommunications and Signal Processing*, 409-412.
- Kavanagh, C. (2012). New consonantal acoustic parameters for forensic speaker comparison. Nepublikovaná disertační práce. York: University of York.
- Kirchhübel, C. & Howard, D. M. (2011). Investigating the acoustic characteristics of deceptive speech. *Proceedings of 17th ICPhS*, 1094-1097.
- Kitamura, T. & Akagi, M. (2007). Speaker Individualities in Speech Spectral Envelopes and Fundamental Frequency Contours. In C. Müller (Ed.), *Speaker Classification I*, vol 4441, 157-176. Berlin: Springer, Berlin, Heidelberg.
- Künzel, H. J. (2000). Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics*, 7, 149-179.
- Lindh, J. (2006). Preliminary Descriptive F0-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. In: *Proceedings of Interspeech 2007*, 2025-2028.
- Markham, D. (1999). Listeners and disguised voices: the imitation and perception of dialectal accent. *Forensic Linguistic* 6(2), 289-299.
- Masthoff, H. (1996). A report on a voice disguise experiment. *Forensic Linguistics*, 3, 160-167.
- 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.
- Neuhauser, S. & Simpson, A. P. (2007). Phonetic correlates of accent authenticity in voice disguise. *Proceedings of IAFPA 2007*.
- Neuhauser, S. (2008). Voice disguise using a foreign accent: phonetic and linguistic variation. *The International Journal of Speech, Language and the Law*, 15, 131-159.
- Nolan, F. (1999). Speaker Recognition and Forensic Phonetics. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences*, 744-767. Oxford: Blackwell Publishers.
- Nolan, F. & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *Speech, Language and the Law*, 12, 143-173.
- Orchard, T. L. & Yarmey, D. (1995). The Effects of Whispers, Voice-Sample Duration, and Voice Distinctiveness on Criminal Speaker Identification. *Applied Cognitive Psychology*, 9, 249-260.
- Pettorino, M.; Hemmerling, D.; Vitale, M. & De Meo, A. (2018). Towards a Speech-Test for Parkinson's Disease detection: a Diachronic Study on Michael J. Fox. *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, 400-404.
- Pfizinger, H. R. (1996). Two approaches to speech rate estimation. *Proceedings of SST '96*, 421-426.

- Pollák, P., Volín, J. a Skarnitzl, R. (2007). HMM-based phonetic segmentation in Praat environment. *Proceedings of XIIIth "Speech and computer – SPECOM 2007"*, 537–541.
- Růžičková, A. (2016). Strategie maskování hlasu u českých mluvčích. Nepublikovaná bakalářská práce. Praha: Fonetický ústav FF UK.
- Schindler, C. & Draxler, C. (2013). Using Spectral Moments as a Speaker Specific Feature in Nasals and Fricatives. *Proceedings of Interspeech 2013*, 2793-2796.
- Skarnitzl, R. (2014). Forezní fonetika. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvčího*, 11-20. Praha: Univerzita Karlova v Praze, Filozofická fakulta.
- Skarnitzl, R. & Hývlová, D. (2014). Statistický popis hodnot základní frekvence. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvčího*, 49-64. Praha: Univerzita Karlova v Praze, Filozofická fakulta.
- Skarnitzl, R.; Lazárková, D.; Nechanský, T. & Šturm, P. (2014). Vokalické formanty. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvčího*, 21-48. Praha: Univerzita Karlova v Praze, Filozofická fakulta.
- Skarnitzl, R.; Šturm, P. & Volín, J. (2016). *Zvuková báze řečové komunikace*. Praha: Nakladatelství Karolinum.
- Skarnitzl, R. & Vaňková, J. (2017). Fundamental frequency statistics for male speakers of Common Czech. *Acta Universitatis Carolinae – Philologica 3, Phonetica Pragensia XIV*, 7–17.
- Van Summers, W.; Pisoni, D. B.; Bernacki, R. H.; Pedlow, R. I. & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America*, 84, 917-928.

- Vaňková, J. & Bořil, T. (2014). Telefonní přenos. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvího*, 104-115. Praha: Univerzita Karlova v Praze, Filozofická fakulta.
- Volín, J. & Bořil, T. (2014). Základní frekvence v konturách a průbězích. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvího*, 65-76. Praha: Univerzita Karlova v Praze, Filozofická fakulta.
- Volín, J. (2017). Appeal and disrepute of the so-called global rhythm metrics. *AUC Philologica*, 2017/3, 79-94.
- Weingartová, L. (2013). Rhythm metrics for speaker identification in Czech. *AUC Philologica 1/2014, Phonetica Pragensia XIII*, 33-42.
- Weingartová, L.; Bořil, T. & Vaňková, J. (2014). Spektrální sklon. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvího*, 77-94. Praha: Univerzita Karlova v Praze, Filozofická fakulta.
- Weingartová, L. & Volín, J. (2014). Temporální charakteristiky. In R. Skarnitzl (Ed.), *Fonetická identifikace mluvího*, 95-103. Praha: Univerzita Karlova v Praze, Filozofická fakulta.
- Weingartová, L. (2015). *Identifikace mluvího v temporální doméně řeči*. Nепublikovaná disertační práce. Praha: Fonetický ústav FF UK.
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*, 2016.

PŘÍLOHA Č. 1

Text pro předčítání maskovaným i nemaskovaným hlasem

Před několika dny kanadští lékaři zveřejnili zprávu, že bydlet u rušné komunikace s sebou nese problémy s pamětí a riziko demence. Nyní s poklesem pozornosti přicházejí Američané. Jejich závěry se ale týkají opeřenců.

