Univerzita Karlova v Praze
Filozofická fakulta
Fonetický ústav

Diplomová práce

Lenka Weingartová

Ukazatele identity mluvčího v oblasti temporálních modulací řečového signálu

Speaker identity indicators in the domain of the temporal modulation of the speech signal

Praha, 2011

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

Poděkování

Největší dík patří vedoucímu práce, Janu Volínovi, jehož bezmezná trpělivost, ochota a inspirace mě hnala kupředu. Dále děkuji Radku Skarnitzlovi a Hynku Bořilovi za odborné konzultace, Janu Harvalíkovi za pomoc v prvních krůčcích praatového skriptování, Františku Vlasákovi a Petře Pecharové za rady ohledně statistiky a spoustu užitečných připomínek. Nakonec bych ráda vyjádřila vděk také svým rodičům za jejich podporu.

Prohlášení

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 20. 1. 2011

Abstrakt

Tato diplomová práce se zabývá rozpoznáváním mluvčího, a to konkrétně v oblasti temporálních změn v řečovém signálu. Po krátkém úvodu do forenzní fonetiky podává přehled přístupů a faktorů, které napomáhají nebo naopak zabraňují úspěšnému rozpoznání. Následně jsou představeny současné přístupy k temporální struktuře řeči a především k metodám její analýzy. Praktickou část práce pak tvoří experiment, který zjišťuje přínos některých temporálních ukazatelů k rozpoznávání mluvčího. Tyto ukazatele jsou %V (poměrné zastoupení vokalických intervalů ve větě), ΔV a ΔC (směrodatná odchylka vokalických, respektive konsonantických intervalů v rámci věty), VarcoV a VarcoC (normalizace předchozích ukazatelů vzhledem k průměrnému trvání daných intervalů) a indexy párové variability (PVI) pro vokalické i konsonantické intervaly, normalizované i nenormalizované. Kromě toho je k zachycení lokálních změn tempa a obzvláště závěrového zpomalování použit ukazatel LAR (převrácená hodnota vzdálenosti středů dvou následujících vokalických intervalů). Zatímco první zmíněné ukazatele nejsou v rozlišení mluvčích příliš úspěšné, LAR se zdá být velmi dobrým nástrojem pro zachycení individuálních rysů mluvčích. Pro praktické využití tohoto ukazatele bude ale potřeba další výzkum, zejména na větším vzorku mluvčích.

Klíčová slova: forenzní analýza, rozpoznávání mluvčího, identita mluvčího, temporální charakteristiky, temporální struktura, artikulační tempo

Abstract

This diploma thesis aims to contribute to the field of speaker recognition in the domain of temporal changes in the speech signal. After a brief introduction into forensic phonetics, it gives an outline of approaches and factors which help or hinder successful recognition. The focus is then shifted to the temporal structure of speech and approaches to its analysis currently in use. The practical section of this thesis consists of an experiment designed to assess the contribution of certain temporal measures to speaker recognition. The variables used here are %V (the proportion of vocalic intervals within a sentence), ΔV and ΔC (the standard deviation of the duration of vocalic/consonantal intervals within a sentence), VarcoV and VarcoC (the previous variables normalised for average interval duration) and the Pairwise Variability Indices, both vocalic and consonantal, raw and normalised. Beside these, another variable is used to capture the local articulation rate and especially final deceleration in the utterances – LAR (the inverse of the distance between successive midpoints of the vocalic intervals). Whereas the first mentioned variables are not very successful in distinguishing the speakers, LAR seems very well suited for capturing speaker idiosyncrasies, although further research (with more speakers above all) will be needed to evaluate the effectiveness of this measure for practical use.

Keywords: forensic analysis, speaker recognition, speaker identity, temporal characteristics, temporal structure, articulation rate

Obsah

1. Úvod

Rozpoznávání mluvčího je technika identifikace osoby podle jejího hlasu. Využívá akustické prvky řečového proudu odlišující mluvčí od sebe navzájem. Často bývá zaměňováno s čím dál tím více používaným rozpoznáváním řeči. Nejde však o jeden a ten samý problém – při rozpoznávání mluvčího nezjišťujeme, co bylo řečeno, ale kdo to říká. Obsah výpovědi je pak vlastně „šum“ zastírající námi požadované informace. Proto lze rozpoznávání řeči označit za úlohu do jisté míry opačnou.

Hypotéza, ležící v základech rozpoznávání mluvčího, je tvrzení, že existují řečové rysy, které jsou pro daného mluvčího jedinečné a mluvčího podle nich můžeme jednoznačně odlišit. K úspěšnému rozpoznání je nutné tyto informace specifické pro jednotlivé mluvčí extrahovat z proudu řeči a podle nich rozhodnout, která osoba mluví. Úlohu rozpoznat mluvčího provádí náš mozek spolu se sluchovým aparátem s většími či menšími úspěchy téměř každodenně.

Když se však pokusíme totéž naučit software, zjistíme, že tento problém zdaleka není triviální. Protože i přesto, že jednotliví mluvčí se od sebe liší tělesnou anatomií – v našem případě zejména tvarem vokálního traktu – a naučenými vzorci chování, které se projevují například v průměrné výšce hlasu či stylu řeči, není ve většině případů vůbec jednoduché získat jednoznačné akustické koreláty těchto až příliš obecně formulovaných parametrů. Z hudební teorie byl přejat pojem „hlasový rejstřík“, o kterém se zdálo, že by mohl sloužit při popisu komplexních hlasových charakteristik, ale kvůli nedostatku jasné definice tohoto pojmu a různících se použitích se ve fonetice neujal (Laver, 1980, str. 93-94).

Rozpoznávání mluvčího se obvykle rozděluje na dvě podoblasti: verifikaci a identifikaci mluvčího. Verifikací je míněna odpověď ano či ne na otázku, zda skutečně mluví daná osoba. V praxi má její využití nesmírný potenciál, zejména v bezpečnostních systémech – povolení či omezení přístupu určitému člověku nebo skupině lidí, například do zabezpečených budov, databází, přístup k bankovnímu účtu přes telefon a

podobně. Protože v těchto případech mluvčí obvykle spolupracuje (musí například vyslovit správné heslo) a chce být rozpoznán, vykazují verifikační úlohy mnohem vyšší úspěšnost než identifikační. Dalším důvodem jsou také kontrolované podmínky (např. neexistence šumu) či vysoká kvalita nahrávacích zařízení.

Identifikace pak znamená zjištění identity neznámého mluvčího. Při verifikaci se vždy porovnává jedna promluva vůči jedné šabloně a výsledkem je shoda či neshoda. Při identifikaci je takových šablon více a hledá se ta s nejlepší shodou. Ačkoliv je nasnadě, že identifikace bude technicky mnohem náročnější, její výhodou (a zároveň i následným zdrojem nesnází) je, že spolupráce mluvčího není nutná, obvykle stačí jakýkoliv vzorek řeči. Při identifikaci je klíčovým faktorem velikost populace – čím větší, tím rychleji vzrůstá pravděpodobnost chyby. Proto se nejčastěji rozpoznává takzvaně v uzavřené množině, kde je rozpoznávaný vzorek řeči konfrontován jen s několika dalšími. Identifikace bývá součástí forenzních postupů při určování totožnosti osob, ale je možné ji využít například také k rozlišení jednotlivých mluvčích v zaznamenané konverzaci nebo při zpracovávání informací pro zpravodajské služby.

Přístupy k rozpoznávání mluvčího je také možné podle využitých prostředků rozdělit na sluchově-percepční (rozpoznávání provádí posluchač) a automatické (rozpoznávání nebo alespoň jeho část provádí počítač). Oběma přístupům bude v této práci věnována samostatná kapitola.

V praxi se automatické (méně často i sluchové) rozpoznávání mluvčího často kombinuje ještě s dalšími biometrickými způsoby identifikace, např. analýzou tváře, optickou biometrií (rozpoznávání podle duhovky nebo sítnice), daktyloskopií, apod. Výhodou biometrických atributů je, že na rozdíl od hesel, klíčů nebo čipových karet nejdou ztratit ani je nelze zapomenout – a těžko se falšují.

K technické stránce automatického rozpoznávání mluvčího jsou nejčastěji využívanými algoritmy skryté Markovovy modely, gaussovské

směsi, algoritmy rozpoznávání vzorů, neuronové sítě, maticové reprezentace nebo rozhodovací stromy. Jde o statistické algoritmy, které mají za úkol porovnávat vstupní a referenční data (čili vzorek řeči rozpoznávaného mluvčího s nějakými dalšími) a rozhodnout, do jaké míry se shodují.

Lingvistiku však daleko více zajímají základní data, jež jsou těmito metodami dále analyzována a klasifikována – které parametry vedou k úspěšnému rozpoznání a do jaké míry jsou užitečné? Které naopak můžeme rovnou zanedbat? Kromě praktické využitelnosti jde o poznatky, jež nám mohou prozradit mnoho nového o lidské řeči a její produkci i percepci.

1.1 Struktura práce

V úvodu této práce se věnuji historii rozpoznávání mluvčího a forenzní fonetiky vůbec. V kapitole 3 jsou rozebrány různé přístupy k rozpoznávání a popsány užitečné faktory, nebo naopak ty, které identifikaci ztěžují. Dále jsou představeny některé současné fungující metody. Kapitola 4 je věnována temporální struktuře řeči a zejména způsobům, jak ji analyzovat a kvantifikovat na jiném než percepčním základě. Praktickou část práce pak tvoří experiment, při kterém jsou temporální ukazatele využity k diskriminaci mezi třemi mluvčími ženského pohlaví. Následuje shrnutí výsledků, jejich interpretace a diskuse.

Přílohy tvoří skripty pro fonetický program Praat využité k extrakci dat, nahrávky ve formátu .wav, soubory typu TextGrid a tabulky s extrahovanými daty ve formátu .xls.

2. Stručný vhled do historie rozpoznávání

Rozpoznávání mluvčího sluchem je schopnost, která se v lidském mozku musela vyvinout ještě dávno před tím, než vůbec vznikla řeč jako taková. K ilustraci tohoto tvrzení nejlépe poslouží následující citát Harryho Holliena (2002):

> V jeskyni byla opravdu tma. Dřevo docházelo a uhlíky v ohni už ani nedoutnaly. Samorost byl ztahaný jako pes, byl pryč dva dny a vracel se s prázdnýma rukama. Jednooká by se měla co nejdřív postarat o oheň, jinak bude muset jít a vypůjčit si uhlíky od té bandy dole pod kopcem. Právě přemítal, zda má pro něj Jednooká něco k jídlu, když ucítil, jak mu vstávají chlupy na krku a zádech. Někdo nebo něco bylo uvnitř v jeskyni. Samorost se vrhl ke kraji skalní římsy; kde jen nechal svou sekeru? Náhle však zaslechl: „Gaval, gaval, uga," a málem se zhroutil úlevou. Byl to Zkroucená čelist. „Grun," odpověděl Samorost rychle, aby nedostal ránu do hlavy on. Doufejme, že stařík přináší alespoň tlustého králíka nebo nějakého pěkného hada. (str. 17, překlad autorka)

Jak vidno, rozpoznávání mluvčího byla pro naše předky schopnost životně důležitá, přestože si nejspíš dlouho neuvědomovali, že k něčemu takovému vůbec dochází. Evoluci pak vděčíme za to, že se tato schopnost zachovala i v současné populaci.

Důležitost správného rozpoznání a potřebnost solidní vědecké metody se začala projevovat až v novověku s nástupem forenzních věd. Při zajišťování důkazů bylo často potřeba se spolehnout pouze na svědectví, jež bylo získáno sluchem. Dá se však takovému svědectví věřit? A do jaké míry?

Sluchové svědectví mělo vždy u soudů jiné postavení než svědectví očité (je tomu tak ostatně i dnes). Hollien nicméně uvádí případ z roku 1907 v USA – první velký případ zahrnující sluchovou identifikaci – kdy znásilněná dívka poznala po hlase útočníka a soudce na základě tohoto svědectví odsoudil podezřelého (Hollien, 2002, str. 19). Tento proces se pak v USA stal precedentem pro mnoho dalších, což podnítilo další výzkum rozpoznávání.

Největší rozmach zažila identifikace mluvčího ve dvacátém století s rozvojem relevantních technologií. Již desítky let na tomto problému pracují například Bellovy laboratoře v New Jersey. Dnes se ve větší či menší míře rozpoznáváním zabývají pracoviště po celém světě.

Historie automatického rozpoznávání mluvčího je v porovnání se sluchově-percepční metodou velmi krátká (sahá teprve něco přes padesát let do minulosti) a samozřejmě spjatá s rozpoznáváním řeči jako takové. První pokusy o rozpoznávací přístroje se začaly objevovat v padesátých letech minulého století a rozpoznávaly vyřčená čísla s pomocí změn základní frekvence během promluvy.

Spolu s technologickým pokrokem a s velkým komerčním potenciálem, který byl hlavně u verifikace mluvčího bryskně rozpoznán, se začaly tyto obory velmi rychle rozvíjet. Lepší nahrávací zařízení zvýšilo kvalitu nahrávek, které mohly být dále detailněji analyzovány, zlepšila se extrakce rysů ze signálu. Zároveň bylo díky vývoji nového softwaru možné implementovat komplexnější rozhodovací algoritmy.

Hollien ovšem uvádí smutný fakt, že práce často narážela na problémy pragmatičtějšího charakteru – mnoha slibným výzkumným projektům bylo komerčními sponzory zastaveno financování, protože nevykazovaly dostatečně rychlý pokrok (Hollien, 2002, str. 136).

Zpočátku bylo nutné prozkoumat, které řečové rysy jsou pro rozpoznání klíčové a které naopak lze zanedbat. Velmi slibnými se ukázala vokalická a nazální spektra či spektrální sklon hlasivkového tónu, jak prokázala například studie provedená v roce 1972 v Boltových, Berankových a Newmanových laboratořích (Wolf, 1972). V jejich dalších výzkumech pak byl zahrnut i vliv zhoršené kvality signálu – např. šumem nebo přenosem přes telefon – a použity vektory zahrnující krátkodobá spektra získaná z LPC koeficientů, kepstrální koeficienty a další.

V 80. letech se pak objevily nové techniky zpracování signálu a nové rozhodovací algoritmy – vektorová kvantizace, skryté Markovovy modely a neurální sítě. První z nich se používala například v Bellových laboratořích (Soong, Rosenberg, Rabiner & Juang, 1985), s úspěšností vyšší než 98%.

Neurální sítě použil ve své práci například Hattori (1992), který také dosáhl vysoké úspěšnosti (95,7–100%) – zlepšovala se s počtem provedených tréninkových iterací. Webb a Rissanen (1993) zkoumali využití skrytých Markovových modelů pro rozpoznávání a také vliv velikosti populace na jejich úspěšnost a dosáhli průměrného výsledku 98,7% správných rozpoznání, přičemž s přibývajícími mluvčími toto procento klesalo. Z dalších algoritmů byly vyzkoušeny ještě například spojité pravděpodobnostní akustické mapy s využitím funkce hustoty pravděpodobnosti (Tseng, Soong & Rosenberg, 1992), které vykázaly dokonce ještě vyšší úspěšnost.

Tento přehled vypadá sice nesmírně optimisticky, nicméně je třeba si uvědomit, že většina těchto výzkumů probíhala v laboratořích, se špičkovým vybavením a téměř naprostou kontrolou všech podmínek. Při praktickém použití například ve forenzní realitě se takto vysoká úspěšnost rozhodně nedá předpokládat. Podle Holliena by optimálním řešením bylo vytvořit týmy složené z inženýrů, fonetiků a forenzních specialistů, kteří by se přímo zaměřovali na forenzní problémy (Hollien, 2002, str. 140).

Nevýhodou těchto přístupů pro lingvistiku je ale jejich neprůhlednost. V současnosti se například pro rozpoznávání velmi často využívají takzvané kepstrální koeficienty, získané z Fourierovy transformace krátkých úseků signálu, jejich následného zlogaritmování a nakonec inverzní Fourierovy transformace. Koeficienty se pak použijí jako vstupy pro některý ze statistických modelů. Ovšem k jakým akustickým jevům přesně se tyto koeficienty vztahují, je velmi obtížné popsat, z výsledků pouze vyplývá, že nějakým způsobem nepřímo zachycují vlastnosti vokálního traktu člověka.

Jako každý vědní obor, ani rozpoznávání mluvčího se nevyhnulo přehmatům a slepým uličkám. Ta nejmarkantnější z nich, která získala velkou oblibu v USA v padesátých a šedesátých letech minulého století, byly tzv. hlasové otisky. Název jasně napovídá analogii s otisky prstů – ve skutečnosti ale šlo o metodu vizuálního porovnávání spektrogramů a vyhledávání charakteristických rysů, které by se daly využít k rozpoznání

mluvčího. (Vzhledem k použití spektrogramů se tedy jedná o první pokus o poloautomatické rozpoznávání.) Tento směr výzkumu nicméně trpěl nedostatečnou vědeckou průkazností a jeho zastánci se opírali pouze o své vlastní experimenty (Hollien, 2002, str. 121). Přestože otec tohoto přístupu, Lawrence Kersta, inženýr z Bellových laboratoří, prohlašoval, že metoda porovnávání hlasových otisků má 99% úspěšnost (Kersta, 1962), a přirovnával jedinečnost hlasových otisků k otiskům prstů, mnoho fonetiků té doby se proti ní postavilo. Ovšem ještě v roce 1983 se o ní Francis Nolan zmiňuje jako o dlouho a vášnivě diskutované, nicméně v praxi stále používané metodě v USA, Kanadě, Itálii a Izraeli (Nolan, 1983, str. 18-25).

Harry Hollien je na druhou stranu jejím dlouholetým odpůrcem. Na podporu svých argumentů uvádí například několik případů, kdy se příznivci hlasových otisků ošklivě spletli s málem fatálními důsledky pro obžalovaného (Hollien, 1990, kap. 10), a neváhá hlasové otisky označit za nesmysl (Hollien, 2002, str. 24). Doddington k tomu ještě dodává, že je zbytečné používat takto spornou metodu, když lepších výsledků se dá dosáhnout za pomoci lidských posluchačů (Doddington, 1985, str. 1657). Mezi fonetiky i inženýry ale panuje shoda, že budoucnost rozpoznávání mluvčího leží právě v automatické nebo alespoň poloautomatické identifikaci.

3. Přístupy k rozpoznávání mluvčího

Jak bylo zmíněno výše, v dnešní době máme k dispozici dva základní způsoby, jak více či méně spolehlivě identifikovat mluvčího. Za prvé samozřejmě ten, který dennodenně využíváme – rozpoznávání lidským sluchem, tzv. sluchově-percepční identifikace. U soudů se sluchová svědectví využívají stejně jako svědectví očitá (i když mají menší váhu) a většinou se ještě kombinují s profesionální fonetickou analýzou.

Druhým možným způsobem je rozpoznávání s pomocí stroje. Většina současných forenzních fonetiků obvykle kombinuje oba přístupy (Hollien, 2002, str. 13).

V následující části se budu věnovat sluchově-percepční analýze, jejím možnostem a hranicím, následovat bude analýza automatická.

3.1 Sluchově-percepční rozpoznávání

Při sluchově-percepční identifikaci probíhá rozpoznávání lidským sluchovým aparátem, a tedy hlavní roli zde hraje posluchač. Tento druh rozpoznávání vykonáváme každý den při telefonických rozhovorech a různých jiných příležitostech, aniž bychom si toho byli plně vědomi. V kriminalistice, při vyšetřování, je ale často potřeba sluchově-percepční identifikace cílená, ať již prováděná laickým svědkem nebo profesionálním fonetikem.

Doddington (1985, str. 1656) na základě svých zkušeností uvádí, že lidští posluchači jsou schopni podávat dobré rozpoznávací výkony, a to i za zhoršených podmínek, jako je šum nebo degradace spektra.

V České republice se prozatím drtivá většina všech odborných posudků pro vyšetřování opírá právě o sluchově-percepční analýzu (J. Volín, osobní komunikace, 2. února 2010).

3.1.1 Faktory pomáhající při rozpoznání

Shrňme si nejdůležitější řečové faktory, které nesou nějakou část informace nutnou pro úspěšné rozpoznání mluvčího. Tento výčet si neklade

nároky na úplnost, slouží jen k seznámení s nejdůležitějšími rysy, na které se rozpoznávání mluvčího zaměřovalo v průběhu let.

Asi nejvíce slyšitelným a taktéž objektivně jednoduše změřitelným faktorem je **výška hlasu** a její změny. Z každodenní zkušenosti například víme, že ženy mají obvykle vyšší hlasy než muži. Výšku hlasu lze poznat téměř bezprostředně poté, co mluvčího uslyšíme (ať již prostým sluchem nebo s pomocí automatické analýzy), a její akustický korelát – základní frekvence (F_0) – se dá snadno extrahovat i z krátkého úseku promluvy. Na základní frekvenci se během let soustředila velká část výzkumu zabývajícího se suprasegmentálními vlastnostmi jazyka a jejich vlivem na rozpoznávání mluvčího, protože je (poměrně) neměnná při přenosu a snadno se ze signálu automaticky extrahuje (Nolan, 1983, str. 121). Mezi vědci panuje shoda, že základní frekvence je při rozpoznávání jedním z nejdůležitějších parametrů (viz např. Matsumoto, Hiki, Sone & Nimura, 1973, str. 435; Markel, Oshika & Gray, 1977, str. 337; Doherty & Hollien, 1978; Nolan, 1983, str. 124; Hollien, 2002, str. 165).

Základní frekvence má však jednu neoddiskutovatelnou vadu. Pokouší-li se někdo o vědomou změnu hlasu, obvykle je její hlavní součástí právě změna výšky hlasu. To může rozpoznání značně ztížit, a je tedy nutné spolehnout se v takovém případě spíše na jiné faktory.

Jeden z problematičtějších rysů je **hlasitost**. Naměřená intenzita hlasu jako akustický korelát hlasitosti se může lišit například se vzdáleností mluvčího od mikrofonu a samozřejmě ji lze také snadno vědomě měnit. Nicméně, mluvčí s hlasitostí pracují každý jinak a může být prospěšné změny hlasitosti pro účely rozpoznávání také začlenit.

Třetím prozodickým rysem je **rytmus**. Mezi akustické koreláty rytmu využívané pro rozpoznávání Hollien (2002) zařazuje trvání hlásek, slabik, slov či vět, slabičné tempo, délka a četnost pauz, a podobně. Výhodou je snadné získávání a analýza těchto parametrů, které většinou nebývají narušeny ani přenosem signálu. V dalších částech této práce ale uvidíme, že analýza temporálních charakteristik se nemusí omezovat jen na takto přímočaré ukazatele.

Dalším zdrojem užitečných informací jsou **segmentální charakteristiky** řeči. I laický svědek zaregistruje například artikulační vady nebo zvláštnosti, profesionál je pak schopen rozeznat jemné nuance, jedinečné pro mluvčího. Hollien (2002, str. 60) uvádí jako nejdůležitější artikulační rysy produkci konsonantů a vokalické formanty. Mimo řečové vady mohou artikulační zvláštnosti vznikat také například působením dialektu.

Kromě neobvyklé výslovnosti některých hlásek nebo hláskových skupin mohou leccos o identitě mluvčího prozradit i správně vyslovené hlásky, zejména nazály. Nolan (1983, str. 75-77) cituje několik experimentů, které prokázaly užitečnost nazál při identifikaci – viz např. Glenn & Kleiner (1968), Wolf (1972) nebo Höfker (1977). Přitom ještě lepší úspěšnost, než studovat vlastnosti izolovaných nazál, vykazuje nazální koartikulace (například před předními a zadními vokály), jak prokázali Su, Li & Fu (1974).

Tato výlučnost nosových hlásek je způsobena zejména rezonancemi (resp. antirezonancemi) v nosní dutině, jejíž tvar je dán a mluvčí jej nedokáže nijak ovlivnit.

Kromě výšky hlasu existuje ještě další faktor, který si při každodenním rozpoznávání obvykle uvědomujeme – a to tzv. **barva hlasu**. Barva je souhrn vlastností, díky kterému poznáme například zvuk klavíru od trubky, a je dán tvarem, materiálem a velikostí nástroje. Lidé se mezi sebou také liší délkou a tvarem vokálního traktu, což způsobuje, že při vytvoření tónu hlasivkami jsou jeho jednotlivé harmonické frekvence ve zbytku vokálního traktu různě utlumovány nebo zesilovány a tvoří tak hlas, charakteristický pro mluvčího.

Pro různé typy fonace, které je mluvčí potenciálně schopen produkovat a které také mohou charakterizovat jeho promluvy, se vžil název **fonační nastavení**. Základem pro jejich popis se stal deskriptivní systém Johna Lavera (1980).

Barva hlasu i fonační nastavení se dají zachytit dlouhodobými spektry, získávanými zprůměrováním krátkodobých spekter z delších částí promluvy (Nolan, 1983, str. 130).

Dalšími idiosynkratickými rysy, které lze v promluvách vysledovat, může být například dialekt, cizí přízvuk, použitá slovní zásoba, neobvyklé umisťování slovního přízvuku a podobně. Přestože by se mohlo zdát, že náš arzenál pro rozpoznání mluvčího je obsáhlý, Hollien na příkladu dvou bratrů s velmi podobnými hlasy varuje před ukvapenými rozhodnutími na základě distinktivních řečových zvláštností, které mohou svádět k jasnému závěru – ano, je to tentýž mluvčí – ale přitom zakrývat jemné rozdíly, které prozrazují opak (Hollien, 2002, str. 47-48).

3.1.2 Problémy při rozpoznávání

V následující části se budu věnovat faktorům, které obvykle rozpoznávání mluvčího ztěžují a kladou na něj další nároky.

