

Univerzita Palackého v Olomouci

Filozofická fakulta

Katedra bohemistiky

**VYBRANÉ KAPITOLY Z FORENZNÍ
FONETIKY**

(komentovaný překlad)

Selected Chapters on Forensic Phonetics

(an annotated translation)

Magisterská diplomová práce

Olomouc 2019

Autor práce: Bc. Jiří Mánek

(česká filologie – anglická filologie)

Vedoucí práce: doc. PhDr. Božena Bednaříková, Ph.D.

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a uvedl v ní všechny použité zdroje a literaturu.

V Olomouci dne 20. 8. 2019

.....
Jiří Mánek

Poděkování:

Tímto děkuji doc. PhDr. Boženě Bednaříkové, Ph.D., za vedení této práce, doc. Mgr. Radku Skarnitzlovi, Ph.D., za pomoc s výběrem textů k překladu a PhDr. Petru Pořízkovi, Ph.D., za podnětné připomínky k terminologii.

Obsah

Úvod	5
1 Vybrané problematické jevy dle jazykových rovin	7
1.1 Rovina foneticko-fonologická	7
1.2 Rovina morfologická	8
1.3 Rovina lexikální	9
1.4 Rovina syntaktická a textová	10
1.5 Chyby ve výchozím textu	11
2 Překladová část 1 (Činnost forenzního fonetika)	13
2.1 Úvod.....	13
2.2 Profilování mluvčího	17
2.3 Srovnávání mluvčích	24
2.4 Závěr	34
3 Překladová část 2 (Poslechově-akustické vs. automatické metody ve forenzněfonetické praxi)	37
3.1 Úvod.....	37
3.2 Stručný přehled vývoje rozpoznávání mluvčích ve forenzním prostředí ...	39
3.3 Metody používané v současnosti	40
3.4 Přehled klíčových parametrů ve fonoskopii	45
3.5 Maskování.....	53
3.6 Modulační teorie řeči	54
3.7 Stručná rekapitulace.....	57
3.8 Předkládání důkazů u soudu	59
3.9 Shrnutí a závěry	63
Glosář	69
Závěr	74
Anotace	75
Resumé	76
Seznam literatury	77

Úvod

Forenzní fonetika (neboli také *fonoskopie* a *audioexpertiza*) je mladým aplikovaným oborem značně interdisciplinární povahy, jenž čerpá poznatky mimo jiné z matematiky, statistiky, akustiky, anatomie a práva, aby je následně uplatnil kupříkladu při identifikaci pachatele trestné činnosti jako výhružných telefonátů či únosů. Při tom se potýká s četnými komplikacemi ve formě špatně analyzovatelných audiozáznamů, z nichž vychází, či kupříkladu maskovacích strategií, jimiž se dotyčná osoba snaží zastříit svou totožnost.

Náplní této práce je komentovaný překlad dvou forenzněfonetických textů, jež byly vybrány po konzultaci s doc. Mgr. Radkem Skarnitzlem, Ph.D., jedním z mála odborníků u nás, který se touto oblastí dlouhodobě zabývá. Prvním z textů je *The forensic phonetician* (zde překládáno jako *Činnost forenzního fonetika*) německého forenzního fonetika Michaela Jessena, který může sloužit jako vhodný úvod do problematiky, neboť popisuje stěžejní oblasti forenzní fonetiky včetně základních zkoumaných parametrů, druhů analýz a možných komplikací, s nimiž se odborník ve forenzním prostředí může setkat.

Druhý z textů, jehož autorem je švédský forenzní fonetik Anders Eriksson, nese název *Aural/Acoustic vs. Automatic Methods in Forensic Phonetic Case Work* (zde překládáno jako *Poslechově-akustické vs. automatické metody ve forenzněfonetické praxi*) a Jessenovu kapitolu do značné míry rozšiřuje o některé pokročilejší koncepty jako věrohodnostní poměr, modulační teorii řeči či Daubertův standard.

Diplomovou práci doplní poznámkový aparát, který se bude vztahovat zejména k problematickým, translatickým a/nebo forenzněfoneticky relevantním pasážím a konkrétním výrazům – zpravidla terminologické povahy – ve výchozích textech. Vybrané poznatky z něj budou posléze extrahovány a ve formě syntetizujících komentářů prezentovány v první kapitole této práce. Z konkrétních, a to zvláště relevantních termínů bude následně vytvořen anglicko-český glosář, který bude umístěn za cílovými texty.

Evidence znalců a tlumočnicků Ministerstva spravedlnosti České republiky v současnosti obsahuje v oboru forenzní fonetiky / fonoskopie pouze dva záznamy, konkrétně jde o PhDr. Marii Hes Svobodovou, Ph.D., a Ing. Zdeňka Švendu, Ph.D. Třetím odborníkem v této oblasti je zmíněný doc. Mgr. Radek Skarnitzl, Ph.D., který coby editor stál u zrodu v tuto chvíli jediné obsáhlejší oborové publikace s názvem *Fonetická identifikace mluvčího*. Tato diplomová práce má za cíl následovat jejího příkladu, zpřístupnit českému čtenáři

některé hlavní teze forenzní fonetiky, přiblížit mu stav dosavadního bádání za hranicemi ČR, a alespoň malou měrou tak přispět k rozšíření povědomí o této disciplíně.

1. Vybrané problematické jevy dle jazykových rovin

Překlad každého výrazu, slovního spojení či větné konstrukce s sebou nese specifické překladatelské řešení, jež vybízí k bližší analýze. My se však zaměříme pouze na jevy, které jsou zvláště relevantní pro zde překládané dvě kapitoly, eventuálně obecně z hlediska forenzní fonetiky.

Jednou ze systémových odlišností mezi angličtinou a češtinou je tendence k nominálnosti (angličtina) vs. tendence k verbálnosti (čeština)¹, která se v případě odborného textu do jisté míry stírá v důsledku toho, že zvýšeným užitím pasiva se čeština kloní spíše k verbonominálnosti a že významovou hutnost, s níž se v případě odborného textu musí potýkat, řeší zpravidla jmennými prostředky. Ty mohou tvořit i delší sekvence, tzv. nominální řetězce, kde za zmínku mimo jiné stojí různé postavení jejich členů: „Český překlad často obsahuje navíc alespoň předložku, která signalizuje typ vztahu mezi určeným a určujícím, kdežto v angličtině, především psané, jazykový signál být nemusí. Pro interpretaci je samozřejmě důležité vzájemné postavení juxtaponovaných substantiv (určující + určený, premodifier + head), např. *club discussion* je *debata v klubu*, kdežto *discussion club* je *diskuzní klub*.“ (Knittlová, 2000, s. 39) Ačkoli použití předložky není v každé situaci nutné či adekvátní, zvýšenou pozornost je v souvislosti s postavením jednotlivých členů třeba věnovat v případě delších řetězců jako *landline phone recorded speech samples*, zde přeloženo jako *vzorky řeči získané pomocí pevné linky*. Nominální charakter angličtiny je v případě odborného textu ještě posílen a verba, jejichž procentuální zastoupení tím pádem klesá, navíc podléhají časté repetici, jež v češtině může být pocíťována jako poměrně závažný stylistický nedostatek. Kupříkladu opakující se konstrukci *Figure x shows*² je proto žádoucí překládat s využitím co největší variability, kupříkladu jako: *obrázek x ilustruje / z obrázku x je patrné/zjevné / na obrázku x můžeme vidět/pozorovat / na obrázku x si všimněme* atd. Jsou však případy, kdy na repetici musíme z důvodu nedostatku vhodných synonym rezignovat (zvláště v případě termínů; o nich podrobněji v oddílu věnovanému rovině lexikální).

1.1 Rovina foneticko-fonologická

Začneme-li na úrovni hlásek, nelze opomenout odlišnosti ve fonémickém inventáři češtiny a angličtiny, který se však do jisté míry liší v závislosti na konkrétní varietě jazyka (například

¹ Podrobnější výklad nabízí např. Mathesius (1913, s. 325–339).

² Kde *x* je neznámá.

neznělá velární frikativa /x/ či vibranta /r/ jsou součástí fonologického systému skotské angličtiny, nikoli však současné RP). Jazykové variety zde uvádíme zejména z toho důvodu, že je zmiňují oba překládané texty (kupříkladu v souvislosti s britskou či australskou angličtinou). Jde o problematiku natolik obsáhlou, že není v našich silách adekvátně ji zde přiblížit, a zájemce tedy odkazujeme například na třísvazkové dílo *Accents of English* (Wells, 1982a, 1982b, 1982c).

Podobně obtížné by bylo postihnout veškeré rozdíly v segmentální a suprasegmentální oblasti mezi angličtinou a češtinou, což mimoto rovněž není cílem této práce. Některé dílčí jevy, jako je zápis [h], který v českém úzu odpovídá znělé glotální frikativě, zatímco v IPA³ koresponduje s glotální frikativou neznělou, uvádíme v poznámkovém aparátu. Poměrně důslednou komparaci obou jazyků v rámci foneticko-fonologické roviny provedla Skaličková (1982). Na některé problémy spjaté s relevantní terminologií (a jejím překladem) jsem poukázal ve své bakalářské práci (Mánek, 2016), mimoto vzhled do zvukového plánu (nejen) češtiny nabízí například Hála (1975), Romportl (1981), Petr et al. (1986) Palková (1997), Krčmová (2008) či Skarnitzl, Šturm a Volín (2016). Z hlediska forenzní fonetiky je pro tuto chvíli podstatné zejména to, že „pro mluvenou formu češtiny se jako nejhodnotnější znaky individuálního projevu mluvčího ukazují frázování, intonace, rytmus a vokalická kvalita a kvantita. Nejnižší individuální hodnotu má (často v důsledku snížené technické kvality analyzovaného materiálu) způsob artikulace konsonantů a konsonantických skupin, pokud se ovšem nejedná o specifickou vadu řeči.“ (Svobodová, 1997)

1.2 Rovina morfologická

V tomto oddílu můžeme alespoň v krátkosti poukázat na frekventovanou monomorfémičnost angličtiny (jako v případě *we will see*), kterou v případě češtiny lze obvykle vyjádřit prostředky polymorfematickými, přičemž díky rozvinuté flexi může i jediný morfém nést více informací (např. *uvidíme*, kde u koncovky plyne náležitost k 1. osobě plurálu a zároveň k indikativu). Jde však o dvousečnou zbraň, neboť ačkoli z ní plynou překladateli určité výhody (a to především s ohledem na větší variabilitu slovosledu), často na něj klade požadavek vyjádřit některé gramatické kategorie, které angličtina vyjádřit nemusí. Především jde o kategorii rodu, kdy například výraz *author* lze překládat buď jako *autor*, nebo jako *autorka*. Ani *propria* v angličtině (a jiných jazycích) zpravidla nereflektují

³ *The International Phonetic Alphabet (mezinárodní fonetická abeceda).*

Dostupné z: <https://www.internationalphoneticassociation.org/content/full-ipa-chart>.

rod, možná s výjimkou křestních jmen u antroponym, u nichž však existuje komplikace dvojího druhu: 1) některá jména (nejen) v angličtině jsou obourodá (*Alexis, Gale, Claude* atp.) a 2) překladatel je ne vždy má k dispozici, neboť v rejstříku sekundární literatury jsou často vyjádřeny pouhou iniciálou (jako *Nolan, F.* či *McDougall K.*). I v případě, že se překladatel rozhodne cizí antroponyma, zejména z ekonomických důvodů, nepřechylovat⁴, v důsledku potřeby mimo jiné vyjádřit shodu podmětu s přísudkem pak mnohdy nezbyvá jiná možnost než pohlaví dotyčné osoby dohledat.

Dalším jevem, s nímž se při překladu z angličtiny často setkáme, je rozšířená konverze. Ta se sice uplatňuje i v češtině, ale v podstatně menší míře (viz zejména Bednaříková, 2009).

1.3 Rovina lexikální

V souvislosti s rovinou lexikální nelze opomenout problematiku terminologie. V případě forenzní fonetiky má značně interdisciplinární charakter a vedle fonetiky a soudní/kriminalistické praxe zasahuje i do oblasti anatomie, akustiky, práva, matematiky, statistiky, informačních technologií ad. To překlad do značné míry znesnadňuje, neboť se od překladatele vyžaduje alespoň částečná obeznámenost se všemi uvedenými. Pokud jde o fonetickou a forenzněfonetickou terminologii, opíráme se zde především o práce Řeřichy (1998), Skarnitzla *ed.* (2014), Krčmové (2014) a Čermáka (2011). Zároveň se pro tuto práci ukázal jako přínosný slovník *Lingea Lexicon 5* a některé specializované slovníky jako IATE⁵. Přesto v této sféře narážíme na nedostatek vhodných překladových protějšků, který pramení zejména ze skutečnosti, že forenzní fonetika v českém prostředí prozatím není dostatečně etablovanou disciplínou (pro konkrétní případy viz poznámkový aparát u cílových textů).

Může být s podivem, že i u odborného textu tohoto typu se můžeme setkat s kulturně-historickým aspektem, jehož úzus je třeba zohlednit – *King Charles I.* (oddíl 3.2) je tak namísto překládat jako *král Karel I. Stuart* – vč. přízviska, aby nedošlo k záměně s rakouským císařem (za povšimnutí stojí i verzálka v označení funkce v angličtině, tj. *King*, která se v češtině nepoužívá). Při volbě překladového protějšku sériového vraha známého

⁴ Jako i v této práci.

⁵ Dostupné z: <https://iate.europa.eu/home>.

jako *Yorkshire Ripper* (úvod oddílu 2) bychom podobným způsobem měli vzít v potaz vžitě označení *Yorkshirský rozparovač* používané českými médii.⁶

Specifickou kategorií jsou výrazy vyplývající z fatické funkce jazyka, jíž „se udržuje nebo modifikuje kontakt mluvčího a posluchače, resp. jejich pozornost, vyladěnost apod.“ (Čermák, 2011, s. 274), o nichž se hovoří v oddílu 2 v souvislosti se srovnáváním mluvčích. Jedná se zpravidla o výplňkové výrazy typu *you know*, *like* či *I mean*, ale i hezitační zvuky *uh*, *um* (zde *ehm*, *hm*), které je však příhodnější řadit k fonetickým jevům.

1.4 Rovina syntaktická a textová

O angličtině se často hovoří v souvislosti s „pevným“ (SVO⁷) slovosledem. Měli bychom se však vyvarovat klamného závěru, že všechny věty v anglicky psaném textu musí tím pádem být vystavěny na tomtéž principu. Huddleston a Pullum (2002, s. 1365–1372) uvádějí několik možných konstrukcí (viz obrázek 1), přičemž pasivum bychom mohli dále rozdělit na krátkou, deagentní variantu (*a house was built*) a dlouhou, s rematizovaným agens (*a house was built by Martin*). Ačkoli jejich užití i v angličtině podléhá určitým pravidlům, často je lze využít také v situacích, kde by české opisné pasivum bylo vnímáno přinejmenším jako silně příznakové, kupříkladu „[...] it is understood by us in the broad sense proposed by Laver (1980)“ se jeví jako příhodnější překládat pomocí aktiva: „[...] zde chápeme v širším slova smyslu tak, jako ji popisuje Laver (1980)“.

Obrázek 1 Přehled anglických syntaktických konstrukcí. V pravém sloupci jejich výchozí protějšek.

i	PREPOSING	<i>This one she accepted.</i>	<i>She accepted <u>this one</u>.</i>
ii	POSTPOSING	<i>I made without delay <u>all the changes you wanted</u>.</i>	<i>I made <u>all the changes</u> <u>you wanted</u> without delay.</i>
iii	INVERSION	<i><u>On board</u> were two nurses.</i>	<i><u>Two nurses</u> were on board.</i>
iv	EXISTENTIAL	<i><u>There is a frog</u> in the pool.</i>	<i><u>A frog</u> is in the pool.</i>
v	EXTRAPOSITION	<i><u>It is clear</u> that he's guilty.</i>	<i><u>That he's guilty</u> is clear.</i>
vi	LEFT DISLOCATION	<i><u>That money I gave her,</u> <u>it must have disappeared.</u></i>	<i><u>That money I gave her</u> <u>must have disappeared.</u></i>
vii	RIGHT DISLOCATION	<i><u>They're still here,</u> <u>the</u> <u>people from next door.</u></i>	<i><u>The people from next door</u> <u>are still here.</u></i>
viii	CLEFT	<i><u>It was you</u> who broke it.</i>	<i><u>You</u> broke it.</i>
ix	PASSIVE	<i><u>The car was taken</u> by Kim</i>	<i><u>Kim</u> took <u>the car</u>.</i>

⁶ Například <https://ct24.ceskatelevize.cz/archiv/1453928-yorkshirsky-rozparovac-sutcliffe-dostal-za-13-vrazd-doivoti> (cit. 2. 8. 2019) či <https://max.iprima.cz/novinky/yorkshirsky-rozparovac-zvysuje-skore-vrazdy-v-yorkshiru-1980> (cit. 2. 8. 2019).

⁷ Subjekt – Verbum – Objekt.

Vedle právě „krátkého“ a „dlouhého“ pasivu je v anglicky psaných odborných textech – forenzněfonetické texty zde překládané nevyjímaje – frekventovaným syntaktickým jevem tzv. „existenciální“ a „existenciálně-lokativní větný typ“ (*there + be*), o němž Dušková (2009) píše: „*there* však není pouze formální prostředek zaplňující pravidelné místo podmětu, nýbrž tvoří konstitutivní složku větného typu existenciálního a existenciálně-lokativního, [...] např. *there is no exception* – není (neexistuje) žádná výjimka. V češtině toto *there* nemá formální protějšek; české věty existenciální mají pouze jeden podmět odpovídající anglickému vlastnímu podmětu v postavení postverbálním.“ Tuto konstrukci musíme v češtině řešit zpravidla opisem; oproti angličtině máme však v češtině navíc k dispozici reflexivní pasivum, jež absenci existenciální konstrukce alespoň do jisté míry kompenzuje.

Kromě reflexivního pasiva⁸ může překladatel využít do značné míry „volného“ českého slovosledu, který je však „volný jen v tom smyslu, že není vázán rozložením větných členů; zpravidla je ale vázán tím, že odpovídá AČ⁹, přesněji řečeno stupnici (nebo hierarchii) výpovědní dynamičnosti (VD); česká věta zpravidla začíná vlastním tématem (kontrasivním nebo jiným), za ním jsou další části tématu (časové a místní kulisy aj.), pak následují vlastní části rématu (ohniska) – nejdříve sloveso, pak další části a nakonec vlastní ohnisko (nositel intonačního centra, nejdynamičtější člen věty).“ (Sgall, 2011, s. 103)

1.5 Chyby ve výchozím textu (VT)

Zvláštní podoblastí překladatelské činnosti je řešení překlepů, nepřesností a chyb, kterých se autor mohl, spíše nevědomky, dopustit. V oddílu 3.7 se například pravděpodobně odkazuje na nesprávný zdroj a v oddílu 3.6.1 není jasné, co měl autor na mysli v pasáži: „inventory of meaningful speech sounds“. *Speech sound* neboli *hláska* (ve významu foném) je totiž dle definice „minimální systémová jednotka nesémantická (resp. „podsémantická“), avšak funkční, schopná rozlišit význam (tj. v kombinaci s jinými ve vyšších jednotkách).“ (Čermák, 2011, s. 116)

Závěrem tohoto oddílu můžeme taktéž zmínit jednu úsměvnou chybu v oddílu 3.4.1, kde se dočteme, že hodnota základní frekvence se mimo jiné odvíjí od faktorů, jako je „age, smoking habits,“ či „decease“ (= *smrt*). Patrně jde o chybu způsobenou hláskovou

⁸ Pro podrobnější výklad viz také Huddleston a Pollum (2005, s. 245–265).