I když se města od sebe v ledasčem liší, hluk z dopravy mají společný. Navzdory tomu na rorýse, rehka, vrabce, hrdličku, straku, holuba hřivnáče i jiné druhy už mimo město téměř nenarazíme. Ptáků, kterým je s námi dobře, je stále víc. Dokonce i těch, u nichž to čekal jen málokdo. O sovách panuje představa, že vyvádějí mladé v hlubokém lese. Z této pohádky nás před časem vyvedli pušticí pruhovaní. V centrech měst se i těmto obrům s téměř metrovým rozpětím křídel daří lépe než za městem. A nejde o ojedinělé případy. Kamerové systémy se staly svědkem, že sovy nechodí do měst jen lovit, ale že zde úspěšně hnízdí a vyvádějí četné městské potomstvo. Znalcům antiky to asi připomnělo rčení „nosit sovy do Athén“. To bývala narážka na množství tamních škol a koncentrace učenců v dobách, kdy Athény byly centrem vzdělanosti. Základ průpovědi ale neměl podtext jízlivosti. Jak ve městě, tak na skalnatých svazích Akropole bývalo sov habaděj.

Zatímco antika si těchto ptáků cenila a dávala je do městského znaku a na mince, naše křesťanská kultura jim přiřkla věci nečisté a zlé. Houkáním měly například věštít smrt. Ještě dnes slyšíme „nesýčkůj“ ve smyslu „nestraš“ a „nepřivolávej neštěstí, zlo a nepříjemnosti“.

I jestřábovým dravcům město nabízí lákadlo v podobě přemnožených holubů. Přesto tito ptáci mají s hnízděním v městských aglomeracích problém a potomstvo zde vyvedou jen zřídka. Ne že by jim vadil hluk nebo přítomnost lidí. Jejich problém je jejich potrava. Měštští holubi, kterým se přezdívá létající krysy, jsou semeništěm všemožných parazitů a mikrobů. Stávají se tak pro hnízdící dravce pohromou. Jejich mláďata útoky parazitů a mikrobiální infekce natolik oslabují, že výsledkem je pokles populace v celém širokém okolním regionu.

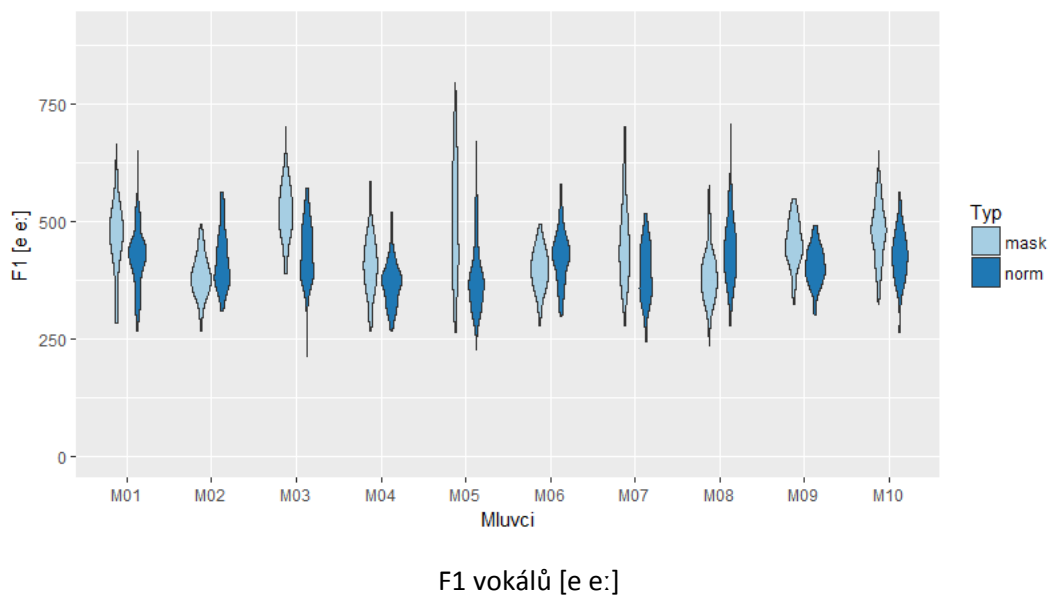
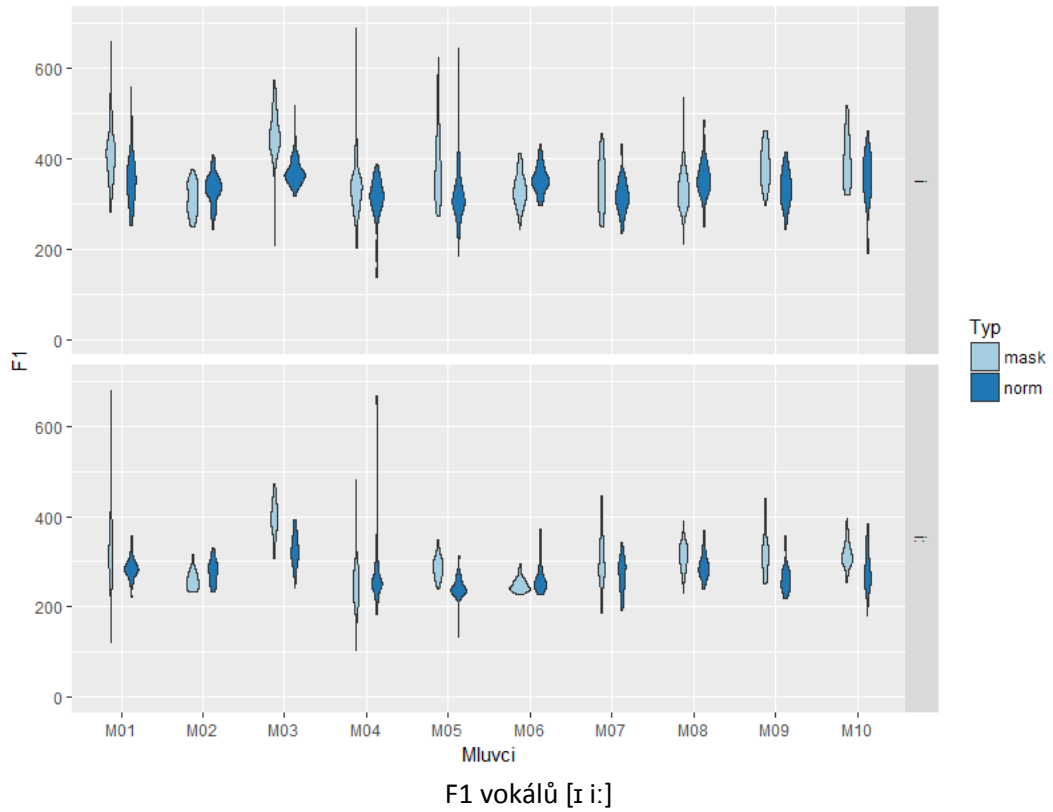
PŘÍLOHA Č. 2