V první řadě je nutno zmínit **paměť**. Je zřejmé, že stejně, jako je tomu u paměti vizuální, postupem času zaslechnuté vjemy zapomínáme. První velký experiment na tomto poli podnikla McGehee (1937, citováno z Hollien, 2002), ve kterém prokázala, že spolehlivost lidského mozku pamatovat si hlasy s časem výrazně klesá – po 5 měsících je dokonce menší než náhodná (pokud ovšem subjekt hlas mluvčího dobře nezná, viz níže). Experiment s podobnými výsledky zopakovala v roce 1944. Následovaly další studie, které oslabování paměti na hlasy potvrzují (jejich přehled viz Hollien, 2002, str. 29-30), ale autoři se plně neshodují na tom, jak přesně křivka „zapomínání“ vypadá. Podle Holliena (2002, str. 31) je příčinou takto nejednotných výsledků přílišné množství proměnných, které do procesu vstupují.

Z toho plyne, že v případě auditivního svědectví je nutné dát pozor na to, jak je to dávno, co svědek hlas mluvčího slyšel, protože po jistém časovém úseku už jeho výpověď může být zcela neprůkazná.

Další problém může nastat, pokud jsou vzorky promluv daného mluvčího **vzdálené v čase**, tzn. nahrané každý jindy. Výzkumu na toto téma není mnoho, ale studie provedená Hollienem a Schwartzovou (2001)

naznačuje, že úspěšnost rozpoznání výrazně klesá až s dvacetiletým rozdílem jednotlivých vzorků promluv. S tím jsou konzistentní i výsledky novější studie Lawsona et al. (2009), kteří zkoumali nahrávky 32 mluvčích během tří let a vliv jejich vzdálenosti v čase na úspěšnost rozpoznávání s pomocí gaussovských směsí. Došli k závěru, že tento vliv je statisticky zanedbatelný. V praktické části práce se tímto problémem budu také okrajově zabývat.

Zkušenost posluchače je faktor, který má samozřejmě vliv pouze při sluchově-percepční analýze. Očekávali bychom, že posluchač, který je na rozpoznávání trénován či má fonetické zkušenosti, bude při rozpoznávání úspěšnější než ostatní. Důkaz, že toto tvrzení skutečně platí, lze najít ve studii Holliena a Schwartzové z roku 2000, ve které posluchači z řad profesionálních fonetiků se zkušenostmi s identifikací mluvčího měli vyšší úspěšnost, a to i za ztížených podmínek (vzorky vzdálené v čase, mluvčí měli podobné hlasy). Viz také Hollien, 2002, str. 34-36.

Podobnost hlasů mluvčích je další z faktorů, který může podstatně snížit úspěšnost identifikace, například pokud mluvčí jsou příbuzní (sourozenci, rodič a dítě, apod.). Ve výše zmíněné studii (Hollien & Schwartzová, 2000) najdeme experiment, ve kterém byla úspěšnost sluchově-percepční analýzy porovnávána také pro vzorky promluv, které zněly podobně. Počet správně rozpoznaných mluvčích zůstal poměrně vysoký (95%) pro vzorky získané současně. V případě vzorků promluv získaných s časovým odstupem klesla úspěšnost na méně než polovinu.

Také je zajímavým faktem, že automatický rozpoznávací systém si téměř nikdy nesplete jednovaječná dvojčata, zatímco lidským posluchačům se to stane téměř vždy, jak uvádí Rosenberg (1973).

Jak pro sluchovou, tak i pro automatickou analýzu je důležitá **délka vzorku** promluvy. Samozřejmě, čím delší je vzorek, tím více „nápověd“ se v něm objeví. A čím kratší, tím těžší bude rozpoznání. Minimální délka požadovaná pro úspěšnou identifikaci bývá většinou 30 sekund (Hollien, 2002, str. 41).

Při použití identifikace v praxi se často nesetkáváme s bezchybnými, studiovými nahrávkami, a přitom je **kvalita vzorku** pro úspěšné rozpoznání naprosto stěžejní. Může být ovlivněná mnoha způsoby – další mluvčí hovořící současně, šepot, použití neobvyklého hlasového rejstříku, jakýkoliv šum – nejen aperiodický zvuk, ale cokoliv, co maskuje signál (v nejhorším případě na podobných frekvencích jako řeč), a v neposlední řadě malá šířka pásma, zapříčiněná obvykle telefonickým přenosem nebo kvalitou mikrofonu. Nedostatečná kvalita vzorku může identifikaci učinit velmi těžkou až zcela nemožnou a jen některé ruchy lze odstranit s pomocí filtrování.

Znalost či neznalost mluvčího může mít na úspěšné sluchové rozpoznání také vliv. Pokud je posluchač s hlasem mluvčího skutečně dobře obeznámen (je to možné předem otestovat), mívá při jeho identifikaci o něco málo větší úspěch. Studií na toto téma je poměrně mnoho, jejich přehled viz např. Hollien, 2002, str. 45-46.

Přejděme k problémům vznikajícím na straně mluvčího. Jeden z nejzajímavějších problémů je bezpochyby úmyslná **změna hlasu**. Je-li úspěšná, může správné rozpoznání zcela znemožnit. Jak bylo již zmíněno výše, jen málokteré rysy nelze vědomě změnit (např. velikost a tvar nosní dutiny) – a navíc je možné ke změně hlasu použít i elektronické přístroje. Dokonce obyčejný šepot stačí ke značnému ztížení identifikace (Orchard & Yarmey, 1995). Hollien (2002, str. 49) uvádí mezi nejčastějšími způsoby maskování hlasu mechanické překážky ve vokálním traktu (tužky, roubíky) nebo změny způsobu fonace (například nazalizovaná řeč). Vyčerpávající přehled výzkumu na toto téma lze najít tamtéž.

Hollien také považuje za nesmírně důležité zjistit, zda se zkoumaný mluvčí skutečně snaží změnit svůj hlas a podle toho postupovat při identifikaci – vyhledávat nekonzistence a místa, kde je maskování oslabeno. Jedinou výhodou při boji s tímto problémem je fakt, že dlouhodobá konzistentní změna hlasu je pro mluvčího nesmírně obtížná – podvědomé řečové chování by totiž musel neustále vědomě modifikovat.

Další nesmírně rozsáhlou a dosud ještě málo prozkoumanou oblastí je **vliv emocí** na rozpoznávání mluvčího. Je-li mluvčí ve stresu nebo pod vlivem silných emocí, jeho hlasový projev se mění, což může identifikaci negativně ovlivnit. Shrnutí informací, známých o těchto jevech, lze najít v Hollienově práci *The Acoustics of Crime* (1990). Stručně zmiňme jen zjištění, že obvykle s vyšším stresem roste F_0 a klesá plynulost řeči.

Také je zkoumán vliv tzv. **Lombardova efektu,** tedy změn řečových parametrů při hovoru v hlučném prostředí (zejména hlasitosti, dále také výšky hlasu nebo tempa), například ve studii Goldenberga, Cohena a Shalloma (2006), kteří konstatovali, že Lombardská řeč může snížit úspěšnost automatického rozpoznávání až o 10%. Navrhují proto při rozpoznání použít pouze takové rysy, které nepodstupují změnu, nebo transformovat tyto změněné rysy zpět.

Všimněme si ještě vlivu **přízvuku** nebo **dialektu** mluvčího na rozpoznávání. Samozřejmě může hodně pomoci, jak v pozitivní, tak i v negativní identifikaci („ten to není"), ale opět bychom se neměli nechat zmást a nevyvozovat ukvapené závěry, protože silný přízvuk může maskovat jemné rozdíly, obzvlášť pro netrénovaného posluchače (Hollien, 2002, str. 55). Na druhou stranu Thompson (1987), Goggin, Thompson, Strube & Simental (1991), Schiller & Köster (1996) nebo Doty (1998) popisují vliv cizího přízvuku na sluchově-percepční rozpoznávání a shodují se, že je spíše nepatrný.

Poslední subkategorií tohoto výčtu budou problémy týkající se posluchače (a tedy opět pouze sluchově-percepční metody rozpoznávání). Je evidentní, že pokud jeho **sluch** není v pořádku, bude narušena i jeho schopnost identifikace. Zjistit to lze například standardním sluchovým (SRT) testem.

Dalo by se také předpokládat, že ne všichni posluchači mají stejné **nadání** pro rozpoznávání mluvčího. A skutečně, Hollien (2002, str. 57) cituje mnoho studií, které tuto hypotézu potvrzují. Čím je to způsobeno, ale dosud není jasné, málokdo se této problematice věnuje.

Posledním faktorem je **věk posluchače**. Na toto téma také existuje poměrně rozsáhlá literatura, například Saito, Asakawa, Shimura & Imaizumi (1995) došli k závěru, potvrzovaném i ostatními, že s věkem dítěte schopnost sluchové identifikace mluvčího poměrně výrazně roste, až přibližně v deseti letech dosáhne úrovně srovnatelné s dospělými. U starých lidí lze s úbytkem sluchu očekávat opět pokles.

3.2 Automatické a poloautomatické rozpoznávání

Hovoříme-li o automatickém nebo poloautomatickém (případně semiautomatickém) rozpoznávání mluvčího, máme na mysli postup, při kterém jednu nebo více částí rozpoznávání ponecháváme na počítači (zejména extrakci rysů a jejich následné porovnávání).

Při automatickém rozpoznávání bohužel nelze využít takzvané vysokoúrovňové řečové rysy, které mohou k rozpoznání velmi dobře sloužit posluchačům. Jde o dialekt, styl řeči, využití specifické slovní zásoby, obsah promluvy, smích a podobně. Většinu těchto charakteristik by bylo složité až nemožné ze signálu automaticky získat, čímž je strojové rozpoznávání limitováno na nízkoúrovňové rysy – akustické charakteristiky přímo se prezentující v řečovém signálu.

V oblasti automatického a poloautomatického rozpoznávání mluvčího se mnohem více překrývají oblasti identifikace a verifikace (protože použité algoritmy mohou být stejné), nicméně daleko větší množství výzkumu se věnuje verifikaci, a to zejména závislé na textu (tj. mluvčí, který je verifikován, musí vyslovit předem daný text, který se porovnává s referenční nahrávkou stejného textu). Její komerční využití je nasnadě – bezpečnostní systémy, domovní zámky, telefonické bankovnictví, omezení přístupu k datům nebo do určitých míst a podobně.

Verifikace je také jednodušší v tom smyslu, že je nutné porovnávat jen jednoho mluvčího s jedním, zatímco při identifikaci porovnáváme jednoho s mnoha. Také podmínky jsou obvykle kontrolované, může být využit kvalitní mikrofon či nahrávací zařízení a mluvčí většinou

spolupracuje (chce být správně rozpoznán). Není tedy divu, že výsledky dosažené při verifikaci bývají mnohem lepší než při identifikaci.

Při verifikaci závislé na textu je mluvčí porovnáván s uloženým profilem (který byl vytvořen někdy dříve při zaznamenávání osoby do systému), obvykle tím způsobem, že se vytvoří referenční soubor řečových parametrů pro všechny uživatele systému a při verifikaci se pak porovnávají, buď na ekvivalentních bodech v promluvě (čili vzorky řeči je nutné časově vyrovnat, aby si odpovídaly), nebo lze tyto řečové parametry porovnávat statisticky (na základě jejich dlouhodobých průměrů). Obě tyto metody mají srovnatelně dobré výsledky (Furui, 1981).

Problémy, které se vynořují v souvislosti s verifikací závislou na textu, závisí jednak na použitých algoritmech a také na nekonzistenci uživatelů. Je žádoucí, aby systém byl odolný vůči změnám tempa či mluvního úsilí, aby rozpoznal uživatele, i pokud má hlas ovlivněný zdravotním stavem nebo emocemi, ale zároveň aby nevpustil podvodníka, imitujícího hlas jiného mluvčího. Tyto dvě proměnné jdou proti sobě a je nutné mezi nimi najít vhodný kompromis.

Ve studii Bellových laboratoří z roku 1972 (Lummis & Rosenberg) byla odolnost verifikačního systému vůči imitátorům testována – pravděpodobnost jejich úspěšného přijetí se výrazně snížila po úpravě systému ve prospěch spektrálních rysů a na úkor prozodických (které se dají snadněji napodobit).

3.2.1 Příklady automatických rozpoznávacích systémů

V následující části textu ilustruji principy fungování automatických a poloautomatických systémů na čtyřech příkladech. První tři z nich budou popsány pouze rámcově, třetímu věnuji celou jednu podkapitolu.

První fungující – a v tomto případě verifikační – systém popisuje Doddington (1985, str. 1661-1662). Jeho účelem bylo omezit vstup neoprávněných osob do hlavního počítačového centra v ředitelství společnosti Texas Instruments v Dallasu. Verifikační algoritmus používá jen jediný vektor rysů, a to výstup čtrnáctikanálového filtru s frekvencemi rovnoměrně rozloženými mezi 300 a 3000 Hz. Jde o verifikaci závislou na

textu, a tak je nutné vstupní rámečky časově vyrovnat s referenčními, což se děje za pomoci zjednodušeného dynamického borcení časové osy. Systém vykazoval míru chybného odmítnutí 0,9% a míru chybného vpuštění 0,7%.

Tvůrci tohoto systému objevili při jeho zavádění několik zajímavých tendencí. Za prvé, mnozí mluvčí byli v průběhu několika prvních pokusů často chybně odmítnuti, což bylo pravděpodobně způsobeno tím, že mluvčí se nového systému zpočátku obávali a jejich hlas se díky tomu měnil. Za druhé, ukázalo se, že většina chybných zamínutí se vázala k malému procentu stále těch samých mluvčích. Bohužel, Doddington nepodává k této záležitosti žádné podrobnější vysvětlení. Logické možnosti jsou tři – buďto je chyba přímo ve verifikačním algoritmu, ve vnějších okolnostech nebo v osobě mluvčího, což se mi zdá jako nejpravděpodobnější alternativa. Může jít o mluvčí s hlasovými problémy, například chronickým onemocněním dýchacího traktu, kvůli kterému dochází často k výrazné změně řečových rysů. Také mohou existovat mluvčí, kteří se snaží systém obelstít a úmyslně mění hlas nebo pronášejí nesprávná slova, přestože jsou ke vstupu oprávněni.

Dalším zajímavým příkladem, zabývajícím se tentokrát již forenzní identifikací nezávislou na textu, je americký program vedený R. Rodmanem (viz např. Klevans & Rodman, 1997, nebo Hollien, 2002, str. 141-146). Tento systém je založený na vyhledávání a porovnávání tzv. „izofonemických sekvencí“, což jsou krátké stejné úseky řeči (např. jeden vokál či diftong), které se ve vstupním vzorku vyhledají, vystříhají a zpracují. Nejspolehlivější jsou, pokud jich je ve vzorku řeči hodně – jejich hodnoty se zprůměrují a tento průměr pak charakterizuje mluvčího. Izofonemické sekvence jsou následně zpracovány algoritmem používajícím diskrétní Fourierovu transformaci a funkci hustoty pravděpodobnosti a výstupem tohoto procesu je množina bodů ve dvoudimenzionálním prostoru tvořící jakousi „cestu“. Porovnávají se pak křivky těchto „cest“ mezi sebou (experimentálně bylo ověřeno, že variabilita mezi promluvami jednoho mluvčího je menší než variabilita mezi mluvčími).

Při identifikaci v uzavřené množině (tj. systém porovnával konečný a „rozumně malý“ počet mluvčích) uvádí autoři chybovost 0%, pro otevřenou množinu (tj. žádný z porovnávaných vzorků nemusel být shodný se vstupním) pak 10-20%. Systém se nicméně vyvíjí dále a jeho úspěšnost se zvyšuje.

Inspirující je také polský projekt pod vedením W. Majewského (Majewski & Basztura, 1996, citováno z Hollien, 2002, str. 146-154), který se drží zásady, že je dobré zkusit všechno, co by mohlo přispět k úspěšnému rozpoznání. Používají tedy nesmírně komplexní postup zahrnující spektra, kepstra, LPC koeficienty, základní frekvenci, frekvence formantů, poměr průchodů nulou a také temporální rysy (bohužel se mi nepodařilo zjistit které). Spolu s těmito automatickými metodami využívají ještě sluchově-percepční metody a vizuální porovnávání sofistikovaných spektrografických zobrazení (vzpomeňme na hlasové otisky výše). Každá z těchto procedur je prováděna trénovanými odborníky, kteří jsou pokaždé jiní, aby nedocházelo k ovlivnění. K vyhodocení se používá statistický algoritmus pro rozpoznávání mluvčích v otevřených množinách – prvním krokem je přiřazení zkoumaného vzorku k jednomu ze známých mluvčích a druhým krokem pak verifikace této dvojice. Pokud je verifikace neúspěšná, vzorek nenáleží do množiny známých mluvčích.

3.2.2 SAUSI

SAUSI je zkratka pro *Semi-Automatic Speaker Identification* a jde o systém vyvíjený pod vedením Harryho Holliena v IASCP (Institute for Advanced Study of Communication Processes) na Floridské univerzitě. Ve své současné podobě používá k rozpoznávání vektory (tj. soubory rysů), které se v minulosti osvědčily jako nejúspěšnější – s různými vektory experimentují Hollien a jeho spolupracovníci již zhruba 50 let. Motivací tohoto projektu je co nejlépe simulovat to, jak lidé rozpoznávají mluvčí – poslouchají některé řečové rysy, ukládají si do paměti idiosynkratické elementy a následně je pak vyvolávají (Hollien, 2002, str. 161-162). Toto tvrzení je však poněkud diskutabilní, protože jak sám Hollien uvádí dále, byly vektory rysů vytvářeny a kombinovány výhradně na základě

statistické úspěšnosti při rozpoznávání. Nicméně to určitě nepopírá, že tyto vektory fonetickou realitu nějak odrážejí.

SAUSI používá v současnosti tyto čtyři vektory rysů:

- **LTS (*long-term spectra*)**: Velmi stabilní vektor, který na základě dlouhodobých spekter odráží barvu hlasu neboli témbr. Podle Holliena (2002, str. 163) tento vektor vydrží i zkreslení šumem, malou šířkou pásma či emočními změnami. Do určité míry funguje i v případě úmyslné změny hlasu. Pro vytvoření křivky dlouhodobého spektra používá SAUSI až 40 parametrů. Rozhodnutí, zda dvě křivky náleží k jednomu mluvčímu, se provádí pomocí porovnání jejich rozdílů (euklidovská a Hammingova vzdálenost).
- **SFF (*speaking fundamental frequency*)**: Jde o vektor popisující základní frekvenci s pomocí následujících parametrů: geometrický průměr F_0, poměr fonace k celkové délce vzorku řeči, směrodatná odchylka všech vyprodukovaných F_0 a půltónové intervaly s informacemi, kolikrát byla frekvence v daném intervalu použita. Na vyhledávání F_0 se používá systém FFI (fundamental frequency indicator), vyvinutý také na IASCP.
- **TED (*time-energy distribution*)**: Tento vektor odráží prozodické informace v řečovém signálu. Sám o sobě sice není příliš úspěšný, vylepšuje ale funkčnost celého procesu v kombinaci s ostatními (Hollien, 2002, str. 166). Obsahuje tyto rysy: počet a délku pauz (intervalů ticha), poměr řeči a pauz, slabičné tempo, poměr řeči a celkového času promluvy, celkový čas promluvy, ve kterém je přítomna akustická energie, a další.
- **VFT (*vowel-formant tracking*)**: Reflektuje vokalické formanty. Formantové parametry jsou získávány už při sběru informací pro LTS – po získání spektra v každém rámečku se aplikuje program, který najde a vyextrahuje hodnoty pro první tři vokalické formanty. VFT pak používá hodnoty reprezentující

polohu F_1, F_2 a F_3, jejich geometrický průměr a směrodatné odchylky.

Jak už bylo zmíněno, tyto vektory fungují nejlépe v kombinaci. Za jejich současnou podobou stojí desítky let experimentování s různými řečovými rysy a jejich úspěšností při rozpoznávání mluvčího.

V praxi se procedury SAUSI použijí třikrát za sebou – pokud možno na jiné vzorky řeči stejných mluvčích, což zajistí vyšší váhu zjištěných výsledků. Program SAUSI byl testován také za podmínek degradujících signál – limitovaná šířka pásma, šum nebo různá akustická prostředí – s celkovou úspěšností 96%. V laboratorních podmínkách uvádějí autoři úspěšnost ještě o něco vyšší, až stoprocentní (Hollien, 2002, str. 190).

Mohlo by se zdát, že v automatické rozpoznávání mluvčího směřuje k naprosté dokonalosti a vše je jen otázka času (a peněz). Doddington však varuje, že úspěšnost automatických rozpoznávačů neporoste lineárně se zlepšováním vybavení, metod extrakce rysů a rozhodovacích algoritmů. Podle něj je automatické rozpoznávání (a zejména identifikace nezávislá na textu) v praxi omezeno nedostatkem kontroly nad podmínkami, a tedy se může stát, že v některých případech rozpoznání prostě nebude možné, bez ohledu na to, jak dokonalou technikou budeme oplývat (Doddington, 1985, str. 1663).

4. Problematika temporální struktury

Přistupme nyní k detailnějšímu zkoumání hlavního tématu této práce, tedy temporálních rysů řeči.

Otázka analýzy temporální struktury řeči je v současné době stále otevřená, zejména v souvislosti s rytmem a rytmickou typologií jazyků (viz zejm. Abercrombie, 1967). Rozdělení jazyků do tříd, na jazyky se slabičnou izochronií (syllable-timed), taktovou izochronií (stress-timed), případně morovou izochronií (mora-timed) neztratilo svou přitažlivost ani po půl století.

Rytmus v jazycích je percepčně velmi dobře zakotvený, intuitivně cítíme změny v rytmu, dokonce i odlišné způsoby rytmizace jiných jazyků. Nazzi a Ramus (2003) nacházejí vliv rytmických tříd na percepci jazyků dokonce už u malých dětí. Problém však nastává v okamžiku, kdy se tyto faktory pokusíme analyzovat, kvantifikovat a nějak akusticky reprezentovat. Stávající rytmické modely jsou totiž založené zejména na percepci a nelze jimi vysvětlit, jak naše percepční systémy rytmus z řečového signálu extrahují (Ramus, Nesporová & Mehler, 1999, str. 269).

Důležitým rysem temporální struktury řeči, který percepci rytmu umožňuje, je střídání kontrastních jednotek – přízvučných a nepřízvučných slabik či vokalických a konsonantických intervalů. Tyto kontrasty a jejich změny jsou základem pro vnímání pravidelností či nepravidelností v proudu řeči.

Při pokusech o zachycení temporálních rysů a popis tempa řeči je nutné rozlišovat dva přístupy k věci. Měří-li se hláskové nebo slabičné tempo, jde v podstatě pouze o jednotky, ve kterých je vyjádřena rychlost produkce řeči (hlásky za sekundu, respektive slabiky za sekundu). Tyto ukazatele však temporální strukturu řeči zachycují jen velmi hrubě. Pfitzinger (1998) konstatuje, že ani jeden z nich dostatečně nekoreluje s výsledky percepčních testů, a na svých datech ukazuje, že lepší úspěšnost má lineární kombinace obou faktorů. Postupně se od takto jednoduchého způsobu upouští, protože na slabičné i hláskové tempo má vliv velké množství dalších faktorů, které není možné kontrolovat.

Také je potřeba si uvědomit, že mluvčí obvykle neprodukují hlásky či slabiky přesně v jejich „slovníkové“ podobě, často redukují či vynechávají. Při měření pak vyvstává problém, zda slabičné a hláskové tempo postavit na tom, co mluvčí říci chtěl, či na tom, co skutečně řekl.

Naopak měří-li se artikulační nebo mluvní tempo, jde o to, co je do měření zahrnuto (tyto termíny jsou tedy zcela jiného charakteru). Artikulačním tempem je obvykle míněna pouze doba, kdy mluvčí artikuluje, a pauzy či nelingvistické zvuky jsou z měření vyřazeny. Mluvním tempem se označuje tempo v době artikulace i s pauzami. Pokud je třeba měřit jinou kombinaci, používá se termín modifikované mluvní tempo, u kterého je předem definováno, co pod něj spadá. Zastřešujícím termínem je pak tempo řeči.

Přístupem k temporální struktuře řeči z jiné strany je snaha o zachycení konfigurace rysů, tj. jak jsou (případně mohou být) kontrastivní jednotky uspořádány a jaký je jejich vzájemný vztah – nezávisle na slabičném či hláskovém tempu mluvčího.

Protože percepčně nejvýznamnějšími částmi promluv jsou jádra slabik – v opozici ke konsonantům v prétuře nebo kodě – přišli Ramus, Nesporová a Mehler (1999) s ukazateli založenými na trvání a variabilitě vokalických a konsonantických intervalů. Vokalickými intervaly označili dobu trvání vokálu, diftongu, případně sledu vokálů, konsonantické intervaly byly pak komplementární. Navrhli měřit tyto tři proměnné:

- **%V**: procento vokalických intervalů ve větě
- **ΔV**: směrodatná odchylka trvání vokalických intervalů v každé větě (indikuje variabilitu vokalických intervalů)
- **ΔC**: směrodatná odchylka trvání konsonantických intervalů ve větě

Podle autorů jsou tyto ukazatele přímými fonetickými koreláty fonologických a fonotaktických vlastností jazyků (například složitosti slabičné struktury). Využili je k rozlišení mezi rytmickými třídami jazyků a podle jejich výsledků jsou především %V a ΔC schopné mezi těmito kategoriemi dobře rozlišovat. ΔV nebylo možné přímočaře interpretovat,

protože jak sami konstatují, do hry vstupuje velké množství dalších faktorů, například fonologická distinkce délky, vokalická redukce nebo kontextově podmíněné prodlužování vokálů. To vše mohlo výsledky silně ovlivnit.