⁹ Aktuální členění.

a grafickou podobností výrazů *disease* (= nemoc) a *decease*, která je však zdrojem poměrně markantního posunu ve významu.

Jak již bylo avizováno, další problematické jevy, jež z překladu výchozích textů¹⁰ vyplývají, budou diskutovány v poznámkovém aparátu překladové části této práce, která bezprostředně následuje.

¹⁰ Dále jen VT.

2. Překladová část 1

Činnost forenzního fonetika

Expertní identifikace mluvčího ve forenzním prostředí

Michael Jessen

2.1 Úvod

Forenzní identifikace mluvčího je stěžejní úlohou vědní disciplíny označované jako *forenzní fonetika a akustika* či *forenzní analýza zvuku a řeči*.¹¹ První označení souvisí s názvem organizace *International Association for Forensic Phonetics and Acoustics* (IAFPA, viz její internetové stránky: www.iafpa.net), která každoročně pořádá konferenci s mezinárodní účastí a má mimo jiné zastoupení v časopise *International Journal of Speech, Language and the Law*. Přestože IAFPA dala jasně najevo, že fonetiky před jinými profesemi v organizaci, jako odborníky na řečové technologie¹², neupřednostňuje, fonetikové zde mají spolu s lingvisty zpravidla nejsilnější zastoupení. Druhý pojem, *forenzní analýza zvuku a řeči*, je ve vztahu ke všem vědním disciplínám, z nichž vychází (fonetika), zcela neutrální a užívá se v souvislosti se sdružením odborníků v rámci ENFSI (*European Network of Forensic Science Institutes*, viz www.enfsi.eu), kde jsou techničtí a IT pracovníci zastoupeni přinejmenším stejně hojně jako fonetikové a lingvisté; částečně proto, že forenzní analýza řeči a zvuku zahrnuje kromě identifikace mluvčího i mnoho dalších podoblastí, které mohou významně čerpat z např. z poznatků zvukových techniků. Tyto zahrnují jednak „čištění“ zvukové nahrávky¹³, tj. úpravu zvuku s cílem zvýšit srozumitelnost nekvalitního záznamu řeči (pomocí pokročilých metod filtrace a dalších metod zpracování signálu), a jeho autentizaci, tj. detekování známek manipulace s danou nahrávkou. Jelikož se uvedenými podoblastmi zabývá i IAFPA, není důvod pojmy *forenzní fonetika a akustika* a *forenzní analýza zvuku a řeči* rozlišovat, a budeme je tedy považovat za synonymní. Třetím výrazem, rovněž souznačným, je *fonoskopie* (v angličtině *Forensic Speech Science* – stejně jako byl

¹¹ Ve výchozím textu (dále jen VT) *Forensic Phonetics and Acoustics a Forensic Speech and Audio Analysis*; v českém prostředí se však nejčastěji setkáme s pojmy *fonoskopie*, *forenzní fonetika* či *audioexpertiza*.

¹² Ve VT *speech engineers*, jejichž náplní práce je zejména řečová syntéza (Pořízka, 2019, osobní komunikace).

¹³ Ve VT *audio enhancement*; doslova jde o zvyšování kvality audiozáznamu.

v roce 2007 pojmenován první studijní program v této oblasti na britské Univerzitě v Yorku, viz www.york.ac.uk/study/postgraduate-taught/courses/msc-forensic-speech-science/¹⁴).

Forenzní identifikaci mluvího lze rozdělit na několik dílčích podoblastí. Klasifikaci, která se v praxi osvědčila, uvádíme na obrázku 2.1.

Obrázek 2.1 Různé úlohy forenzní identifikace mluvího

	<i>Zvuková nahrávka neznámého mluvího¹⁵ je k dispozici</i>	<i>Zvuková nahrávka není k dispozici, ale existuje svědek</i>
Podezřelý ¹⁶ je k dispozici	Srovnávání mluvích [pokud podezřelý spolupracuje nebo jsou k dispozici jeho dřívější nahrávky a lze je použít]	Svědék zná podezřelého: standardní svědecká výpověď
Podezřelý není k dispozici	Profilování mluvího ¹⁷ a/nebo zveřejnění nahrávky neznámého mluvího v médiích	Svědék nezná podezřelého: hlasová konfrontace Účast odborníka pouze sporadická; možný vývoj: hlasový portrét ¹⁸ za použití řečové syntézy

Pokud existuje audiozáznam jak neznámého mluvího (tj. pachatele v případech jako únos, stalking¹⁹ nebo obchodování s drogami), tak i podezřelé osoby, je možné provést tzv. srovnání²⁰ mluvích a použít jej u soudu coby důkazní materiál. Setkáme se i se synonymním označením – *srovnání hlasů*. Pokud je podezřelý ochoten spolupracovat, lze, mj. pod taktovkou forenzního experta, pořídit zvukový záznam jeho řeči a s ním dále pracovat; kupříkladu je možné mluvený projev neznámého mluvího transkribovat do textové podoby a následně požádat podezřelého, aby jej přečetl nebo ve vhodných úsecích zopakoval. Tím je zaručena identita obou textů, výhodná pro některé navazující úkony jako

¹⁴ Původní hypertextový odkaz byl zastaralý, a byl proto nahrazen aktuálním. Přesto nelze vyloučit, že implementací budoucích změn na webových stránkách univerzity nedojde k podobné situaci.

¹⁵ Synonymem je *sporný mluví*; v případě jeho audiozáznamu mluvíme o *sporné nahrávce* (Skarniztl, 2014, s. 16).

¹⁶ Synonymem je *srovnávací mluví*; v případě jeho audiozáznamu mluvíme o *srovnávací nahrávce* (tamtéž).

¹⁷ Neboli *typování pachatele* (Svobodová, Voříšek, 2014, s. 140).

¹⁸ Ve VT *acoustic phantom picture*; ustálený český ekvivalent se nepodařilo dohledat, lze proto navrhnout označení *hlasový portrét (podezřelé osoby)*, případně *akustické portrétování*.

¹⁹ Neboli *nebezpečné pronásledování*.

²⁰ Pro jeden konkrétní případ je zde používán dokonavý vid, pro označení podoblasti identifikace mluvího vid nedokonavý.

měření formantů samohlásek. (Formanty jsou rezonanční frekvence, které jsou dány tvarem vokálního traktu²¹ a které měříme v hertzech (Hz); nejnižší rezonanční frekvence se nazývá první formant – obvykle vyjádřen zkratkou F1 – a nejvyšší rezonanční frekvenci ještě často využívanou ve forenzní praxi je třetí formant, F3.) Identita textů však s ohledem na forenzní identifikaci mluvčího není podmínkou, a čtení nebo (tím spíše) opakování nadto může mít za následek nepřírozenou prozódii, což s sebou nese určité komplikace. Zvukový záznam podezřelého by tedy měl obsahovat i vzorek řeči, který byl pronesen co možná nejspontánněji. Pokud však podezřelý nespolupracuje a nesouhlasí s tím, aby byl jeho hlas nahráván, je na právním systému dané země a na okolnostech konkrétního případu, zda je možné využít dřívější audiozáznamy dotyčné osoby, například z policejních výslechů nebo z telefonních odposlechů. Další forma nespolupráce ze strany podezřelého nastává, když souhlasí s nahráváním, ale následně se snaží – ať už zcela nepokrytě, či jen sotva znatelně – svůj hlas maskovat. V tomto případě se musí odborník z pohledu forenzní fonetiky rozhodnout, zda je ještě takový důkaz použitelný. Metodologie srovnávání mluvčích zahrnuje širokou škálu zvukových i akustických parametrů a budeme se jí zabývat ve třetím oddílu.

Máme-li k dispozici zvukový záznam neznámého mluvčího, ale zatím se nepodařilo nalézt žádné svědky, je stále možné na základě této nahrávky sestavit profil mluvčího (synonymními pojmy jsou analýza hlasu a hlasové profilování). Ten obvykle vyžaduje policie v probíhajícím vyšetřování, při hledání podezřelého. Užitečné jsou zde například informace týkající se věku, pohlaví, oblasti původu, společenského postavení a vlivu cizího jazyka. Profilováním mluvčích se budeme zevrubněji zabývat ve druhém oddílu. V situaci, kdy je vyžadován profil mluvčího, je rovněž možné zveřejnit část nahrávky neznámé osoby prostřednictvím masmédií jako TV, rádia či internetu. Obvykle se tak ovšem postupuje jen u zvláště exponovaných případů; částečně kvůli značnému zatížení odborných pracovníků, kteří musí analyzovat reakce veřejnosti s následným srovnáním mluvčích.

Některé forenzní případy začínají profilováním mluvčích a končí jejich srovnáváním. Nejlepším příkladem je pravděpodobně případ Yorkshirského rozparovače²², kde obě fáze dělí 30 let. V raném stádiu případu byl sestaven profil volajícího, který se za Yorkshirského

²¹ Ve VT *vocal tract*; Skarnitzl (2014, s. 19) užívá i označení *řečové ústrojí*. V české literatuře se můžeme setkat s protějškem *hlasové/fonační ústrojí* (např. Řeřicha, 1998, s. 52), který ovšem dle Krčmové (2009, s. 41) neznamená totéž; nezahrnuje totiž nadhrtanové prostory, které jsou součástí ústrojí *modifikujícího/artikulačního*. Vokální trakt je v tomto ohledu termín nadřazený, který obě ústrojí slučuje.

²² Označení používané českými médii, ve VT *Yorkshire Ripper*.

rozparovače – který v letech 1975–1980 zavraždil v Leedsu, Bradfordu, Huddersfieldu a Manchesteru 13 žen – vydával. Později vyšlo najevo, že šlo o podvodné telefonáty, a pachatele se v roce 2005 podařilo dopadnout pomocí analýzy DNA. Srovnání hlasu podezřelého a hlasu volajícího ze sedmdesátých let odhalilo výrazné podobnosti, a tedy že se pravděpodobně jedná o tutéž osobu. Pro detailní popis případu Yorkshirského rozparovače viz zejména Ellis (1994) a French et al. (2006).

Může se stát, že ačkoli neexistuje nahrávka neznámé osoby, svědek slyšel její hlas. V některých případech, jako je loupež nebo znásilnění, je mimoto svědek zároveň obětí. Tehdy je podstatné jak z vědeckého, tak právního hlediska, zda svědek znal pachatele již před spácháním daného trestného činu, či nikoli. Pokud ano, vyžaduje se od něj tzv. *identifikace známého mluvčího*, v opačném případě pak *identifikace neznámého mluvčího*. Identifikace známého mluvčího vstupuje do důkazního řízení ve formě standardního svědectví. Jedna z hlavních výzev tkví v průkaznosti – tedy zda v souladu s vědeckými poznatky o rozpoznatelnosti mluvčího člověkem²³ obecně je takové svědectví důvěryhodné, či zda nemůže být v důsledku nepříznivých podmínek jeho hodnověrnost zpochybněna. Mezi takové podmínky řadíme příliš krátké promluvy, velká vzdálenost a časový odstup, rušné prostředí a nestandardní způsoby mluvy jako křik (pro nedávnou případovou studii a odkazy na další literaturu viz Blatchford a Foulkes 2006). Případy identifikace neznámého mluvčího vyžadují odlišnou metodologii, kupříkladu s pomocí *hlasové konfrontace*²⁴, která se též označuje jako *voice parade* (viz Nolan 2003).

Čtvrtý případ uvedený v tabulce 2.1 se váže k situaci, kdy sice existuje svědek zločinu, ale není k dispozici ani podezřelý, ani audiozáznam. Ačkoli k tomu v praxi běžně dochází, odborníci jsou jen zřídka požádáni o součinnost, alespoň v Německu; pro uvedený scénář totiž neexistuje zavedený forenzní postup. Velmi žádoucí by byl určitý ekvivalent tzv. portrétování (podle výpovědi svědků) používaného ve vizuální sféře (pro tuto tezi viz Nolan 1983: 208). Velmi slibný je vývoj současných technologií v řečové syntéze pod výrazy „hlasová transformace“ a „hlasová konverze“²⁵ (obecný přehled nabízí Stylianou 2008).

²³ Ve VT *human speaker perception*; pozor zde musíme dávat na tzv. „bracketing paradox“ (Spencer, 1988, s. 663–682) a nevnímat toto spojení jako *[human] speaker perception* (rozpoznatelnost lidského mluvčího), ale jako *[human speaker] perception*.

²⁴ Ve VT *voice line-up*.

²⁵ Oba termíny se vztahují k úpravě hlasu jednoho (tzv. *výchozího*) mluvčího tak, aby zněl jako jiný (tzv. *cílový*) mluvčí. Tento proces je založen na modifikaci charakteristik vokálního traktu a specifické prozodie daného mluvčího.

Srovnávání a profilování mluvčích (viz tabulka 2.1 vlevo), spadá do oblasti, jíž odpovídá název této kapitoly, tj. expertní identifikace mluvčího (tento termín užívají např. Künzel 1995 či Broderers 2001). Toto označení stojí v opozici pojmu *identifikace rodilým mluvčím*, který se váže ke scénářům popsaným v tabulce 2.1 vpravo. Ačkoli tento druh identifikace provádí jedinci bez patřičného vzdělání v řečové analýze, jeho proces je řízen odborníky, kteří mají na starost plánování a realizaci úkonů jako hlasové konfrontace. Alternativou termínu *expertní identifikace mluvčího* je *technická identifikace mluvčího* (Nolan 1983, 1997). Jak Nolan (1997) zdůrazňuje, adjektivum *technická* je v tomto případě třeba chápat v širším slova smyslu – nezahrnuje pouze instrumentálně orientované prostředky, např. pro spektrální analýzu, ale i postupy na přístrojích nezávislé (jako fonetickou transkripci, založenou na poslechové analýze). V této kapitole budeme operovat pouze s prvním termínem, který vyzývá k bližšímu zamyšlení nad potřebnými kvalifikacemi, které by odborníkům na forenzní identifikaci mluvčích neměly chybět. Vzhledem k tomu, že tato problematika závisí na různých metodách používaných v procesu identifikace, budeme se jí věnovat až v závěru této kapitoly.

2.2 Profilování mluvčího

V oblasti profilování mluvčích je vhodné rozlišovat mezi klasifikací mluvčích a rozpoznáváním jejich nápadných rysů, do něhož lze zapojit i neobornou veřejnost. Klasifikací mluvčích (pro detailní popis viz Müller 2007) rozumíme proces odvozování „třídy“ nebo „kategorie“, do níž daný mluvčí – na základě zvukového materiálu – náleží. Relevantní jsou zejména, jak jsme již nastínili výše, oblast původu, věk, pohlaví a sociální status. Jedním z přínosů klasifikace mluvčích v probíhající policejní vyšetřování je, že informace extrahované z nahrávky neznámé osoby může policie použít přímo, bez nutnosti záznam podrobněji analyzovat. Pokud vzorek řeči například vykazuje známky cizineckého²⁶ (ruského) přízvuku, a navíc dialektu typického pro okolí Stuttgartu na jihozápadě Německa (jak tomu bylo v případě organizovaného pašování drog v roce 2005), policie se může ihned zaměřit na populaci s těmito charakteristikami.

Vzorek řeči však může obsahovat i další charakteristiky mluvčího, zvláštní a nápadné i z pohledu laika, přestože takové rysy nemusí mít nic společného s klasifikací mluvčích.

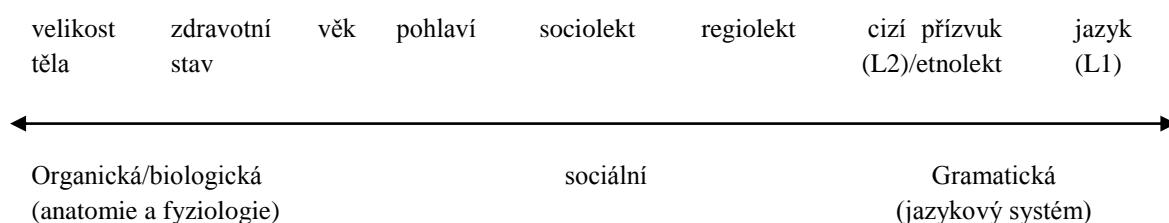
²⁶ Jako výhodné se jeví rozlišovat mezi termíny *cizinecký* a *cizí přízvuk*. Cizineckým přízvukem se myslí přízvuk netuzemský, zatímco cizí přízvuk nemusí nutně pocházet ze zahraniční – spíše jde o přízvuk neobvyklý v daném prostředí, např. ostravský přízvuk v Praze (Skarnitzl, 2019, osobní komunikace).

Hlas může být například výrazně vysoký nebo hluboký, s třepenou²⁷ či dyšnou fonací, drsný či se zjevnou nazalitou, tempo řeči příliš rychlé či pomalé a artikulace nezvykle pečlivá či nedbalá atp. Pokud má policie v rukou podobný popis, event. pokud je zpřístupněn širší veřejnosti (spolu s informacemi z klasifikace mluvího, např. o přítomnosti specifického přízvuku), někdo z vyšetřovatelů nebo veřejnosti může najít vhodného kandidáta, který by popisu odpovídal. Profilování hlasu by se proto nemělo omezovat pouze na klasifikaci mluvích, mělo by se provádět i za účelem vymezit takové rysy řeči, které budou nápadné i laické veřejnosti.

Dalším úkolem nejen profilování, ale i srovnávání mluvích může být hodnocení vlivu přechodných stavů, například stresu, citového hnutí nebo vlivu drog a alkoholu.

Důležitá terminologická poznámka, jež vyplývá z této diskuse a na niž narazíme i v oddílu o srovnávání mluvích, se vztahuje k rozdílu mezi klasifikací mluvích individuálními rysy mluvího. Jevy využívané při klasifikaci mluvích už ze své podstaty nemohou být individuální; naopak charakterizují celé množiny mluvích. Naproti tomu výška hlasu či tempo řeči se u každého liší, což ovšem neznamená, že můžeme mluví identifikovat pouze na základě těchto individuálních charakteristik. Tím se teprve otevírá problém, zda a eventuálně jak budeme v budoucnu schopni identifikovat mluvího pouze na základě nahrávky jeho řeči (problematikou se hlouběji zabývá Nolan 1997). Informace o tempu řeči a výšce hlasu přesto hrají důležitou roli při rozlišování jednotlivých mluvích.

Obrázek 2.1 Domény forenzní klasifikace mluvího



Obrázek 2.1 znázorňuje oblasti klasifikace mluvích, které se již uplatňují v praxi, popř. čekají na intenzivnější využití v příštích letech (velikost těla); zde jsou seřazeny podle toho, zda vycházejí primárně z organických/biologických, či sociálních faktorů, případně zda jsou dány jazykovým systémem.

²⁷ Synonymem je *chraptivá fonace*.