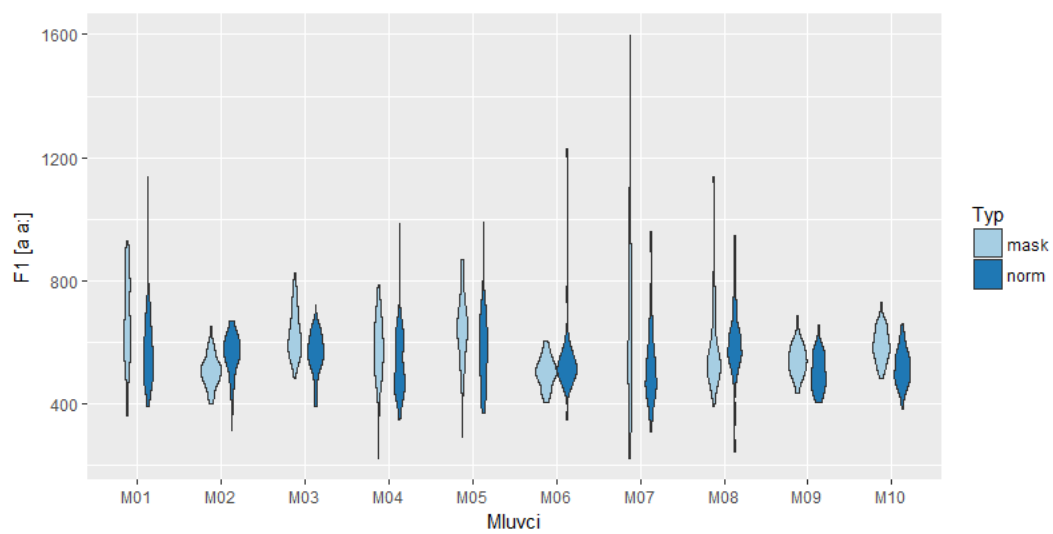
Charakteristiky nemaskovaných a maskovaných projevů jednotlivých mluvčích

| Mluvčí | Nemaskovaný projev | Maskovaný projev |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| M01 | - hlubší hlas - variabilní melodie | - tlačená fonace, chrapot - značný důraz na přízvučné slabiky - podobné melodické průběhy uvnitř i na konci nádechových úseků – klesání - vyšší tempo |
| M02 | - poměrně vysoký hlas - u neukončujících melodémů ve finální pozici stoupání, u ukončujících klesavých melodémů klesání, jinak poměrně plochá intonace - výrazné koncové dloužení V - posun artikulace vokálů mírně dopředu | - snížení hrtanu, dojem hlubšího hlasu - kořen jazyka posunutý dozadu - klesání i v neukončujících melodémech |
| M03 | - vysoký hlas - variabilní melodie | - dojem vyššího hlasu - nazalizace - zvýšený kořen jazyka |
| M04 | - hlubší hlas - klesavá intonace na konci nádechových úseků - poměrně vysoké tempo | - zvýšený kořen jazyka - plochá melodie - tlačená fonace, napjatý hrtan (dojem většího mluvního úsilí) - delší pauzy |
| M05 | - průměrná výška hlasu - variabilní melodie | - vysoká f0 - odlišné intonační kontury, melodie méně variabilní, často delší setrvávání na jedné úrovni - vyšší tempo |
| M06 | - hlubší hlas - mírná nazalizace - výrazné [s](možný vliv podmínek při nahrávání) | - snížení hrtanu, kořen jazyka vzadu - dojem většího vzrůstu a hlubšího hlasu |
| M07 | - průměrná výška hlasu - variabilní melodie - u neukončujících melodémů ve finální pozici stoupání, u ukončujících klesavých melodémů klesání - poměrně nízké tempo | - vyšší f0 - zvýšený hrtan - nazalizace - výrazné protahování přízvučných vokálů a vokálů ve finální pozici - výrazně variabilní melodie |
| M08 | - vyšší hlas - variabilní f0 - pečlivá artikulace | - imitace ruského přízvuku - výrazné přízvukování - sevřená čelist - laryngalizace - chrapot (místy tlačená fonace) |
| M09 | - hlubší hlas - nepříliš variabilní melodie - poměrně vysoké tempo - dojem artikulace s fixovanou čelistí | - vyšší f0 - nazalizace - zvýšení hrtanu - na konci nádechových úseků plochá intonace |
| M10 | - vyšší hlas - plošší intonace - nazalizace | - zvýšení f0 - variabilnější melodie - místy „skřípot“ (tlačená fonace) |