Tyto ukazatele popisují temporální strukturu globálně – tzn. na úrovni úseků, vět, případně na ještě vyšší (hodnoty těchto tří proměnných se následně ještě průměrují a výsledkem je jedna hodnota pro mluvčího, případně pro celý jazyk).

Lowová, Grabová a Nolan (2000) při svém výzkumu rytmických rozdílů mezi varietami angličtiny (singapurská vs. britská) navrhli využít proměnnou, která by lépe reflektovala lokální povahu rytmu (to znamená vztahy jednotek na úrovni taktů a nižší). Tato proměnná zachycuje vztah po sobě následujících prvků a přitom zachovává hypotézu, že pro percepci rytmu je důležité střídání vokalických a konsonantických intervalů v promluvě. Nazvali ji index párové variability – **PVI** (z angl. *Pairwise Variability Index*). Tento index vyjadřuje rozdíl v trvání vždy dvou po sobě jdoucích vokalických intervalů, normalizovaný vzhledem k průměrnému trvání těchto intervalů.[1]

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

Vzorec 1: Výpočet normalizovaného indexu párové variability (nPVI), kde dk je trvání k-tého intervalu a m je počet intervalů.

Grabová a Lowová (2002) následně navrhly i nenormalizovanou verzi tohoto indexu, rPVI (z angl. „raw“).

[1] Zde budu stejně jako Grabová a Lowová (2002), White a Mattys (2007a i 2007b) a další používat pro normalizovaný index zkratku *nPVI*.

$$rPVI = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m-1) \right]$$

Vzorec 2: Výpočet nenormalizovaného indexu párové variability (rPVI), kde dk je trvání k-tého intervalu a m je počet intervalů.

Gibbon a Gutová (2001) ovšem podotýkají, že normalizované PVI nabývá hodnot v poněkud neintuitivním rozsahu 0–200. Vynecháním dvojky ve vzorci získáme proměnnou s polovičními hodnotami, tedy v rozsahu 0–100. Gibbon a Gutová ji nazvali rytmický poměr (*Rhythm Ratio*). Na nPVI je lineárně převeditelná (nPVI = RR*2).

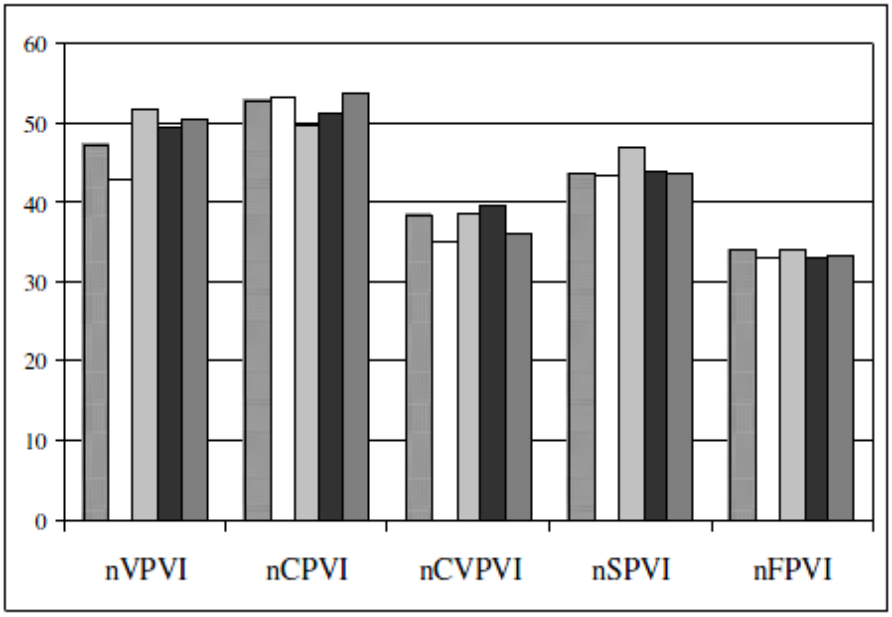
Lowová et al. (2000) docházejí k závěru, že vokalické nPVI je lepším ukazatelem temporální struktury než intervalové ukazatele Ramuse et al. a lépe koreluje s etablovanými rytmickými třídami jazyků. Grabová a Lowová pak konstatují, že normalizace je při zkoumání rozdílů mezi jazyky dokonce žádoucí, protože nenormalizované rPVI je příliš citlivé na změny celkového tempa mezi mluvčími i mezi výpověďmi. Důležitost normalizace při rozlišování jazyků potvrzuje i Ramus (2002), jehož nenormalizované intervalové ukazatele (výše zmíněné %V, ΔV a ΔC) silně reflektují specifika jednotlivých mluvčích.

Jsou to ale právě individuální rozdíly mezi mluvčími, které leží ve středu pozornosti této práce a které zde budou zachycovány s pomocí temporálních ukazatelů. Jak jsem zmínila výše, normalizované ukazatele jsou mnohem robustnější, méně citlivé ke specifickým rozdílům mezi mluvčími, a tedy se dá se očekávat, že budou pro tento výzkum méně vhodné.

Rozšířit PVI také na konsonantické intervaly vyzkoušeli kromě Grabové a Lowové rovněž Asuová a Nolan (2006) při porovnávání rytmických rozdílů mezi estonštinou a angličtinou, na základě hypotézy, že konsonantické PVI lépe zachytí rytmicky relevantní rozdíly mezi jazyky ve struktuře slabiky (tj. počty konsonantů v prétuře a kodě). V případě

konsonantických intervalů ovšem Grabová a Lowová polemizují s vhodností normalizace, která by mohla tyto rozdíly zastřít.

Z grafu pro hodnoty všech nPVI pro pět estonských mluvčích ze studie Asuové a Nolana (2006) je ovšem vidět, že i normalizované ukazatele individuální rozdíly do určité míry reflektují, nejsilněji pak vokalické nPVI.



Graf 1: Hodnoty normalizovaných PVI pro pět estonských mluvčích. První dvě hodnoty jsou normalizované vokalické a konsonantické PVI, zbylé tři jsou párové indexy vyslovených slabik, lingvistických slabik a stop.
Převzato z Asuová & Nolan (2006).

Dellwo (2006) pak navrhl normalizovat také intervalové ukazatele Ramuse et al. a označil je **VarcoV** a **VarcoC**. Jde o směrodatnou odchylku trvání vokalických, respektive konsonantických intervalů dělenou průměrným trváním těchto intervalů.

V obou svých studiích z roku 2007 porovnávali White a Mattys PVI, intervalové ukazatele (%V, ΔV a ΔC) a také normalizované intervalové ukazatele (VarcoV a VarcoC), to vše opět za účelem zachytit rozdíly mezi postulovanými rytmickými třídami jazyků. Z párových indexů použili nPVI-V (normalizovaný index vokalických intervalů) a rPVI-C (nenormalizovaný index konsonantických intervalů), jak navrhovaly

Grabová a Lowová. Shledávají zejména ΔV a ΔC silně ovlivněné řečovým tempem jednotlivých mluvčích, přičemž opět normalizované ukazatele lépe vystihovaly celkové rozdíly mezi jazyky a rytmickými třídami.

Ke stejnému závěru dochází i Yoon (2010), také podle jeho výsledků (zkoumal deset mluvčích jednoho dialektu americké angličtiny) normalizace rozdíly mezi mluvčími smazává. Nejlepším ukazatelem specifičnosti mluvčích se na jeho materiálu ukázalo být %V, rPVI-C a VarcoV, nejhorším pak nPVI-V. Své výsledky s ohledem na ΔV a ΔC Yoon nepopisuje.

Některé další studie ovšem vrhají stín na schopnost těchto ukazatelů jednoznačně popsat rytmické třídy jazyka. Loukinová, Kochanski, Rosner, Keanová a Shihová (2010) porovnávali 15 různých temporálních ukazatelů (mezi nimi i všechny výše zmíněné) a jejich schopnost diskriminovat jazyky a rytmické třídy. Zaznamenali při tom značné variace uvnitř jednotlivých jazyků a došli k závěru, že jediný ukazatel k rozlišení nestačí, je jich potřeba kombinovat více. Zároveň konstatují, že jejich výsledky nejsou vůbec konzistentní s hypotézou rytmických tříd, protože rytmus je multidimenzionální systém, na který má vliv velké množství faktorů.

Ještě kritičtěji se k problému staví Arvanitiová (2009), která prohlašuje, že zkoumat rytmus jen za pomoci ukazatelů tempa je zcela pochybené, protože rytmus je jevem složeným z mnoha dalších parametrů. A samotné tempo je také v řeči ovlivňováno jinými faktory: kontrastivní délkou vokálů, kontextuálními efekty v důsledku znělosti a pozice ve slabice, prodlužováním v souvislosti s přízvukem, kontextově a prozodicky určenou redukcí vokálů, prodlužováním na konci fráze, atd.

V případě jejích dat ukazatele tempa ve skutečnosti s rytmickými třídami vůbec nekorelují (resp. tato korelace je pod hranicí náhody). Rovněž zaznamenává výrazný vliv mluvčího, způsobu řeči (čtená, spontánní) a výběru materiálu. Variabilita mezi jednotlivými mluvčími byla v jejích výsledcích dokonce vyšší než variabilita mezi jazyky.

Podobný názor zastávají také Wiget, White, Schupplerová, Grenonová, Rauchová a Mattys (2010) – zkoumali přímo vliv mluvčího, materiálu a také anotátora na rytmické ukazatele v angličtině a došli k závěru, že výsledky pro daný jazyk mohou být ovlivněny každou z těchto proměnných a je proto třeba interpretovat je velmi obezřetně. Největší rozdíly mezi hodnotami ukazatelů byly způsobeny rozdílností textu (je ale nutné si uvědomit, že od každého mluvčího měli k dispozici pouze pět vět, což je pro jakoukoliv generalizaci žalostně málo – a vysoký vliv textu je v takovém případě naprosto očekávatelný), ovšem variace mezi mluvčími byly také signifikantní. Tyto variace byly přitom nejmarkantnější u %V a VarcoV a nejmenší u nPVI-V, podobně jako u Yoona. Na druhou stranu, efekt anotátora se ukázal být malý (pokud všichni anotátoři postupovali podle dohodnutého protokolu) a použití automatického značkovacího programu mělo výsledky srovnatelné.

Rozdíly mezi mluvčími v rámci jednoho jazyka (v tomto případě němčiny) obzvláště u ukazatelů %V, ΔC a PVI (nespecifikováno) zachytili také Dellwo a Koreman (2008). Zároveň zaznamenali velmi malý vliv i drastických změn tempa jednoho mluvčího na hodnoty těchto ukazatelů.

To vše nahrává záměru této práce – zdá se, že i když využití intervalových ukazatelů pro rozlišení mezi jazyky je přinejmenším diskutabilní, mělo by být možné je úspěšně využít k identifikaci mluvčího, a to obzvláště ty nenormalizované.

Při analýze dat se bude třeba pozastavit i nad problémem fonologické distinkce délky vokálů, která by mohla mít podstatný vliv na vokalické proměnné. Ovšem Podlipský, Skarnitzl a Volín (2009, Tabulka 1) ukazují, že poměr průměrné délky dlouhých a krátkých vokálů v češtině je pouze 1,29 (pro i) až 1,79 (pro a) a v nefinálních pozicích dokonce ještě nižší.

Všechny dosud zmíněné ukazatele ovšem popisují tempo globálně (ať už na úrovni mluvčího či jazyka), přestože do PVI je vnesen i prvek lokálních změn. Výsledkem je však stále jedno číslo, což není pro popis lokálních změn řečového tempa vůbec dostačující. Návrh, jak tyto lokální

změny kvantifikovat přímo, podává Volín (2009). Měří lokální artikulační tempo (označené $LAR_{pk\text{-}pk}$) jako převrácenou hodnotu vzdálenosti dvou po sobě následujících vokalických jader (ozn. $Dur_{pk\text{-}pk}$).

$$LAR_{pk-pk} = \frac{1}{Dur_{pk-pk}}$$

Vzorec 3: Výpočet proměnné $LAR_{pk\text{-}pk}$, kde $Dur_{pk\text{-}pk}$ je vzdálenost středů dvou po sobě následujících vokalických intervalů.

Tato proměnná (případně její klouzavým průměrem vyhlazená křivka) je podle Volína dobrým ukazatelem temporálního značení prozodických předělů v češtině, které jsou téměř vždy doprovázeny borcením metra výpovědi.[2]

Cílem této práce je zjistit, jak – a zda vůbec – jsou výše zmíněné proměnné schopné zachytit rozdíly mezi jednotlivými mluvčími stejného jazyka (v tomto případě češtiny). Zároveň bude možné prozkoumat, jak se na hodnotách těchto proměnných projevuje vzdálenost promluv v čase a zda je tento rozdíl srovnatelný s rozdíly mezi mluvčími. Pro forenzní využití závěrů by bylo vyhovující, kdyby alespoň některé proměnné byly schopné různost mluvčích spolehlivě zachytit.

[2] Protože nemůže dojít k záměně, budu tuto proměnnou v dalším textu označovat pouze zkratkou LAR.

5. Experiment

5.1 Metoda a materiál

Pro tuto práci byly použity nahrávky z Pražského fonetického korpusu (Skarnitzl, 2010). Materiál se skládal ze čtených rozhlasových zpráv od tří dospělých žen, rodilých mluvčích češtiny, z nichž žádná neměla vadu řeči a všechny hovořily spisovnou češtinou bez nářečních prvků. Od každé mluvčí byly k dispozici dvě různé nahrávky, obě v rozmezí minimálně 1 rok od sebe. Všechny nahrávky jsou přiloženy na datovém CD (viz příloha 2).

Tabulka 1 zobrazuje údaje o použitých nahrávkách. Jejich průměrná délka je 3 a čtvrt minuty, počet slov se pohybuje v rozmezí 409–577, s průměrem 467 slov. Nahrávky obsahují průměrně 46 nádechových úseků. Ve čtvrtém sloupci je zobrazen počet pauz, tj. počet tichých intervalů delších než 120 ms, které ale neobsahují nádech. (Pauzy s nádechem nebyly do analyzovaného materiálu vůbec zařazeny.) Pátý sloupec uvádí počet přeřeknutí, spočítaný jako počet nedokončených slov, která mluvčí následně opravila. Jiný druh dysfluencí se v materiálu nevyskytl. Časový rozdíl mezi nahrávkami je ve dvou případech mezi jedním a dvěma roky a ve třetím dokonce větší než pět let.

Nahrávka	Poč. NU	Počet slov	Počet pauz	Poč. přeřeknutí	Trvání (min)	Datum nahrání
GVA02	40	410	5	0	2,8	17.10.2004
GVA07	39	409	0	0	2,73	31.3.2010
MSA01	51	577	5	2	4,05	10.10.2004
MSA03	45	460	5	0	3,57	11.1.2006
SSA02	55	496	3	0	3,28	11.2.2004
SSA03	47	452	3	3	3,26	5.1.2006

Tabulka 1: Údaje o nahrávkách jednotlivých mluvčích. Zkr. NU označuje nádechové úseky

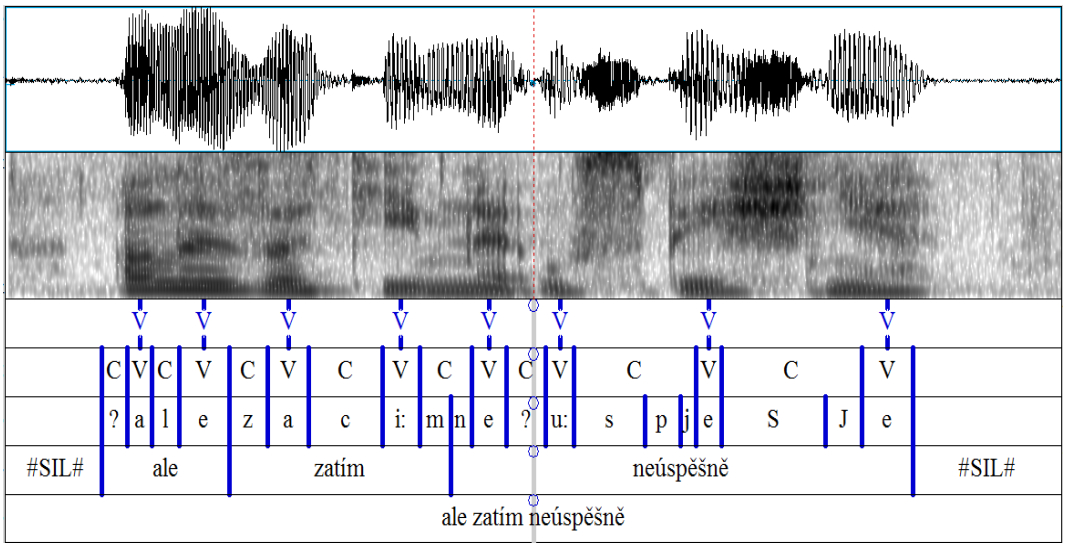
Každá nahrávka byla rozstříhána na nádechové úseky, které byly následně očíslovány podle odstavců ve čteném textu. Tyto úseky budou pro další analýzu brány jako základní.

Nahrávky byly segmentovány a anotovány v programu Praat (Boersma & Weenink, 2010) na úrovni slov a hlásek s pomocí automatického algoritmu Prague Labeller (Pollák, Volín & Skarnitzl, 2008). Posléze bylo nutné chybně umístěné hranice segmentů ručně opravit. Jako vodítko pro určování hranic hlásek byly použity zásady navržené Machačem a Skarnitzlem (2009), využívající vizuální prozkoumání spektrogramu i oscilogramu a pozorný poslech. Hranice vokalických segmentů byly určovány s pomocí nástupu (a ukončení) zřetelné formantové struktury, která se projevila ve spektrogramu jako sekvence tzv. formantových sloupků a v oscilogramu nárůstem, respektive úbytkem energie v amplitudě. Pokud začátek či konec hlásky nebyl zřetelný a hlásky se prolínaly, byla hranice umístěna do středu této přechodové oblasti. Nakonec byly hranice s pomocí skriptu automaticky umístěny do nejbližších průchodů nulou (viz příloha 1a).

Následně byla vytvořena další anotační vrstva, ve které byly jednotlivé hlásky označeny jako vokály nebo konsonanty a seskupeny do vokalických, respektive konsonantických intervalů, nezávisle na hranicích slov (skript tvoří přílohu 1b). Slabikotvorné konsonanty byly přitom označeny jako vokály, protože jde o znělostní jádro slabiky, které může nést přízvuk, a tedy je důležité pro percepci tempa. Vokalický interval pak trvá od zahájení vokálu (případně slabikotvorného konsonantu, diftongu nebo sledu vokálů) do jeho ukončení. Konsonantický interval je vše mezi tím (vyjma pauz).

Jako poslední bylo pro měření lokálních změn artikulačního tempa s pomocí proměnné LAR potřeba najít a označit středy vokalických intervalů. To bylo opět provedeno automaticky s pomocí skriptu v programu Praat (viz příloha 1c).

Soubory formátu TextGrid se všemi těmito údaji jsou k dispozici na přiloženém datovém CD (viz příloha 2).



Obrázek 1: Editační okno v programu Praat. První tři textové vrstvy zobrazují středy vokalických intervalů, vokalické a konsonantické intervaly a jednotlivé hlásky. Na čtvrté vrstvě jsou anotována slova a na páté nádechové úseky.

Pauzy a přeřeknutí jsem se při analýze rozhodla zcela vynechat a zabývat se pouze artikulačním tempem. Tyto jevy sice mohou poskytnout zajímavé informace o řečovém chování mluvčího, nicméně v materiálu jich bylo příliš málo na vyvození jakýchkoliv obecnějších závěrů. Při konstrukci skriptů jsem tedy postupovala tak, že intervaly obsahující pauzy byly zcela vynechány, jako by v materiálu vůbec nebyly (podobně řešily tento problém i Grabová a Lowová (2002) a další). Přeřeknutí byla z materiálu vyřazena ručně (kvůli jejich nízkému počtu a vysoké pracnosti automatického rozpoznání skriptem) – interval obsahující přeřeknutí byl na vrstvě konsonantických a vokalických intervalů označen zvláštní značkou a skripty jej pak přeskakovaly stejně jako pauzy.

Extrakce hodnot trvání vokalických a konsonantických intervalů, jejich směrodatných odchylek a výpočet příslušných ukazatelů tempa byl opět proveden s pomocí skriptů (viz přílohy 1d-1g), které tato data nakonec převedly do tabulek ve formátu MS Excel, kde následně probíhala jejich další analýza.

Při takto nevelkém množství materiálu by mohla nastat situace, kdy texty jednotlivých mluvčích by se od sebe natolik lišily, že by to mělo

zásadní vliv na měřené ukazatele. Všechny totiž více či méně na textu závisí. K zachycení reprezentativnosti textu byl tedy u každého nádechového úseku spočten poměr počtu konsonantů a vokálů (nikoliv intervalů, ale skutečně pronesených hlásek), který alespoň hrubě zachycuje složitost slabičné struktury v daném úseku (skript tvoří přílohu 1j). Při analýze pak byla sledována korelace tohoto poměru C/V a jednotlivých ukazatelů.

Rovněž byl změřen také počet dlouhých vokálů a jejich procentuální zastoupení v daném úseku. To bylo následně použito k prozkoumání vlivu distinkce délky vokálů na vokalické ukazatele.

Dle mého názoru je prozkoumat vliv textu nutné (ať už tímto nebo jiným způsobem) pokaždé, kdy analyzovaný materiál není opravdu obsáhlý. Pokud by byl totiž zjištěn statisticky signifikantní rozdíl mezi texty jednotlivých mluvčích, nalezené rozdíly v těchto temporálních ukazatelích by mohly být způsobeny textem a nikoliv individuálními projevy mluvčích – a nešlo by pak vyvodit žádné zobecnitelné závěry.

Proměnná LAR byla z materiálu extrahována také s pomocí praatového skriptu (viz přílohy 1h a 1i). Hodnoty byly vypsány do tabulky; ke každému úseku příslušelo o jednu hodnotu méně, než v něm bylo obsaženo vokalických intervalů. K vyhlazení křivky tempa byl stejně jako u Volína (2009) použit tříbodový harmonický klouzavý průměr, který lépe reflektuje celkovou tendenci křivky dat a zmírňuje dopad extrémních hodnot (obzvláště vysokých).

Šest posledních hodnot proměnné LAR, respektive čtyři poslední hodnoty tříbodového klouzavého průměru byly využity k popisu a analýze závěrového zpomalování na konci úseku. Vyřazeny byly ty nádechové úseky, které obsahovaly méně než sedm vokalických intervalů, a zároveň ty, které mezi těmito sedmi intervaly obsahovaly úsekový prozodický předěl – protože změna tempa na konci tohoto předělu by zkreslila celkové zpomalování na konci nádechového úseku. Výhodnější by bývalo bylo řídit se přímo délkou posledního prozodického úseku než jen počtem vokalických intervalů, ale v materiálu nebyla vrstva prozodických úseků anotována.

5.2 Analýza dat

Analýza získaných dat byla provedena v tabulkovém editoru MS Excel s pomocí statistických a matematických funkcí. Byla používána varianta t-testů pro nekorelovaná měření s homogenním rozptylem a Pearsonův korelační koeficient (kde není řečeno jinak). V několika případech byl využit i Spearmanův korelační koeficient, spočítaný programem Analyse-it pro MS Excel.

Globální proměnné (%V, ΔV, ΔC, VarcoV, VarcoC a všechny PVI) byly měřeny pro každý nádechový úsek zvlášť (tzn. každý nádechový úsek byl reprezentován jednou hodnotou proměnné) a následně zprůměrovány (aritmetickým průměrem) pro danou nahrávku a posléze i pro obě nahrávky od jedné mluvčí. Ke spočtení směrodatné odchylky těchto průměrů byla použita funkce pro výběrovou směrodatnou odchylku.

Soubory dat pro jednotlivé nahrávky a mluvčí byly následně porovnávány t-testy: obě nahrávky od jedné mluvčí mezi sebou, a také každá mluvčí jako celek s ostatními.

V případě všech proměnných bylo také prozkoumáno, do jaké míry závisí na textu, a tedy do jaké míry korelují s poměrem konsonantů a vokálů, případně s procentem dlouhých vokálů (což se týkalo pouze vokalických proměnných).

Aby byl odstraněn nežádoucí vliv textu, byla provedena také analýza upraveného vzorku dat po vyřazení úseků s okrajovými hodnotami poměru C/V (většími nebo naopak menšími než ±2 směrodatné odchylky od průměru pro všechny mluvčí).

Zkoumán byl také vliv délky úseku – kvanitifkovaný jako trvání artikulace v sekundách. Lepším řešením by pravděpodobně bylo založit trvání spíše na počtu taktů nebo prozodických úseků, ty ale nebyly v materiálu anotovány. A počet slov mohl být ošidný kvůli jejich různé délce.

Protože všechny nebo téměř všechny hodnoty trvání by se vešly do intervalu ±2 směrodatných odchylek, odstranila jsem hodnoty pod 5. a

nad 95. percentilem (což znamenalo vyřadit 4–5 nejnižších a nejvyšších datových bodů).

U některých proměnných byly na základě těchto výsledků dále provedeny i podrobnější analýzy a pokusy o zjištění příčin rozdílů (nebo naopak chybějících rozdílů).

Proměnná LAR byla využita k deskripci a zkoumání závěrového zpomalování na konci nádechového úseku. Úseky byly rozděleny podle koncového melodému – na ukončující klesavý, ukončující stoupavý, případně neukončující.