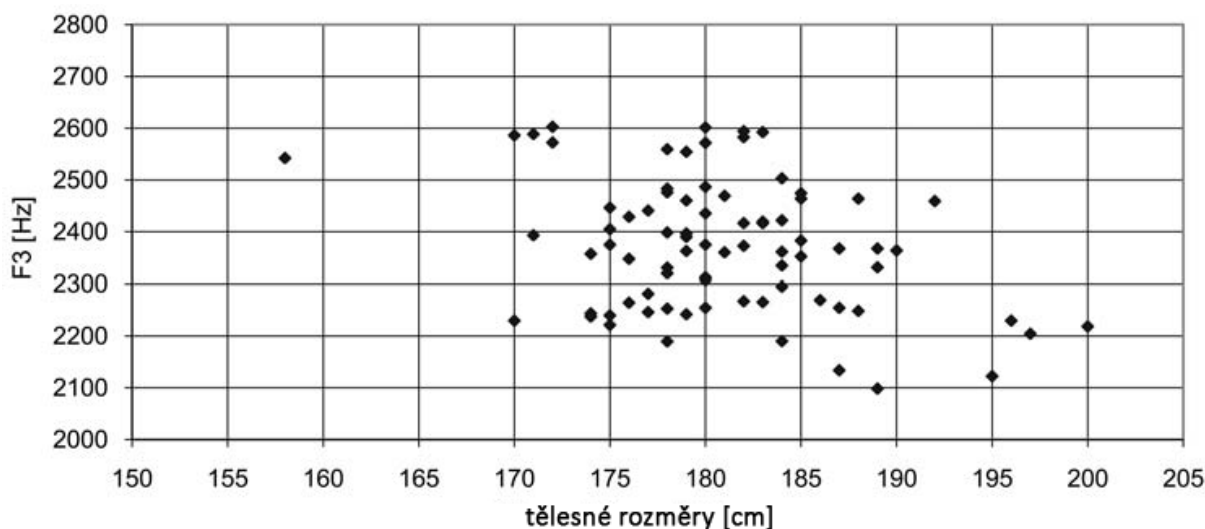
Existuje určitá korelace mezi **velikostí těla** (tj. výškou mluvčího a v menší míře i jeho hmotností) a délkou vokálního traktu (vzdáleností mezi hrtanem a rty). Délku vokálního traktu lze zase přibližně odvodit z formantů samohlásek. Tato korelace však není příliš vysoká; proto zatím není možné přesně vypočítat tělesné rozměry podle formantových frekvencí. Zde je zapotřebí dalšího výzkumu (viz Rendall a kol. 2005). Přesto lze říci, že velmi nízké formantové frekvence pravděpodobně nebudou pocházet od vzrůstem malé osoby a zdrojem velmi vysokých frekvencí formantů pravděpodobně nebude nikdo abnormálně vysoký (Greisbach 1999). Na základě laboratorních dat (Jessen et al., 2005), které shromáždil německý *Bundeskriminalamt*²⁸ (BKA), můžeme Greisbachovo tvrzení potvrdit (obrázek 2.2). Z obrázku 2.2 je patrné, že existuje záporná korelace mezi rozměry těla a průměrnou frekvencí třetího formantu (podle metody LTF²⁹ na obrázku 2.5). Přestože je korelace ve středním pásmu formantových frekvencí a tělesné výšky velmi slabá, jiná situace nastává v krajních hodnotách. Formanty v horním pásmu, nad 2500 Hz, odpovídají mluvčím vzrůstem malým až středně vysokým, zatímco nízké formanty, pod 2200 Hz, korespondují s osobami nadprůměrně vysokými. Průměrné hodnoty druhého formantu (zde neuvedeny) se podobají hodnotám třetího formantu.

Příkladem z praxe může být případ organizovaného obchodu s drogami v jednom německém městě poblíž polských hranic, kde byl neznámý mluvčí komplici přezdíván *der Kurze* (hovorově „malý“). Formantové frekvence jeho hlasu byly ve srovnání s mužskou³⁰ populací vyšší a podezřelý, jehož hlas vykazoval podobné hodnoty a u něhož se později ukázalo, že jde právě o neznámého mluvčího, byl skutečně malého vzrůstu.

²⁸ Spolkový kriminalistický úřad.

²⁹ Dlouhodobá formantová distribuce (*long-term formant distribution*) .

³⁰ Ve VT *male adult*; v odborné literatuře se setkáme i se zdánlivým pleonasmem *dospělý muž / dospělá žena*, přičemž muž/žena v tomto kontextu nemusí nutně konotovat nic jiného než opačné póly pohlavní dichotomie (= nebere se v úvahu věk dotyčné osoby, a v zásadě tak o pleonasmus nejde).



Obrázek 2.2 Záporná korelace mezi tělesnými rozměry [cm] a průměrnou frekvencí třetího formantu [Hz] ve vzorku 81 německy hovořících dospělých mluvčích mužů

Ukazatele **zdravotního stavu** v řeči lze nalézt především v oblasti jazykové, řečové a hlasové patologie. Poruchy zvláště relevantní pro forenzní praxi jsou ty dlouhodobého charakteru – na rozdíl od kupř. laryngitidy, jež má rychlý průběh. Příkladem zdravotního stavu, který našel uplatnění ve forenzní praxi, je koktavost a vadné tvoření sykavek – označované jako sigmatismus³¹ – přestože omezené pásmo propustnosti u telefonátů představuje pro analýzu sigmatismu výraznou překážku. Künzel (1987: 65, 71) uvádí případ levicového terorismu, v němž se hlas pachatele vyznačoval pisklavým zvukem coby příznačným rysem nezvyklého tvoření postalveolární neznělé sykavky, patrným i na spektogramu. Zdravotní specifika, která nespádají do řečové a jazykové patologie a která se podařilo forenzně využít, zahrnují například překážku v dýchacích cestách, částečně spojenou s obezitou. Další relevantní oblastí je vliv psychiatrických/psychických poruch na řeč (Darby 1981). Při hodnocení řeči ovlivněné zdravotním stavem jedince se doporučuje konzultovat s kvalifikovanými lékaři a odborníky na patologii řeči a jazyka.

Co se týče oblasti věku, většinu trestných činů v našich případech páchají muži ve věku 20 až 40 let; někdy až 50, ale jen zřídka starší. K většině změn v řeči ovšem dochází před a po zmíněném intervalu, tj. v dětství, v pubertě a v přechodu do stáří. Právě těmito obdobími výzkumníci věnují největší pozornost, zatímco k forenzně nejrelevantnějším obdobím je dostupné pouze skromné množství informací. Nejvíce podkladů lze najít k vlivu stáří na hladinu základní frekvence (F0). U mužů F0 klesá až do 40. roku života, od cca 50 let má

³¹ Neboli *šišlání*.

následně rostoucí tendenci. U žen dochází k postupnému poklesu počínaje ranou dospělostí, někdy bývá pokles urychlen v období menopauzy a opět může dojít k mírnému nárůstu F0 ve stáří či pozdním stáří (Baken a Orlikoff 2000: 173–76, Schötz 2006: 83). Tuto tendenci ve své rozsáhlé studii potvrdila Schötz a dále zjistila, že s věkem rovněž úzce souvisí tempo řeči – s přibývajícím věkem se postupně snižuje, jak u mužů, tak u žen (Schötz 2006: 110). Našla také další významnou korelaci – zvýšení rozsahu intenzity hlasu – zde je však zapotřebí dalšího výzkumu. Ve forenzní praxi se s doménou věku (v rámci klasifikace mluvčích) nejčastěji operuje tehdy, když se na základě celkového percepčního dojmu odhaduje stáří konkrétního mluvčího. Z výzkumu vyplývá, že průměrný rozdíl mezi odhadovaným a kalendářním věkem je zhruba šest let, přičemž odborníci z forenzní oblasti dosahují v přesnosti odhadu o něco lepších výsledků než laikové (Braun 1996). Koreláty věku, zejména v mládí a stáří, vychází především z biologických faktorů. Můžeme však narazit i na takové, které jsou motivovány faktory sociálními a na něž má obvykle vliv pohlaví a společenské postavení (Foulkes a Docherty 2006). Motivace může vycházet i z jazykového systému, jestliže reflektuje hláskové změny. Wells (1999) například uvádí, že realizace [h]³² ve výslovnosti slov jako *white* u vzorku mluvčích narozených mezi lety 1933 a 1973 a později slábla, čím později se subjekt narodil. Bližší vzhled do problematiky věku s ohledem na produkci řeči a její percepci nabízí Linville (2001) a Schötz (2006, 2007). Další doménou, jíž se budeme zabývat, je pohlaví.³³ Rozlišování pohlaví neznámého mluvčího na zvukové nahrávce (dospělého jedince nedůchodového věku) zpravidla nepředstavuje problém, protože jak fonetické, tak i laici jsou ve většině případů schopni na základě poslechu pohlaví správně určit. Je to dáno tím, že ženy mají zpravidla mnohem vyšší průměrnou výšku hlasu než muži. Význam výšky hlasu pro identifikaci pohlaví byl prokázán i v percepčních experimentech (Coleman 1976). Vysvětlení můžeme najít v odlišných rozměrech hlasivek, které jsou u mužů přibližně o 60 % delší než u žen (Titze 1994: 173). Odborník z oblasti forenzní fonetiky je schopen objektivně kvantifikovat výšku hlasu prostřednictvím analýzy základní frekvence (F0). Podle výzkumu zaměřeného na čtené texty, kterého se účastnilo 100 mužů a 50 žen (a to německy hovořících v dospělém věku), mužský hlas vykazuje průměrnou výšku zhruba 115 Hz a ženský okolo 210 Hz (Künzel 1989). Důležitější než průměrná hodnota u velkého množství mluvčích je však *distribuce* F0

³² Ačkoli se tento znak v české transkripci používá pro znělou glotální frikativu, v mezinárodní fonetické abecedě (IPA) se používá pro její neznělý protějšek (znělá se značí jako [ɦ]).

³³ Následuje pasáž, kterou zde uvádím pouze v poznámce (týká se rozlišení dvou anglických ekvivalentů českého výrazu *pohlaví* – *sex/gender*): Výraz *sex* běžně odkazuje k biologickým aspektům dichotomie muž–žena, zatímco *gender* se zpravidla užívá ve spojitosti se sociokulturními aspekty této dichotomie.

mužů a žen v odhadované populaci. Rozsah střední hodnoty F0 od nejnižší položených hlasů po ty nejvyšší je u mužů zhruba 80 až 170 Hz (Künzel 1989, Jessen et al., 2005) a u žen zhruba 165 až 260 Hz (Künzel 1989, Simpson a Ericsson 2007). Tuto informaci lze využít v případech, kdy je obtížné percepčně určit pohlaví na nahrávce, kde je neznámý hlas na typickou mluvčí ženu příliš nízký, a na typického mluvčího muže příliš vysoký. V takovém případě odborník určí, zda a do jaké míry zjištěná F0 více odpovídá promluvě muže, či ženy. Distribuce F0 uvedené výše byly získány z laboratorních nahrávek neutrální mluvy. U hlasité či emotivní řeči se mohou obě distribuce prolínat výrazněji (viz také obrázek 2.4).

Setkáme se však i s případy, kdy parametr výšky hlasu k rozlišení pohlaví mluvčího nelze použít. Příčinou může být její nespolehlivost v důsledku maskování³⁴ hlasu (např. při třepené fonaci) či šepotu (který se rovněž používá jako maskovací strategie). Nejen v těchto případech lze využít i jiné fonetické signály pohlaví – jeden z významných rozdílů mezi pohlavími například spočívá v odlišné délce vokálního traktu (a jejími akustickými důsledky), který je u mužů asi o 12 % delší než u žen (Fitch a Giedd 1999). Vliv této délky je patrný, změříme-li frekvence formantů F1 až F3 (formanty vyšší než F3 v nahrávkách telefonních hovorů obvykle nejsou k dispozici). Ženský hlas má totiž zpravidla vyšší formantové frekvence než mužský (Hillenbrand et al. 1995). Frekvence formantů mají pro určování pohlaví podobný význam jako F0 (Lehiste a Meltzer 1973; Mullennix et al. 1995). Protože je však obtížné percepční vlivy délky vokálního traktu zanést do fonetické transkripce či jej odlišit od vnímané výšky hlasu, hrají v otázce určování pohlaví (jako i v otázce klasifikace mluvčích) důležitou roli akusticko-fonetické metody (pro zevrubnější popis viz Jessen 2007b).

Ačkoli byly charakteristiky pohlaví, které jsme zde dosud adresovali, povahou organické/biologické, ne všechny řečové rozdíly mezi muži a ženami mají původ zde. Stojí za zmínku, že biologicky determinované odlišnosti lze maximalizovat i minimalizovat působením sociokulturních faktorů. Johnson (2006) například uvádí, že rozdíly ve formantech mužského a ženského hlasu mohou být znatelnější u mluvčích ruského původu než u mluvčích dánského původu, což je pravděpodobně způsobeno spíše sociokulturními než systémovými odlišnostmi mezi oběma jazyky. Dalším důkazem, že se charakteristiky pohlaví neomezují pouze na oblast anatomie a fyziologie, jsou rozdíly v mládí – mezi chlapci a děvčaty – kde jsou organické rozdíly příliš malé na to, aby objasnilly veškeré pozorovatelné

³⁴ Neboli *deformace hlasu* (Svobodová a Voříšek, 2014, s. 142).

odlišnosti v řeči (Whiteside 2001). Patrný může být i vliv sexuální orientace, kdy někteří mluvčí přejímají charakteristiky opačného pohlaví (Munson a Babel 2007).

Kategorizovat faktory na sociální a biologické je značně problematické a jinak je tomu v případě obdobné klasifikace příčin (taktéž na biologické a sociální). Bylo prokázáno, že ženy mají větší vokální prostor, což znamená, že se od sebe u žen různé samohlásky v daném jazyce akusticky výrazněji liší (Diehl et al. 1996). U žen je rovněž patrná tendence častěji realizovat závěrové souhlásky ve finální pozici slova a obecně souhlásky méně redukovat (Byrd 1994). Jak větší vokální prostor, tak sníženou redukcí konsonantů lze chápat jako formy preciznější artikulace. Biologické vysvětlení zní, že vyšší důraz na artikulační preciznost u žen je kompenzací za horší srozumitelnost, jež je způsobena méně zřetelnou strukturou formantů u vyšších hlasů (Diehl et al. 1996). Navíc z důvodu menšího vokálního traktu ženy snáze zaujímají cílovou artikulační pozici³⁵ (Simpson 2001). Sociální výklad argumentuje, že preciznější artikulace žen je zvláštním případem tendence přiblížit se mluvou standardu (mluvit „správně“ v normativním smyslu) a dosáhnout vyšší úrovně prestiže, jak uvádí poznatky z oblasti sociolingvistiky (pro podrobnosti viz Cheshire 2002). Závěrem podotkneme, že v případech, kdy určování pohlaví mluvčího představuje problém, lze čerpat nejen z uvedených, ale i jiných charakteristik pohlaví rozpracovaných v odborné literatuře.

Jako příklad určování pohlaví v praxi uveďme případ z roku 2003, kdy letiště v německém Düsseldorfu obdrželo několik výhružných telefonátů, že nebudou-li v určitý den zrušeny všechny přílety a odlety, dojde k odpálení bomby. Případ nebyl brán na lehkou váhu, protože volající tvrdil, že je napojen na teroristickou organizaci. Veškerý letový provoz tak skutečně na celý den ustal. Určit pohlaví dotyčného nebylo snadné, protože mluvil velmi hlasitě (jak muži, tak ženy mají při hlasité mluvě vyšší průměrnou F0, tudíž je obtížné je rozlišit). Vliv rezonančních frekvencí vokálního traktu také nepřinesl jasné závěry. Z několika kratších pasáží bylo nicméně možné zjistit, že se pravděpodobně jedná o ženu – její hlas vykazoval nižší rezonanční frekvence kvůli zaokrouhlení rtů coby maskování. Později se ukázalo, že tento závěr byl správný. Nebylo zjištěno napojení na žádnou teroristickou organizaci, motivem dané osoby byla podle všeho nechuť onoho osudného dne odletět na dovolenou s přítelem, s nímž se před incidentem pohádala.

³⁵ Tzn. artikuluji snáze než muži.

Pro **sociolekt**, **regiolekt (dialekt)** a **cizí přízvuk/etnolekt** zde nezbyváá prostor, mimoto se o nich zájemce o problematiku může dočíst v široké škále učebnic a příruček (např. Chambers et al. 2002; Doughty a Long 2003). Specifika všech tří uvedených navíc silně závisí na konkrétním jazyce a zemi. Kupříkladu německá regionální, sociální a etnická specifika se budou lišit od britských. Odborníci, kteří se chtějí věnovat případům v dané zemi a daném jazyce, musí být důvěrně obeznámeni s lokální situací, a obecný přehled bez konkretizace místa (země) a jazyka tedy není příliš k užitku. Pohlaví a další domény profilování mluvčích, kterým se v této práci věnujeme, jsou naopak univerzálnější povahy, což je dáno především jejich biologickými aspekty. Pro více informací o forenzní klasifikaci mluvčích viz Künzel (2004), French a Harrison (2006) a Jessen (2007b).

Pojem „jazyk“ na obrázku 2.1 vpravo je relevantní zejména tehdy, není-li jazyková klasifikace příslušných mluvčích, kterou poskytla třetí strana, správná nebo dostatečně přesná. K tomu dochází u nahrávek z oblastí, kde se setkává vícero jazyků (např. Kavkaz), nebo z oblastí s komplikovanou multilingvální situací (např. berberské jazyky vs. arabština v severní Africe).

2.3 Srovnávání mluvčích

V oddílu o profilování mluvčích jsme zmínili rozdíl mezi klasifikací mluvčích a určováním jejich individuálních charakteristik, kterého je schopna i laická veřejnost. Ve sféře srovnávání mluvčích se nabízí podobné dělení. Zaprvé, klasifikace mluvčích je součástí srovnávání mluvčích a zároveň i jejich profilování. Zadruhé, srovnávání mluvčích přináší informace o individuálních rysech daného mluvčího, ale na rozdíl od procesu profilování jej provádí experti, nikoli laici. Situaci shrnuje obrázek 2.3.

Obrázek 2.3 Úkoly profilování a srovnávání mluvčích



Mezi oběma strategiemi určování charakteristik konkrétního mluvčího může docházet k průniku. Jak při profilování, tak při srovnávání mluvčích může být například zjištěno

nezvykle vysoké tempo řeči. Je však zřejmé, že rozsah analyzovaných charakteristik bude širší a výsledky přesnější při srovnávání mluvčích než při jejich profilování. Charakteristiky konkrétního mluvčího, které vyžadují akustická měření, jsou kupříkladu užívány pouze ve sféře srovnávání mluvčích.

Klasifikace mluvčích může hrát významnou roli při srovnávání mluvčích, a to ve dvou případech. Zaprvé může vést k pádnému důkazu hovořícímu v *neprospěch* totožnosti dvou mluvčích; k podobné situaci dochází, pokud jsou u dvou mluvčích zjištěny jednoznačně odlišné hodnoty v jedné či více oblastech klasifikace. Jestliže je kupříkladu odhadovaný věk jednoho mluvčího přibližně 20 let a druhého přes 50 (a nahrávky pochází ze současnosti), s velkou pravděpodobností nepůjde o tutéž osobu. Obdobná situace nastává v případě dialektové neshody, ačkoli bychom měli vzít v potaz možný bidialektismus. Zadruhé, pokud je nějaký rys v populaci vzácný, výrazně to hovoří *ve prospěch* totožnosti dvou mluvčích. Nejlepším příkladem jsou nejrůznější fyziologické zvláštnosti, o jejichž míře výskytu se lze dočíst v lékařské literatuře. Podobně příznačný může být dialekt, sociolekt i cizí přízvuk, to se však odvíjí od kontextu daného případu. Bavorské nářečí bude kupříkladu vzácnější u hranic s Dánskem než v běžném bavorském městě. Jde však o kontextovou „a priori“ informaci (Rose 2002), kterou forenzní expert nemá vždy k dispozici, a i když má, není jasné, nakolik by měla být součástí analýzy.

Tabulka 2.2 Individuální rysy důležité při srovnávání mluvčích.