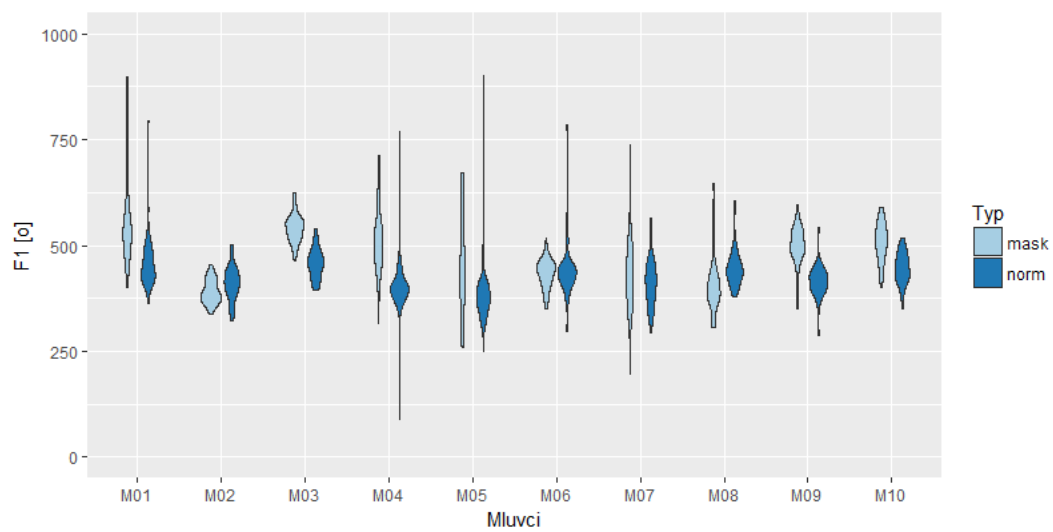
PŘÍLOHA Č. 3

Hodnoty formantů (Hz) vyčíslené pro jednotlivé vokalické kvality

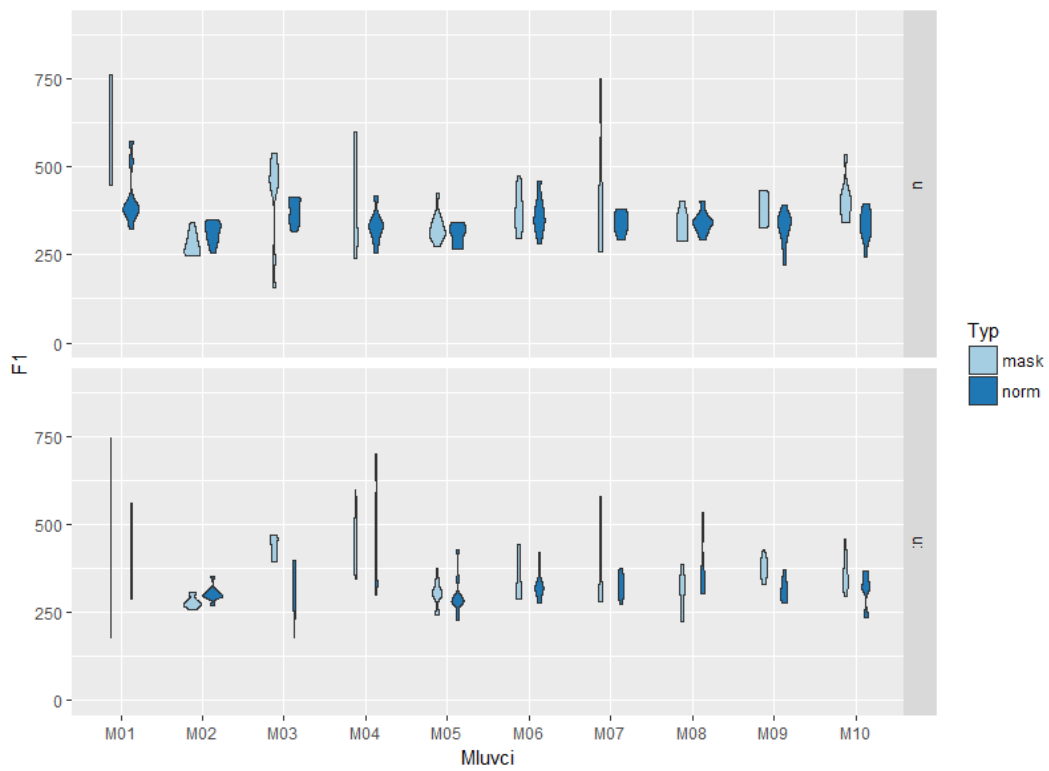




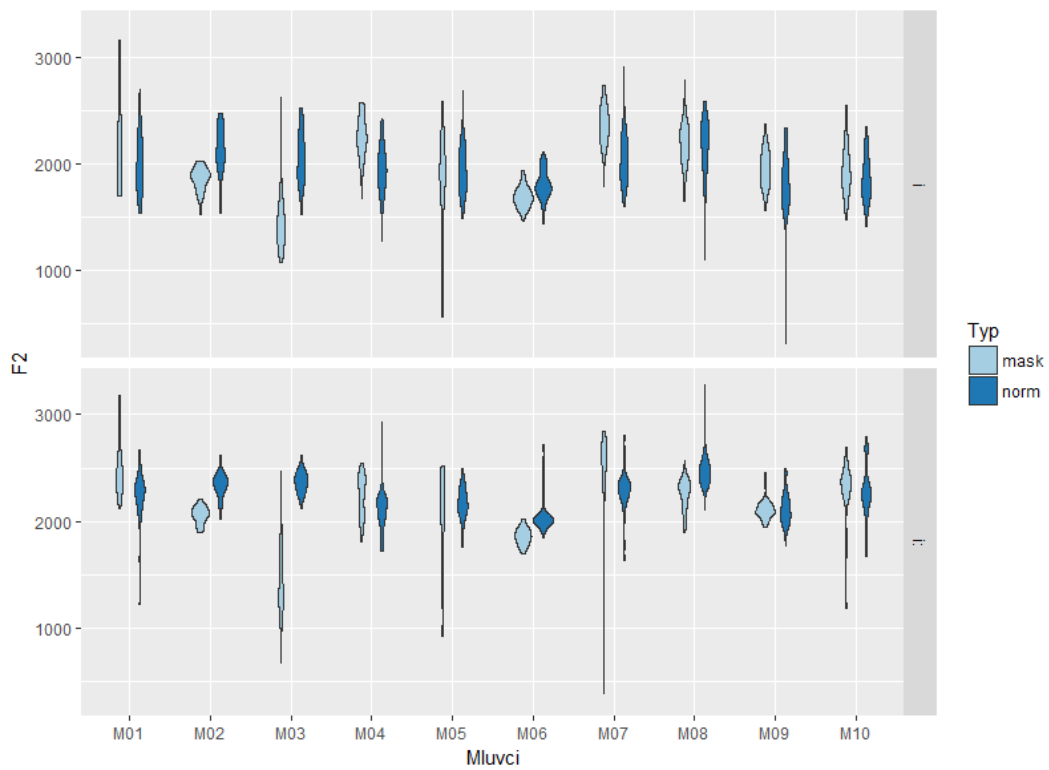
F1 vokálů [a a:]