Z naměřených hodnot LAR jsem využila 6 posledních hodnot, 4 u tříbodového klouzavého průměru. Opět by bylo lepší řídit se přímo délkou posledního prozodického úseku – ale nebyly v materiálu anotované a z časových důvodů již nebylo možné je dodatečně anotovat. Z analýzy byly tedy vyřazeny ty úseky, které mezi sedmi posledními slabikami obsahovaly prozodický předěl, pauzu nebo přeřeknutí. Úseky s dysfluencí v jiné části byly v materiálu ponechány.

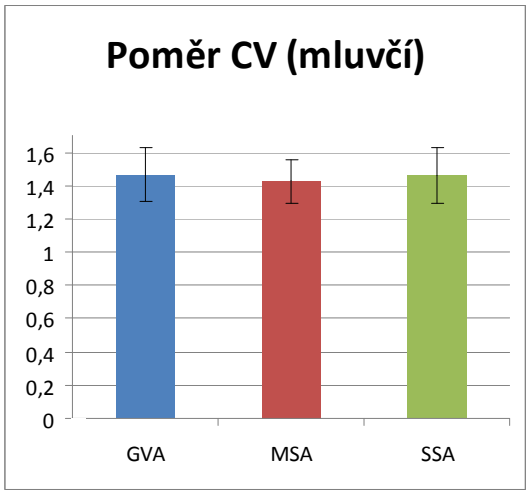
Pro tyto hodnoty byl následně spočítán koeficient lineární regrese – ten představuje gradient klesání (případně stoupání) pomyslné křivky hodnot LAR.

Tabulky s extrahovanými daty a provedenými analýzami jsou k dispozici na přiloženém datovém CD (viz přílohy 3a-i).

6. Výsledky

6.1 Poměr konsonantů a vokálů

T-testy poměru konsonantů a vokálů v nádechových úsecích ukázaly, že rozdíly mezi mluvčími (ani mezi dvěma nahrávkami od jedné mluvčí) nejsou v tomto ohledu statisticky signifikantní. Tedy lze s vysokou pravděpodobností říci, že text je reprezentativní a nevykazuje ani u jedné nahrávky výrazné odchylky v poměru konsonantů a vokálů.



	Průměr	SD
GVA	1,46	0,16
MSA	1,42	0,13
SSA	1,46	0,17

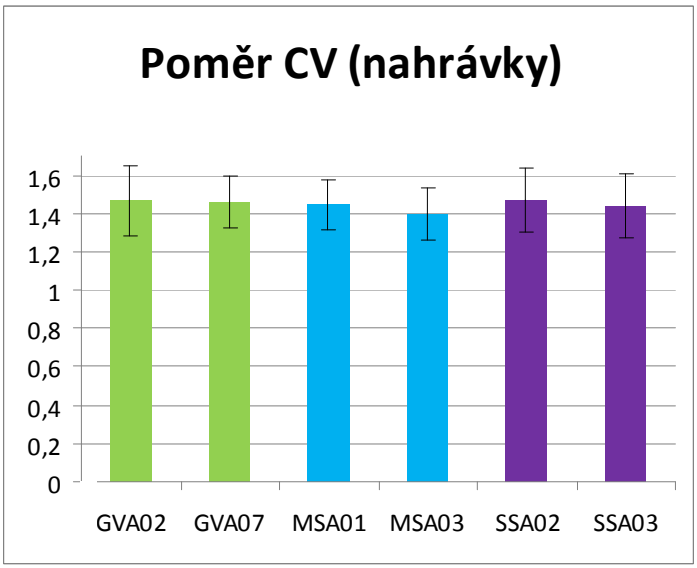
Graf 2: Průměrné hodnoty poměru konsonantů a vokálů pro jednotlivé mluvčí.

Graf 2 zobrazuje hodnoty poměru C/V – tj. počtu konsonantů v jednom úseku děleného počtem vokálů v tomtéž úseku – zprůměrované pro jednotlivé mluvčí jako celek. Chybové úsečky reprezentují (stejně jako v dalších grafech) směrodatnou odchylku dat. Tabulka vpravo uvádí zdrojová data pro daný graf. Ve druhém sloupci je vždy průměrná hodnota daného ukazatele a ve třetím její směrodatná odchylka.

Jak vidíme v tabulce níže, hodnoty se od sebe liší pouze nepatrně a na hladině spolehlivosti $\alpha = 0{,}05$ nebyl žádný z rozdílů statisticky signifikantní.

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 1,8; p>0,05	t(196) = 1,58; p>0,05	t(179) = 0,25; p>0,05

Tabulka 2: Výsledky t-testů pro dvojice mluvčích a hodnoty poměru C/V.



	Průměr	SD
GVA02	1,47	0,18
GVA07	1,46	0,14
MSA01	1,45	0,13
MSA03	1,40	0,14
SSA02	1,47	0,17
SSA03	1,44	0,17

Graf 3: Průměrné hodnoty poměru C/V pro jednotlivé nahrávky zvlášť.

Graf ukazuje zprůměrované hodnoty poměru C/V pro jednotlivé nahrávky každé mluvčí. Opět je vidět, že data jsou poměrně kompaktní.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 0,26; p>0,05	t(94) = 1,81; p>0,05	t(100) = 1; p>0,05

Tabulka 3: Výsledky t-testů pro dvojice nahrávek od jedné mluvčí a hodnoty poměru C/V.

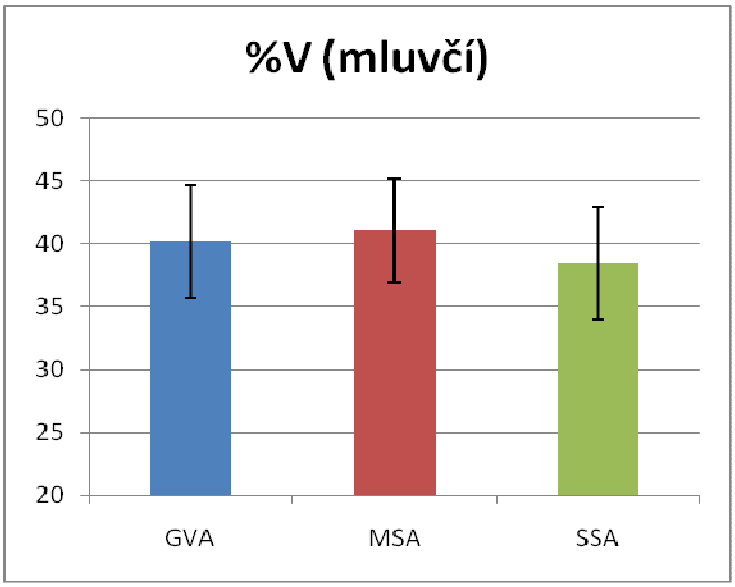
V tabulce jsou uvedeny výsledky t-testů při porovnávání dvou nahrávek od jedné mluvčí mezi sebou. Na stejné hladině spolehlivosti opět nebyl zaznamenán žádný statisticky signifikantní rozdíl.

Není tedy žádný důvod k zamítnutí nulové hypotézy – jednotlivé nahrávky se od sebe v poměru C/V výrazně neliší. Vliv tohoto poměru na měřené ukazatele by tedy také neměl být signifikantní.

Pokud by byl výsledek opačný, bylo by potřeba k naměřeným datům přistupovat nesmírně obezřetně. Na proměnných by se pak mohl projevovat spíše vliv textu než individuální řečové chování mluvčích a nalezené rozdíly by nemusely vůbec reflektovat rozdíly mezi mluvčími.

6.2 Ukazatel %V

První měřenou temporální proměnnou bylo %V, tedy procento vokalických intervalů v nádechovém úseku.



	Průměr	SD
GVA	40,21	4,47
MSA	41,09	4,11
SSA	38,43	4,47

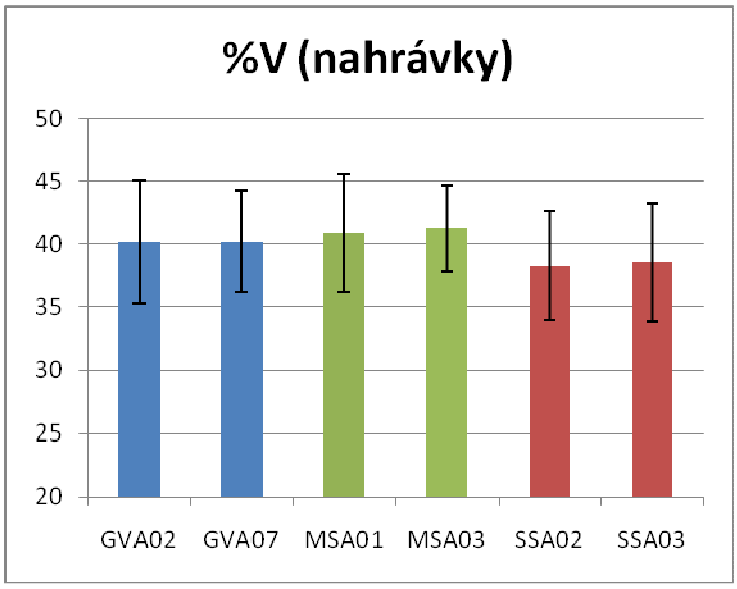
Graf 4: Průměrné hodnoty ukazatele %V pro jednotlivé mluvčí.

V grafu jsou zobrazeny hodnoty průměrného %V pro jednotlivé mluvčí, které byly spočítány aritmetickým průměrem hodnot pro jednotlivé nádechové úseky.

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 1,35; p>0,05	t(196) = 4,36; p<0,05	t(179) = 2,67; p<0,05

Tabulka 4: Výsledky t-testů pro dvojice mluvčích a hodnoty ukazatele %V.

Rozdíly mezi soubory dat se projevily jako statisticky signifikantní ($p<0,05$) ve dvou případech: mezi MSA a SSA (největší) a mezi GVA a SSA.



	Průměr	SD
GVA02	40,20	4,91
GVA07	40,22	4,04
MSA01	40,92	4,65
MSA03	41,28	3,43
SSA02	38,27	4,31
SSA03	38,61	4,68

Graf 5: Průměrné hodnoty ukazatele %V pro jednotlivé nahrávky.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 0,01; p>0,05	t(94) = 0,44; p>0,05	t(100) = 0,37; p>0,05

Tabulka 5: Výsledky t-testů pro dvojice nahrávek od jedné mluvčí a hodnoty %V.

Rozdíl mezi dvěma nahrávkami od jedné mluvčí nebyl statisticky signifikantní ani v jednom případě.

Při bližším zkoumání hodnot ukazatele %V jsem si všimla, že u každé mluvčí se najde několik úseků, které svým %V vybočují z poměrně kompaktního rozložení ostatních. Tyto okrajové hodnoty byly u dvou mluvčích hned úvodní fráze „Dobré ráno“, v obou případech silně ve prospěch vyššího %V. U třetí mluvčí byla extrémní hodnota (velmi nízké %V) způsobena textem s velkým množství konsonantických shluků („především v příhraničních oblastech“) – jde o úsek s výrazně nejvyšším poměrem C/V v celém materiálu.

Po odstranění těchto hodnot a zopakování t-testů byly tytéž rozdíly stále signifikantní. I přesto by mohlo být žádoucí takto nestandardní úseky z materiálu vyřadit, obzvlášť pokud se nejedná o větší množství dat.

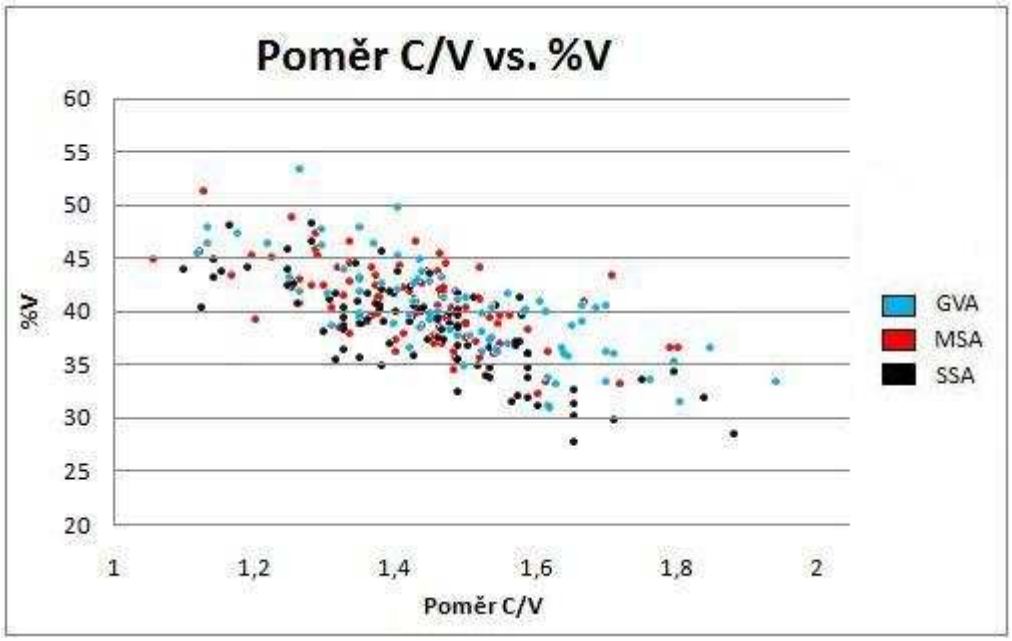
Je očividné, že proměnná %V triviálně závisí na textu, a to na poměru vokálů v textu – otázka je, do jaké míry. Spočítané korelace %V s poměrem C/V jsou záporné (což se dalo očekávat – čím větší poměr konsonantů v úseku, tím menší bude procento vokalických intervalů) a středně silné až vysoké.

	1. nahrávka	2. nahrávka	celkem
GVA	-0,74	-0,75	-0,75
MSA	-0,6	-0,6	-0,58
SSA	-0,74	-0,84	-0,79

Tabulka 6: Pearsonovy korelační koeficienty pro hodnoty %V a poměr C/V u jednotlivých nahrávek a celkem pro mluvčí.

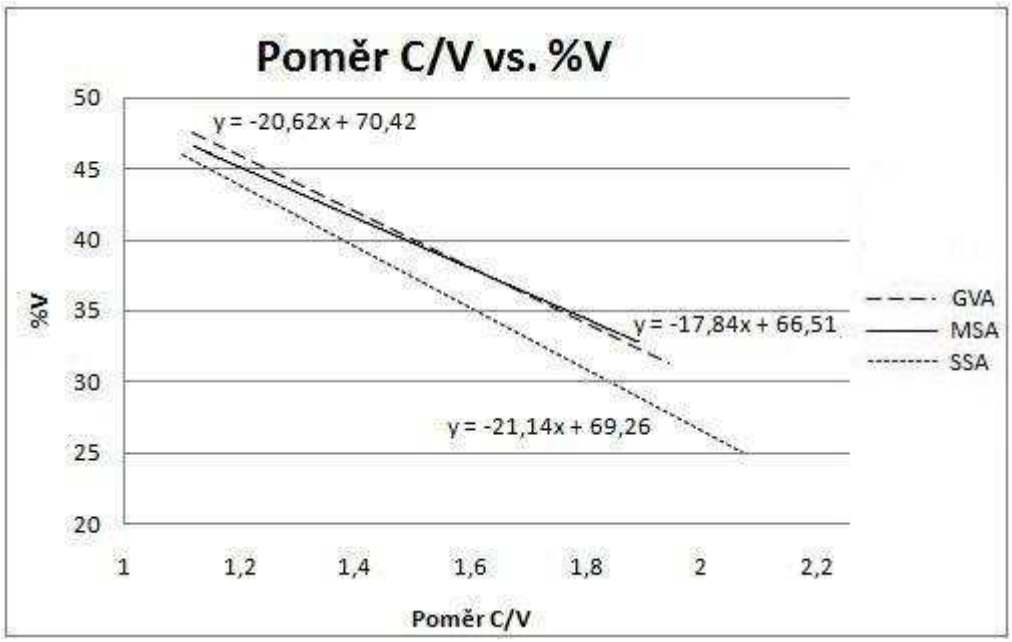
Za zmínku stojí fakt, že korelace pro jednotlivé nahrávky jsou si velmi podobné, kromě nahrávek SSA. Tam je větší rozdíl způsoben výše zmíněným vybočujícím úsekem s velmi vysokým poměrem C/V a velmi nízkým %V. Pokud jej vynecháme, klesne korelace u 2. nahrávky SSA na -0,77 a celý průměr na -0,76.

Ideálním případem ke zkoumání tohoto vlivu by byl stejný text u všech tří mluvčích. Toho by bylo možné dosáhnout jedině kontrolovaným experimentem v laboratorních podmínkách. Bohužel při použití v praxi to lze jen těžko očekávat.



Graf 6: Závislost hodnot %V na hodnotách poměru C/V pro všechny tři mluvčí.

Zobrazíme-li hodnoty %V a poměru C/V pro všechny tři mluvčí do grafu, zjistíme, že výsledná množina je poměrně homogenní. Necháme-li ale spočítat a nakreslit lineární spojnice trendů pro každou mluvčí zvlášť, zjistíme, že jejich sklony se o něco liší.



Graf 7: Lineární spojnice trendů závislosti hodnot %V na hodnotách poměru C/V pro všechny tři mluvčí.

Graf ukazuje sklony spojnic trendů hodnot %V a poměru C/V. Celková tendence (která by ovšem mohla být jazykově specifická) je jasně vidět, projevují se však i odlišnosti u jednotlivých mluvčích. Největší rozdíl je vidět u mluvčí MSA, sklon přímky je nejmenší. To znamená, že tato mluvčí má menší tendenci ke krácení vokálů v úsecích s vyšším poměrem konsonantů. Mluvčí SSA a GVA mají gradient podobný, nicméně z polohy spojnice lze vyvodit, že mluvčí SSA má obecně nižší %V a tedy kratší vokály než GVA – což je ostatně vidět i v grafech 4 a 5.

Přestože poměr C/V a %V spolu do jisté míry korelují, rozdíly mezi mluvčími v poměru C/V nejsou statisticky signifikantní, zato v hodnotách proměnné %V ano (ve dvou ze tří případů). S největší pravděpodobností tedy rozdíly mezi %V jednotlivých mluvčích nejsou způsobeny pouze stavbou textu (a tedy poměrem konsonantů vůči vokálům), ale existují také individuální rysy, které se do hodnot %V promítají a mluvčí od sebe odlišují.

Dalším prověřovaným problémem byla fonologická distinkce délky vokálů a její vliv na %V. Kvantifikovala jsem ji jako procento fonologicky dlouhých vokálů (bez ohledu na jejich skutečné trvání) z celkového počtu vokálů v daném úseku. Všimněme si, že jde o úplně jiný druh vztahu než u proměnné %V, ta počítá trvání vokalických intervalů vzhledem k celkovému trvání artikulace v úseku.

Zásadní vliv fonologické distinkce délky vokálů se neprokázal, korelace %V s procentem dlouhých vokálů byla u všech tří mluvčích pouze nízká až střední.

	1. nahrávka	2. nahrávka	celkem
GVA	0,34	0,21	0,29
MSA	0,66	0,07	0,48
SSA	0,4	0,3	0,35

Tabulka 7: Pearsonovy korelační koeficienty pro hodnoty %V a procento dlouhých vokálů v úseku u jednotlivých nahrávek a celkem pro mluvčí.

Z tabulky 7 vidíme, že korelace %V a procenta dlouhých vokálů je ve všech případech výrazně nižší než korelace s poměrem C/V (viz tabulka 6). Zajímavý je zejména velký rozdíl u mluvčí MSA. Jeden z faktorů, který k němu přispívá, je opět úvodní úsek „Dobré ráno“. Odstraníme-li oba dva tyto úseky z obou nahrávek, korelace u první nahrávky se sníží na 0,57, u druhé se zvýší na 0,08, celková pak bude 0,37. Přesto rozdíl zůstává opravdu významný.

Podíváme-li se na průměrné procentuální zastoupení dlouhých vokálů v textech, pohybuje se ve všech třech případech kolem 21% a rozdíly mezi mluvčími nejsou v tomto ohledu statisticky signifikantní ($p>0,05$). Zajímavé je, že přestože mluvčí MSA má v textu dlouhých vokálů nejméně (19,9%), její %V je nejvyšší.

Všechny zmíněné skutečnosti vedou nakonec k závěru, že vliv distinkce vokalické délky na hodnotu %V je spíše nepodstatný.

Pro další analýzu této proměnné jsem se rozhodla vynechat úseky, které mají velmi vysoký nebo velmi nízký poměr C/V, a to konkrétně menší nebo větší než rozdíl dvou směrodatných odchylek od průměrného poměru C/V pro všechna data. Pro každou mluvčí tak bylo odstraněno kolem šesti úseků, aby byl minimalizován nežádoucí vliv textu na hodnoty %V.

Výsledky t-testů, ve kterých nebyly zahrnuty tyto úseky s příliš vysokým nebo nízkým poměrem C/V se od původních liší pouze nepatrně. Signifikantní byl opět rozdíl mezi MSA a SSA: $t(183) = 4,02$; $p<0,05$ a GVA a SSA: $t(166) = 2,23$; $p<0,05$

Z dat jsem dále vyřadila příliš krátké a příliš dlouhé úseky (brána byla pouze délka artikulace, tj. celkový součet vokalických a konsonantických intervalů v úseku) a ponechány byly ty s délkou mezi 5. a 95. percentilem. Rozdíly se projevily mezi stejnými mluvčími jako v případě kompletních dat – MSA a SSA: $t(176) = 4,86$, $p<0,05$ a GVA a SSA: $t(161) = 3$, $p<0,05$.

Vyzkoušela jsem ještě rozdělit data podle délky artikulace v úseku na nadprůměrně a podprůměrně dlouhé – vzhledem k průměru každé mluvčí. Při porovnání pouze úseků s nadprůměrnou délkou se statisticky signifikantní rozdíl objevil pouze mezi mluvčími MSA a SSA: $t(86) = 2{,}45$, $p<0{,}05$. Při porovnání podprůměrně dlouhých úseků se znovu objevily rozdíly jak mezi MSA a SSA: $t(95) = 3{,}23$, $p<0{,}05$, tak i mezi GVA a SSA: $t(87) = 2{,}01$, $p<0{,}05$. K rozdílu mezi GVA a SSA tedy zřejmě více přispívají krátké úseky.

U rozdílu mezi mluvčími MSA a SSA, který byl vzhledem k %V největší, jsem ještě vyzkoušela odstranit iniciální nádechové úseky (tj. ty, které byly první v odstavci textu). Rozdíl mezi oběma mluvčími zůstal stále statisticky signifikantní: $t(145) = 3{,}01$, $p<0{,}05$. Po odstranění finálních úseků v odstavcích byl výsledek podobný: $t(145) = 3{,}16$, $p<0{,}05$.

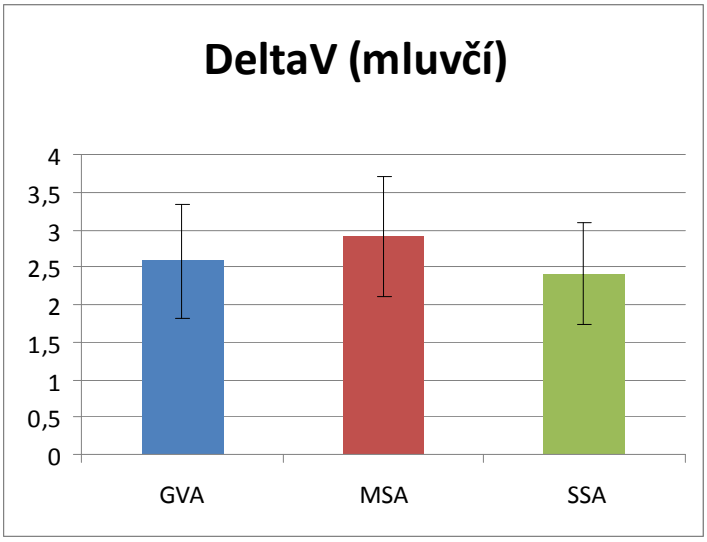
Tedy lze s velkou pravděpodobností říci, že %V opravdu reflektuje globální tendenci mluvčích nějak nakládat s vokály (příp. konsonanty), která se projevuje v každém úseku (a tedy ne například jen zpomalováním na závěr odstavce nebo promluvy). Pro bližší zkoumání lokálních změn (tedy zejména finálního zpomalování v rámci úseku) už %V nestačí, bude potřeba použít proměnnou LAR.

Na závěr jsem poslechově a vizuálně prozkoumala některé vybrané úseky mluvčích MSA a SSA – buďto ty, které měly stejný poměr C/V ale rozdílné %V, nebo takové, kde mluvčí použily stejná slova – a jako nejpravděpodobnější příčina rozdílu v %V mi připadá důkladnější artikulace konsonantů a konsonantických skupin (zejména na rozhraní slov) u druhé mluvčí.

6.3 Ukazatel ΔV

Proměnnou ΔV se měří směrodatná odchylka trvání vokalických intervalů v rámci nádechového úseku. Jde tedy o ukazatel variability vokalických intervalů.

Rozdíly v rámci jednoho mluvčího byly u této proměnné v některých případech dokonce větší než mezi mluvčími. V jednom případě byl statisticky významný rozdíl i mezi oběma nahrávkami jedné mluvčí.