Organické	Idiolektální	Habituální
1. Průměrná základní frekvence (F0)	4. Individuální aspekty sociolektu, regiolektu a cizího přízvuku	7. Artikulační tempo ³⁷
2. (Dlouhodobé) frekvence formantů	5. Lingvisticko-fonetické detaily	8. Variabilita F0
3. Barva hlasu ³⁶	6. Forenzně-lingvistické rysy	9. Dysfluentní chování

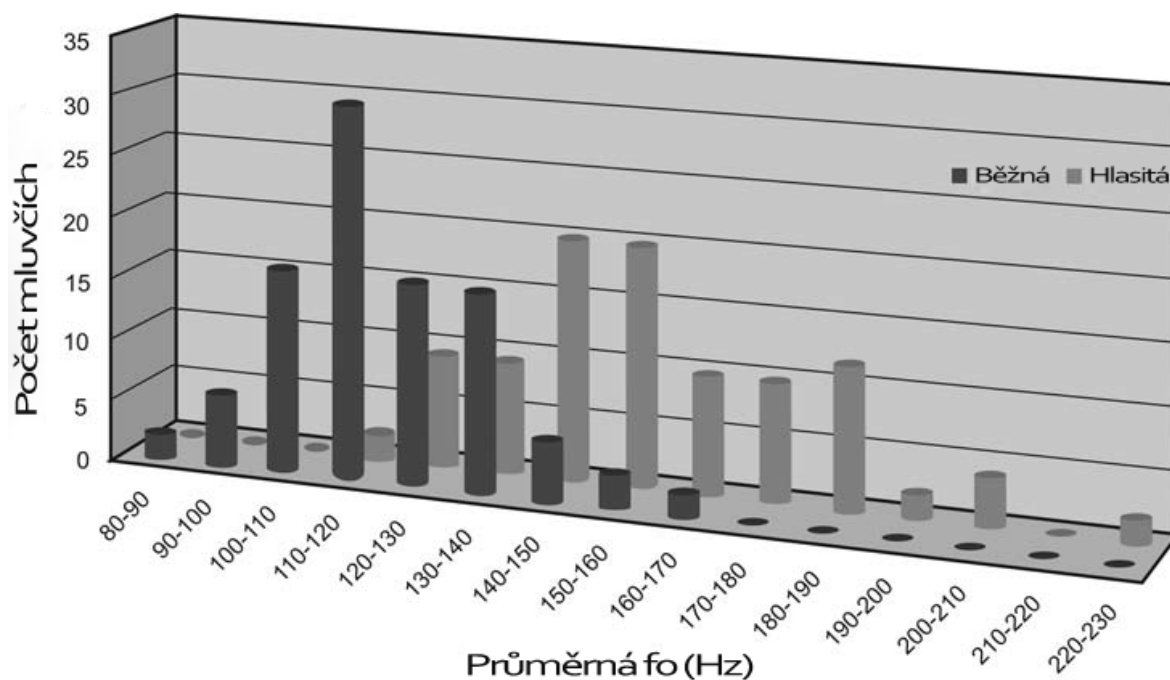
³⁶ Ve VT *voice quality*; Skarnitzl (2014, s. 18) uvádí i doslovný překlad *kvalita hlasu*, který je však v češtině méně uzuální; možným ekvivalentem je i (*hlasový*) *tébr*, *timbre* či *hlasové zabarvení* (srov. *rejstřík* v: Weingartová, Bořil a Vaňková, 2014, s. 83).

³⁷ Není totéž, co mluvní tempo; jde o termín mluvnímému tempu podřazený, nezahrnující mj. frekvence a délky pauz (Weingartová a Volín, 2014, s. 96).

Pokud jde o individuální charakteristiky mluvčího z pohledu odborníka, tabulka 2.2 uvádí kompilaci takových, které se často uplatňují při srovnávání mluvčích. Návrh klasifikace v tabulce 2.2 se nepatrně liší od klasifikace *Stimme, Sprache, Sprechweise*, „hlas, jazyk, způsob mluvy“, užívanou BKA (Künzel 1987, 1995; Gfroerer 2006). Nejpodstatnější rozdíl spočívá mezi kategoriemi „idiolektální“ a „jazykové“. V dřívějším pojetí BKA byly totiž kategorie dialekt, sociolekt a cizí přízvuk řazeny do společné kategorie „jazykové charakteristiky“, zatímco současné pojetí tyto kategorie vnímá jako součást klasifikace mluvčích. Výhodou současného přístupu je jasnější vymezení obou konceptů, tj. klasifikace mluvčích a individuálních charakteristik mluvčího (obrázek 2.3).

Výšku hlasu a její akustický korelát, **průměrnou základní frekvenci (F0)**, jsme již zmínili coby významný korelát pohlaví a uvedli jsme, že na něj má z hlediska anatomie vliv především délka hlasivek. Podobná motivace platí i v rámci téhož pohlaví, kde jsou mezi mluvčími patrné výrazné rozdíly. Výše jsme poukázali na již klasickou Künzelovu (1987, 1989) studii o rozdílech v F0 mezi mluvčími. Novější studii, opět na materiálu němčiny, představili s konzistentními výsledky a s větším rozsahem mluvních stylů Jessen et al. (2005). Hodnoty průměrné základní frekvence spontánní řeči (získané v laboratorních podmínkách) a spontánní řeči ovlivněné Lombardovým efektem³⁸ (mluvčímu je do sluchátek puštěno 80 dB bílého šumu, což má za následek vyšší hlasitost řeči) jsou uvedeny na obrázku 2.4 (přehled dalších studií o Lombardově efektu nabízí Jessen et al. 2005).

³⁸ Ve VT *Lombard condition*.



Obrázek 2.4 Histogram průměrné F0 pro spontánně pronesenou řeč 100 německy hovořících mužů. Výsledky ukazují řeč s Lombardovým efektem (světlé sloupečky, zadní řada) a řeč běžně hlasitou (tmavé sloupečky, přední řada).

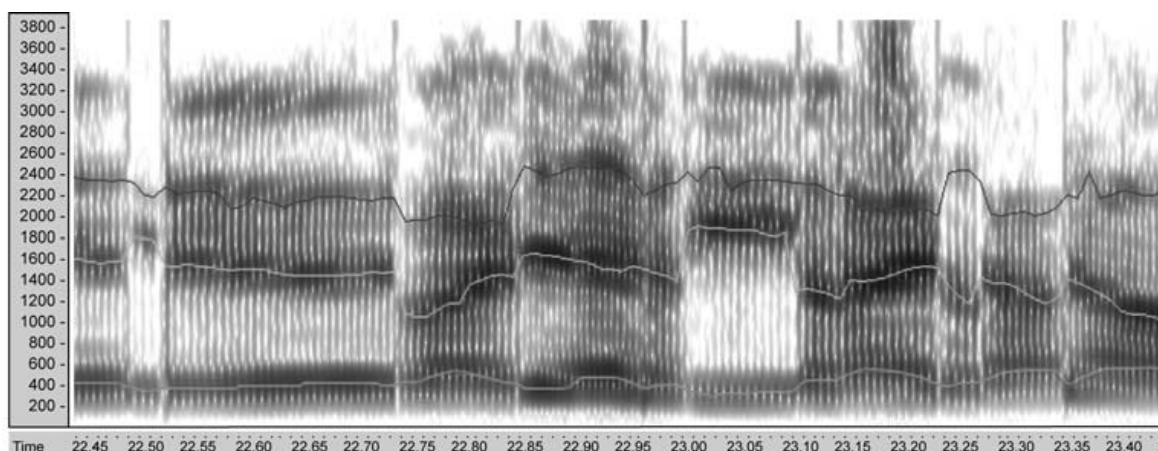
Obrázek 2.4 mimo jiné ukazuje, že průměrná F0 běžně hlasité řeči činí u mužů velmi často přibližně 115 Hz, tj. jde o typickou hodnotu, zatímco hodnoty 85 či 165 Hz najdeme pouze vzácně. Vzhledem k odlišnostem v četnosti je důkaz, že je podezřelá osoba totožná s neznámou, pádnější tehdy, spadají-li do téhož intervalu s nízkou typičností. Hůře prokazatelné naopak je, spadají-li do téhož typického intervalu (pro vysvětlení uvedeného příkladu z hlediska bayesovské³⁹ statistiky viz Jessen 2008b; pro další příklady bayesovského přístupu viz Rose 2002). Na obrázku taktéž můžeme pozorovat posun v distribuci směrem k vyšším hodnotám F0, je-li řeč pronášena hlasitěji. Na hlasitou mluvu ve forenzní praxi narazíme často; nárůst F0 s ní spojený je případem tzv. intraindividuální variability, jevu ve forenzní analýze řeči rovněž nijak neobvyklého. Soudní expert, který s ním přijde do kontaktu, by si měl v odborné literatuře a na základě reálných případů nastudovat jeho příčiny a možný rozsah.

Výzkum průměrné F0 proběhl i na 100 mužích hovořících britskou angličtinou (Hudson et al. 2007). Průměrná hodnota F0 zde činila v případě spontánní řeči nahrávané v laboratorních podmínkách 106 Hz, zatímco při v zásadě identickém mluvním stylu

³⁹ Statistický koncept, který zohledňuje i unikátní, neopakující se jevy, a lze jím vypočítat např. pravděpodobnost identity pachatele a podezřelé osoby (Bořil a Weingartová, 2014, s. 128).

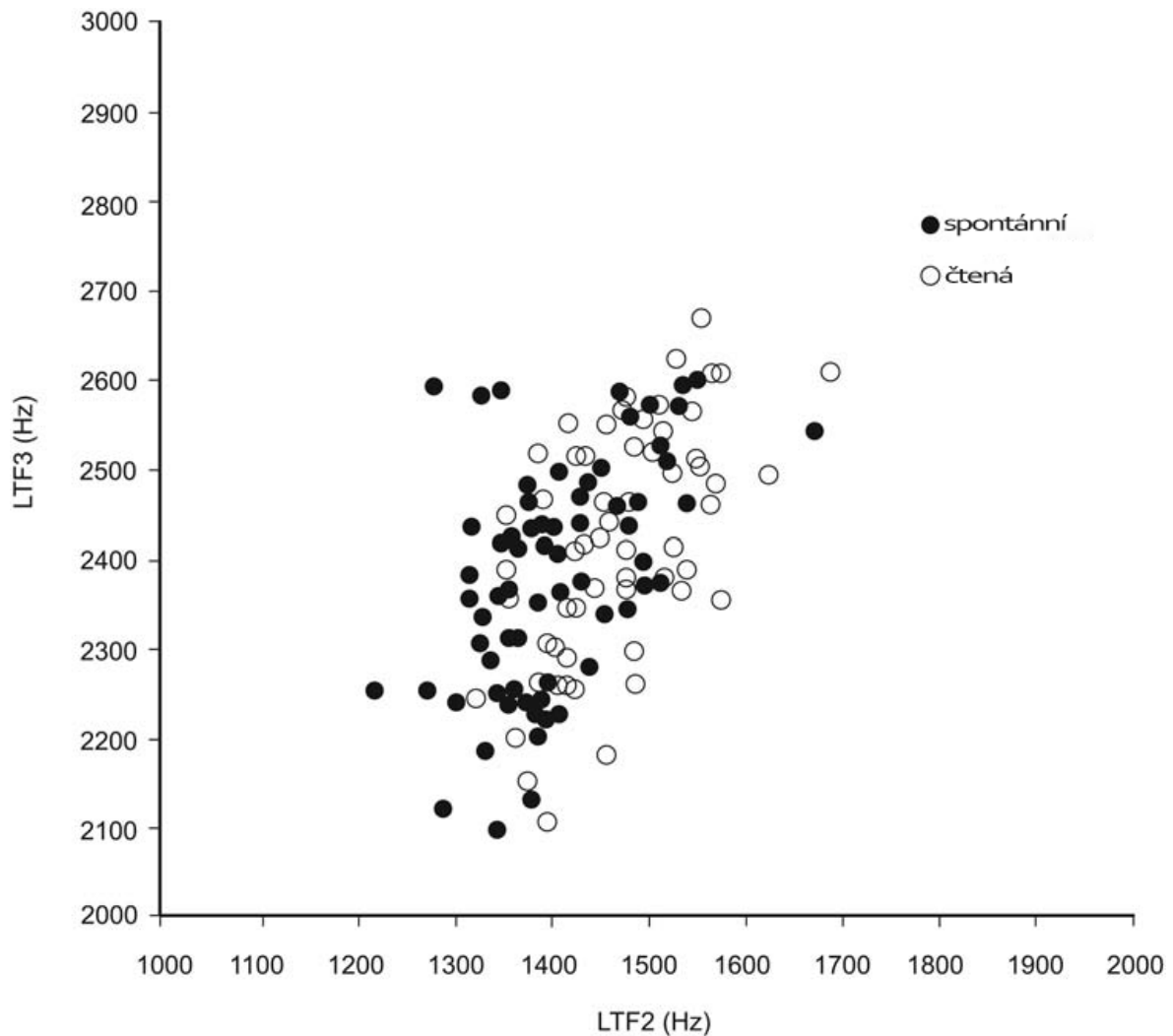
v němčině dosahovala 120 Hz (Jessen et al. 2005). Možným vysvětlením tohoto nesouladu je, že průměrná F0 se u každého jazykového společenství částečně liší (srov. sociokulturní vlivy na akustické koreláty pohlaví v oddílu 2). Podobné případy závislosti průměrné F0 na konkrétním jazyce uvádí Braun a Wagner (2002).

Již výše jsme zmiňovali **frekvence formantů** (jež jsou dány délkou vokálního traktu) coby důležité koreláty při rozlišování mužů a žen. Totéž platí v případě F0 v rámci jednotlivých pohlaví. Existují různé způsoby, jak zachytit individuální rysy mluvčího s ohledem na strukturu formantů. V naší laboratoři často používáme metodu dlouhodobé formantové distribuce (LTF⁴⁰), kterou navrhli Nolan a Grigoras (2005). V souladu s ní signál řeči upravíme tak, že ponecháme pouze průběhy vokálů, v nichž je patrná struktura formantů. Následně aplikujeme automatickou detekci formantů a všechny případné chyby manuálně opravíme. Výsledné formantové stopy exportujeme a použijeme některé statistické metody, zejména průměrování. Fáze před exportováním je znázorněna na obrázku 2.5. Kvalita a kvantita forenzního materiálu se může lišit, a v některých případech proto není analýza LTF možná.



Obrázek 2.5 Ilustrace metody LTF: spektrogram (na ose x čas v sekundách; na ose y frekvence v hertzích) se stopami formantů F1, F2 a F3 (zdola nahoru) na vokalických intervalech řečového signálu. Příklad pochází z případu vydírání jistého majitele diskotéky.

⁴⁰ Z anglického *long-term formant distribution*.

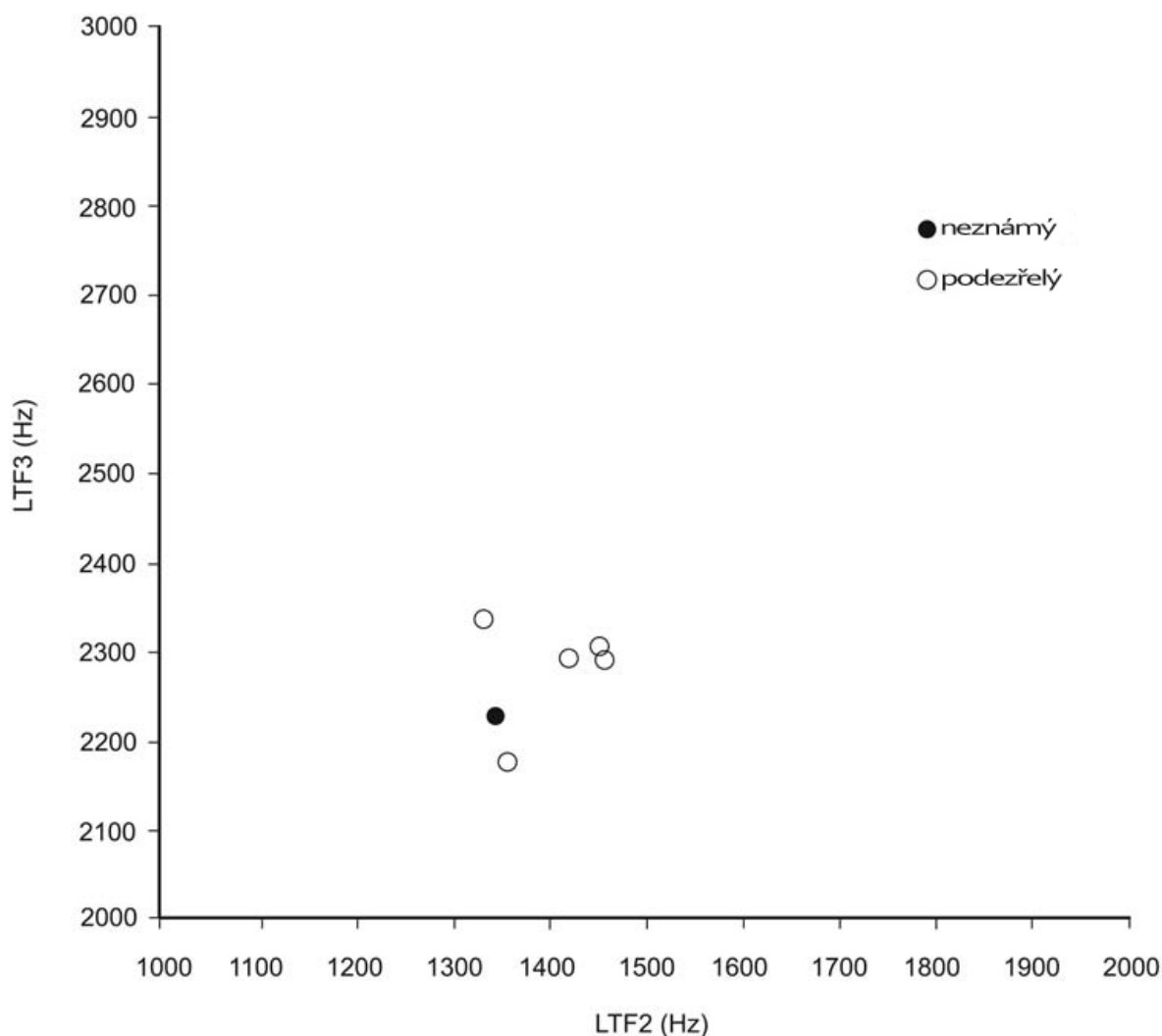


Obrázek 2.6 Referenční data pro průměrné hodnoty LTF [Hz] v spontánní (plné kroužky) a čtené řeči (prázdné kroužky) u 71 německy hovořících mužů. Na ose x dlouhodobý druhý formant, na ose y dlouhodobý třetí formant (podle Moos 2008a).

Na obrázku 2.6 můžeme vidět průměrnou LTF 71 ze 100 německy hovořících mužů, jejichž audiozáznamy byly součástí korpusu použitého při výzkumu parametrů F0 – viz Jessen et al. (2005) (Moos 2008a, 2008b). Měření LTF vycházela z verze korpusu, který byl přenášen prostřednictvím mobilního telefonu. Obrázek 2.6 navíc poskytuje dobrý vhled do možností metody LTF s ohledem na intraindividuální variabilitu, přesněji do jaké míry ji lze postihnout. V souladu s uvedeným obrázkem si například můžeme představit mluvčího, jehož LTF2 by činila 1300 Hz a LTF3 například 2200 Hz – takové hodnoty by byly v populaci poměrně vzácné. Z obrázku je nadto patrné, že LTF2 a LTF3 do jisté míry korelují, což koresponduje s výsledky výzkumu zjednodušeného modelu vokálního traktu (úvod do problematiky nabízí Reetz a Jongman 2009). Dále můžeme upozornit na mírné odchylky mezi čtenou a spontánní řečí (vyšší hodnoty najdeme u řeči čtené). Ačkoli dosud

pravděpodobně neproběhl podobný výzkum LTF s obdobně početným vzorkem mluvčích na jiných jazycích, zkušenosti z praxe zatím naznačují, že i u osob hovořících jinými jazyky je patrný v zásadě totožný formantový prostor.