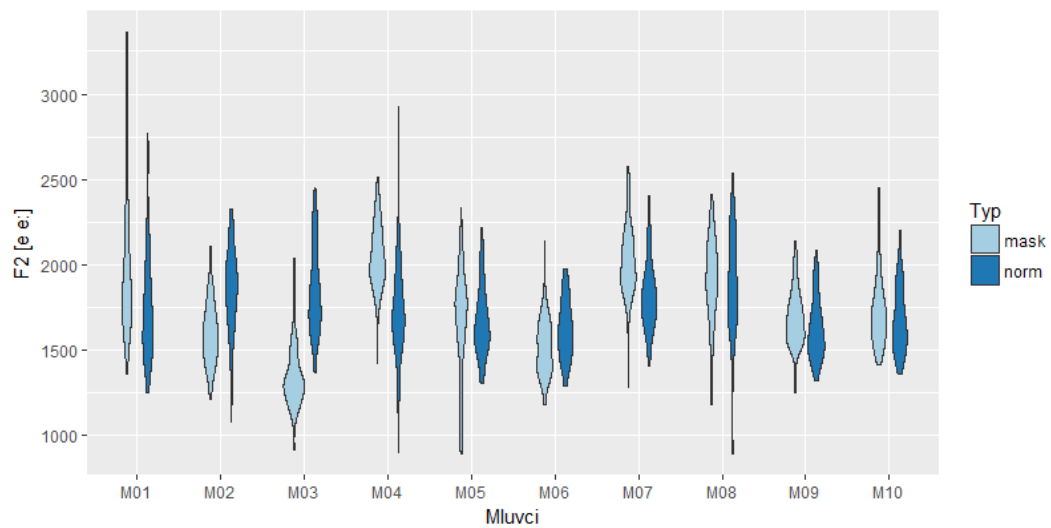
F1 vokálu [o]



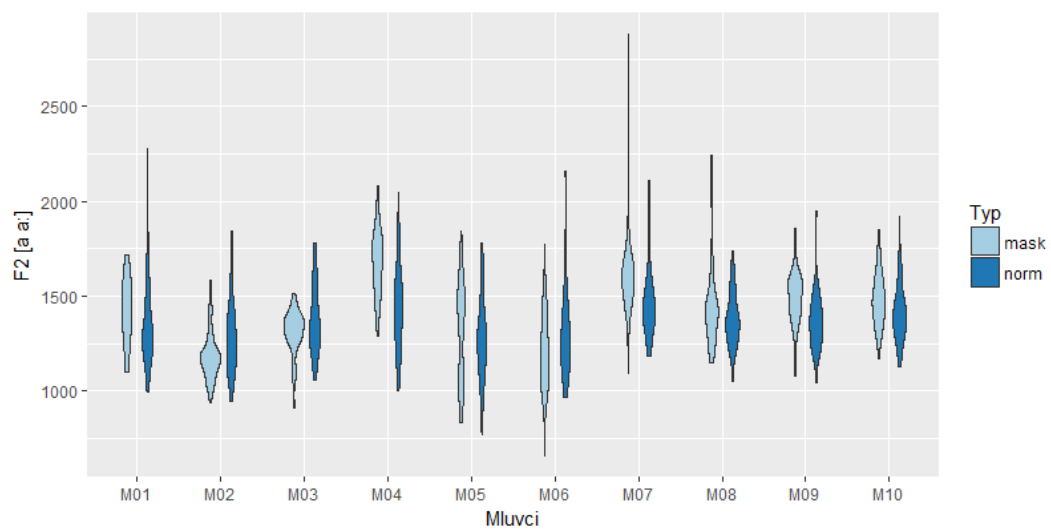
F1 vokálů [u u:]



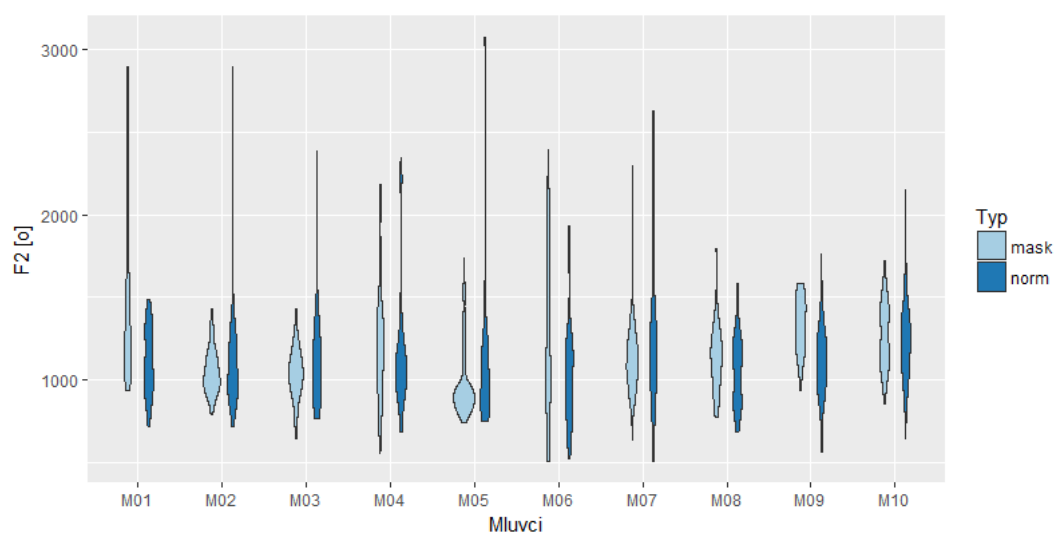
F2 vokálů [ɪ i:]