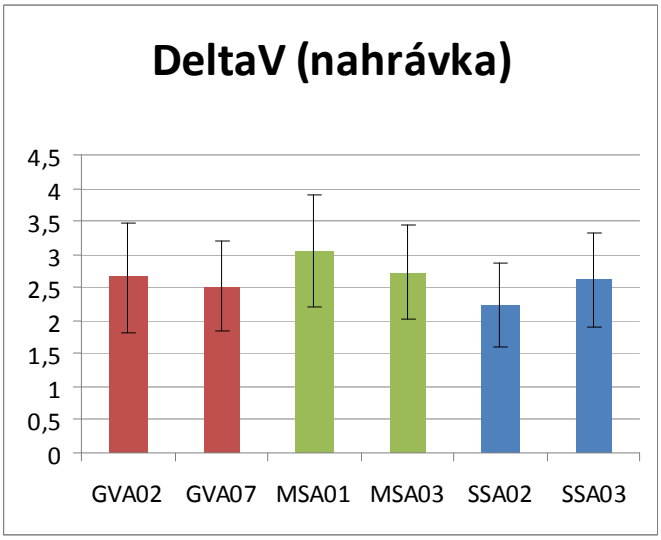
	Průměr (*100)	SD (*100)
GVA	2,58	0,76
MSA	2,90	0,80
SSA	2,41	0,69

Graf 8:Průměrné hodnoty ukazatele ΔV pro jednotlivé mluvčí (vynásobené stem).

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 2,68; p<0,05	t(196) = 4,58; p<0,05	t(179) = 1,56; p>0,05

Tabulka 8: Výsledky t-testů pro dvojice mluvčích a hodnoty ukazatele ΔV.

Z grafu 8 a příslušné tabulky je vidět, že rozdíly mezi mluvčími branými jako celek byly signifikantní opět jen ve dvou případech (ale tentokráte v jiných) – mezi GVA a MSA a mezi MSA a SSA.



	Průměr (*100)	SD (*100)
GVA02	2,64	0,83
GVA07	2,51	0,68
MSA01	3,04	0,85
MSA03	2,73	0,72
SSA02	2,24	0,63
SSA03	2,62	0,71

Graf 9: Průměrné hodnoty ukazatele ΔV pro jednotlivé nahrávky od každé mluvčí.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 0,77; p>0,05	t(94) = 1,93; p>0,05	t(100) = 2,9; p<0,05

Tabulka 9: Výsledky t-testů pro obě nahrávky od jedné mluvčí a hodnoty ukazatele ΔV.

ΔV ukázala také rozdíl mezi oběma nahrávkami mluvčí SSA. U mluvčí MSA byl rozdíl okrajově významný – jen těsně větší než 0,05.

Po vyřazení úseků s vysokým a nízkým poměrem C/V se rozdíly projevily u stejných dvojic dat, jen u mluvčí MSA se rozdíl zvětšil a pravděpodobnost klesla pod hladinu α: $t(89) = 1{,}99$, $p<0{,}05$.

Korelace s poměrem C/V se nedala příliš očekávat, a také žádná nebyla nalezena. Hodnoty korelačních koeficientů pro mluvčí byly u GVA -0,03, u MSA 0,02 a u SSA -0,09. Použila jsem Spearmanův koeficient, protože linearita korelace nebyla příliš pravděpodobná.

Ovšem mohl by se zde výrazně projevit především vliv distinkce vokalické délky. Počet dlouhých vokálů v úseku by mohl přímo ovlivňovat hodnoty ΔV. Zde by o lineární závislost nešlo zcela určitě – mnoho krátkých stejně jako mnoho dlouhých vokálů by variabilitu snižovalo a

nejvyšší by se dala očekávat kolem padesátiprocentního zastoupení dlouhých vokálů. Proto byl pro spočítání korelace použit opět Spearmanův a nikoliv Pearsonův korelační koeficient.

Nicméně ani tak nebyla žádná významná korelace nalezena. Spearmanovy koeficienty byly pro GVA 0,38, pro MSA -0,1 a pro SSA 0,31.

Statisticky signifikantní rozdíl v procentuálním zastoupení dlouhých vokálů u různých nahrávek nebyl nalezen nikde, jak už jsem uvedla výše. Rozdíly v ΔV mezi nahrávkami tedy nejsou způsobeny ani vlivem odlišného počtu dlouhých vokálů v každé nahrávce.

Procentuální zastoupení fonologicky dlouhých vokálů v textu neovlivňuje tedy příliš ani proměnnou ΔV – korelace jsou tak slabé, že nulovou hypotézu nelze zamítnout. Zastoupení dlouhých vokálů v textu nemá tudíž na variabilitu vokalických intervalů nejspíše žádný dopad.

Vyzkoušela jsem stejně jako u %V také odstraňovat iniciální a finální úseky v odstavcích. Po vyřazení iniciálních úseků zůstávají rozdíly téměř stejné, po vyřazení finálních byla ovšem situace zajímavější – rozdíly se objevily mezi všemi třemi dvojicemi mluvčích (v rámci jedné mluvčí zůstal rozdíl u SSA). Je možné, že závěrové dloužení na konci odstavců a promluv rozdíly ve variabilitě vokalických intervalů stírá. Ale také to mohla být statistická chyba zapříčiněná příliš malým vzorkem.

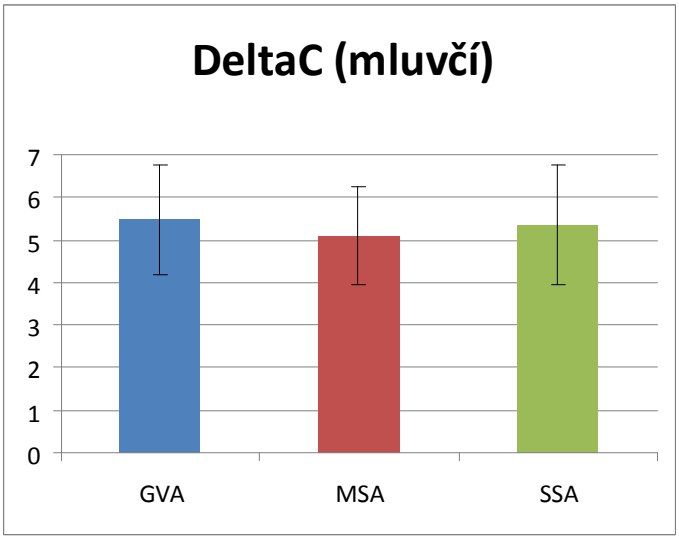
Po odstranění příliš dlouhých a příliš krátkých úseků (opět mezi 5. a 95. percentilem) nastala jediná změna – rozdíl mezi nahrávkami mluvčí MSA se stal signifikantním: $t(84) = 2{,}1$, $p<0{,}05$. Jinak vše zůstalo.

K rozdílu mezi oběma nahrávkami MSA přispívá zejména několik neobvykle vysokých hodnot ΔV u MSA01. Dva ze tří úseků s nejvyšším ΔV totiž obsahují vlastní jméno s neobvyklou fonotaktikou (pro češtinu) a dlouhým vokalickým shlukem (Howard). Stejný trend se projevil i u VarcoV, viz níže.

Na podobný problém narazíme u SSA – první z nahrávek má výrazně nižší průměrné ΔV než druhá. Zde se ale žádná takto zjevná příčina nalézt nedá.

6.4 Ukazatel ΔC

ΔC je směrodatná odchylka trvání konsonantických intervalů. Podle předchozích výsledků by měla spíše reflektovat složitost slabičné struktury v daném jazyce (tj. množství a délku konsonantických shluků) než individuální přístup mluvčího.

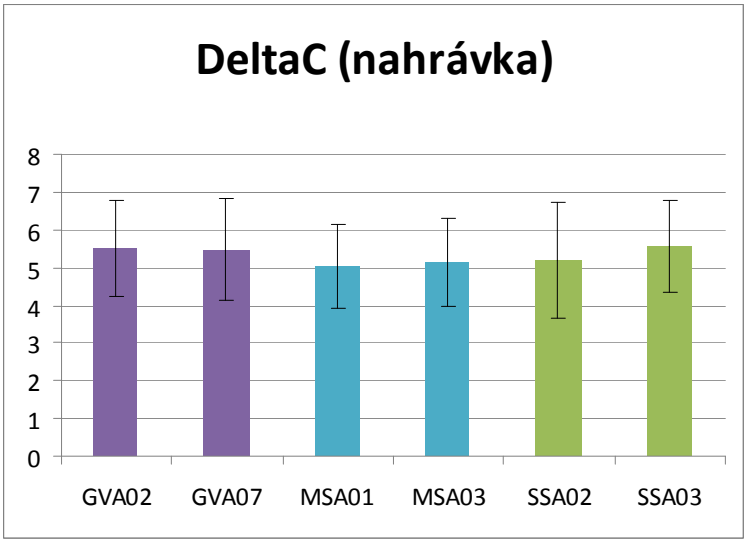


	Průměr (*100)	SD (*100)
GVA	2,58	1,30
MSA	2,90	1,14
SSA	2,41	1,40

Graf 10: Průměrné hodnoty ukazatele ΔC pro jednotlivé mluvčí (vynásobené stem).

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 2,16; p<0,05	t(196) = 1,51; p>0,05	t(179) = 0,56; p>0,05

Tabulka 10: Výsledky t-testů pro dvojice mluvčích a hodnoty ukazatele ΔC.



	Průměr (*100)	SD (*100)
GVA02	5,50	1,26
GVA07	5,48	1,34
MSA01	5,04	1,12
MSA03	5,16	1,17
SSA02	5,20	1,53
SSA03	5,57	1,23

Graf 11: Průměrné hodnoty ukazatele ΔC pro jednotlivé nahrávky každé mluvčí (vynásobené stem).

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 0,06; p>0,05	t(94) = 0,52; p>0,05	t(100) = 1,34; p>0,05

Tabulka 11: Výsledky t-testů pro dvojice nahrávek od každé mluvčí.

Ukázalo se, že proměnná ΔC není příliš vhodná pro diskriminaci mluvčích. Rozlišila od sebe s pravděpodobností $p<0,05$ pouze mluvčí GVA od MSA. Ani rozdíly mezi dvěma nahrávkami od jedné mluvčí se v hodnotách ΔC neprojevily.

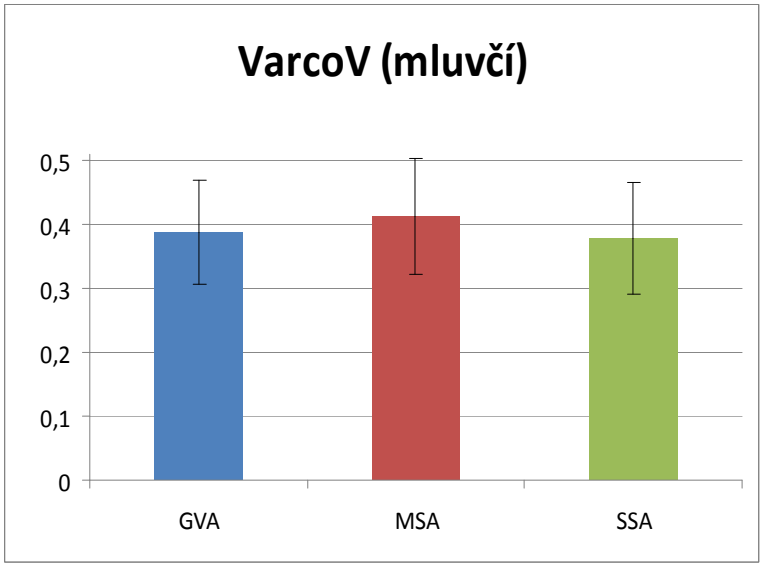
Tato proměnná by mohla být také silně ovlivněná textem, tj. poměrem C/V – je možné, že idiosynkracie mluvčích na ni nemají až takový vliv. Korelační koeficienty jsou středně vysoké: pro GVA je 0,52, pro MSA 0,49 a pro SSA 0,53. Opět byl místo Pearsonova koeficientu použit Spearmanův, protože stejně jako u ΔV se nedá předpokládat jednoznačná lineární závislost obou proměnných (tj. čím více konsonantů, tím větší variabilita konsonantických intervalů).

Po odstranění extrémů poměru C/V nebyl už žádný z rozdílů mezi dvojicemi statisticky signifikantní. Stejný výsledek se objevil i po odstranění příliš krátkých a dlouhých úseků.

Lze tedy říci, že ΔC idiosynkratické chování mluvčích nezachycuje a k rozlišování mezi mluvčími je nevhodná.

6.5 Ukazatel VarcoV

Ukazatel VarcoV je vlastně ΔV normalizované vzhledem k průměrnému trvání vokalických intervalů. Směrodatná odchylka trvání vokalických intervalů se průměrným trváním těchto intervalů vydělí. Normalizace by měla zmenšit vliv různého tempa mezi mluvčími i mezi nádechovými úseky.

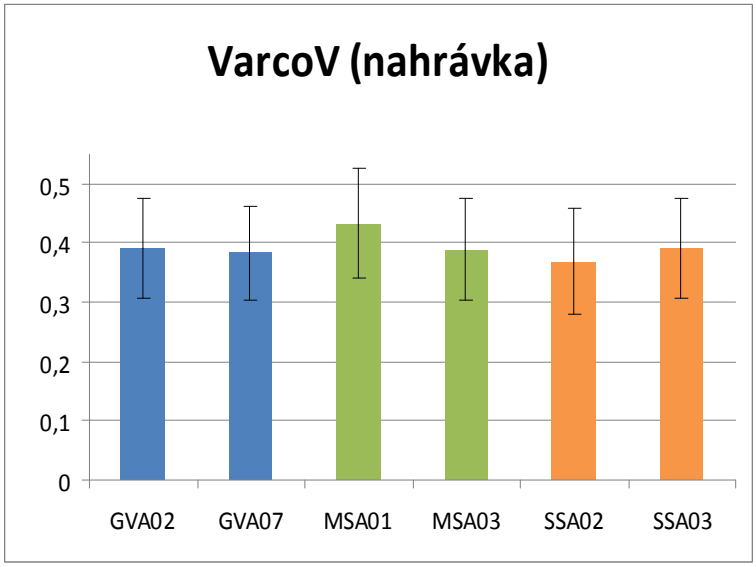


	Průměr	SD
GVA	0,39	0,08
MSA	0,41	0,09
SSA	0,38	0,09

Graf 12: Průměrné hodnoty ukazatele VarcoV pro jednotlivé mluvčí.

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 1,88; p>0,05	t(196) = 2,67; p<0,05	t(179) = 0,71; p>0,05

Tabulka 12: Výsledky t-testů pro dvojice mluvčích a hodnoty VarcoV.



	Průměr	SD
GVA02	0,39	0,08
GVA07	0,38	0,08
MSA01	0,43	0,09
MSA03	0,39	0,09
SSA02	0,37	0,09
SSA03	0,39	0,08

Graf 13: Průměrné hodnoty ukazatele VarcoV pro jednotlivé nahrávky každé mluvčí.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 0,45; p>0,05	t(94) = 2,42; p<0,05	t(100) = 1,22; p>0,05

Tabulka 13: Výsledky t-testů pro VarcoV a dvojice nahrávek každé mluvčí.

Jak vidíme z tabulek 12 a 13, VarcoV od sebe odlišilo jen MSA a SSA, a zároveň také obě nahrávky mluvčí MSA. K obojímu přispěla stejně jako u ΔV vysoká hodnota VarcoV u nahrávky MSA01, kde dva ze tří úseků s nejvyšším VarcoV obsahují slovo s vokalickým shlukem pro češtinu netypickým.

Po odstranění extrémů poměru C/V se zvýraznil rozdíl mezi GVA a MSA, který byl před tím jen okrajově signifikantní: $t(163) = 2{,}04$, $p<0{,}05$. Vše ostatní zůstalo nezměněno.

Korelace VarcoV s poměrem C/V jsou stejně jako u ΔV naprosto zanedbatelné. Spearmanovy korelační koeficienty jsou u GVA 0,06, u MSA 0,11 a u SSA 0,03. Korelace s procentem dlouhých vokálů jsou o něco vyšší, ale stále ještě jde o nízkou korelaci: GVA 0,32, MSA 0,31 a SSA 0,23.

Data bez okrajových hodnot poměru C/V jsem ještě podrobila zkoumání s ohledem na trvání úseku. Po odstranění příliš dlouhých a

krátkých úseků (do 5. a nad 95. percentilem) se situace výrazně změnila – rozdíly u VarcoV zmizely, signifikantní zůstal jen rozdíl mezi oběma nahrávkami MSA. Zahrnutím střídavě krátkých a dlouhých úseků do t-testů bylo zjištěno, že k rozdílu u dvojic GVA/MSA a MSA/SSA přispívají extrémně krátké i dlouhé úseky přibližně stejnou měrou.

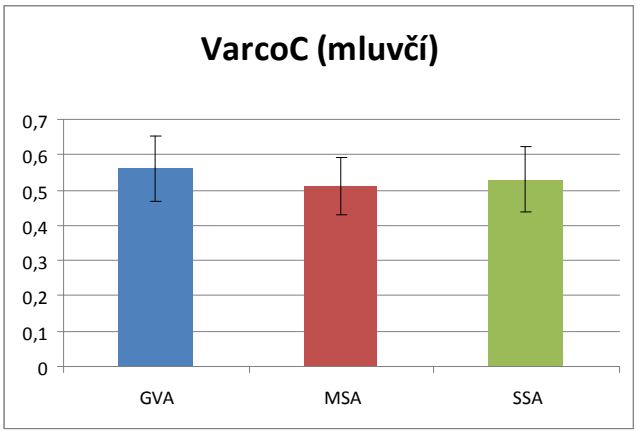
Na VarcoV má tedy vliv mnoho faktorů a rozdíly mezi mluvčími příliš spolehlivě nezobrazuje. Naopak vyzdvihla rozdíl v rámci jedné mluvčí, způsobený do velké míry textem.

Při měření vokalických proměnných ΔV a VarcoV je tedy potřeba dát pozor na obsah textu, který může výsledky zkreslit, obzvlášť při malém objemu dat. Přestože globální poměr C/V má na tyto proměnné jen malý vliv, lokální jevy je v případě mých dat výrazně ovlivnily a zkreslily výsledky.

VarcoV je tedy stejně jako ΔC vhodné pro zobecňování na celý jazyk, individuální rysy příliš nezachycuje.

6.6 Ukazatel VarcoC

Podobně jako VarcoV je i VarcoC směrodatná odchylka trvání konsonantických intervalů normalizovaná vzhledem k průměrnému trvání těchto intervalů.



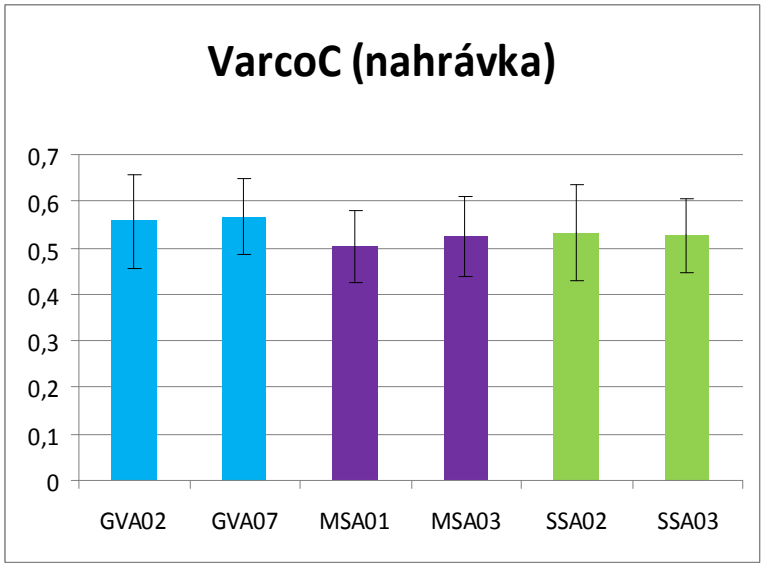
	Průměr	SD
GVA	0,56	0,09
MSA	0,51	0,08
SSA	0,53	0,09

Graf 14: Průměrné hodnoty ukazatele VarcoC pro jednotlivé mluvčí.

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 3,7; p<0,05	t(196) = 1,35; p>0,05	t(179) = 2,28; p<0,05

Tabulka 14: Výsledky t-testů pro dvojice mluvčích a hodnoty VarcoC.

V hodnotách VarcoC se projevil statisticky signifikantní rozdíl mezi dvěma dvojicemi mluvčích – GVA/MSA a GVA/SSA.



	Průměr	SD
GVA02	0,56	0,1
GVA07	0,57	0,08
MSA01	0,50	0,08
MSA03	0,52	0,09
SSA02	0,53	0,1
SSA03	0,53	0,08

Graf 15: Průměrné hodnoty VarcoC pro jednotlivé nahrávky každé mluvčí.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 0,5; p>0,05	t(94) = 1,16; p>0,05	t(100) = 0,31; p>0,05

Tabulka 15: Výsledky t-testů pro dvojice nahrávek od každé mluvčí zvlášť.

Co se týče jednotlivých nahrávek, žádná dvojice nebyla natolik odlišná, aby pravděpodobnost klesla pod hladinu 0,05.

Výraznější vliv poměru konsonantů a vokálů se také u této proměnné nedal příliš očekávat – a vypočítané Spearmanovy korelační koeficienty to potvrdily: u mluvčí GVA byla korelace 0,09, u mluvčí MSA 0,24 a u mluvčí SSA téměř nulová.

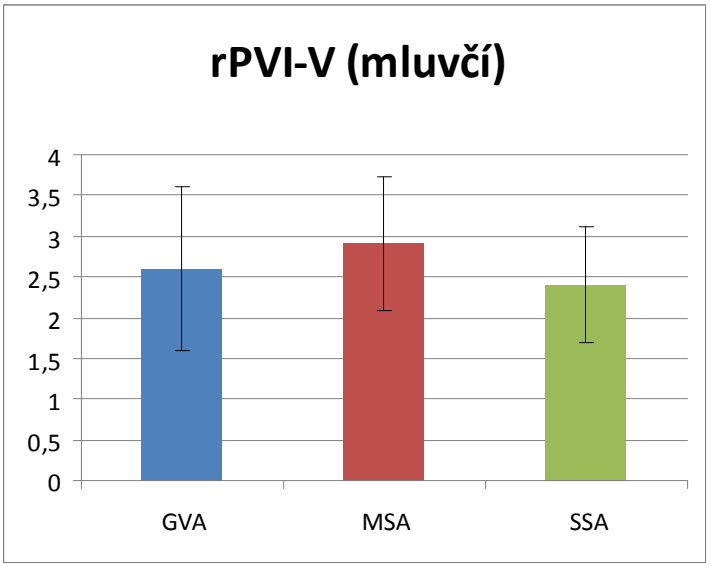
Po odstranění extrémních hodnot poměru C/V zůstaly všechny rozdíly stejně signifikantní jako předtím. Po vyřazení extrémů trvání přestal ale být statisticky signifikantní rozdíl mezi GVA a SSA. Zahrnutím střídavě krátkých a dlouhých úseků bylo zjištěno, že na signifikanci rozdílu mají vliv spíše extrémně krátké úseky – při zahrnutí předtím vyřazených krátkých úseků se signifikance na hladině 0,05 opět objevila, při znovuzahrnutí dlouhých nikoliv.

6.7 Ukazatele PVI

Index párové variability vyjadřuje rozdíl v trvání po sobě následujících vokalických intervalů – jeho normalizovaná verze pak rozdíl ještě dělí průměrným trváním těchto intervalů.

Nejprve se budu věnovat nenormalizovanému indexu vokalických intervalů, tedy rPVI-V.

6.7.1 rPVI-V

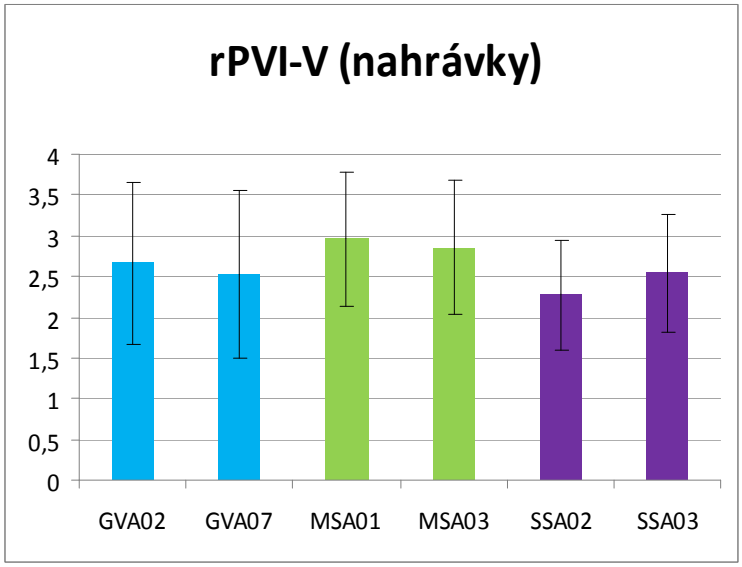


	Průměr (*100)	SD (*100)
GVA	2,60	1,01
MSA	2,91	0,82
SSA	2,40	0,71

Graf 16: Průměrné hodnoty rPVI-V pro všechny tři mluvčí (vynásobené stem).

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 2,25; p<0,05	t(196) = 4,67; p<0,05	t(179) = 1,57; p>0,05

Tabulka 16: Výsledky t-testů pro hodnoty nenormalizovaného vokalického PVI a dvojice mluvčích.



	Průměr (*100)	SD (*100)
GVA02	2,67	0,99
GVA07	2,53	1,03
MSA01	2,96	0,82
MSA03	2,86	0,83
SSA02	2,27	0,68
SSA03	2,55	0,72

Graf 17: Průměrné hodnoty nenormalizovaného vokalického PVI pro každou nahrávku zvlášť (vynásobené stem).

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 0,63; p>0,05	t(94) = 0,6; p>0,05	t(100) = 1,98; p>0,05

Tabulka 17: Výsledky t-testů pro hodnoty nenormalizovaného vokalického PVI a dvojice nahrávek od každé mluvčí.