Nyní si můžeme ilustrovat, jak lze důkazní materiál týkající se LTF využít v praxi. V prosinci roku 2007 jistý mladík anonymně zavolal na policejní stanici poblíž Kasselu v Německu s varováním před plánovanou střelbou v nedaleké škole. Policii se podařilo najít podezřelého, který souhlasil s pořízením zvukového záznamu a následně pětkrát přečetl text, který policie transkribovala z původního telefonátu. V tu chvíli bylo namísto získat i vzorky spontánního projevu, ale policie nekonzultovala postup s odborníky. Později proběhla měření LTF, jakož i dalších fonetických parametrů.



Obrázek 2.7 Průměrné hodnoty LTF [Hz] získané z konkrétního případu; jedna hodnota se vztahuje k neznámému (= spornému) mluvčímu (plné kroužky), zbylé hodnoty ke každému z pěti čtení transkribovaného textu, o němž byl požádán podezřelý (prázdné kroužky).

Z obrázku 2.7 vyplývá, že hodnota LTF2/F3 neznámého mluvčího je v mezích pěti hodnot získaných z audiozáznamu podezřelého. LTF podezřelého vykazovaly tendenci dosahovat vyšších hodnot, což je v souladu se závěry, k nimž dochází Moos (2008a, 2008b), tj. že hodnoty LTF čtené řeči bývají v průměru o něco vyšší než u řeči pronesené spontánně. Kromě prolínání hodnot neznámého a podezřelého mluvčího obrázek 2.7 poukazuje také na to, že hodnoty zjištěné v tomto případě jsou ve srovnání s distribucí na obrázku 2.6 poněkud podprůměrné. Jak překryv hodnot (vysoká míra podobnosti), tak distribuce pod populačním průměrem (relativně nízká typičnost) hovoří pro identitu obou mluvčích.

Kromě metody LTF máme k dispozici i jiné způsoby, jak zachytit formantovou informaci individuálního mluvčího, například měřením formantových frekvencí ve středu vokálu (viz Rose 2002) nebo dynamických vlastností formantů⁴¹ (viz McDougall 2006 a Morrison 2008). Formantové informace patří v oblasti srovnávání mluvčích k těm nejprůkaznějším, přesto je zapotřebí dalšího výzkumu, než budeme schopni v úplnosti pochopit a využít jejich potenciál (Rose 2006).

Barvu hlasu uvedenou v tabulce 2.2 zde chápeme v širším slova smyslu tak, jako ji popisuje Laver (1980), tj. v souvislosti nejen s nastavením hrtanu (třepeň, dyšná, tlačena a drsná⁴² fonace), ale i s nastavením nadhrtanových⁴³ prostor, včetně nazálnosti/nenazálnosti, zaokrouhlení rtů⁴⁴ a pozice dolní čelisti. Vzhledem k frekvenčnímu omezení telefonních nahrávek, jež jsou ve forenzní praxi běžné, a jejich snížené kvalitě v důsledku přítomnosti šumu a zkreslení signálu není akustické měření barvy hlasu v praxi snadné. V tomto případě se tedy opíráme zejména o poslechovou analýzu. Zevrubnější náhled do problematiky s vazbou na forenzní identifikaci mluvčího nabízí Köster a Köster (2004) či Nolan (2005).

Základní frekvenci hlasu, frekvence formantů a barvu hlasu v tabulce 2.2 klasifikujeme jako **organické**, neboť vychází z organických odlišností mezi dvěma mluvčími, např. v délce vokálního traktu či ve velikosti hlasivek. To ovšem neznamená, že na rysy, které sem řadíme, mají organické faktory výlučný vliv. Viděli jsme kupříkladu, že na průměrnou F0 či na strukturu formantů mohou mít dopad i faktory sociokulturní. Barva hlasu je sem zařazena spíše provizorně; na základě dalšího bádání může vyjít najevo, že příhodnější by byla třetí kategorie, „habituální“. Tento problém se odvíjí mj. od konkrétních typů barvy hlasu. Drsný

⁴¹ Skarnitzl et al. (2014, s. 21) hovoří o *formantových trajektoriích*.

⁴² Podle: Rendár (2015, s. 10)

⁴³ Ve VT *supralaryngeal*.

⁴⁴ Těž *labializace*.

hlas má kupříkladu výrazně organický základ, zatímco zaokrouhlování rtů náleží do habituální oblasti. Neměli bychom opomenout ani obecné rozdíly v barvě hlasu mezi mluvčími různých jazyků a dialektů (pro konkrétní údaje a úvod do problematiky viz Braun a Wagner 2002).

Ačkoli je „idiolekt“ (viz druhý sloupec tabulky 2.2) coby termín v lingvistice již dlouho zavedený, stále skýtá prostor pro diskuzi – alespoň v souvislosti se svým významem pro identifikaci mluvčích (některými aspekty se zabývá Nolan 1997; Jessen 2008b či, v této publikaci, Grant⁴⁵). Do tohoto konceptu zde řadíme tři typy jevů. Zaprvé jde o případy, kdy mluvčí *kombinuje* aspekty několika regionálních, sociálních a multilingválních variet. Toto pojetí vychází z myšlenky, že coby mluvčí můžeme v průběhu života měnit geografickou oblast i sociální prostředí, přičemž si po zbytek života neseme určité rysy každého z nich, a to v do značné míry unikání kombinaci (za ozřejnění této interpretace patří dík Francisu Nolanovi). Na idiosynkrazii mluvčího má vliv i osvojení druhého jazyka (Oksaar 1987), neboť v něm daný jednotlivec dosáhl určité úrovně, a to určitým způsobem – roli zde hrají vloh, motivace, délka expozice danému jazyku či moment prvního kontaktu s ním. S idiolektem souvisí možnost, zdali mluvčí neuplatňuje pouze některé rysy dialektu, a jiné nikoli. Zvláštním případem je zde koncept *míry* dialektu, tj. zda mluvčí vykazuje silnější či slabší známky daného dialektu (či jiné variety). Míra dialektu je rovněž patrná z frekvence výskytu (resp. důslednosti dodržování) či síly fonetické/fonologické (či jiné) charakteristiky dialektu.

Druhé pojetí idiolektu se týká lingvistických a fonetických prostředků nedialektové povahy – s dialekty či jinými varietami souvisí nanejvýš částečně. Příkladem může být epenteze ploziv (vlození závěrové souhlásky) ve slovech jako *hen^lce* nebo *el^lse*. Ačkoli bylo již dříve prokázáno, že se tento druh epenteze odvíjí od daného dialektu (Fourakis a Port 1986), závisí též na konkrétním mluvčím, a to i v rámci téhož dialektu (Yoo a Blankenship 2003; Jessen 2008b). Jako další příklad potenciálních fonetických prostředků idiolektálního charakteru, na něž často narazíme ve forenzní literatuře, můžeme uvést glotalizaci hlásek /p, t, k/ ve finální pozici nebo také velarizaci postvokálního /l/ v angličtině (Nolan 1983) a spirantizaci (změnu plozivy ve frikativu) intervokálního /b/ či realizaci glotalní plozivy⁴⁶ v němčině (Künzel 1987). Další příklady segmentálních jevů příznačných pro konkrétního mluvčího a nezávislých na dialektech a sociolektech, předkládá Moosmüller (1997), přičemž

⁴⁵ Konkrétně jde o 33. kapitolu nazvanou „Text messaging forensics“, jejímž autorem je Tim Grant.

⁴⁶ Neboli *rázu*.

zdůrazňuje, že podobné idiosynkrazie je namísto hledat v prozodicky slabých pozicích a neformální mluvě.

Třetí pojetí idiolektu důkladněji než forenzní fonetika rozpracovává forenzní lingvistika; vztahuje se k idiosynkratickým aspektům slovní zásoby a syntaxe (Coulthard 2004). Příkladem jevů, které se vyskytují v mluveném jazyce, jsou výplňková slova jako *vlastně*⁴⁷. Samozřejmě nelze říci, že by takové výrazy byly příznačné pro jednoho konkrétního mluvčího, ale jejich výběr a frekvence výskytu jsou charakteristické pouze pro omezenou skupinu mluvčích.

Třetí kategorii v tabulce 2.2 nazýváme **habituální**. V určitém smyslu ji definujeme záporně: řadíme zde takové charakteristiky mluvčího, které postrádají zřejmý organický základ a zároveň nemají spojitost s jazykovými konvencemi, jež vyžaduje či očekává daný jazykový systém nebo sociální skupina. Příkladem jevů ze zmíněné tabulky, které jsou ve forenzní praxi frekventované, jsou artikulační tempo (pro úvod do problematiky a populační statistiku viz Jessen 2007b) a variabilita F0 (pro hlubší vzhled viz Jessen et al. 2005). Dalším habituálním jevem je dysfluentní chování (zejména používání hezitačních zvuků⁴⁸ typu *ehm*, *hm*). To podle analýzy korpusu BKA zmíněného výše přináší o konkrétním mluvčím řadu potenciálně užitečných informací (Tschäpe et al. 2005), což koresponduje s tvrzením, že hezitační zvuky mluvčí netvoří vědomě (Corley a Steward 2008). Rozdíly mezi mluvčími v užívání hezitačních zvuků nachází i de Leeuw (2007).

Zatímco zde adjektivum „idiolektální“ používáme pro označení idiosynkrazií úzce spjatých s jazykovým systémem (gramatikou), termín „habituální“ používáme naopak pro ty idiosynkrazie, které leží mimo oblast gramatiky (ta například nespecifikuje tempo řeči). Idiolektální rysy jsou nadto často segmentální povahy, zatímco rysy habituální převážně prozodické, byť najdeme i výjimky. Taktéž narazíme na jevy obtížně zařaditelné – za bohatý zdroj odlišností mezi mluvčími je kupříkladu považována koartikulace (vzájemné působení sousedících hlásek) (Nolan 1983). Nabízí se otázka, zda je idiolektální, protože zahrnuje (segmentální) fonetické jevy, které jsou spjaté s jazykovým systémem (např. podoba

⁴⁷ Ve VT *you know*; podobné výrazy se překládají jen obtížně, přibližným ekvivalentem může být *vlastně*, *prostě* či *jako*.

⁴⁸ Ekvivalentem *fillers* jsou v češtině zpravidla *výplňková/parazitní slova*. O nich se však mluví v předcházejícím odstavci, a to v souvislosti s odlišnou, idiolektální, kategorií. *Fillers* se však nevztahují výhradně k lexikálním prostředkům: „A word or sound filling a pause in an utterance or conversation (e.g. er, well, you know)“ (<https://en.oxforddictionaries.com/definition/filler>, cit. 17. 5. 2019) a Jessen s tímto termínem operuje právě v souvislosti s prostředky nelexikálními. Proto jsou zde *fillers* překládány jako *hezitační zvuky*. Za zmínku stojí i poslední uvedený příklad, kterým OED doplňuje definici *fillers*, tj. výraz *you know*, který by tak dle Jessena spadal do sféry idiolektu.

koartikulace se mezi jazyky může různit), nebo habituální, neboť 1) vychází z volnosti jak přejít z jednoho artikulačního cíle na druhý a 2) lze ji lokalizovat na škále od habituálně pečlivého způsobu mluvy po nedbalý. Uvedený příklad má ilustrovat, že klasifikace v tabulce 2.2 slouží pouze jako stručný návod jak rozlišit různé druhy individuálních rysů mluvího (další klasifikace nabízí Nolan 1983 a Rose 2002).

2.4 Závěr

V této kapitole jsme si představili stručný přehled profilování a srovnávání mluvích – dvou podoblastí forenzní identifikace mluvích. Zdůraznili jsme, že obě sféry vyžadují účast odborníka, kterému nechybí fonetické ani lingvistické vzdělání. V oblasti fonetiky se neobejdeme bez fonetické transkripce založené na poslechové analýze ani bez akustických měření, která se týkají i oblastí, které nespádají do gramatického systému jazyka (jako jsou individuální rysy související s barvou hlasu či artikulačním tempem). Znalost lingvistiky naproti tomu najde upotřebení při identifikaci dialektu či jiné jazykové variety a mj. hraje významnou roli při řešení forenzních případů, které zahrnují přítomnost cizího jazyka, kde je rovněž zapotřebí navázat spolupráci s konzultanty⁴⁹ a dalšími lingvisty. Pokud se u odborníka v průběhu případu projeví jakékoli mezery ve vzdělání, měl by vědět, kde hledat relevantní odbornou literaturu, ať už fonetickou, či lingvistickou, a musí jí i rozumět. BKA čas od času oslovují zahraniční forenzní instituce, jež pro své pracovníky poptávají školení v oblasti forenzní identifikace mluvích a osvědčení, že jsou způsobilí tuto činnost vykonávat samostatně. V takových případech poukazujeme na to, že výcvik tohoto druhu je možný pouze tehdy, má-li zájemce patřičné vzdělání v lingvistice, fonetice nebo v oboru blízce příbuzném, a zároveň se předpokládá určitá obeznámenost v rámci forenzní fonetiky. Je pravda, že některé aspekty metodologie týkající se identifikace mluvích se lze naučit poměrně snadno, například měření průměrné F0 či dlouhodobých formantů. Tyto znalosti však dostačují pouze, nejedná-li se o komplikovaný případ. Jestliže kupříkladu máme k dispozici nekvalitní nahrávku, kde třetí formant není jasně rozpoznatelný a není zřejmé, zda je určitý akustický jev součástí řečového signálu či neřečového šumu, je zapotřebí mnohem hlubšího povědomí o produkci řeči a jejích akustických důsledků. Nebo pokud si například navzdory množství důkazů svědčících o identitě dvou mluvích neodpovídají distribuce formantů, je žádoucí opřít se o poznatky z oblasti intraindividuální variability.

⁴⁹ Ve VT *informants*; obvyklým českým ekvivalentem jsou *informátoři* – ne však v tomto případě. Jedním z významů anglického *informant* je totiž „A person from whom a linguist or anthropologist obtains information about language, dialect, or culture.“ (<https://en.oxforddictionaries.com/definition/informant>, cit. 17. 5. 2019).

Případy mohou být složité z nejrůznějších důvodů a názorem autora této kapitoly je, že akademická znalost fonetiky a lingvistiky je pro identifikaci mluvcích klíčová. Ne všichni odborníci na forenzní analýzu řeči a zvuku však toto stanovisko sdílí. Ačkoli na tvrzení explicitního charakteru narazíme pouze vzácně, v některých zemích patrně převládá názor, že problém forenzní identifikace mluvcích je čistě technického rázu, a lze jej tedy vyřešit zakoupením vhodných nástrojů na analýzu řeči a/nebo využitím nástrojů na automatické rozpoznávání mluvcích. Ty znamenají velký pokrok, ale je vhodné je vnímat spíše tak, že rozšiřují paletu jiných foneticko-lingvistických metod, avšak nenahrazují je. Na problematiku automatického rozpoznávání mluvcích v této kapitole nezbyvá prostor, ale základní poučení a odkazy na doporučenou literaturu lze najít v Jessen (2008a).

Doporučená literatura

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

(Učebnice, která poskytuje vhled do celého spektra forenzní fonetiky a akustiky – nejen z oblasti identifikace mluvcích – a která obsahuje i nemálo odkazů na literaturu z dob, kdy se tato disciplína teprve začínala formovat.)

Jessen, M. (2008) 'Forensic phonetics', *Language and Linguistics Compass*, 2: 671–711. (Nedávná sonda do současného stavu forenzní identifikace mluvcího s důrazem na teorii a metodologii srovnávání mluvcích, včetně jejich automatického rozpoznávání.)

Nolan, F. (1983) *The Phonetic Bases of Speaker Recognition*, Cambridge: Cambridge University Press.

(Stěžejní monografie, která pokládá mnohé terminologické a konceptuální základy forenzní identifikace mluvcího a rovněž obsahuje relevantní fonetické experimenty.)

—(1997) 'Speaker recognition and forensic phonetics', in W.J. Hardcastle and J. Laver (eds) *The Handbook of Phonetic Sciences*, Oxford: Blackwell. (Zásadní práce, která zdůrazňuje důležitost forenzní praxe coby oblasti aplikované fonetiky a vnímá forenzní fonetiku jako etablovanou součást fonetiky.)

Rose, P. (2002) *Forensic Speaker Identification*, London: Taylor and Francis. (Moderní učebnice forenzní identifikace mluvcích, která klade důraz na akustickou fonetiku a bayesovský přístup, čímž poskytuje zajímavý náhled na možný vývoj této disciplíny.)

Webové přílohy

Zdroje uvedené níže jsou doplněny o konferenční powerpointovou prezentaci, kterou lze shlédnout na webové stránce přidružené k této publikaci, dostupné na <http://www.forensiclinguistics.net/>.

Jessen, M. (2008b) ‘Categorical v. Continuous Variations Between Speakers’, paper presented at the 17th Annual Conference of the International Association for Forensic Phonetics and Acoustics, Lausanne. Supported by: *IAFPA 2008 Jessen.ppt*

Moos, A. (2008b) ‘Long-Term Formant Distribution (LTF) Based on German Spontaneous and Read Speech’, paper presented at the 17th Annual Conference of the International Association for Forensic Phonetics and Acoustics, Lausanne. Supported by: *IAFPA 2008 Moos.ppt*

Tschäpe, N., Trouvain, J., Bauer, D. and Jessen, M. (2005) ‘Idiosyncratic Patterns of Filled Pauses’, paper presented at the 14th Annual Conference of the International Association for Forensic Phonetics and Acoustics, Marrakesh. Supported by: *IAFPA 2005 Tschaepe et al.ppt*

3. Překladová část 2

Poslechově-akustické vs. automatické metody ve forenzněfonetické praxi

Anders Eriksson

Abstrakt Tato kapitola se zabývá analýzou řeči ve forezním kontextu. Zohledňuje jak přístupy tzv. poslechově-akustické, tak i metody automatické a následně popisuje jejich aplikaci ve forezní praxi. Ta představuje mnohé výzvy, které v laboratorních podmínkách, odkud aplikované metody vzešly, nenajdeme. Forezní praxe se dotýká různých podoblastí, jako je srovnávání hlasů, profilování mluvčího, fonetická transkripce nebo „čištění“ špatně srozumitelných nahrávek, ale v této kapitole se detailněji zaměříme pouze na srovnávání hlasů. Účelem fonoskopie je předložit důkaz průkazný u soudu. S tím jsou spojeny další komplikace jako výběr vhodného formátu – toho, jakým způsobem bude důkazní materiál prezentován. Porovnáme klady a zápory tradičního přístupu, který pracuje s hodnotící stupnicí, eventuálně s věrohodnostním poměrem, a to s ohledem na současnou diskuzi v rámci forezních věd o předkládání důkazních materiálů.