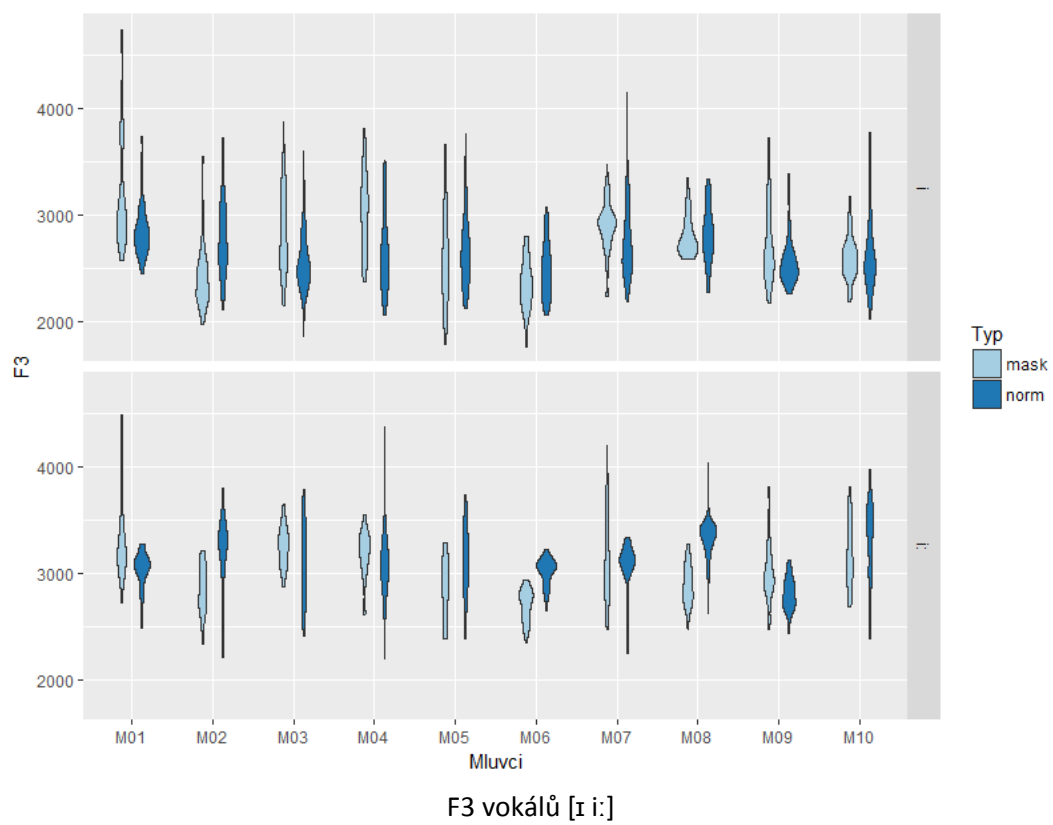
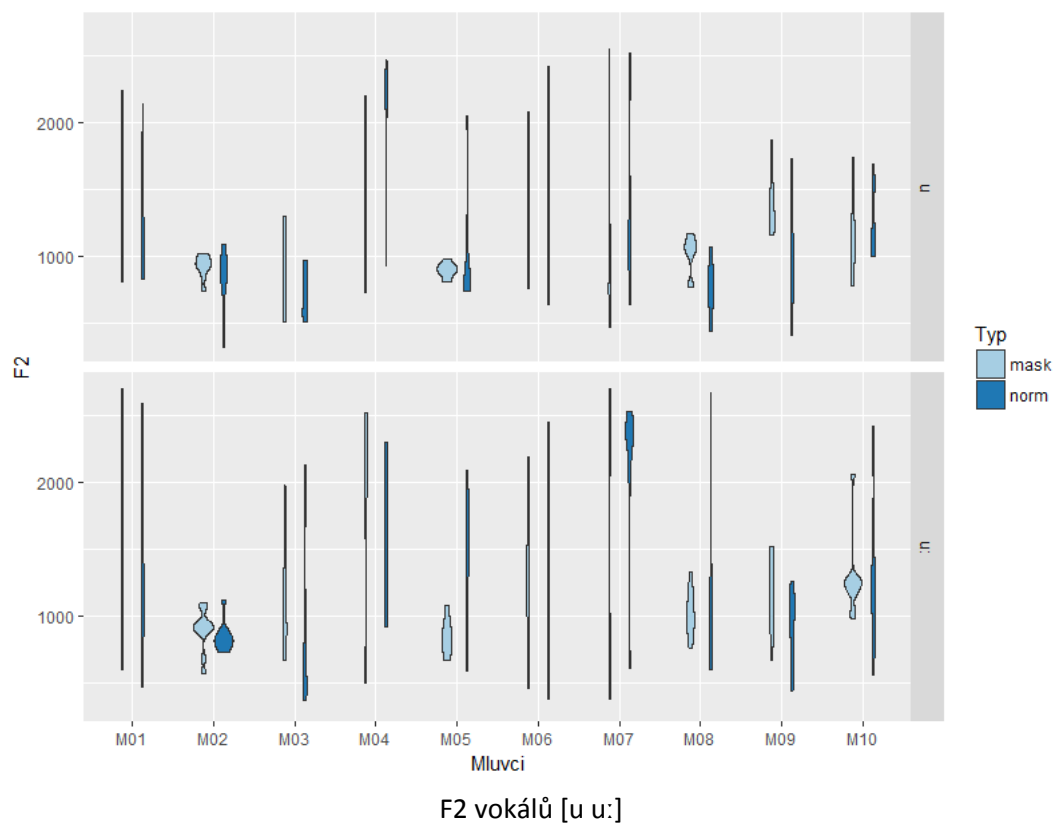
F2 vokálů [e e:]