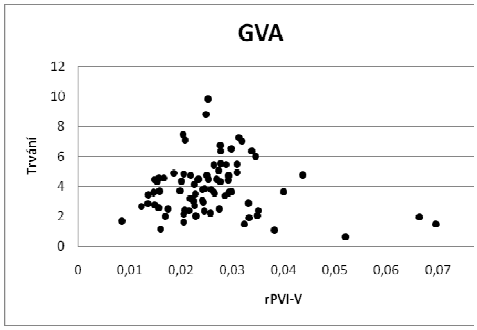
T-testy našly v hodnotách ukazatele rPVI-V signifikantní rozdíl mezi dvojicemi mluvčích GVA – MSA a MSA – SSA. U dvojic nahrávek každé mluvčí rozdíl ani jednou nepřekročil hranici 95% spolehlivosti, i když v případě SSA jen velmi těsně.

Korelace s poměrem C/V se neobjevila. Pearsonovy i Spearmanovy korelační koeficienty se pohybovaly okolo nuly. Po odstranění extrémů poměru C/V se signifikance rozdílů při t-testech nezměnila. Lze tedy prohlásit, že na tento ukazatel nemá slabičná struktura textu pravděpodobně žádný vliv.

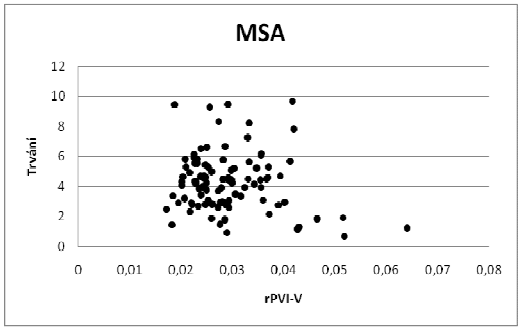
Na druhou stranu, korelace s dlouhými vokály už se našla o hodně vyšší, Spearmanův koeficient pro GVA je 0,4, pro MSA 0,45 a pro SSA 0,34 (protože by opět byla pravděpodobnější nelineární korelace). Ovšem stále je to příliš nízké pro dokázání nějakého většího vlivu.

Všimla jsem si při tom zajímavé věci – u každé mluvčí jsou hodnoty rPVI-V umístěny v poměrně kompaktním chumlu, ale několik hodnot je neobvykle vysokých a vybočujících. Kromě mluvčí SSA jde o velmi krátké úseky (artikulace pod dvě vteřiny a čtyři až deset slabik), ve kterých malé množství měřených intervalů zřejmě způsobuje tak vysokou variabilitu. U SSA má nejvyšší rPVI-V úsek dlouhý přes 4 vteřiny a nevykazuje žádné zvláštní znaky, které by rozdíl v hodnotě ukazatele mohly zapříčinit.

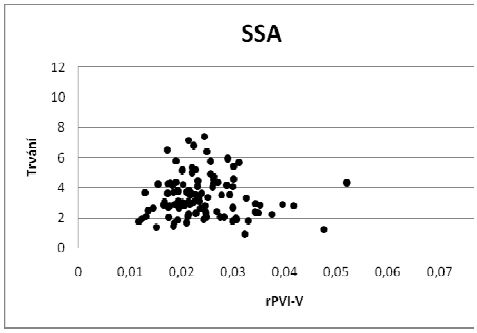
rPVI-V v závislosti na trvání úseku v sekundách zobrazují grafy 17a-c.



Graf 18a: Hodnoty rPVI-V pro mluvčí GVA v závislosti na trvání úseku.



Graf 18b: Hodnoty rPVI-V pro mluvčí MSA v závislosti na trvání úseku.



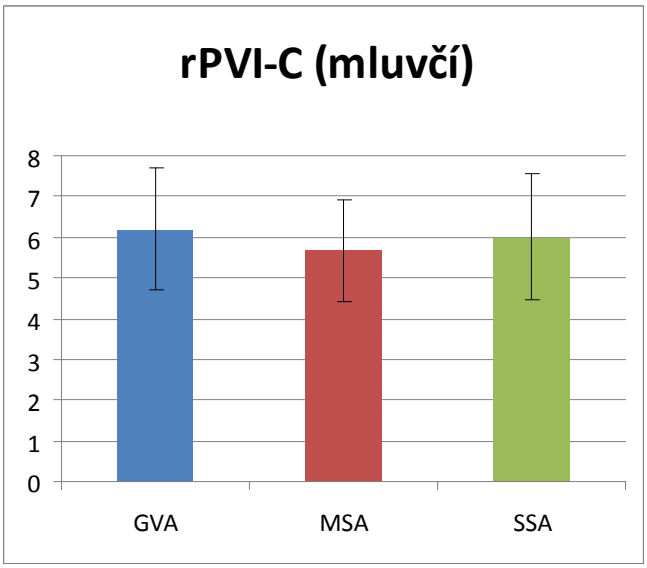
Graf 18c: Hodnoty rPVI-V pro mluvčí SSA v závislosti na trvání úseku.

Nabízí se jasná možnost – odstranit krátké úseky, a podívat se, co to s rozdíly udělá. Odstraňovat okrajově dlouhé by nemělo být nutné, jejich rPVI-V hodnoty jsou vesměs kolem průměru.

Po odstranění hodnot pod 5. percentilem zůstalo vše stejné jako před odstraněním, rozdíl mezi GVA a MSA stále nebyl signifikantní (a dokonce se ještě zmenšil). Stejně to dopadlo i když jsem odstranila hodnoty pod 10. percentilem a následně úseky s trváním pod 2 sekundy. Trochu lepšího výsledku bylo dosaženo po odstranění přímo daných vybočujících hodnot – nicméně stále ne dost na to, aby byl rozdíl signifikantní na hladině $\alpha = 0{,}05$. Mezi těmito dvěma mluvčími tedy rozdíl v nenormalizovaném vokalickém indexu párové variability není dost velký.

6.7.2 rPVI-C

Do rPVI-C – tedy nenormalizovaného indexu párové variability konsonantických intervalů – jsem s ohledem na výsledky Yoona (2010) a dalších vkládala naději. Překvapivě se ale ukázalo, že v případě mého materiálu jde o jednu z proměnných s nejmenší schopností diskriminace mezi mluvčími.

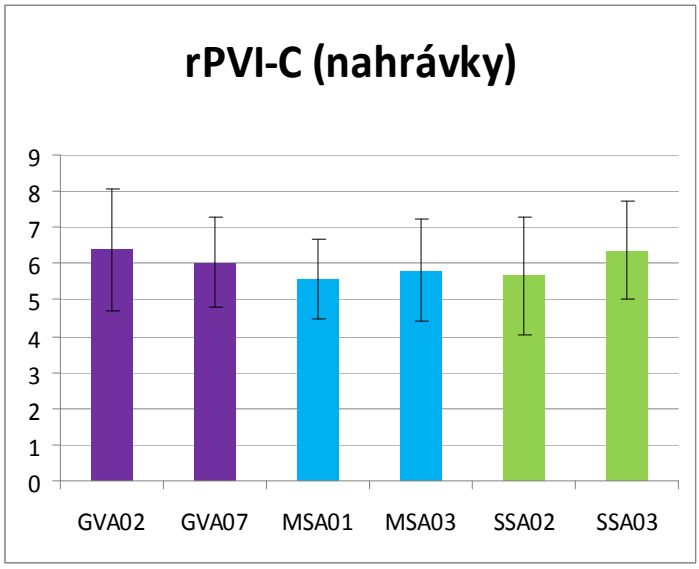


	Průměr (*100)	SD (*100)
GVA	6,21	1,48
MSA	5,68	1,25
SSA	6,00	1,55

Graf 19: Průměrné hodnoty rPVI-C pro jednotlivé mluvčí.

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 2,54; p<0,05	t(196) = 1,56; p>0,05	t(179) = 0,92; p>0,05

Tabulka 18: Výsledky t-testů pro hodnoty rPVI-C a dvojice mluvčích.



	Průměr (*100)	SD (*100)
GVA02	6,38	1,69
GVA07	6,03	1,23
MSA01	5,57	1,11
MSA03	5,81	1,4
SSA02	5,68	1,64
SSA03	6,37	1,37

Graf 20: Hodnoty rPVI-C pro jednotlivé nahrávky každé mluvčí.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 1,05; p>0,05	t(94) = 0,96; p>0,05	t(100) = 2,3; p<0,05

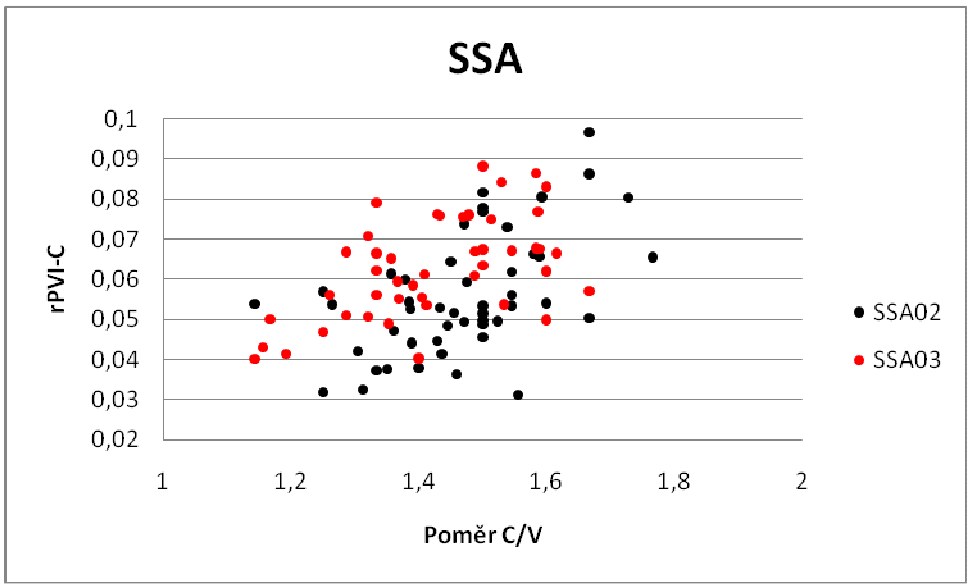
Tabulka 19: Výsledky t-testů pro hodnoty rPVI-C a obě nahráky od každé mluvčí.

Hodnoty rPVI-C od sebe odlišily pouze mluvčí GVA a MSA, ale zároveň také obě nahrávky mluvčí SSA. Z grafu 20 je ale vidět, že průměry všech nahrávek jsou poměrně rozptýlené.

Korelace s poměrem C/V se na rozdíl od rPVI-V pohybovala o něco výše, GVA 0,48, MSA 0,3 a SSA 0,47. To naznačuje, že na tento ukazatel má složitost slabičné struktury nějaký, i když možná jen malý, vliv. Statistická signifikance rozdílů se po vyřazení okrajových hodnot poměru C/V z t-testů nezměnila.

Při porovnávání dat ještě navíc bez okrajově krátkých a dlouhých úseků došlo k jediné změně – rozdíl mezi GVA a MSA přestal být signifikantní a zůstal pouze rozdíl mezi oběma nahrávkami SSA.

Při zkoumání příčin rozdílu mezi SSA02 a SSA03 bylo zjištěno, že v obou nahrávkách se vyskytuje vždy jedna hodnota výrazně vyšší než ostatní – jedna z nich je způsobená složitou slabičnou strukturou s vysokým poměrem C/V (text: „především v příhraničních oblastech“) a druhá epentetickým vokálem na rozhraní slov následovaným drobnou pauzou v délce přibližně 40 ms – obojí velmi prodloužilo délku daného konsonantického intervalu. Nicméně tyto dva úseky nemají, zdá se, na celkový stav rPVI-C žádný vliv, protože rozdíl v této proměnné mezi oběma nahrávkami se projevuje i po jejich nezahrnutí do testů.

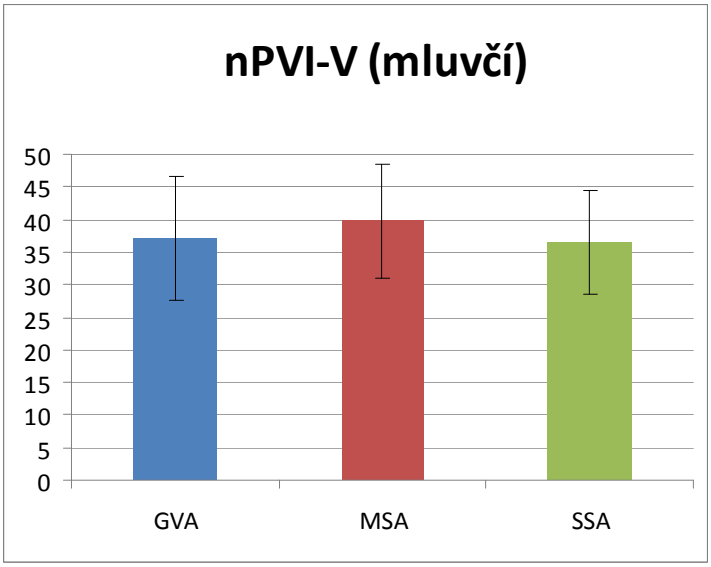


Graf 21: Hodnoty rPVI-C v nahrávkách SSA02 a SSA03, v závislosti na poměru C/V.

Prozkoumáme-li každou nahrávku zvlášť, po odstranění úseků s extrémními hodnotami C/V, zjistíme, že rPVI-C koreluje s poměrem C/V o něco výrazněji, než když bereme v úvahu obě nahrávky společně. U obou je Pearsonův korelační koeficient roven přibližně 0,54. V grafu 21 je jasně vidět, že úseky SSA03 mají častěji nižší poměr C/V. Je tedy možné, že rozdíl v textu mezi oběma nahrávkami, přestože sám o sobě není statisticky významný, přispívá k utvoření rozdílu v rPVI-C.

6.7.3 nPVI-V

Normalizujeme-li rPVI-V vzhledem k průměrnému trvání měřené dvojice intervalů, dostaneme tento ukazatel, který by měl být méně náchylný k ovlivnění zejména tempem promluvy.



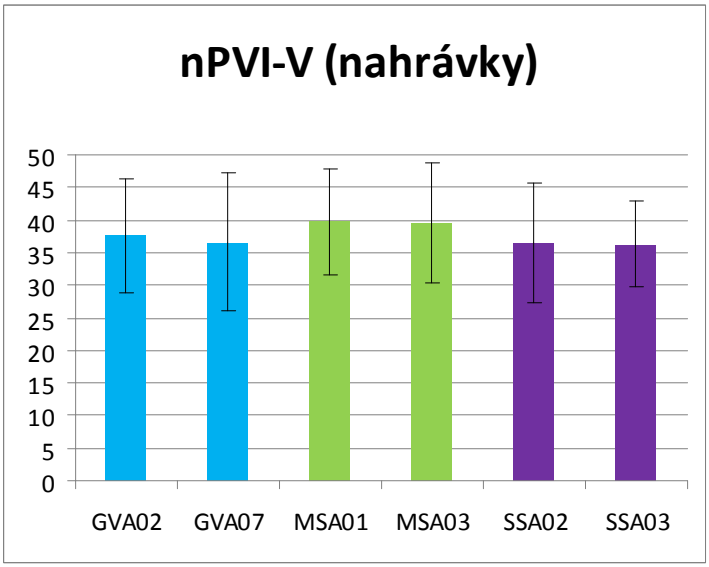
	Průměr	SD
GVA	37,1	9,57
MSA	39,73	8,67
SSA	36,44	8

Graf 22: Průměrné hodnoty ukazatele nPVI-V pro jednotlivé mluvčí.

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 1,9; p>0,05	t(196) = 2,77; p<0,05	t(179) = 0,5; p>0,05

Tabulka 20: Výsledky t-testů pro hodnoty nPVI-V a dvojice mluvčích.

V hodnotách tohoto ukazatele se projevil statisticky signifikantní rozdíl pouze mezi MSA a SSA. U dvojice GVA – MSA pak zůstal rozdíl jen těsně nad hladinou spolehlivosti.



	Průměr	SD
GVA02	37,58	8,65
GVA07	36,61	10,51
MSA01	39,86	8,12
MSA03	39,57	9,35
SSA02	36,52	9,11
SSA03	36,34	6,58

Graf 23: Průměrné hodnoty nPVI-V pro jednotlivé nahrávky každé mluvčí.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 0,45; p>0,05	t(94) = 0,16; p>0,05	t(100) = 0,11; p>0,05

Tabulka 21: Výsledky t-testů pro hodnoty nPVI-V a dvojice nahrávek od každé mluvčí.

Z tabulek 20 a 21 je ale vidět, že normalizace proměnné rPVI-V rozdíly mezi mluvčími značně zmenšila – a o rozdílech v rámci jedné mluvčí to platí dvojnásob, tam se hodnoty téměř vyrovnaly.

Po odebrání okrajových hodnot poměru C/V došlo ke zvětšení rozdílu mezi GVA a MSA a zároveň téměř ke smazání rozdílu mezi GVA a SSA. Výsledky ukazuje tabulka 22. Hodnoty rozdílů v rámci mluvčích se nijak výrazně nezměnily.

GVA/MSA	MSA/SSA	GVA/SSA
t(163) = 2,25; p<0,05	t(185) = 2,57; p<0,05	t(168) = 0,005; p>0,05

Tabulka 22: Výsledky t-testů pro dvojice mluvčích a ukazatel nPVI-V po vyřazení úseků s extrémně vysokým nebo nízkým poměrem C/V.

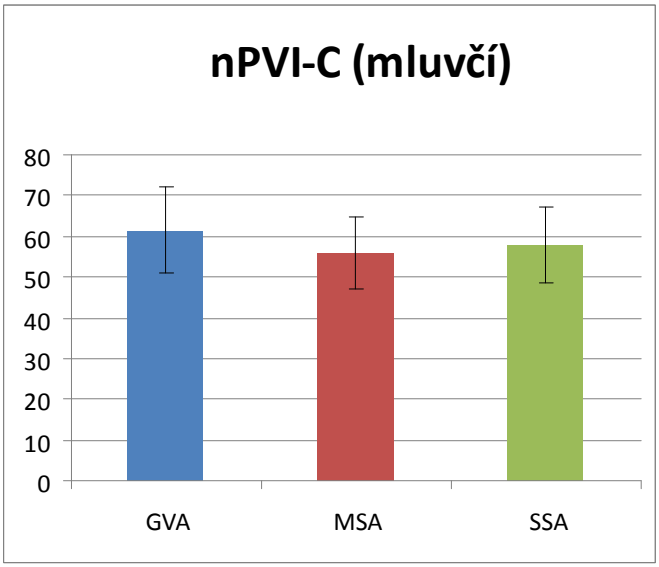
Vzhledem k výsledkům u rPVI-V se ani zde žádná korelace s poměrem C/V nedala očekávat, což potvrdily i následné výpočty koeficientů.

Z upravených dat byly opět dále vyřazeny příliš dlouhé a příliš krátké úseky. Následkem této změny byly smazány úplně všechny rozdíly – jejich významnost stoupla nad hladinu alfa. GVA/MSA: $t(155) = 1{,}94$; $p>0{,}05$ a MSA/SSA: $t(176) = 1{,}91$; $p>0{,}05$. Střídavým zahrnutím původně vyřazených krátkých a dlouhých úseků jsem usoudila, že k rozdílu přispívaly všechny, u MSA/SSA pak mnohem více krátké. Situace je to velmi podobná jako u VarcoV.

Tento ukazatel opět příliš nekoreluje ani s ukazatelem procenta dlouhých vokálů v úsecích, korelace jsou pro GVA 0,38 a pro MSA a SSA 0,32. Spearmanovy koeficienty jsou velmi podobné.

6.7.4 nPVI-C

Poslední z globálních ukazatelů je stejně jako předchozí normalizovanou verzí rPVI-C, tj. indexu variability po sobě jdoucích párů konsonantických intervalů.



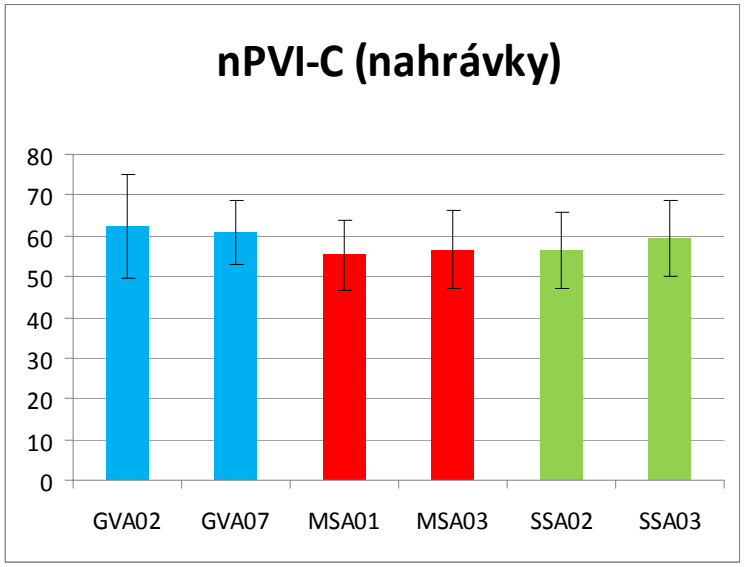
	Průměr	SD
GVA	61,57	10,52
MSA	55,97	9,06
SSA	57,86	9,49

Graf 24: Průměrné hodnoty ukazatele nPVI-C pro jednotlivé mluvčí.

GVA/MSA	MSA/SSA	GVA/SSA
t(173) = 3,79; p<0,05	t(196) = 1,44; p>0,05	t(179) = 2,49; p<0,05

Tabulka 23: Výsledky t-testů pro ukazatel nPVI-C a dvojice mluvčích.

Statisticky významný rozdíl se objevil mezi dvojicemi mluvčích GVA a MSA a také GVA a SSA.



	Průměr	SD
GVA02	62,23	12,63
GVA07	60,89	7,9
MSA01	55,37	8,62
MSA03	56,64	9,58
SSA02	56,55	9,43
SSA03	59,39	9,43

Graf 25: Průměrné hodnoty ukazatele nPVI-C pro jednotlivé nahrávky každé mluvčí.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(77) = 0,57; p>0,05	t(94) = 0,69; p>0,05	t(100) = 1,51; p>0,05

Tabulka 24: Výsledky t-testů pro ukazatel nPVI-C a dvojice nahrávek od každé mluvčí zvlášť.

V rámci jednotlivých mluvčích se rozdíl neobjevil nikde, jak je jasně vidět z tabulky 24.

Vliv textu kvantifikovaný jako poměr počtu konsonantů a počtu vokálů s touto proměnnou příliš nekoreluje. U mluvčí GVA byl korelační koeficient 0,12, u mluvčí MSA 0,16 a u SSA 0,27.

Po zopakování t-testů bez okrajových hodnot poměru C/V zůstávají rozdíly signifikantní u stejných dvojic. Tentýž výsledek dostaneme i po odstranění okrajově krátkých a dlouhých úseků.

6.8 Ukazatel LAR

Proměnná LAR není představována jen jednou hodnotou pro každý nádechový úsek, jako tomu bylo u předchozích proměnných, nýbrž množinou hodnot, které vyjadřují převrácenou hodnotu vzdálenosti středů následujících vokalických intervalů. Každý nádechový úsek má pak tolik hodnot LAR, kolik má vokalických intervalů, mínus jedna. V drtivé většině případů pak koresponduje i s počtem slabik.

Každý nádechový úsek jde zachytit a popsat křivkou hodnot LAR. K vyhlazení extrémů této křivky byl použit tříbodový klouzavý harmonický průměr.

Závěrové zpomalování bylo kvantifikováno s pomocí posledních šesti hodnot LAR (čtyř pro vyhlazená data) – to znamená posledních sedm slabik úseku. Úseky, které byly kratší nebo obsahovaly v této části dysfluenci či úsekový prozodický předěl, byly z analýzy vyřazeny, protože by mohly zcela zamlžit hledaný trend. Z těchto dat byla nakonec vypočítána směrnice lineární regrese, která popisuje pokles či růst hodnot. Právě tyto směrnice pak byly použity k finální analýze a diskriminaci mluvčích.

Celkem úseků s minimálně 7 slabikami bylo 272, z toho 199 končilo melodémem ukončujícím klesavým, 72 melodémem neukončujícím a 1 melodémem ukončujícím stoupavým. Dále pak byly vyfiltrovány úseky, které v této poslední části obsahovaly dysfluenci nebo prozodický úsekový předěl. Zbylo jich 115 pro melodém ukončující klesavý a 44 pro melodém neukončující.

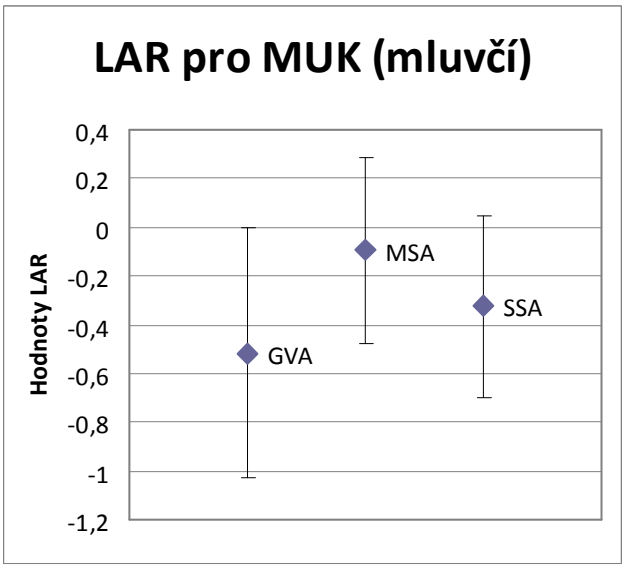
Dále jsem analyzovala primárně úseky s melodémem klesavým, sekundárně pak i s melodémem neukončujícím.