3.1 Úvod

Až donedávna se většina forezních případů opírala o tzv. poslechově-akustické metody. V průběhu posledního desetiletí se však do popředí postupně začínají dostávat metody automatické. Ačkoli nemáme k dispozici přesné údaje o tom, jak velký podíl případů zahrnoval ten či onen postup, většina případů se v současnosti řeší výhradně metodami poslechově-akustickými. Kombinace obou přístupů našla uplatnění v kriminalistických ústavech v Německu a ve Francii či u nezávislých analytiků v různých částech světa. V této kapitole představíme klady a zápory obou, přičemž vycházíme mj. z dlouholetých praktických zkušeností získaných při spolupráci se švédskou policií. Důraz bude kladen na aplikaci a využitelnost poznatků ve skutečné forezní praxi spíše než na výzkum, z něhož čerpají.⁵⁰

V této kapitole se zaměříme na forenzněfonetickou praxi, jež zahrnuje tzv. identifikaci mluvčích, přestože, jak uvidíme, tento termín není příliš vhodný. Analýza řeči ve forezním prostředí však čítá i jiné podoblasti, na něž bychom neměli zapomínat. Jednou z nich je transkripce audiozáznamů. Může se zdát, že jde o triviální činnost, kterou zvládne každý

⁵⁰ Kontaktní údaje/vizitka A. Erikssona v zápatí VT v této práci vypouštíme.

i bez forenzního či lingvistického/fonetického vzdělání, což ovšem není ani zdaleka pravda. Při zmíněném druhu transkripce často hraje zásadní roli, zda víme, co je v daný moment řečeno, přičemž si nemůžeme dovolit hádat. Pokud není osoba pověřená transkripcí schopna s jistotou určit, jaká slova, výrazy nebo číselné údaje v nahrávce zaznívají, nesmí tuto skutečnost zatajit. U neškolených jedinců je patrná výrazná tendence zapisovat ne to, co je ve skutečnosti řečeno (včetně nesprávné výslovnosti, přerázení, heziticí a repeticí), jako spíše to, co na základě vlastních jazykových znalostí coby rodilých mluvčích „očekávají“ že daný mluvčí „asi“ říká; bez ohledu na fakt, že v některých případech je zhruba nemožné s jistotou určit, co přesně bylo proneseno. Správně transkribovat obtížnou nahrávku vyžaduje dobrý teoretický základ i praktické zkušenosti, a to z mnoha důvodů: jedním je nezřídka špatná kvalita záznamu. Dalšími faktory mohou být překrývající se promluvy, nedbalá výslovnost, držení telefonu daleko od úst, rušné prostředí, špatný přenos telefonního signálu ad. Uvedme alespoň dva příklady, kde důsledně provedená transkripce sehrála klíčovou roli. První případ se týkal nelegálního obchodu se zbraněmi mezi zločineckými gangy. S pomocí odposlechu mobilního telefonu jednoho z pachatelů se podařilo nahrát konverzaci, k níž došlo během předání. Ta se ovšem s příslušnou zásilkou nijak nesouvisela – došlo k tomu, že pachatel s někým telefonoval, zatímco ve stejnou chvíli přímo na místě komusi dával pokyny ohledně nakládky jistých automatických pušek. Přestože dotyčný v průběhu mluvil do telefonu z větší vzdálenosti, což negativně ovlivnilo příjem, podařilo se z obsahu extrahovat dostatečné množství informací a na základě nich potvrdit, že skutečně šlo o obchod se zbraněmi. Další případ se týkal podezření z nezákonného obchodu s cennými papíry, kde byla zvuková kvalita záznamu z odposlouchávaného telefonu poměrně vysoká, jenže mluvčí pro různé skutečnosti opakovaně použili krycí označení. Bylo proto nezbytné důsledně transkribovat vše řečené, byť by to nezasvěcenému místy téměř jistě nedávalo smysl. Tyto případy mají ilustrovat, že fonoskopie často obnáší více než jen rozhodování, zda vzorek řeči A a B pochází od téže osoby.

V krátkosti zmiňme také proces čištění nekvalitních zvukových souborů. Filmy a detektivní seriály v nás budí dojem, že stačí pouhých několik kliknutí myši, a nahrávka dříve nesrozumitelná je rázem křišťálově čistá. Ve skutečnosti to tak není. Co si málokdo uvědomuje (včetně nemalého počtu policistů), je fakt, že čištění s sebou téměř vždy nese i odstranění potenciálně užitečných informací. Záznam je pak sice příjemnější na poslech, ale filtrováním můžeme mimoděk zkreslit výsledky následné akustické analýzy. V určitých případech, kdy potřebujeme především co nejlépe rozumět tomu, co je v nahrávce řečeno,

má čištění signálu své opodstatnění. Nejlepších výsledků (kdy dochází k odstranění potenciálně užitečné informace pouze v minimu případů) dosáhneme tehdy, pokud je rušivý signál⁵¹ striktně periodický a má malou šířku pásma. To lze ilustrovat na případu loupeže, jíž padl za oběť jistý taxikář. Celý incident byl zaznamenán bezpečnostním systémem nainstalovaným ve voze. Na nahrávce se s pravidelnou frekvencí ozývalo velmi nepříjemné úzkopásmové pípání. Podobné zvuky lze poměrně snadno odstranit, aniž bychom podstatně zdeformovali zbylou část signálu. Díky tomuto postupu bylo následně mnohem snazší konverzaci transkribovat. Uvedený příklad je dokladem, že čištění signálu může být dobrým pomocníkem, ačkoli spíše ve zvláštních případech. Ve většině běžných případů bývá jeho pozitivní dopad minimální, často bývá spíše nežádoucí.

3.2 Stručný přehled vývoje rozpoznávání mluvcích ve forenzním prostředí

Zbytek kapitoly se bude věnovat rozpoznávání mluvcích neboli jejich srovnávání, což má našemu pojetí blíže. Začneme ale stručným nástinem historie rozpoznávání mluvcích ve forenzním kontextu, abychom lépe pochopili, v jaké fázi se právě ocitáme a jaké kroky k tomu vedly. Málakoho asi překvapí, že dřívější rozpoznávání mluvcích se opíralo výhradně o poslechovou analýzu. Z dnešního pohledu se tedy jednalo o důkazní materiál ve formě sluchové identifikace mluvcího svědkem⁵². To prakticky vzato znamená, že důkazní materiál se opírá pouze o paměť svědka, kde je uložen hlas pachatele. První zaznamenaný případ, kde sluchová identifikace mluvcího svědkem v dnešním slova smyslu sehrála u soudu klíčovou roli, pochází z roku 1660. Týkal se jistého Williama Huleta, který byl obžalován z královraždy, k níž mělo dojít popravou krále Karla I. Stuarta⁵³. Jistý svědek totiž vypověděl, že zaslechl popravčího, jenž měl zahalenou tvář, a „podle jeho řeči“ poznal Huleta. Hulet byl následně porotou shledán vinným z velezrady a byl odsouzen k trestu smrti. Sluchová identifikace svědkem spadá do forenzního srovnávání hlasů, k němuž má zájemce o problematiku k dispozici celou řadu pramenů. Cílem této publikace je však popsat analýzu audiozáznamů, takže se této oblasti budeme věnovat jinde.

Příchodem nahrávacích zařízení se pro fonoskopii otevřely zcela nové možnosti. Významným milníkem byl zejména vývoj nových technologií pro analýzu, obzvláště

⁵¹ Ve VT *noise*, který ve fonetice často překládáme jako *šum*; zde to však možné není, protože šum je neperiodické povahy. Jinými slovy: „Laická představa o tom, že stačí od signálu se šumem jednoduše odečíst samotný šum, získaný např. z místa těsně před nebo po záznamu, je chybná. Šum je signálem náhodným, a tudíž nelze determinovat přesný tvar jeho vlny.“ (Svobodová a Voříšek, 2014, s. 141).

⁵² Ve VT *earwitness evidence*; překlad podle: Skarnitzl, 2014, s. 13. Ustálený český ekvivalent v kratší formě pravděpodobně neexistuje.

⁵³ Ve VT *King Charles I.*

spektrografu. Základní teze byly publikovány ve třicátých letech [56] a první komerčně dostupný přístroj (Sonagraph) byl vyroben na počátku let čtyřicátých. Jeho přínos pro fonetickou analýzu je nesporný, ale představa, že by se identifikace mluvčího prováděla na základě vizuální analýzy spektrogramu – pro tuto techniku se vžilo označení *voiceprinting* – nebyla realistická. Iluze zřejmě vznikla kvůli povrchovým podobnostem mezi spektrogramem a otisky prstů. Koncept *voiceprintingu*, tzn. rozpoznávání mluvčího založené na vizuálním porovnávání spektrogramů, přesto v průběhu následujících let sehrál významnou roli při nejednom vyšetřování – byť důkazní materiál z něj plynoucí nebyl vždy přípustný u soudu. Rovněž je třeba říci, že přední fonetické té doby nepovažovali tuto metodu za spolehlivou a že přesnost, o níž mluvili její stoupenci, se v žádné studii nepodařilo prokázat. Podíváme-li se na uvedený vývoj prizmatem teorie vědy, můžeme mluvit o posunu v paradigmatu, zásadním obratu v přístupu od ryze sluchové identifikace směrem k metodám akustickým. V současné době *voiceprint*, neboli otisk hlasu⁵⁴, používají pouze někteří soukromí detektivové a, co je zajímavé, i FBI – kromě toho je však jeho význam pro forenzní praxi zanedbatelný. Podobně vypadal vývoj v Sovětském svazu, přičemž existují i důkazy o využití automatických (nebo alespoň poloautomatických) metod. Prozatím o tom víme jen velmi málo, ale zajímavé líčení těchto snah v rámci poloautomatického identifikace mluvčího nabízí Solženicyn ve svém románu *V kruhu prvním*⁵⁵.

3.3 Metody používané v současnosti

Jak jsme již zmínili v úvodu, ve forenzní praxi dnes stejně jako po dvě nebo tři předcházející dekády převládá kombinace poslechového a akustického přístupu. Ten je využíván větší částí evropských i mimoevropských fonetiků, a to včetně většiny členů jediné profesionální organizace operující na poli forenzní analýzy řeči, *Mezinárodní asociace forenzní fonetiky a akustiky* (IAFPA). Jak ovšem bylo rovněž uvedeno, do popředí se začínají dostávat metody automatické. Oba přístupy si stručně přiblížíme v následujících dvou oddílech. Zevrubnější výklad bude následovat v oddílech pozdějších.

3.3.1 Metody poslechové a akustické analýzy

Poslechovými a akustickými metodami jsou myšleny lingvistické závěry, k nimž odborník dochází na základě poslechu nahrávky řeči, v kombinaci s měřením akustických jevů,

⁵⁴ Doslovný překlad ze: Skarnitzl, 2014, s. 13; analogicky k *fingerprint* – otisk prstu.

⁵⁵ Ve VT *The First Circle* by Solzhenitsyn; zohlednit je zde zapotřebí jednak český zápis jména autora, ale i vžitě označení příslušného titulu.

kterých zpravidla využívají fonetické při popisu řeči. S poslechovou analýzou souvisí faktory jako přechyby, patologické vady řeči, idiosynkrazie, dialekt, cizinecký přízvuk ad; s akustickou analýzou potom formanty, průměrná základní frekvence a směrodatná odchylka F0, tempo řeči⁵⁶ apod.

Proces poslechově akustické analýzy můžeme schematicky rozdělit do následujících kroků:

1. Pečlivý poslech řečového materiálu

Tím získáme zhruba přehled základních faktorů, jako je délka trvání užitečného materiálu či zvuková kvalita, ale také nám umožňuje identifikovat rysy vhodné k detailnější analýze.

2. Transkripce materiálu

Není mi známo, do jaké míry transkripci využívají jiná pracoviště, ale v naší praxi v Göteborgu se velmi osvědčila. Jednak významně zpřehlední celý materiál, a zároveň lze s její pomocí vyhledat a automaticky extrahovat data pro různé typy analýz, kupř. analýzu formantů. Nabízí se řada odlišných, ale rovnocenných možností – naše pracoviště používá skripty balíčku pro akustickou analýzu programu Praat.

3. Jazyková analýza

Jazykově laděnou analýzou vzorků řeči zaměřenou na oblast dialektu, sociolektu, přízvuku, gramatiky ad. můžeme o mluvčím zjistit mnoho užitečných informací. U tohoto typu analýzy je nejméně jasné jak docílit jejího úplného nahrazení automatickými metodami.

4. Akustická analýza

Akustická analýza vždy zahrnuje analýzu nejčastěji využívaných parametrů jako formantů, základní frekvence, intonačních kontur atd. Uplatňuje se zde ale i mnoho dalších parametrů, které blíže popíšeme v oddílu 3.4.

5. Shrnutí poznatků ve formě studie

Tato oblast je zpravidla, dle mého názoru, Achillovou patou poslechově-akustického přístupu. Pokud bychom se drželi postupu navrhovaného Nolanem⁵⁷ (s. 10, [42]), analýzu

⁵⁶ Terminologická poznámka: Weingartová a Volín (2014, s. 96) používají *tempo řeči* jako „zastřešující termín“ pro mluvní (speech rate) i artikulační tempo (articulation rate).

⁵⁷ Tato část v původním textu chybí; celá věta v orig. zní „If we were to follow the guidelines suggested by (p. 10, [42]) the analysis should be performed in two stages—feature extraction and feature comparison.“ Nolan je ale později v textu zmiňován, a zejména je autorem publikace, na niž se odkazuje v hranaté závorce.

bychom měli provádět ve dvou fázích – nejprve extrakcí s následným porovnáním rysů. Kroky 1–4 mohou představovat fázi extrakce. Ze stručného přehledu výše je patrné, že zmíněné kroky nepředstavují větší komplikaci.

Mnohem problematičtější je však krok poslední, tedy porovnání rysů. Podle Nolana bychom každého porovnávaného mluvčího měli znázornit jedním bodem ve vícerozměrném prostoru tak, že hodnoty získané analýzou každého parametru použijeme jako souřadnice. Na základě těchto znázornění bychom následně měli využít metriku vzdálenosti, díky níž dokážeme odhadnout vzdálenost mezi mluvčími. Tato metoda se však využívá pouze zřídka, a i když se pro ni rozhodneme, neexistuje konsenzus jak přesně postupovat.

Jiným způsobem, jak shrnout výsledky, je prostřednictvím věrohodnostního poměru, což je však v současné době poměrně neobvyklé. Jedním z důvodů je, že pokud máme použít všechny nebo alespoň několik parametrů popsaných v oddílu 3.4, jednoduše nemáme k dispozici dostatek údajů o distribuci hodnot příslušných parametrů v relevantní populaci. Někteří forenzní fonetikové jsou proto názoru, že věrohodnostní poměr není v současnosti reálně využitelný. Coby alternativa se jeví metoda navrhaná ve stanovisku Spojeného království [19]. V čem se liší uvidíme v oddílu 3.8.3.

Poslechově-akustický přístup je do značné míry subjektivní, a výsledky analýz se tak mohou u jednotlivých odborníků lišit. Nemusí nutně znamenat, že jeho závěry nejsou správné. Na druhou stranu ale není jasné, jak lze jejich správnost měřit a porovnávat, což je dle mého názoru zásadní slabinou tohoto přístupu.

Většina odkazů v oddílu 3.4, v němž si přiblížíme silné i slabé stránky nejružnějších parametrů uplatňovaných ve forenzní praxi, se vztahuje ke studiím a případům z praxe, kde byl využit právě přístup poslechově-akustický, což odráží jeho dominantní (alespoň doposud) postavení.

Zájemci o hlubší sondu do zmíněného přístupu mají k dispozici řadu textů a učebnic, z nichž mohou čerpat. Solidní základ nabízí například následující tituly: [3, 23, 42, 49].

3.3.2 Automatické metody

Výraz „automatické“ v „automatické rozpoznávání mluvčího“ (ASR) přísně vzato neznamená nic jiného, než že rekognici neprovádíme my, manuálně, ale necháme ji na počítači (či na jiném přístroji). Plně či poloautomatickými metodami by hypoteticky bylo

možné provést analýzu všech parametrů popsaných v oddílu 3.4, i kdyby mělo jít pouze o to ušetřit čas.

V současnosti je však namístě systémy automatického rozpoznávání mluvího označovat jako systémy automatického rozpoznávání hlasu či – jak je nazýváme my – systémy automatického srovnávání hlasů, neboť rekognice se vztahuje spíše k jedincům než ke vzorkům řeči, s nimiž tyto systémy pracují. Výraz „mluvčí“ je taktéž zavádějící, a to ze stejného důvodu. Proces rozpoznávání se totiž přísně vzato netýká mluvího-jedince. K tomu by totiž bylo zapotřebí množství dalších důkazů (jako otisky prstů, DNA, výpověď očitého svědka apod.), jež forenzní fonetik nemá k dispozici. „Mluví“ se začal používat proto, že existuje také „rozpoznávání řeči“, týkající se něčeho zcela jiného.

Je třeba mít na paměti, že většina systémů ASR, které dnes známe, neporovnává řeč, ale hlasy, přičemž rozhodně nejde o totéž. Rozličné techniky uplatňované v ASR jsou popsány v jiných kapitolách této publikace; v tuto chvíli stačí říci pouze tolik, že nejrozšířenější ASR systémy dneška analyzují hlas, nebo přesněji to, co nazýváme hlasový tónbr. K tomu využívají mel-frekvenční keprální koeficienty (MFCC), které vycházejí z výzkumu lidské percepce. Pokud nám tedy jde primárně o srovnání *hlasů* (v tomto konkrétním smyslu), ASR je správnou volbou. Jak jsme však viděli výše, řeč zahrnuje mnohem více než hlas. Mluvího můžeme charakterizovat na základě způsobu mluvy, dialektu, tempa řeči, intonačních kontur, slovního přízvuku, ale i individuálnějších znaků jako přechů, zvláštního výběru slov, nesprávné výslovnosti apod.

Probíhají snahy začlenit některé z těchto faktorů do ASR, ale pokud je nám známo, žádný podobný systém se v současné forenzní praxi nepoužívá. Doposud se nám jeví výhodné rozlišovat mezi hlasem a řečí. Naše chápání „hlasu“ je inspirováno modulační teorií řeči, již se budeme věnovat později. Částečně se liší od toho, jak „hlas“ obvykle nazýváme, ale tónbr je zde důležitým faktorem.

V souvislosti s forenzní fonetikou bychom neměli zapomínat na jeden zvláště důležitý aspekt, jímž je skutečnost, že identifikace jedince, v našem případě mluvího, by vyžadovala metodu, pomocí níž by jej bylo možné vyčlenit z otevřené množiny mluvích. To je v některých oblastech hypoteticky možné. Co se týče DNA a otisků prstů, bylo by možné – alespoň teoreticky – vytvořit kompletní databázi populace. Za předpokladu, že bychom k takové databázi měli přístup a použité analytické postupy by byly dostatečně spolehlivé, o identifikaci jedince (na základě shody s příslušným vzorkem) by se reálně dalo uvažovat.

Je něco podobného proveditelné i v oblasti řeči? Hypoteticky bychom pro ni mohli sestavit populační databázi, ale mějme na paměti, že na rozdíl od databáze DNA či otisků prstů by taková databáze rychle zastarávala. Zatímco se totiž u dotyčného jedince DNA či otisky prstů v průběhu času nemění (s výjimkou případu úrazu rukou, kdy se otisky mohou poškodit), lidský hlas a řeč se kupříkladu vlivem stárnutí, nemoci, kouření či úrazu mění. Hlasové charakteristiky se však mohou změnit i v krátkém časovém horizontu, kupříkladu v důsledku nachlazení, emočního stavu mluvčího, vlivem drog či okolních podmínek jako šumu v pozadí.

Závěrem tedy podotkněme, že výraz „automatická rekognice mluvčího“ patří spíše do oblasti science-fiction. Zaměřit bychom se měli spíše na srovnávání mluvčích, které se dnes v podstatě omezuje jen na srovnávání hlasů. Přílehavějším označením by tedy bylo automatické porovnávání hlasů (AVC).