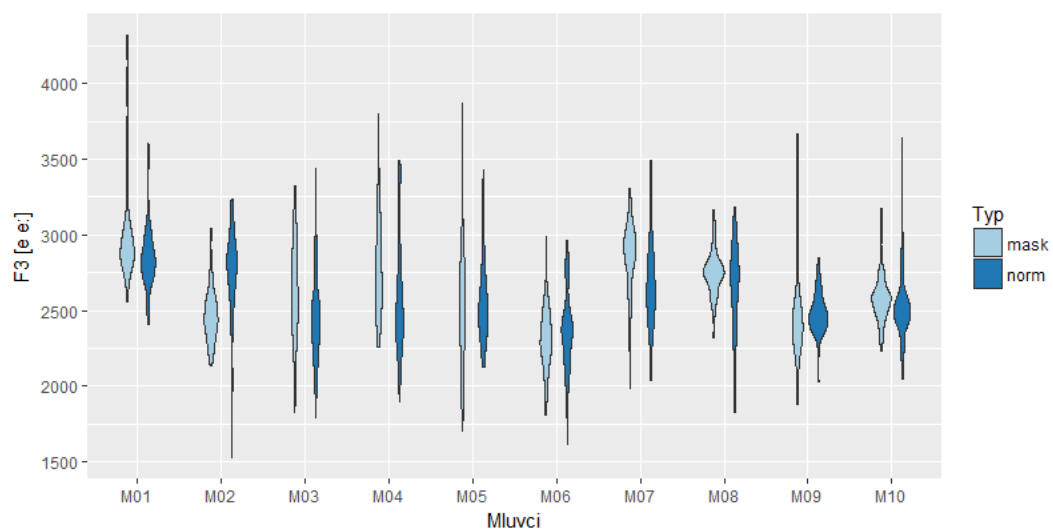


F2 vokálů [a a:]

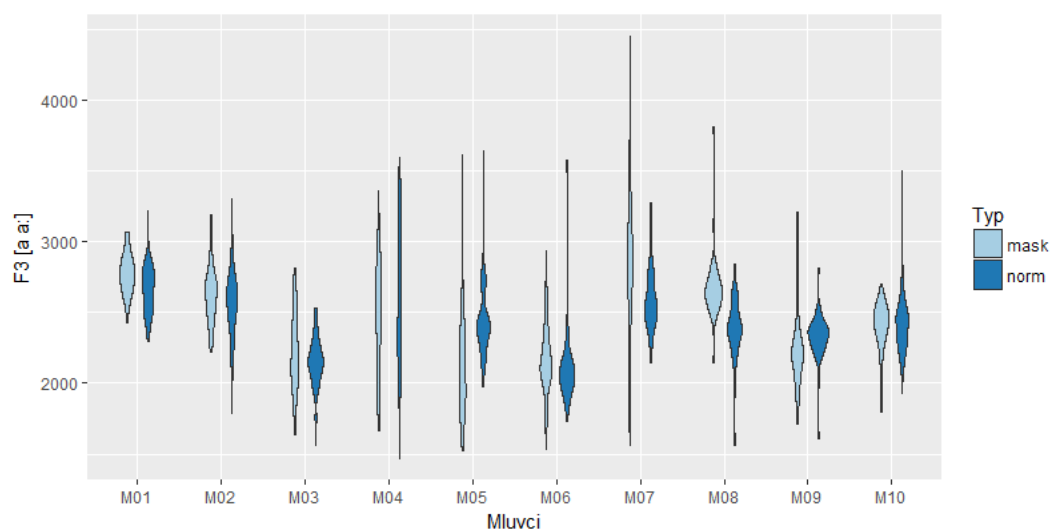


F2 vokálu [o]