6.8.1 LAR pro úseky s melodémem ukončujícím klesavým

Práce s proměnnou LAR pro úseky s klesavým melodémem probíhala ve dvou krocích. Nejprve jsem pro takto určená data (tj. směrnice lineární regrese pro každý úsek) provedla všechny t-testy a vyhodnotila je. Statisticky významné rozdíly se projevily pouze u dvojice GVA – MSA a u vyhlazených dat ještě u dvojice GVA – SSA, což bylo méně, než jsem očekávala. Ovšem příčina mohla ležet nikoliv v neexistujících rozdílech mezi mluvčími, nýbrž v nesprávném určení, zda finální část úseku obsahuje prozodický předěl.

Proto jsem vybrané úseky ještě jednou prošla a vyřadila z nich ty, u kterých jsem si významností předělu v poslední části (zda je taktový nebo úsekový) nebyla zcela jistá. Z celkového počtu úseků jich bylo z analýzy takto vyřazeno 13 (zhruba 10%), zbylo tedy 102 úseků s melodémem ukončujícím klesavým.

Výsledky této druhé analýzy uvádím v grafech a tabulkách níže, nejprve pro nevyhlazené hodnoty LAR.

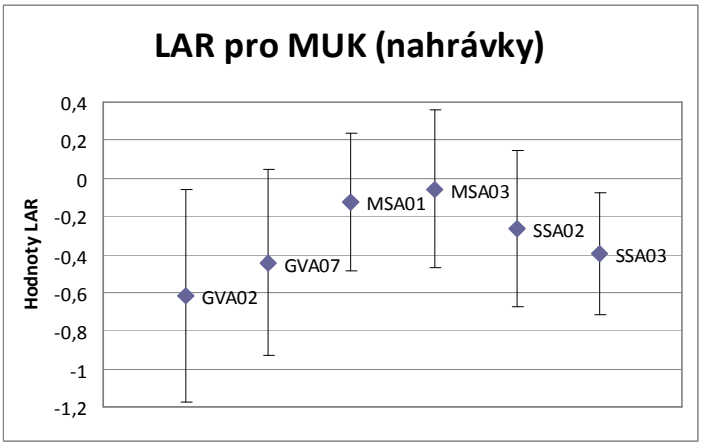


	Průměr	SD
GVA	-0,52	0,52
MSA	-0,10	0,38
SSA	-0,33	0,37

Graf 26: Průměrné hodnoty směrnic lineární regrese ukazatele LAR pro úseky končící melodémem ukončujícím klesavým a jednotlivé mluvčí.

GVA/MSA	MSA/SSA	GVA/SSA
t(65) = 3,84; p<0,05	t(72) = 2,61; p<0,05	t(61) = 1,7; p>0,05

Tabulka 25: Výsledky t-testů pro hodnoty směrnic lineární regrese ukazatele LAR v úsecích s melodémem ukončujícím klesavým.



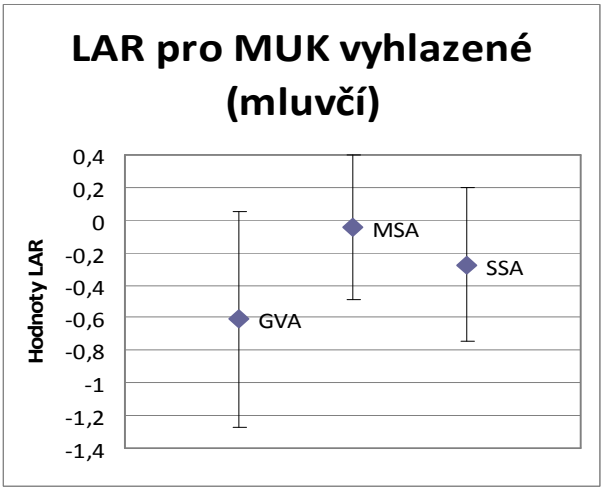
	Průměr	SD
GVA02	-0,62	0,56
GVA07	-0,44	0,49
MSA01	-0,12	0,36
MSA03	-0,06	0,41
SSA02	-0,26	0,41
SSA03	-0,40	0,32

Graf 27: Průměrné hodnoty směrnic lineární regrese ukazatele LAR pro úseky končící melodémem ukončujícím klesavým v jednotlivých nahrávky od každé mluvčí.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(26) = 0,88; p>0,05	t(37) = 0,53; p>0,05	t(33) = 1,06; p>0,05

Tabulka 26: Výsledky t-testů pro hodnoty směrnic lineární regrese ukazatele LAR v úsecích s melodémem ukončujícím klesavým.

Nevyhlazená data od sebe rozpoznala dvě ze tří dvojic mluvčích, rozdíly v rámci jedné mluvčí se neprojevily. U vyhlazených dat byl výsledek optimální – byly rozpoznány všechny tři dvojice mluvčích.



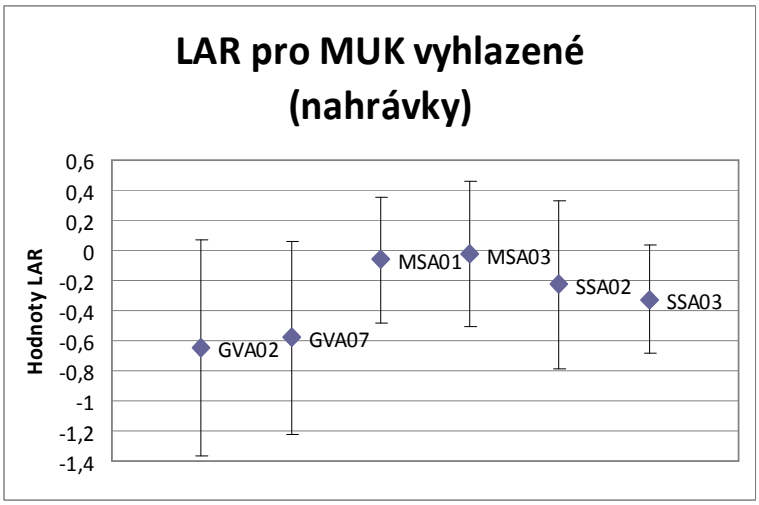
	Průměr	SD
GVA	-0,61	0,66
MSA	-0,05	0,44
SSA	-0,27	0,47

Graf 28: Průměrné hodnoty směrnic lineární regrese ukazatele LAR vyhlazené tříbodovým klouzavým průměrem pro úseky s melodémem ukončujícím klesavým v závěru.

GVA/MSA	MSA/SSA	GVA/SSA
t(65) = 4,16; p<0,05	t(72) = 2,14; p<0,05	t(61) = 2,3; p<0,05

Tabulka 27: Výsledky t-testů pro hodnoty směrnic lineární regrese ukazatele LAR vyhlazené tříbodovým klouzavým průměrem v úsecích s melodémem ukončujícím klesavým.

Z tabulky 27 je jasné, že ukazatel LAR zúžený na posledních 6 hodnot každého úseku konečně dosáhl kýženého úspěchu, dokázal od sebe odlišit všechny tři mluvčí. I přesto, že jak je vidět z chybových úseček, variabilita dat je poměrně značná. Mohou k ní přispívat zejména lingvistické faktory – zda úsek ukončuje celou promluvu nebo jen větu v odstavci, zda ve zkoumané části není slovo s výrazným přízvukem, apod. Nicméně co se týče polohy úseku v textu, dá se předpokládat, že všechny mluvčí měly situaci podobnou. A umisťování přízvuku na závěr nádechového úseku, tedy zdůrazňování posledního slova či taktu by mohlo být právě to individuální chování mluvčí, které se zde projevuje (všimla jsem si toho hlavně u mluvčí MSA).



	Průměr	SD
GVA02	-0,65	0,72
GVA07	-0,58	0,64
MSA01	-0,06	0,42
MSA03	-0,02	0,48
SSA02	-0,23	0,55
SSA03	-0,33	0,36

Graf 29: Průměrné hodnoty směrnic lineární regrese ukazatele LAR vyhlazené tříbodovým klouzavým průměrem pro úseky s melodémem ukončujícím klesavým v jednotlivých nahrávkách od každé mluvčí.

GVA02/GVA07	MSA01/MSA03	SSA02/SSA03
t(26) = 0,25; p>0,05	t(37) = 0,3; p>0,05	t(33) = 0,6; p>0,05

Tabulka 28: Výsledky t-testů pro hodnoty směrnic lineární regrese ukazatele LAR vyhlazené tříbodovým klouzavým průměrem v úsecích s melodémem ukončujícím klesavým.

Rozdíly v rámci jedné mluvčí jsou opět pod hranicí spolehlivosti, a to dost výrazně.

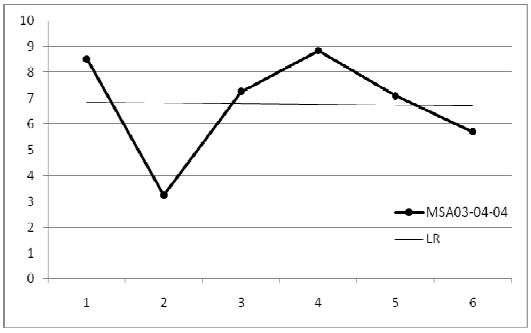
Z grafu 29 je také vidět další zajímavý fakt – porovnáme-li jej s grafem 28, zjistíme, že data vyhlazená tříbodovým kouzavým průměrem mají vesměs vyšší směrodatnou odchylku a tedy vyšší variabilitu než data nevyhlazená. Tento fakt jsem si nedokázala nijak vysvětlit, proto jsem podrobněji prozkoumala pět úseků s největším rozdílem v hodnotě směrnice lineární regrese mezi vyhlazenými a nevyhlazenými daty. Údaje o nich jsou uvedeny v následující tabulce.

Ozn. úseku	LR	LR (vyhl.)	Text posledních 7 slabik	Slabičná struktura textu
MSA03-04-04	-0,03	0,73	(po)pulační politika	CV-CV-CCV-CV-CV-CV-CV
GVA07-06-02	-0,35	-0,99	promile alkoholu	CCV-CV-CV-CVC-CV-CV-CV
GVA07-05-06	-0,51	0,12	ministerstva kultury	CV-CV-CCVC-CCCV-CVC-CV-CV
SSA02-03-02	-0,21	0,41	zdražovat cigarety	CCCV-CV-CVC-CV-CV-CV-CV
SSA03-08-01	-0,28	0,32	německém Oberhofu	CV-CV-CCVC-CV-CV-CV-CV

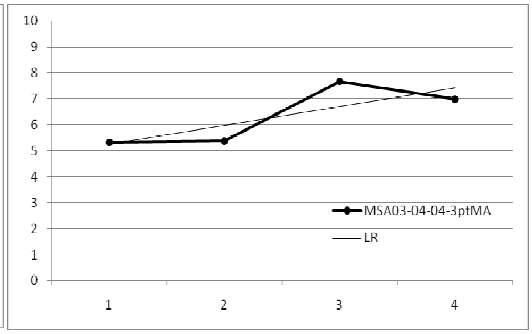
Tabulka 29: Údaje o vybraných úsecích s největším rozdílem mezi směrnicemi lineární regrese pro vyhlazená a nevyhlazená data.

Ve druhém a třetím sloupci tabulky jsou uvedeny hodnoty směrnice lineární regrese pro LAR v daném úseku. Druhý údaj je spočítán z vyhlazených dat. Ve čtvrtém sloupci je text posledních sedmi slabik a v pátém sloupci pak jeho slabičná struktura. Rázy jsou brány jako konsonanty.

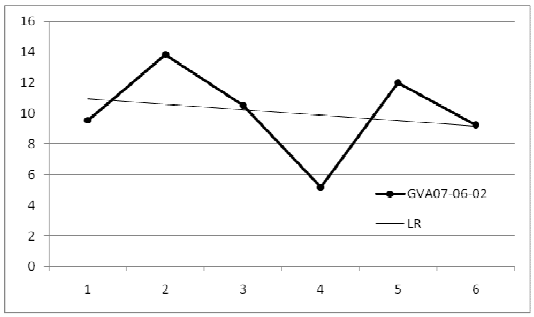
Kromě úseku GVA07-05-06 mají všechny podobnou slabičnou strukturu – CV s jednou složitější slabikou uvnitř (konsonanty před prvním vokálem jsou irelevantní, protože první hodnota LAR je spočítána jako vzdálenost středu prvního vokalického intervalu od druhého). Křivky hodnot LAR v těchto úsecích jsou zobrazeny v grafech níže.



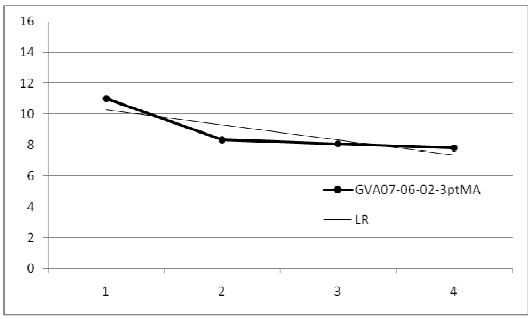
Graf 30a: Křivka hodnot LAR pro úsek MSA03-04-04 a směrnice její lineární regrese.



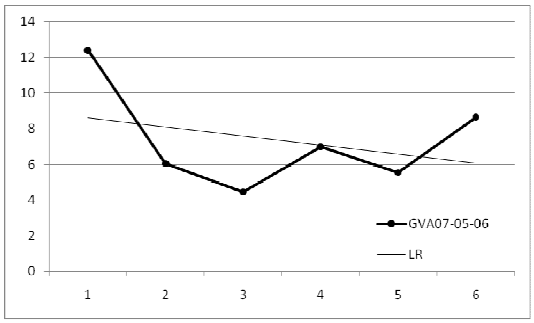
Graf 30b: Křivka vyhlazených hodnot LAR pro MSA03-04-04 a směrnice její lineární regrese.



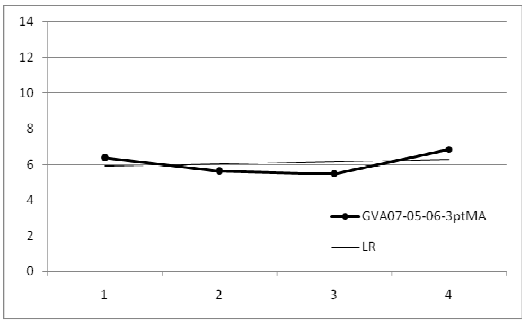
Graf 30c: Křivka hodnot LAR pro úsek GVA07-06-02 a směrnice její lineární regrese.



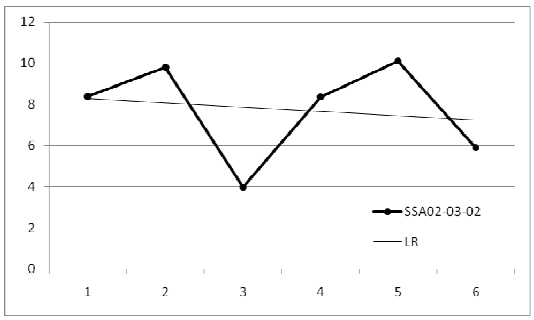
Graf 30d: Křivka vyhlazených hodnot LAR pro GVA07-06-02 a směrnice její lineární regrese.



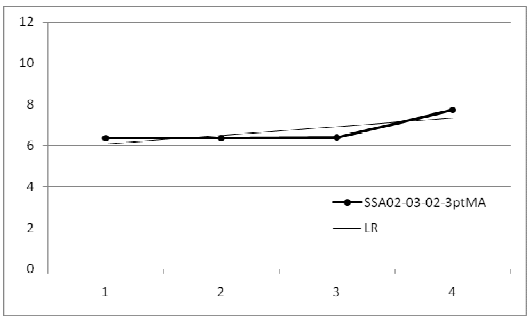
Graf 30e: Křivka hodnot LAR pro úsek GVA07-05-06 a směrnice její lineární regrese.



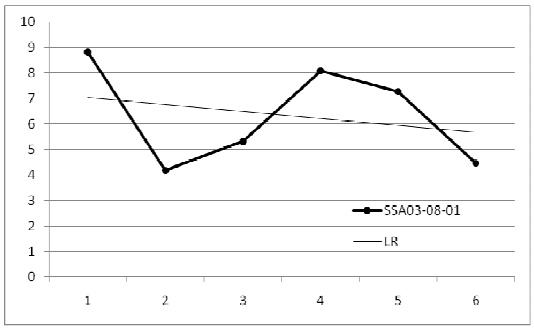
Graf 30f: Křivka vyhlazených hodnot LAR pro GVA07-05-06 a směrnice její lineární regrese.



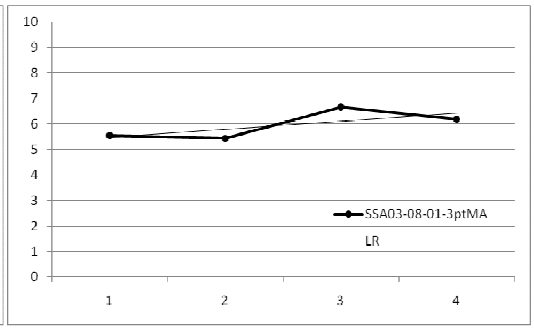
Graf 30g: Křivka hodnot LAR pro úsek SSA02-03-02 a směrnice její lineární regrese.



Graf 30h: Křivka vyhlazených hodnot LAR pro SSA02-03-02 a směrnice její lineární regrese.



Graf 30i: Křivka hodnot LAR pro úsek SSA03-08-01 a směrnice její lineární regrese.



Graf 30j: Křivka vyhlazených hodnot LAR pro SSA03-08-01 a směrnice její lineární regrese.

Propady v nevyhlazených křivkách jsou způsobeny právě konsonantickými shluky, které zvětšují vzdálenost vokalických intervalů od sebe, a tak snižují hodnotu LAR. Rozdíly mezi směrnicemi regrese jsou způsobeny právě těmito propady, které ve vyhlazené křivce už nejsou zdaleka tak výrazné.

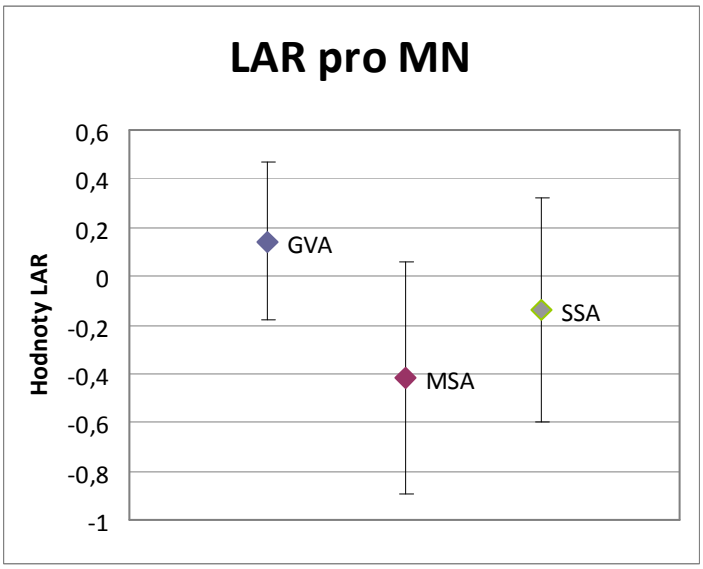
Slabičná struktura takovýchto úseků je ale zcela přirozená, konsonantické shluky se v češtině vyskytují běžně, takže určitě nejde o ojedinělé výjimky. Z grafů 30a-j se zdá, že tyto shluky obecně velikosti gradientů nevyhlazených křivek spíše snižují, a nejspíš proto jsou nevyhlazená data semknutější.

Vliv konsonantických shluků na hodnoty LAR je tedy poměrně velký – při počtu šesti hodnot to je ostatně docela očekávatelné – takže pokud jej chceme minimalizovat, je rozhodně potřeba křivky vyhlazovat.

Směrnice lineární regrese proměnné LAR vyhlazené klouzavým průměrem tedy jsou schopné dobře zachytit rozdíl mezi mluvčími při závěrovém zpomalování a zároveň nezveličovat rozdíly v rámci jedné mluvčí. Je ale třeba velké opatrnosti při určování dat vhodných pro analýzu, jinak významně vzrůstá riziko chyby 2. druhu (nesprávného přijetí nulové hypotézy).

6.8.2 LAR pro úseky s melodémem neukončujícím

Melodémů neukončujících bylo na koncích nádechových úseků výrazně méně. Proto jsem se v tomto případě již nepokoušela analyzovat obě nahrávky téže mluvčí jednotlivě – například u MSA03 byly k dispozici pouze tři hodnoty.



	Průměr	SD
GVA	0,14	0,32
MSA	-0,42	0,48
SSA	-0,14	0,46

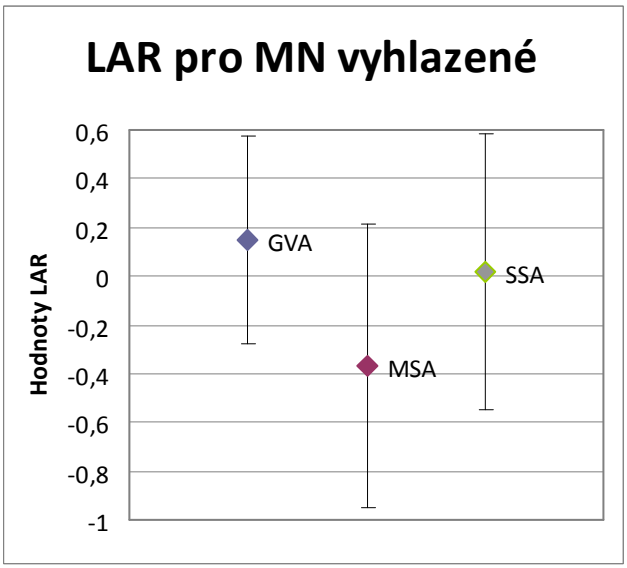
Graf 31: Průměrné hodnoty směrnic lineární regrese ukazatele LAR pro úseky s neukončující melodémem v závěru.

GVA/MSA	MSA/SSA	GVA/SSA
t(22) = 3,47; p<0,05	t(27) = 1,5; p>0,05	t(33) = 2,05; p<0,05

Tabulka 30: Výsledky t-testů pro hodnoty směrnic lineární regrese ukazatele LAR v úsecích s neukončujícím melodémem.

Z tabulky vidíme, že i v neukončujících melodémech se mluvčí od sebe liší – jmenovitě GVA od MSA a GVA od SSA. Vezmeme-li ale hodnoty vyhlazené tříbodovým klouzavým průměrem, rozdíl mezi GVA a SSA se výrazně zmenší, protože průměr mluvčí SSA se posunul blíže k nule.

Směrodatné odchylky pro vyhlazená data jsou také opět o něco větší než pro nevyhlazená.



	Průměr	SD
GVA	0,15	0,43
MSA	-0,37	0,58
SSA	0,02	0,56

Graf 32: Průměrné hodnoty směrnic lineární regrese ukazatele LAR vyhlazené tříbodovým klouzavým průměrem pro úseky s neukončujícím melodémem v závěru.

GVA/MSA	MSA/SSA	GVA/SSA
t(22) = 2,51; p<0,05	t(27) = 1,68; p>0,05	t(33) = 0,77; p>0,05

Tabulka 31: Výsledky t-testů pro hodnoty směrnic lineární regrese ukazatele LAR vyhlazené tříbodovým klouzavým průměrem v úsecích s neukončujícím melodémem.

Tyto výsledky je ale potřeba posuzovat velmi opatrně. Za prvé, objem dat se výrazně zmenšil, což zvyšuje pravděpodobnost statistické chyby, a za druhé, neukončující melodémy se od sebe mohou dosti lišit, například podle toho, jak významný je předěl na konci úseku nebo jaká věta na něj navazuje. Pro podrobnější zkoumání by rozhodně bylo třeba provést ještě lingvistickou analýzu a rozdělit neukončující melodémy podle dalších kritérií.

Samozřejmě se mohla také v těchto datech objevit stejná chyba jako u ukončujících melodémů – nesprávně určená poloha prozodického předělu. Připadalo mi však neužitečné více se těmito melodémy zabývat, protože

jejich množství je oproti ukončujícím zhruba třetinové a navíc zde do hry vstupuje mnohem víc faktorů.

7. Závěr

V této práci jsem prozkoumala deset temporálních ukazatelů a jejich možný přínos k rozpoznání mluvčího. Devět z těchto ukazatelů byly globální proměnné zachycující nějakým způsobem chování vokalických, případně konsonantických intervalů v rámci jednoho nádechového úseku. Poslední proměnná, LAR, zachycovala změny tempa lokálně a tu jsem použila ke zkoumání závěrového zpomalování jednotlivých mluvčích.

Proměnná %V (procento trvání vokalických intervalů v celém úseku) byla schopná od sebe odlišit dvě ze tří dvojic mluvčích. Je ovšem ze všech globálních proměnných nejvíce citlivá na text. Korelace s poměrem počtu konsonantů a vokálů byly středně vysoké až vysoké. Při analýze a případném využití této proměnné k rozpoznání mluvčího (nebo i za účelem popsání temporální struktury celého jazyka) je tedy třeba dát pozor, zda se texty mluvčích od sebe výrazně neliší ve složitosti slabičné struktury.

Variabilitu vokalických intervalů zachycovala proměná ΔV. Ta našla rozdíly mezi dvěma dvojicemi mluvčích a zároveň také u dvou mluvčích mezi jejich dvěma nahrávkami. Všeobecně lze ale říci, že rozdíly mezi mluvčími a mezi jednotlivými nahrávkami byly velmi podobné. Korelace s poměrem vokálů a konsonantů sice byla nízká, nicméně se ukázalo, že některé z nejvyšších hodnot ΔV v celém materiálu byly ovlivněny jedním cizím slovem s neobvyklým vokalickým shlukem. Aby byl odstraněn tento vliv textu na ΔV, musela by cizí slova být z materiálu zcela vyřazena. Vokalické shluky o více než dvou vokálech jsou v češtině velmi neobvyklé, a tedy vnáší do materiálu nežádoucí vliv.