Známým problémem automatických systémů je tzv. nesoulad v okolnostech vzniku, například u zvukové kvality mezi vzorky řeči, a to kvůli rozdílům v přenosovém kanálu. Jestliže byl vzorek řeči získán z telefonního hovoru, ale referenční databáze obsahuje audiozáznamy pořízené přímo, pak systém nepracuje tak dobře jako v případě, že referenční data byla rovněž získána prostřednictvím telefonu. Jednodušší případy variability jsou do jisté míry řešitelné, ale odchylky v kvalitě kanálu jsou v reálné forenzněfonetické praxi mnohem výraznější.⁵⁸ Co když se nám kupříkladu dostane do rukou audiozáznam křičícího bankovního lupiče pořízený pomocí nekvalitního mikrofону ze vzdálenosti tří metrů, nadto z místnosti se silnou ozvěnou (k čemuž dochází poměrně často)? V reálné forenzní praxi se setkáváme s nespočtem obdobných případů a některé z nich znamenají velký problém jak pro poslechovou, tak i akustickou analýzu.

System ASR, který používáme v naší praxi na Univerzitě v Göteborgu pro porovnání hlasů, jsou sady francouzského ALIZE SpkDet, vyvinuté a vydané jako open source software⁵⁹ pod licencí LGPL [4–5]. Výsledky srovnání hlasů se následně spojí s tradiční poslechově-akustickou analýzou a kombinované výsledky jsou následně vyjádřeny pomocí

⁵⁸ V původním znění: „Now simple cases of variation like that can to some extent be taken care of, but the in forensic phonetic case work variation in channel quality is much wider.“ Za určitým členem na začátku druhé věty následuje předložkové spojení („in forensic phonetic case work“). Možným vysvětlením je, že na tomto místě byla omylem vynechána nominální fráze při současném zachování původního členu; pravděpodobnější však je, že došlo k pouhému „prohození pozic“, a druhá věta souvětí měla znít: „but in the forensic phonetic case work variation in channel quality is much wider“.

⁵⁹ Neboli software s volnou licencí (otevřeným zdrojovým kódem).

9bodové slovní hodnotící stupnice⁶⁰ popsané později. Významným vylepšením by bylo zkombinovat výsledky v souladu s bayesovským věrohodnostním poměrem; v současné době však není jasné, jak by jej bylo možné (do procesu) implementovat.

Nejčastěji používaným, event. testovaným systémem v policejních laboratořích je dle dostupných informací španělský *Batvox* od společnosti *Agnitio*. Dosud byl testován policií ve Švédsku a několika dalších zemích. Informace o tom, do jaké míry se uplatňuje ve forenzní praxi, však nejsou k dispozici. *Batvox* má trvale dobré výsledky a nemáme pochyb, že jde o velmi kvalitně zpracovaný systém. Co nás však znepokojuje, je jeho všeobecně známé snadné ovládání, což se může zdát jako rozporuplné tvrzení, ale není tomu tak. V souladu s Daubertovým standardem (viz s. 10, [41]), který říká, že by výpověď znalce měla vycházet z „vědeckých poznatků“, existuje u každého systému příliš nenáročného na obsluhu riziko, že s ním může po základním zaškolení pracovat i osoba neznalá principů jeho fungování, což by rozporovalo s požadavkem na erudici (nejen forezního) odborníka. Znalec se specializací na fonoskopii by měl všechny prostředky, s nimiž pracuje, znát tzv. „do puntíku“ a používat je ryze za účelem ušetřit čas, a nemuset tak celý proces absolvovat pouze s papírem a tužkou. Nemám jediný důkaz, že by byl *Batvox* či kterýkoli jiný obdobný systém zneužíván. Pouze bych tímto rád vyzval k obezřetnosti před možným rizikem.

3.4 Přehled klíčových parametrů ve fonoskopii

V následující části upustíme od rozlišování poslechově-akustických a automatických metod analýzy. Ačkoli je pravda, že větší část automatických metod dosud vycházela z analýzy hlasu, a to zejména s využitím mel-kepstrálních frekvenčních koeficientů (MFCC), není důvod, proč by jak automatické, tak i poloautomatické metody nemohly najít uplatnění i v jiných typech analýz. Poslechově-akustický přístup pak může podobným způsobem být obohacen o oblast hlasové analýzy. Několik studií, na něž odkazujeme níže, pracuje s konceptem věrohodnostního poměru, což hraje významnou roli v případě, že chceme důkazní materiál předložit u soudu – neboť nabízí standardizovaný způsob, jak prezentovat výsledky, a to bez ohledu na typ (použité) analýzy. Výsledky poslechově-akustického přístupu, do značné míry založeného na subjektivním vjemu, naopak podobně standardizovaným způsobem prezentovat nelze. Blíže do problematiky nahlédneme v jiném oddílu, v souvislosti s předkládáním důkazního materiálu u soudu, novými trendy

⁶⁰ Ve VT *verbal 9-point scale*.

ve forenzně-právním kontextu a tím, co někteří nazývají zásadním obratem v oblasti fonoskopie [38, 39].

Jednou ze slabin mnoha běžně používaných akustických parametrů je, že nejsou dostatečně robustní vůči faktorům jako je styl mluvy, kvalita záznamu, šířka pásma kanálu ad. Dále v textu se budeme zabývat nejen krátkým popisem možných komplikací, s nimiž můžeme přijít do kontaktu při srovnávání mluvčích (v souvislosti s uvedenými parametry a ukazateli), ale i jak je řešit.

Následující výklad se zaměří na popis dvou tradičně využívaných parametrů, a to základní frekvence a formantů. Jak za moment uvidíme, v důsledku intraindividuální variability a mnoha dalších faktorů nejsou tyto parametry plně dostačující, a začínají je proto pozvolna vytlačovat přístupy sofistikovanější. Vzhledem k tomu, že však zastávaly význačné postavení v historickém vývoji fonoskopie, a nadto jsou dodnes frekventovaně využívány, zasluhují přinejmenším zmínku. Tím, že si je přiblížíme, si navíc můžeme připravit půdu pro výklad důmyslnějších přístupů.

3.4.1 Základní frekvence (F0)

Základní frekvenci můžeme definovat jako frekvenci hlasivkových kmitů při fonaci. Její užitečnost pro srovnávání mluvčích ve forenzní praxi je již dlouho rozporuplná. Pro popis variability f_0 se často doporučuje analýza dlouhodobé distribuce, konkrétně s ohledem na aritmetický průměr či směrodatnou odchylku [49]. Střední hodnoty i směrodatné odchylky jsou do jisté míry odvislé od délky trvání vzorku řeči, neexistuje však všeobecná shoda, jak dlouhé by tyto vzorky měly minimálně být, aby přinesly věrohodné výsledky. Horii [24] se domnívá, že by měly přesahovat hranici 14 s; podle jiných studií by měly být dlouhé alespoň 60 s [42] až 2 min [3]. Rose [49] uvádí, že hodnoty F_0 „[...] u sedmi jedinců hovořících čínským dialektem se ustálily mnohem dříve než po 6 vteřinách“. Totéž jako se dočteme v [42], což znamená, že minimální požadovaná délka se může jazyk od jazyku lišit. Braun [7] zdůrazňuje, že naměřené hodnoty rovněž závisí na fyziologických faktorech, jako je věk, kouření, nemoc⁶¹, intoxikace, ale i faktorech psychologických, jako je citové rozpoložení mluvčího, ospalost nebo zvýšené mluvné úsilí. Přesto je názoru, že „můžeme-li komunikační chování považovat za ‚normální‘“, 15–20 s by mělo být dostačovat. Taktéž zmiňuje vliv okolního hluku na způsob mluvy. Bez ohledu na uvedené zdroje variability se však základní

⁶¹ V orig. *decease* (= smrt); patrně jde o chybu způsobenou hláskovou podobností výrazů *decease* a *disease* (= nemoc).

frekvence při vyšetřování coby užitečný forenzněfonetický parametr již častokrát osvědčila [2, 3, 22, 29, 42, 49]. Dalším faktorem, který na F0 může působit, je maskování hlasu. Jedna vcelku obvyklá maskovací strategie zahrnuje zvýšení či snížení F0, což koresponduje i s výsledky studie [31], kde mělo 100 subjektů za úkol přečíst text pomocí maskování dle vlastního výběru, přičemž mluvčí velmi často volili právě zvýšení či snížení F0. U mluvčích s nižší F0 (vzhledem k průměru) byla nadto patrná tendence ještě více ji prohloubit a u mluvčích s nadprůměrně vysokou F0 ji naopak zvyšovat. Je zřejmé, že v důsledku toho v zásadě postrádá srovnávání základních frekvencí ve forenzní praxi smysl.

Pro mluvčí je typické kladné sešikmení distribuce F0 (Jassem et al. 1973), což znamená, že hodnoty f0 jsou asymetricky rozloženy kolem střední hodnoty. Tradiční prostředky pro popis hladiny F0, jako jsou průměr či medián, proto mohou vést ke klamným závěrům. Kladné sešikmení je způsobeno zejména tím, že je více prostoru pro variabilitu f0 směrem vzhůru než směrem dolů⁶². Jestliže budeme základní frekvenci měřit v pultónech, docílíme pouze snížení sešikmení, ne jeho úplné eliminace.

Pokud má být parametr F0 odolný vůči některým zdrojům variability (jak jsme pozorovali výše), je třeba hledat alternativní způsob, jak jej vyjádřit. Jednou z možností, kterou používáme my a která dokáže lépe zachytit individuální hladinu F0 mluvčího, je tzv. základní hladina⁶³. Její výhodou je mimo jiné to, že je o poznání robustnější než střední hodnota či medián, a podle studie [35] téměř vůbec nepodléhá změnám v kvalitě kanálu či emotivitě projevu. Detailněji se konceptem základní hladiny budeme zabývat v oddílu 3.6.

Veškeré faktory uvedené výše souvisí s intraindividuální variabilitou, ale ve forenzní praxi se užitečnost toho či onoho parametru odvíjí také od interindividuální variability. I v případě, že se nám u daného mluvčího podaří poměrně přesně určit jeho základní frekvenci, její prediktivní schopnost ještě nemusí být příliš vysoká, jestliže zjištěná hodnota F0 spadá do oblasti, kam spadá i většina dalších mluvčích. Jako možný způsob, jak vyjádřit intraindividuální vs. interindividuální variabilitu, se nabízí tzv. *F-poměr* [5, 25, 28]. Pomocí věrohodnostního poměru, který si osvětlíme později, můžeme zase údaje obou druhů variability sloučit. Pokud bychom měli přístup k dostatečně obsáhlým a vhodným databázím,

⁶² Hlas typického mluvčího muže, který čte větu neutrálním tónem, může mít průměrnou F0 110 Hz, minimální hodnotu 85 Hz a maximální hodnotu 160 Hz. Při energičtějším/živějším čtení zůstává minimální hodnota zhruba totožná, zatímco maximální může přesáhnout i 200 Hz.

⁶³ Podle Skarnitzl s. 55.

dokázali bychom poměrně snadno posoudit, nakolik jsou v daném případě údaje o základní frekvenci k užítku.

3.4.2 Formanty

Formanty lze vymezit jako rezonanční frekvence ve vokálním traktu, odvislé od tvaru a objemu různých dutin traktu. Z tohoto krátkého popisu by mohl leckdo vyvodit závěr, že se frekvence formantů neustále mění podle toho, jak při řeči pohybujeme spodní čelistí, jazykem a rty. Přesto u formantových frekvencí pozorujeme určité pravidelnosti. Za normálních mluvních podmínek jsou pro danou osobu, mluvní tempo a vokál poměrně neměnné. Šva⁶⁴, samohláska tvořená v neutrálním a uvolněném nastavení mluvních orgánů, navíc koreluje s délkou vokálního traktu a uplatňuje se též jako parametr hlasového tónu (ve standardních podmínkách). Traunmüller [58] v souvislosti s tím hovoří o nosné vlně, kterou při řeči modulujeme, díky čemuž dokážeme vyjádřit nejen jazyková sdělení, ale i nejrůznější informace paralingvistické povahy. Pro zevrubnější popis viz oddíl 3.6.

Přestože jsou frekvence formantů při artikulaci hlásek proměnlivé, závisí do určité míry na velikosti a proporcích mluvních orgánů jednotlivce. Kdyby se nám podařilo určit, co má spojitost s „neutrální“ velikostí a proporcemi vokálního traktu na straně jedné, a modulacemi způsobenými artikulací na straně druhé, formantové frekvence by v zásadě mohly sloužit pro identifikaci mluvčího. To však kvůli četným komplikacím, jež na ně působí, nebývá snadné. Jednou z možných komplikací je tempo řeči. Imaizumi a Kiritani (1989) prokázali, že hodnoty dvou nejvyšších formantů, s nimiž se v praxi ještě běžně počítá (F3 a F4), „s měnícím se tempem značně kolísají“.

Taktéž bylo zjištěno, že ke změnám formantových frekvencí dochází i vlivem zvýšeného mluvního úsilí [61]. Při hlasité řeči nebo křiku, což je dle dostupných audiozáznamů mimochodem nejčastější mluvní způsob u bankovních loupeží, můžeme pozorovat nárůst u prvního formantu (F1) [14]. Referenční nahrávky v případech bankovních loupeží však obvykle pochází z výslechů, kde podezřelý hovoří uvolněným tónem hlasu, což jakoukoli akustickou komparaci podstatně znesnadňuje. Ačkoli do určité míry dokážeme odhadnout, jak se frekvence formantů vlivem zvýšeného mluvního úsilí zachovají, stále víme příliš málo na to, abychom se odvážili vyslovit jakékoli konkrétnější prognózy.

⁶⁴ Neutrální vokál, v češtině se objevuje i v zápisu *schwa* (např. Skarnitzl et al., 2014, s. 26).

Dalším faktorem, o němž je známo, že působí na formantové frekvence, je šířka pásma, jež souvisí s telefonními nahrávkami. Výzkumem jejího vlivu se zabývalo hned několik studií, které zahrnovaly poslechovou identifikaci hlasu. První studie tohoto druhu pracovaly s pouhou simulací telefonní kvality zvuku, konkrétně například s filtry pásmové propustnosti [53]. V jiných studiích (např. Rathborn et al. 1981) se již uplatnily vzorky řeči získané pomocí pevné linky⁶⁵. Obecně však lze říci, že významnější vliv telefonní kvality zvuku na přesnost poslechové identifikace hlasu nebyl prokázán. V akustických analýzách byl však její vliv patrný; mimo jiné v souvislosti s výrazným dopadem na první formant [34]; příkladem budiž studie s poněkud poplašným názvem „Pozor na vliv telefonního přenosu“ [32], jež poukazuje na silný vliv telefonem přenášené řeči právě na zmíněný F1. Sporné nahrávky pochází mnohdy z telefonních odposlechů, zatímco referenční vzorky z přímo nahrávaných policejních výslechů. Künzel zastává poměrně striktně stanovisko, že jakoukoli analýzu formantů bychom v takových případech vůbec neměli provádět. Ačkoli ne všichni tento názor sdílí, obecná shoda panuje v tom, že pokud se jedná o kombinaci vzorků nahraných přímo a prostřednictvím telefonu, při formantové analýze by se mělo postupovat se zvláštní obezřetností.

Většina telefonních odposlechů, které jsou nějakým způsobem spjaté s kriminální činností, v současnosti pochází z mobilních telefonů, což dokládají i zprávy forenzních vyšetřovatelů ze Švédska, Spojeného království a Německa, které navíc poukazují na rostoucí tendenci jejich využití (právě) v této souvislosti [43]. Podrobný a odborný popis nejrůznějších způsobů, jakými přenos prostřednictvím mobilního telefonu ovlivňuje kvalitu zvuku, nabízí [21]. Tento přenos ale kromě omezené šířky pásma přináší mnoho dalších komplikací s nepříznivým dopadem na zvukovou kvalitu a přesnost formantové analýzy. Byrne a Foulkes [8] uvádí, že přenos mobilním telefonem má za následek významný, „29% nárůst hodnot prvního formantu oproti přímému nahrávání“, a jiná studie [21] prokázala, že ovlivněny byly dokonce první tři formanty, za určitých okolností poměrně zásadně.

Na formantové frekvence mohou mít vliv i některé typy maskování. Analýza profesionálních i amatérských imitátorů (viz Papcun [44]) odhalila, že jsou schopni se, alespoň do určité míry, přiblížit formantovým hodnotám cílových mluvčích. K podobným závěrům dochází i další studie [15, 17]. Je však namístě podotknout, že v reálné forenzní praxi se s imitacemi příliš často nesetkáme; přesto bychom neměli zapomínat na skutečnost, že snaha „znít“ jako

⁶⁵ Ve VT *landline phone recorded speech samples*.

jiný mluvčí mívá za následek posun formantových hodnot dále od hodnot typických pro daného mluvčího.

Abychom zrekapitulovali nejdůležitější poznatky: ačkoli jsou hodnoty formantů mimo jiné dány fyziologickými vlastnostmi vokálního traktu, a tedy jsou pro jedince do určité míry unikátní, analýzu formantů komplikuje celá řada faktorů, které na hladinu formantových frekvencí mohou mít značný vliv, a tím významně zpochybňují výsledky formantové analýzy ve forenzním prostředí.