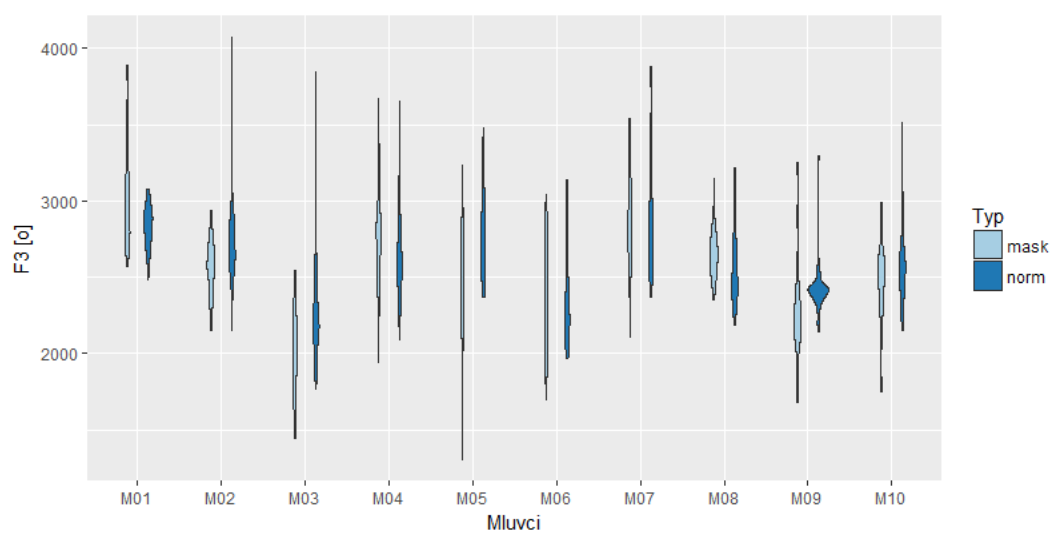




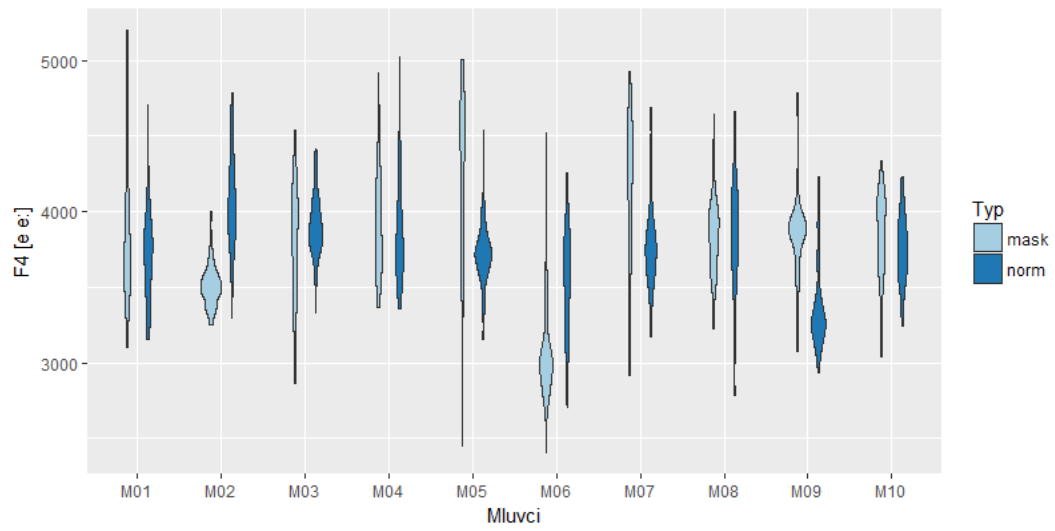
F3 vokálů [e e:]



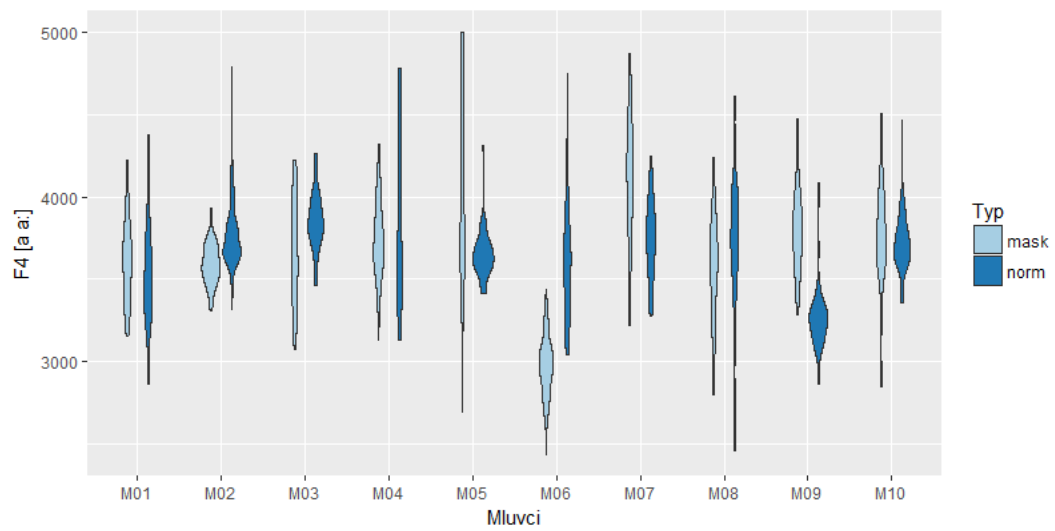
F3 vokálů [a a:]



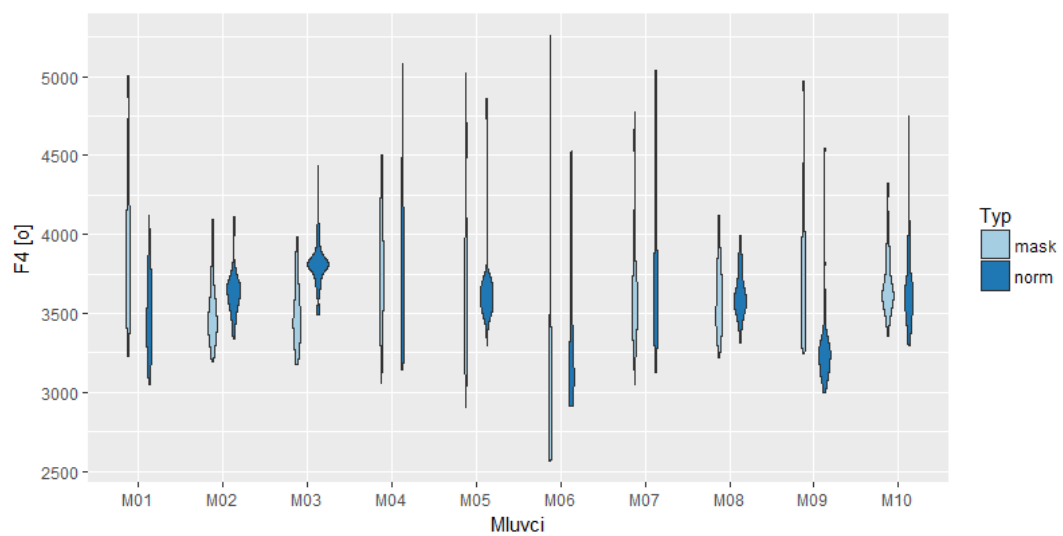
F3 vokálu [o]



F4 vokálů [e e:]



F4 vokálů [a a:]



F4 vokálu [o]