ΔC rozdíly mezi mluvčími nezachytila vůbec, spíše než pro rozpoznání mluvčích je tedy vhodná pro kategorizaci jazyků jako celků.

U proměnné VarcoV, což je ΔV normalizované vzhledem k průměrnému trvání vokalických intervalů v daném úseku, byly výsledky ještě nejednoznačnější než u ΔV. Po odstranění úseků s nestandardně vysokým a nízkým poměrem konsonantů a vokálů a úseků velmi krátkých a dlouhých nebyly nalezeny vůbec žádné rozdíly – až na obě nahrávky od

jedné z mluvčích. Tam se ale mohl stejně jako u ΔV projevit vliv onoho nestandardního vokalického shluku.

Ukazatel VarcoC normalizuje směrodatnou odchylku trvání konsonantických intervalů. Signifikantní rozdíl nalezl mezi dvěma ze tří dvojic mluvčích, jeden z nich ale zmizel po odstranění velmi krátkých úseků. Mezi dvěma nahrávkami jedné mluvčí nebyl významný rozdíl ve VarcoC nikde.

První z indexů párové variability, rPVI-V, našel rozdíly také mezi dvěma dvojicemi mluvčích. Po podrobnějším prozkoumání bylo ale zjištěno, že velmi krátké úseky vnášejí do dat poměrně značnou variabilitu a mohly by existující rozdíly zamlžovat. Po jejich odstranění se o něco zvýraznil rozdíl mezi poslední dvojicí mluvčích, ovšem ne dost, aby překročil hranici 95% spolehlivosti. Druhé dva rozdíly signifikantní zůstaly. Rozdíly mezi jednotlivými nahrávkami jedné mluvčí dost významné nebyly.

Normalizovaná verze tohoto indexu, nPVI-V, našla v podstatě tytéž rozdíly, které ale přestaly být statisticky významné po vyřazení extrémů poměru C/V a okrajově dlouhých a krátkých hodnot.

Co se týče konsonantických intervalů, výsledky pro rPVI-C byly také rozpačité. Rozdíl se objevil jen mezi jednou dvojicí mluvčích a přestal být statisticky významný po stejných úpravách jako výše. Kromě toho našla tato proměnná stabilní rozdíl v rámci jedné z mluvčích. U té k rozdílu zřejmě přispívá vliv textu, protože každá nahrávka zvlášť koreluje poměrně výrazně s poměrem C/V.

Při normalizaci se zvýraznil rozdíl mezi jednou z mluvčích a druhými dvěma, a tedy byl významný rozdíl již mezi dvěma dvojicemi mluvčích. Normalizovaný konsonantický index párové variability, nPVI-C, byl stabilnější, stejné rozdíly vycházely i při manipulaci s daty – odstraňování extrémů poměru C/V a trvání. V rámci jednotlivých mluvčích se signifikantní rozdíl neobjevil.

Následující tabulka shrnuje výsledky všech globálních ukazatelů po odstranění úseků s okrajovým poměrem C/V a okrajově krátkých a dlouhých.

Proměnná	GVA/MSA	MSA/SSA	GVA/SSA	GVA02/ GVA07	MSA01/ MSA03	SSA02/ SSA03
%V	NE	ANO	ANO	NE	NE	NE
DeltaV	ANO	ANO	NE	NE	ANO	ANO
VarcoV	NE	NE	NE	NE	ANO	NE
rPVI-V	ANO	ANO	NE	NE	NE	NE
nPVI-V	NE	NE	NE	NE	NE	NE
DeltaC	NE	NE	NE	NE	NE	NE
VarcoC	ANO	NE	NE	NE	NE	NE
rPVI-C	NE	NE	NE	NE	NE	ANO
nPVI-C	ANO	NE	ANO	NE	NE	NE

Tabulka 32: Statisticky signifikantní rozdíly, které zachytily jednotlivé globální proměnné na hladině spolehlivosti α = 0,05 po vyřazení úseků s extrémním poměrem C/V a trváním.

Ani jeden globální ukazatel tedy neukázal rozdíl mezi všemi třemi mluvčími (nebo mezi jednotlivými nahrávkami všech mluvčích) a pro úspěšné rozpoznání bude nutné tyto ukazatele kombinovat. Ovšem zejména ΔC, VarcoV, VarcoC, rPVI-C ani nPVI-V se neukázaly být nijak zvlášť vhodné pro zachycení rozdílů mezi mluvčími; u %V, ΔC a rPVI-C je třeba dát si pozor na vysokou korelaci se slabičnou strukturou textu. Statisticky signifikantní rozdíly v poměru konsonantů a vokálů u jednotlivých mluvčích povedou téměř jistě k rozdílům také v těchto proměnných.

Je ostatně vidět, že tyto ukazatele byly primárně navrženy pro zobecňování na celý jazyk a zachycení indiviuálních rysů mluvčích je tak u nich jev nežádoucí. Nejvíce se od sebe mluvčí odlišily u proměnné rPVI-V, která je tak pro rozpoznání ze všech nejslibnější.

Vliv fonologické distinkce vokalické délky se u této ani u žádné jiné vokalické proměnné silněji neprojevil. Lze tedy konstatovat, že v případě češtiny se při používání těchto ukazatelů není nutné na problémy s délkou

vokálů vůbec soustředit. Nemůže být ale na škodu analyzovaný materiál nejprve prověřit, zda nejeví výrazné rozdíly v zastoupení dlouhých vokálů.

Co se týče trvání úseku, tak z výsledků vyplývá, že zejména velmi krátké úseky (pod dvě sekundy) vnášejí do testů větší rozdíly a zvyšují tak pravděpodobnost chyby 1. druhu (tedy chybného zamítnutí nulové hypotézy). Je proto žádoucí takovéto úseky z analýzy odstranit.

Rozdíl v čase mezi oběma nahrávkami jedné mluvčí se na těchto ukazatelích nijak výrazně neprojevil. Dokonce mluvčí GVA, jejíž nahrávky byly od sebe vzdáleny více než pět let, byla neobyčejně konzistentní – signifikantní rozdíl se neobjevil v žádném ukazateli. U MSA a SSA se sice sporadicky objevil, ale není jisté, zda na něj neměly vliv ještě další nekontrolované faktory.

Jako poslední byla zkoumána závěrečná zpomalování s pomocí proměnné LAR a gradientu jejích koncových hodnot.

U melodémů neukončujících bylo materiálu příliš málo a jeho variabilita málo kontrolovatelná k provedení hlubší analýzy. Přesto byly nějaké rozdíly nalezeny.

Mnohem spolehlivější se ukázaly gradienty LAR při zachycení individuálního zpomalování na konci úseků s melodémem ukončujícím klesavým. Tam se objevily rozdíly mezi všemi třemi dvojicemi mluvčích a zároveň žádné v rámci jedné mluvčí, což byl požadovaný výsledek. Hodnoty LAR bylo ovšem potřeba nejprve vyhladit tříbodovým klouzavým harmonickým průměrem, jinak byl do materiálu vnášen přílišný vliv slabičné struktury analyzovaných slov.

8. Diskuse

Za první důležitý výsledek této práce pokládám zjištění, že globální intervalové ukazatele Ramuse et al. (1999) a párové ukazatele Lowové et al. (2000) a další jejich varianty nejsou příliš spolehlivými indikátory identity mluvčího. Přesto z nich ale lze nějaké závěry vyvodit – například pokud se neukáže významný rozdíl ani v jednom měřeném ukazateli, s největší pravděpodobností jde o jednoho a toho samého mluvčího. Naopak to ale nemusí platit – rozdílnost mluvčích automaticky neimplikuje rozdíl v temporálních ukazatelích. Nejnadějnější se ukázal být v tomto směru ukazatel rPVI-V, ovšem až po odstranění příliš krátkých úseků, které vnášely do výsledků nežádoucí variabilitu.

Naopak ukazatel LAR převzatý od Volína (2009), respektive gradient jeho závěrečných hodnot vyhlazených tříbodovým klouzavým průměrem byl s to rozpoznat od sebe všechny tři mluvčí a dosáhl tedy požadovaného efektu. S jeho analýzou však byly spojeny určité těžkosti.

Náročnost získávání všech měřených ukazatelů spočívala zejména ve správné anotaci materiálu, která je časově velmi náročná. (Zbytek už byl jen otázkou konstrukce skriptů pro extrakci dat a několika jednoduchých statistických testů.) Přestože pro češtinu už je k dispozici automatický značkovací program Prague Labeller (Pollák et al., 2008), není jeho anotace natolik přesná, aby nemusela být kontrolována a opravována člověkem. A tento lidský prvek může do materiálu vnést nepřesnosti – ovšem podle výsledků Wigeta et al. (2010) je efekt anotátora minimální, pokud postupují podle stejného protokolu. Ten už naštěstí pro češtinu existuje (Machač & Skarnitzl, 2009). Plně automatická anotace materiálu by však ušetřila velké množství času a práce.

Dále by bylo ideální mít materiál segmentován na úrovni hlásek, slov, taktů i prozodických úseků. Dalo by se pak pracovat s délkou úseků podle počtu taktů nebo prozodických úseků a ne jen podle sekund, jako tomu bylo zde. Rovněž pro zkoumání závěrového zpomalování proměnnou LAR by bylo mnohem jednodušší přímo zjistit délku posledního prozodického úseku, než slepě extrahovat posledních šest hodnot.

Pokud je mi ale známo, tak na segmentování taktů nebo prozodických úseků zatím žádná automatická procedura neexistuje – jejich hranice v textu jsou často nejasné až sporné. Pro správnou anotaci je potřeba velmi pečlivý poslech zkušeného experimentátora, jinak může docházet k chybám, které zamlžují výsledky, jako se to stalo zde u prvních testů ukazatele LAR. Tam bylo potřeba pozorně vyřadit úseky, kde se ve zkoumané poslední části vyskytl prozodický úsekový předěl, jinak se zvyšovala pravděpodobnost chyby 2. druhu – nesprávného přijetí nulové hypotézy.

Tento bod postupu je velmi náchylný k chybám a podle mého názoru do něj osoba experimentátora vnáší velkou míru subjektivity – takže považuji za rozumné zařadit do analýzy jen ty úseky, u kterých není opravdu žádných pochyb, že poslední prozodický úsek má požadovanou délku. Což ale na druhou stranu může výrazně zmenšit vzorek.

Také obsah textu všech mluvčích by neměl být ponechán bez povšimnutí. Zejména cizí nebo neobvyklá slova mohou do analýzy vnést nežádoucí vlivy. Při malém množství materiálu ale může mít vliv i slabičná struktura textu, na které všechny měřené ukazatele (byť některé nepřímo) závisejí. Užitečným a rychlým způsobem, kterým jde prozkoumat, zda se texty od sebe v tomto ohledu významně neliší, je spočítání prostého poměru počtu konsonantů a vokálů v jednotlivých úsecích. Čím vyšší poměr, tím více konsonantů a složitější slabičná struktura. Porovnáním těchto hodnot pro jednotlivé mluvčí pak lze zjistit významnost rozdílu v zastoupení konsonantů vůči vokálům v daných textech.

Zabývala jsem se také možným vlivem fonologické distinkce délky vokálů na hodnoty vokalických ukazatelů. Z výsledků vyplývá, že rozdíl mezi trváním realizací fonologicky dlouhých a krátkých vokálů nebyl u mnou zkoumaných mluvčích dostatečný – a ve shodě s údaji ve článku Podlipského et al. (2009) se odvážím tvrdit, že tento závěr lze zobecnit. Není tedy třeba se při měření vokalických ukazatelů na distinkci délky nijak zvlášť ohlížet.

Ve své analýze jsem z materiálu zcela vyřadila pauzy a přeřeknutí. Jejich četnost, podoba a umístění by mohla také vypovídat o identitě mluvčího – materiál by ale v tom případě musel být sbírán přímo s ohledem na tento výzkumný záměr.

V případě námitky, že při takto malém objemu dat by mohly všechny rozdíly být způsobeny nějakými jinými, nekontrolovanými faktory, bych ráda poukázala na fakt, že všechny měřené proměnné měly podobné rozdělení mluvčích – rozdíl mezi GVA a MSA byl obvykle největší, přičemž SSA ležela někde mezi nimi, blíž ke GVA. Je ale pravda, že žádoucí k potvrzení efektivity této metody by byla v první řadě analýza většího počtu nahrávek i mluvčích a také zahrnutí mluvčích obou pohlaví do materiálu.

Připomenout musím také fakt, že všechny zkoumané nahrávky byly záznamem čteného a nikoliv spontánního projevu. Před využitím těchto metod v praxi by bylo třeba v dalším výzkumu ověřit jejich spolehlivost také na spontánní řeč.

Literatura

Abercrombie, D. (1967). *Elements of general phonetics.* Edinburgh: Edinburgh University Press.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm, *Phonetica 66*, 46–63.

Asu, E. L. & Nolan, F. (2006). Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. In *Proceedings of Speech Prosody 2006*, Dresden, Germany.

Boersma, P. & Weenink, D. (2010). Praat: doing phonetics by computer [Počítačový program]. Verze 5.2.01, získáno 9. listopadu 2010 z http://www.praat.org/

Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for C. In P. Karnowski & I. Szigeti (eds.) *Language and Language-processing.* Frankfurt am Main: Peter Lang, 231-241.

Dellwo, V. & Koreman, J. (2008). How speaker idiosyncratic is measurable speech rhythm? [Abstrakt]. *2008 International Association for Forensic Phonetics and Acoustics.* Annual Conference, Lausanne (Switzerland).

Doddington, G. (1985). Speaker Recognition – Identifying People by their Voices. *Proc. IEEE*, 73: 1651-1664.

Doherty, E. T. & Hollien, H. (1978). Multiple-factor speaker identification of normal and distorted speech. *Journal of Phonetics*, *6*, 1-8.

Doty, N. (1998). The Influence of Nationality on the Accuracy of Face and Voice Recognition. *Am. J. Psychol.*, *111*, 191-214.

Furui, S. (1981). Comparison of speaker recognition methods using statistical features and dynamic features. *IEEE Trans. Acoust. Speech, Signal Processing.* ASSP-29, str. 342-350.

Gibbon, D. & Gut, U. (2001). Measuring speech rhythm. *Proceedings of Eurospeech 2001,* Aalborg,Denmark, I: 91-94.

Glenn, J. W. & Kleiner, N. (1968). Speaker identification based on nasal phonation. *JASA*, *43*, 368-372.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In N. Warner, & C. Gussenhoven (eds.),

Papers in laboratory phonology 7, 515–546. Berlin: Mouton de Gruyter.

Goggin, J., Thompson, C., Strube, G. & Simental, L. (1991). The Role of Language Familiarity in Voice Identification. *Memory Cogn.*, *19*, 448-458.

Goldenberg, R., Cohen, A. & Shallom. I. (2006). The Lombard Effect's Influence on Automatic Speaker Verification Systems and Methods for its Compensation. *ITRE 2006:* 233-237.

Hattori, H. (1992). Text Independent Speaker Recognition Using Neural Networks, *Proc. ICASSP-92*, 2: 153-156.

Hollien, H. (1990). *The Acoustics of Crime*. New York: Plenum Press.

Hollien, H. (2002). *Forensic Voice Identification*. London: Academic Press.

Hollien, H. & Schwartz, R. (2000). Aural-Perceptual Speaker Identification: Problems with Noncontemporary Samples. *Forensic Linguistics*, *7*, 199-211.

Hollien, H. & Schwartz, R. (2001). Speaker Identification Utilizing Noncontemporary Speech. *J. Forensic Sci.*, *46*, 63-67.

Höfker, U. (1977). Phoneme-ordering for speaker recognition. *Contributed Papers to the 9th International Congress on Acoustics, Madrid 1977*. Madrid: Spanish Acoustical Society.

Kersta, L. G. (1962). Voiceprint Identification. *Nature*, *196*, 1253-1257.

Klevans, R. & Rodman, R. (1997). *Voice Recognition*. Boston: Artech House Inc.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Lawson, A. D., Stauffer, A. R., Smolenski, B. Y., Pokines, B. B., Leonard, M. & Cupples, E. J. (2009). Long term examination of intra-session and inter-session speaker variability. In: *Interspeech 2009*, 2899-2902.

Loukina, A., Kochanski, G., Rosner, B., Keane, E. & Shih, C. (2010). Rhythm measures and dimensions of durational variation in speech. Nepublikováno. Preprint online na http://kochanski.org/gpk/papers/2010/classifier.pdf.

Low, E. L., Grabe, E. & Nolan, F. (2000) Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English. *Language & Speech* 43 (4), 377-401.

Lummis, R. C. & Rosenburg, A. E. (1972). Test of an automatic speaker verification method with intensively trained professional mimics. *J. Acoust. Soc. Amer.* 51: 131-132.

Machač P. & Skarnitzl R. (2009). *Fonetická segmentace hlásek.* Praha: Nakladatelství Epocha.

Majewski, W. & Basztura, C. (1996). Integrated Approach to Speaker Recognition in Forensic Application. Forensic Linguistics, 3: 50-64.

Markel, J. D., Oshika, B. T. & Gray, A. H. (1977). Long term feature averaging for speaker recognition. *IEEE Trans. Ac., Sp. & Sig. Proc.* ASSP-25, 330-337.

Matsumoto, H., Hiki, S., Sone, T. & Nimura, T. (1973). Multidimensional representation of personal quality of vowels and its acoustical correlates. *IEEE Trans. Aud. & Electroac.* AU-21, 428-436.

McGehee, F. (1937). The Reliability of the Identification of the Human Voice. *J. Gen. Psychology*, *17*, 249-271.

McGehee, F. (1944). An Experimental Study of Voice Recognition. *J. Gen. Psychology*, *31*, 53-65.

Nazzi, T. & Ramus, F. (2003): Perception and acquisition of linguistic rhythm by Infants. *Speech Commun.* 41:233–243.

Nolan, F. (1983, reedice 2009). *The phonetic bases of speaker recognition.* Cambridge: Cambridge University Press.

Orchard, T. & Yarmey, A. (1995). The Effects of Whispers, Voice-Sample Duration and Voice Distinctiveness on Criminal Speaker Identification. *Appl. Cogn. Psychol.*, *9*, 249-260.

Pfitzinger, H. R. (1998). Local speech rate as a combination of syllable and phone rate. In: *Proc. of ICSLP '98*, vol. 3, 1087-1090, Sydney.

Podlipský, V. J., Skarnitzl, R. & Volín, J. (2009). High Front Vowels in Czech: a Contrast in Quantity or Quality? In: *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, 132-135. Brighton: ISCA.

Pollák, P., Volín, J. & Skarnitzl, R. (2008). Phone Segmentation Tool with Integrated Pronunciation Lexicon and Czech Phonetically Labelled Reference Database. In: *Proceedings of 6th International Conference on Language Resources and Evaluation*, vol. 1, 1-5. Paris: ELRA.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In: *Proceedings of Speech Prosody 2002*, 115–120. Aix-en-Provence.

Rosenberg, A. E. (1973). Listener performance in speaker verification tasks. *IEEE Trans. Audio Electroacoust.*, AU-21, 221-225.

Saito, K., Asakawa, K., Shimura, Y. & Imaizumi, S. (1995). Development of Speaker Identification in Young Children. *Ann. Bull, RIPL, Tokyo*, *29*, 55-58.

Schiller, N. O. & Köster, O. (1996). Evaluation of a Foreign Language Speaker in Forensic Phonetics: A Report, *Forensic Linguistics*, *3*, 176-185.

Skarnitzl, R. (2010). Prague Phonetic Corpus: status report. In: R. Skarnitzl (ed.), *AUC Philologica - Phonetica Pragensia XII*, 65-67. Praha: Karolinum.

Soong, F., Rosenberg, A., Rabiner, L. & Juang, B. (1985) A Vector Quantization Approach to Speaker Recognition. *Proc. ICASSP-85*, 387-390.

Su, L-S., Li, K-P. & Fu, K. S. (1974). Identification of speakers by use of nasal coarticulation. *JASA*, *56*, 1876-1882.

Thompson, C. (1987). A Language Effect in Voice Identification. *Appl. Cogn. Psychol.*, *25*, 121-131.

Tseng, B., Soong, F. & Rosenberg, A. (1992) Continuous Probabilistic Acoustic Map for Speaker Identification. *Proc. ICASSP-92*, 1: 161-164.

Volín, J. (2009). Metric warping in Czech newsreading. In: R. Vích (ed.) *Speech Processing - 19th Czech-German Workshop*, 52-55.

Webb, J. & Rissanen, E. (1993). Speaker Identification Experiments Using HMM's, *Proc. ICASSP-93*, 2:387-390.

White, L. & Mattys, S. L. (2007a). Calibrating rhythm: First language and second language studies. Journal of Phonetics 35, 501-522.

White, L. & Mattys, S. L. (2007b). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M.-J. Solé (eds.), *Segmental and prosodic issues in Romance phonology*, 237–257. Amsterdam: John Benjamins.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *J. Acoust. Soc. Am.* 127, 1559–1569.

Wolf, J. J. (1972). Efficient acoustic parameters for speaker recognition. *JASA*, *51*, 2044-2056.

Yoon, T. J. (2010). Capturing Inter-speaker Invariance Using Statistical Measures of Rhythm.
Speech Prosody 2010, 100201:1-4.

Přílohy

Prvních devět příloh tvoří skripty napsané v programu Praat sloužící k extrakci dat z materiálu. Podrobnosti o jejich fungování včetně nutných předpokladů jsou uvedeny v komentářích uvnitř skriptů.

1a. Skript ZeroCrossings

Skript, který posune všechny hranice intervalů v souborech typu TextGrid do nejbližších průchodů nulou v odpovídající nahrávce.
K nalezení je na přiloženém CD ve složce *Skripty* pod jménem *script_ZeroCrossings*.

1b. Skript CVcluster

Vytvoří v souborech formátu TextGrid novou intervalovou vrstvu, na které označí vokály písmenem V a konsonanty písmenem C, a následně je seskupí do vokalických nebo konsonantických intervalů. K dispozici je na přiloženém CD ve složce *Skripty* pod jménem *script_CVcluster*.

1c. Skript PointTier

Tento skript vytvoří v souboru formátu TextGrid novou vrstvu, na které vyznačí středy vokalických intervalů. Nachází se na přiloženém CD ve složce *Skripty* v souboru *script_PointTier*.

1d. Skript %V

Tento skript extrahuje z TextGridů hodnoty ukazatele %V a uloží je do tabulky ve formátu .xls. Na přiloženém CD je uložen ve složce *Skripty* pod jménem *script_%V*.

1e. Skript DeltaV-C

Skript, který ze souborů ve formátu TextGrid extrahuje ukazatele ΔV a ΔC a hodnoty převede do tabulky ve formátu .xls. K dispozici je na přiloženém CD ve složce *Skripty* v souboru *script_DeltaV-C*.

1f. Skript VarcoV-C

Tento skript extrahuje ukazatele VarcoV a VarcoC z TextGridů do tabulky ve formátu .xls. K dispozici je na přiloženém CD ve složce *Skripty* jako soubor *script_ VarcoV-C*.

1g. Skript PVI

Extrahuje ze souborů ve formátu TextGrid ukazatele rPVI-V, rPVI-C, nPVI-V a nPVI-C a uloží je do tabulky ve formátu .xls. Na přiloženém CD jej lze najít ve složce *Skripty* pod jménem *script_ PVI*.

1h. Skript LAR

Vypočítá hodnoty proměnné LAR a jejich tříbodový klouzavý harmonický průměr. Nachází se na přiloženém CD ve složce *Skripty* pod jménem *script_ LAR*.

1i. Skript LAR-final

Tento skript vypočítá hodnoty proměnné LAR a jejich tříbodový klouzavý harmonický průměr pro posledních 7 vokalických intervalů. Na CD je uložen ve složce *Skripty* pod jménem *script_ LAR-final*.

1j. Skript CVratio

Skript, který v každém souboru typu TextGrid zjistí počet konsonantů a vokálu a spočítá jejich poměr. Na přiloženém CD je ve složce *Skripty* pod názvem *script_ CVratio*.

2. Nahrávky a soubory typu TextGrid

Nahrávky všech tří mluvčích a k nim náležející soubory formátu TextGrid jsou uloženy na přiloženém CD ve složce *Nahrávky+TextGridy*.

Třetí část příloh tvoří tabulky s extrahovanými daty a jejich statistickou analýzou, umístěné na CD ve složce *Tabulky*.

3a. ***Slova_trvani_pauzy.xls***

Zde naleznete údaje o počtu slov, trvání, počtu pauz a počtu přeřeknutí v každé z nahrácek.

3b. ***PomerCV.xls***

Zde lze najít údaje o poměru konsonantů a vokálů a jejich analýzu.

3c. ***PercentVocal.xls***

V této tabulce jsou uvedeny hodnoty ukazatele %V a jejich analýza.

3d. ***DeltaV-C.xls***

Tato tabulka obsahuje hodnoty ΔV a ΔC a jejich analýzu.

3e. ***VarcoV-C.xls***

Tabulka hodnot ukazatelů VarcoV a VarcoC a jejich analýza.

3f. ***PVI.xls***

Zde najdete hodnoty a analýzu všech ukazatelů PVI.

3g. ***LAR.xls***

Tabulka uvádí všechny hodnoty ukazatele LAR pro celý materiál.

3h. ***LAR-final.xls***

Tato tabulka obsahuje hodnoty LAR pro posledních 7 vokalických intervalů a jejich analýzu.

3i. ***Trvani_useku.xls***

V této tabulce jsou shrnuta data o všech ukazatelích vzhledem k délce trvání úseků a jejich analýza.