3.4.3 Rytmus a jiná schémata v temporální oblasti

Ve srovnávání mluvčích by se mohly uplatnit i parametry obvykle spojované s rytmem řeči, ale i další parametry reflektující motorickou činnost spjatou s artikulací. Prozatím se jedná o dostatečně neprobádanou oblast, a není tedy namístě činit předběžné závěry, nakořik budou ve sféře srovnávání mluvčích využitelné. První prognózy vypadají přesto poměrně optimisticky. Začneme-li poměrně zeširoka, bylo zjištěno, že pro každého jedince jsou příznačné určité vysoce automatizované motorické činnosti, které se časem stávají poměrně stabilní a neměnné, což bylo prokázáno například u charakteru chůze [6, 65] či psaní na psacím stroji nebo počítači [57]. Pokud by se skutečně ukázalo, že jsou dostatečně stabilní a interindividuální variabilita uspokojivě bohatá, neměl by být problém využít je pro identifikaci osob. To je ostatně i případ chůze (viz kupříkladu dva odkazy výše). Dalším příkladem vysoce automatických motorických činností je artikulace. Pokud by se ukázalo, že je taktéž do jisté míry individuální, ve sféře srovnávání mluvčích by jistě byla přínosem. Jedna studie zaměřená na imitaci [17] například odhalila, že ačkoli se zkoumanému profesionálnímu imitátorovi podařilo napodobit cílové hodnoty průměrné základní frekvence, rytmyzace⁶⁶, globálního tempa řeči a do určité míry také formantových frekvencí, rytmyzace na segmentální rovině měla obecně blíže jeho vlastním řečovým návykům než mluvě cílového mluvčího. Porovnávala se tři ztvárnění totožného textu, a to verze cílového mluvčího a dvě verze imitátora – jedna s imitací a druhá bez (tzn. s přirozeným, neutrálním hlasem imitátora). Všechny tři verze byly transkribovány na segmentální úrovni a u jednotlivých hlásek byl vyznačen jejich nástup, čímž vznikly tři časové sekvence (cíl, imitace, přirozená mluva) pro každou repliku. Odchyly v rytmyzaci u imitace a přirozené verze v porovnání s cílovou nahrávkou byly následně hlásku po hláске zaneseny do grafu

⁶⁶ Ve VT *timing*; na segmentální úrovni taktéž *načasování* (artikulačních fází), u vyšších artikulačních celků *frázování* (Pořízka, 2019, osobní komunikace).

pro celé promluvy, vybrané úseky a vybraná slova. Zatímco byly v obou provedeních imitátora patrné pouze minimální rozdíly, v porovnání s cílovou nahrávkou se jednalo o rozdíly poměrně markantní. Jinými slovy – ačkoli imitátor zvládl napodobit cílového mluvčího v jiných testovaných parametrech, vlastní frázovací návyky nijak znatelně modifikovat nedokázal. Za dvěma studii, které se nesly v podobném duchu, avšak se zaměřením na dynamické vlastnosti formantů, stojí McDougall [36, 37], která dochází k názoru, že při srovnávání mluvčích lze uplatnit i individuální schémata načasování vybraných pohybů při artikulaci. V minulosti proběhlo několik pokusů popsat rytmický charakter jazyků [20, 48]. V experimentech zaměřených na srovnávání mluvčích bylo testováno hned několik parametrů spojených s délkou trvání, které se používají pro klasifikaci jazyků, jako je index párové variability (PVI) či %V (procentuální zastoupení vokálních intervalů v celkovém trvání) [10, 64, 66]. Dellwo a Koreman [11] ve své studii analyzovali nahrávky 10 německy hovořících subjektů, kteří měli pětkrát přečíst krátký text, pokaždé v jiném tempu od normálního až po „nejvyšší možné“, přičemž testované proměnné zahrnovaly %V a PVI (které jsme zmínili výše) a ΔC (směrodatnou odchylku trvání konsonantických intervalů). Všechny zůstaly stabilní bez ohledu na komunikační situaci, a to jak u daného mluvčího, tak s ohledem na vícero mluvčích. Taktéž bychom měli zmínit experiment [12] zabývající se maskováním, v němž subjekt – rodilý mluvčí anglického jazyka – přečetl 29 vět jednak svým přirozeným hlasem, tak hlasem maskovaným ve formě imitace dialektu. Navzdory tomu, že byl v jeho maskované řeči patrný nárůst průměrné základní frekvence, směrodatné odchylky a (hlasové) dyšnosti, velká část testovaných temporálních ukazatelů „nevykazovala mezi přirozeným a maskovaným hlasem žádné změny“. Na jakékoli prognózy, které by souvisely s potenciálním uplatněním v reálné forenzní praxi, je zatím příliš brzy, ale průběžné výsledky jsou více než slibné.

3.4.4 Řečové tempo

Řečové tempo se vztahuje k počtu řečových jednotek za minutu nebo za sekundu. Nejčastěji jej měříme ve *slovech za minutu* (slov/m⁶⁷) či, pokud cílíme na vyšší preciznost, raději ve *slabikách za sekundu* (slab/s). Kromě toho není zcela jasné, jak přistupovat k problematice pauz. Bavíme-li se například o tempu řeči, počítáme i s pauzami. Ve většině případů je však výhodné od nich tempo skutečně pronášené řeči od pauz oddělit. V takovém případě hovoříme o *artikulačním tempu*, jehož typické hodnoty se pohybují kolem 5–7 slabik za

⁶⁷ Ve VT *wpm*; nabízí se ekvivalent *s/m*, což však může být matoucí (mohlo by být např. zaměněno za *slabik za minutu*).

sekundu. Slabičné tempo měříme od jednoho slabičného onsetu⁶⁸ ke druhému, popřípadě od onsetu jednoho vokálu ke druhému. U nahrávek v přiměřeně dobré kvalitě lze slabičné počátky detekovat i automaticky, avšak úměrně zhoršující se kvalitě nahrávky klesá i přesnost detekce. V případě nekvalitních audiozáznamů tak zpravidla není jiná možnost, než detekci onsetů provést manuálně. Lze říci, že tempo řeči odráží způsob mluvy, zatímco tempo artikulace motorické načasování v podobném smyslu, jako jsme mohli vidět v předcházejícím oddílu. Na základě toho, co dnes víme o stabilitě motorického načasování, je možné předpokládat, že by i artikulační tempo mohlo do jisté míry mít individuální charakter, neboť vykazuje pouze mírnou intraindividuální variabilitu. Tím se řadí mezi adepty ke zkoumání s potenciálním využitím ve forenzní praxi. Tuto domněnku již potvrdily závěry několika málo studií. Künzel [30] například analyzoval různé aspekty tempa řeči a jeho variabilitu mezi čtenou a spontánně tvořenou řečí. Ukázalo se, že způsob mluvy neměl na artikulační tempo „téměř žádný vliv“ a samotné tempo artikulace bylo „u jednotlivců pozoruhodně konstantní“. I testovaná rozlišovací schopnost tempa artikulace (při zohlednění jak intraindividuální, tak interindividuální variability) obstála poměrně dobře. U mluvního tempa, které způsob mluvy reflektuje, naproti tomu byla – ne překvapivě – patrná vyšší intraindividuální variabilita a o poznání menší rozlišovací schopnost.

3.4.5 Barva hlasu

Přínosná pro srovnávání mluvcích by mohla být i barva hlasu. Až doposud probíhal výzkum tohoto parametru zpravidla ve spojitosti s patologií řeči, ačkoli může vyjít najevo, že některé jevy s ní spojené by byly přínosem právě i pro oblast srovnávání mluvcích – to však prozatím nelze dostatečně podložit výzkumem. Ve sféře řečové patologie se při popisu barvy hlasu nejčastěji užívají termíny třas a tremolo, přičemž třas značí kolísání základní frekvence mezi jednotlivými periodami a tremolo související změny v amplitudě. Ačkoli jejich možné využití pro srovnávání mluvcích doposud nebylo řádně (pro)zkoumáno, několik málo studií usiluje o nápravu. Wagner [63] kupříkladu dokázal podle třasu rozlišit vzorek mluvcích s hlasovou patologií od mluvcích s normálním hlasem. Farrús a Hernando [18] zkoumali kromě třasu i tremolo a doplnili jimi faktory použité v systému automatické verifikace⁶⁹ mluvcího. Třebaže třas a tremolo samy o sobě nevykazovaly „dostatečně dobré“ (*sic*)

⁶⁸ Neboli *iniciála*.

⁶⁹ Neboli *ověřování/ověření*.

výsledky, jakmile byly implementovány do systému založeného na prozodických a spektrálních parametrech, jeho úspěšnost se zvýšila.

Další možné, byť dosud netestované, deskriptory hlasu čítají mimo jiné tvar hlasivkového pulzu a spektrum hlasivkového zdroje. Neměli bychom opomenout ani třepenou fonaci, o níž můžeme obecně říci, že jde o perceptuální kategorií a že má několikero možných příčin. Pakliže subglotální tlak klesne pod určitou hodnotu, obvykle ke konci fráze, dochází k prudkému snížení základní frekvence na polovinu, což může mít za následek i tzv. diplofonii⁷⁰ (pravidelné kolísání v amplitudě mezi jednotlivými pulzy, přičemž každý následující puls výrazně ztrácí na intenzitě). Míra třepené fonace se u každého liší, ale dosud neproběhl žádný výzkum, který by odpověděl na otázku, zda je její interindividuální variabilita dostatečně velká na to, abychom mohli uvažovat o jejím uplatnění ve forenzní praxi. Většina jiných deskriptorů vychází primárně z poslechové analýzy, a zákonitě jsou tak do značné míry subjektivní. Zmínit můžeme například případ, kdy hlasivky plně nedovírají a propouští proud vzduchu (tzv. *leaky voice*), dále drsnou fonaci (napjaté hlasivky), šepot (neúplná znělost⁷¹ jindy běžně znělých segmentů), dysfonii⁷² (nepravidelné zavírání) a falzet (základní frekvence se zvýší natolik, že má za následek prudký pokles glotálního spektra). Pokud je mi však známo, žádný z uvedených jevů nebyl dosud testován s ohledem na možnou aplikaci ve forenzním prostředí.

3.4.6 Jazykové faktory

V neposlední řadě je třeba poukázat na lingvistické faktory, které čítají dialekt, přízvuk, idiosynkratické výrazy, neobvyklé chyby ve výslovnosti ad. Ve forenzní praxi není ojedinělé, aby právě ony hrály v zásadě rozhodující roli⁷³.

3.5 Maskování

Audiozáznamy maskovaného hlasu ve forenzním prostředí nejsou příliš časté, ale setkat se s nimi můžeme. Jak jsme již nastínili výše, jednou z poměrně obvyklých maskovacích strategií je změna základní frekvence, což mimo jiné koresponduje se závěry studie [31] zmíněné dříve. Künzel dále uvádí, že další nejčastěji uplatňované druhy maskování (v případech, jimiž se zabývá německá federální policie, BKA) jsou falzet, třepená fonace,

⁷⁰ Neboli *dvojhlási*.

⁷¹ Neboli *hlasnost* či *sonorita*.

⁷² Ve VT *hoarse voice*.

⁷³ Jiné kolokace v angličtině a v češtině; ve VT „to carry a rather decisive weight“.

šepot, imitace cizineckého přízvuku a mluvení s ucpaným nosem⁷⁴. Přestože jsou všechny uvedené povahou vcelku primitivní, mohou proces srovnávání mluvcích výrazně zkomplikovat. Pro více podrobností o různých typech maskování a jejich vlivu na identifikaci mluvcích viz [16].

Jako další forma maskování se občas používá tlačaná fonace. Na provedení není příliš složitá, a zároveň významně snižuje úspěšnost procesu srovnávání mluvcích. Negativní vliv má zejména na automatické metody. Pracovali jsme na případě únosu, kdy s ní pachatel hovořil při svých telefonátech, v nichž žádal o výkupné. Referenční nahrávky tvořily dva telefonáty podezřelého jisté známé, které uskutečnil prostřednictvím mobilního telefonu. V případě bylo zjištěno několik podobností ve výrazivu a mluvním způsobu, které se shodovaly s informacemi o podezřelém a na základě nichž jsme pojali důvodné podezření, že hlas mluvcího z referenčních nahrávek a hlas z telefonátů, v nichž mluvčí žádal o výkupné, patří též osobě. Automatická analýza přesto nepřinesla jasné závěry, které by měly šanci obstát u soudu. Na tomto příkladu lze ilustrovat, jak i triviální forma maskování dokáže efektivně eliminovat možnost správného srovnání mluvcích.

Srovnávání mluvcích hovořících maskovaným hlasem lze provádět ve čtyřech krocích, 1) identifikujeme daný druh maskování, 2) aplikujeme dosavadní poznatky o vlivu té či oné maskovací strategie na oblast akustiky, 3) ze vzorku řeči se snažíme co nejlépe odstranit deformaci vzniklou maskováním, 4) na základě „čistého“ audiozáznamu přikročíme ke srovnání mluvcích. V současnosti probíhá výzkum [45–46], který se touto problematikou zabývá, ale zatím není jasné, k jakým závěrům dospěje.

3.6 Modulační teorie řeči

V naší praxi využíváme dva koncepty vycházející z modulační teorie řeči [58], konkrétně definici „hlasu“ inspirovanou konceptem nosné pozice⁷⁵ a parametr základní frekvence, který nazýváme „základní hladina“. Z toho důvodu je namístě tuto teorii v krátkosti přiblížit a vysvětlit proč by ve výkladu forenzního srovnávání mluvcích neměla chybět.

Jak si můžeme povšimnout, v tabulce 3.1 není zmínka o „hlasovém zdroji“. Je to z toho důvodu, že v rámci modulační teorie není hlasový zdroj samostatným parametrem; místo něj užíváme parametr jiný, který mu do jisté míry odpovídá – nosnou pozici. Jde o klíčový pojem, související (ale nikoli totožný) s konceptem nosné vlny. Podle modulační teorie jsou

⁷⁴ Ve VT *pinching one's nose*; odborným pojmem by byla denazalizace.

⁷⁵ Podle: Skarnitzl a Hývlová (2014, s. 54).

veškeré informace přenášené v řečové komunikaci určitou modulací nosné pozice, odkud pramení i její název. Jak ale můžeme nosnou pozici definovat? Když v roce 1994 Traunmüller svou teorií poprvé představil, napsal: „o nosné pozici bychom měli uvažovat tak, že má vlastnosti charakteristické pro ‚neutrální‘ vokál, přibližně [ə]“.

Význam nosné pozice v současnosti spočívá v tom, že jde o výhodný způsob jak nazírat „hlas“, tj. jako vokál šva, který je tvořen uvolněným způsobem, bez zvýšeného mluvního úsilí stejně jako bez zvýšené základní frekvence. Výhodou je, že jednou z možností, jak vyjádřit akustické vlastnosti hlasu, je tím pádem prostřednictvím MFCC zobrazení nosné pozice. V naší praxi se osvědčilo rozlišovat mezi hlasem a řečí, přičemž hlas definujeme jako nosnou pozici; protože však tento termín není v úzu příliš rozšířen, označujeme jej raději jako „hlas“. Podstatné je, že analýza mel-frekvenčních koeficientů se neomezuje pouze na analýzu neutrálně tvořených vokálů šva, ale zahrnuje širokou škálu hlásek (vyjádřených modulacemi nosné pozice v souladu s pojetím modulační teorie). Konkrétní postup s ohledem na srovnávání hlasů, ve významu srovnávání nosných pozic, nabídneme v oddílu 3.6.1.

Z modulační teorie vychází i koncept neutrální, individuální hladiny základní frekvence. Traunmüller [58] hovoří o „stabilním bodě přibližně $1,5 SD^{76}$ pod průměrnou hodnotou F_0 “. Tuto hladinu se podařilo empiricky odvodit na základě zkoumání chování mluvčích a bylo například prokázáno, že v případě osob, u nichž kolísá živost projevu, funguje jako vlastní vektor základní frekvence. Označujeme ji jako *základní hladinu F_0* a lze ji charakterizovat jako frekvenci, k níž se daný mluvčí navrácí po tom, co realizuje prozodicky či jinak motivovaná vychýlení od ní. Odhad její hodnoty činí dle studie [60] 1,43 směrodatné odchylky pod aritmetickým průměrem. Jinými slovy je přibližně o 7 % výše oproti nejnižším frekvencím, které u daného mluvčího můžeme pozorovat. Pokud mluvčí hlasem dále klesá, obvykle to má za následek třepenou fonaci, již se mluvčí zpravidla vyhýbají. V zásadě ovšem neexistuje odpovídající horní hranice, v důsledku čehož dochází ke zkreslení distribuce směrem k vyšším frekvencím.

⁷⁶ Směrodatné odchylky.

Tabulka 3.1 Schematické znázornění různých vlastností řeči dle modulační teorie řeči (jde o adaptaci tabulky z [59])

Vlastnost	Informace	Jevy
<i>Jazyková</i>		
Konvenční, sociální	Dialekt, přízvuk, mluvní způsob, ...	Slova, hlásky, prozodie
<i>Expresivní</i>		
Psychosociální v rámci variability mluvčích	Emoce, postoj	Typ fonace, registr, mluvnický úsilí, tempo řeči, ...
<i>Organická</i>		
Anatomická mimo rámec variability mluvčích	Věk, pohlaví, patologie, ...	Velikost hrtanu, délka hlasivek, ...
<i>Perspektivní</i>		
Prostorová, přenosová	Místo, orientace, kanál, ...	Vzdálenost, zkreslení kanálu, ...

3.6.1 Více o hlasu vs. nosné pozici

Modulační teorie předpokládá, že lze úplně oddělit nosnou pozici od informace (jak lingvistické, tak paralingvistické), která je přenášena modulací nosné pozice. Jak jsme již uvedli výše, k nosné pozici má poměrně blízko neutrálně tvořený vokál šva, což znamená, že neutrálně tvořený /a/ má k nosné pozici dále. Z toho vyplývá, že mel-kepstrální frekvenční koeficienty, které se uplatňují v systémech automatického srovnávání hlasů, jsou „kontaminovány“ sdělením, konkrétně fonologií daného jazyka (souborem kontrastivních hlásek užívaných v tom či onom jazyce). Pakliže má tedy automatické srovnání hlasů přinést spolehlivé výsledky, znamená to, že musíme pracovat jen a pouze s referenční databází, která přesně odpovídá danému jazyku, dialektu a mluvnickému způsobu ze sporných a srovnávacích nahrávek? V zásadě se jedná o empirickou otázku, kterou můžeme zodpovědět až na základě experimentů, které by porovnávaly různé kombinace zkoumaného materiálu a referenčních databází. Pokud by MFCC vyjadřovaly výhradně nosnou pozici, podobné testy bychom nepotřebovali, alespoň ne ve spojitosti s jazykem. Nosná pozice je dána výlučně fyziologickými vlastnostmi konkrétního mluvčího (velikostí a proporcemi mluvnických orgánů). Přesto bychom měli podotknout, že tyto rysy jsou nezávislé na jazyce. Kdybychom měli k dispozici dokonale spolehlivý systém automatického srovnávání nosných pozic a k tomu sporné a srovnávací nahrávky řeči v němčině, nezáleželo by na tom,

zda by referenční databáze byly v němčině či například ve španělštině. U přístupu využívajícího MFCC by to podstatné bylo, ale dosud neproběhl výzkum, který by přesně určil/stanovil, do jaké míry. Neoficiální testy na vzorcích němčiny porovnávané se španělskou databází ovšem naznačují, že jazyková shoda není tak nezbytná, jak by se mohlo zdát (Künzel, osobní komunikace⁷⁷). Dialekt příslušného jazyka by na výsledky měl mít ještě menší vliv. Budeme-li například vycházet z toho, že mezi mluvčími různých španělských dialektů není zásadnějších anatomických odlišností a že inventář hlásek je mezi jednotlivými dialekty přibližně stejný, potom není příliš podstatné, zda pro referenční databázi použijeme ten či onen dialekt. I tato hypotéza již byla testována. Ze studie autorů Moreno et al. prezentované na konferenci IAFPA v roce 2006 například vyplynulo, že rozdíly mezi třemi testovanými španělskými dialekty byly příliš malé na to, aby měly na výsledky rekognice zásadnější vliv. Ve studii byl využit systém Batvox od společnosti Agnitio, který pracuje s MFCC; „podezřelé“ zde tvořili mluvčí z Andalusie a referenční nahrávky obsahovaly andaluský, kastilský a galicijský dialekt. Výsledky se bez ohledu na referenční dialekt v podstatě nelišily, z čehož může vyvodit – byť bychom měli vzít v potaz, že jde pouze o jedinou studii – následující: ačkoli mel-kepstrální frekvenční koeficienty nereflektují pouze nosnou pozici, ale i fonologii daného dialektu, *odlišnosti* mezi jednotlivými hlasy zjištěné automatickým srovnáním hlasů přibližně odpovídají předpokládaným rozdílům v nosných pozicích mezi mluvčími. Dává proto smysl uvažovat o *interindividuálních odlišnostech*, jež vyplynou z automatického srovnání hlasů, jako o *odlišnostech v nosné pozici*. Přitom však nesmíme zapomínat, že MFCC a nosná pozice nejsou totéž.

3.7 Stručná rekapitulace

Nolan [42] ve své publikaci o fonetickém rozpoznávání mluvčích zastává názor, že na parametry užívané v procesu rekognice by se měla vztahovat následující kritéria:

1. Vysoká variabilita mezi jednotlivými mluvčími
2. Nízká intraindividuální variabilita
3. Odolnost vůči maskování či imitaci
4. Dostupnost
5. Robustnost při přenosu
6. Měřitelnost

⁷⁷ Ve VT zkratka *p.c.*, která nabízí několikero možných interpretací; nejpravděpodobnější v daném kontextu se nám však jeví *personal communication/correspondence*.

Jak jsme mohli vidět v přehledu výše, v praxi obvyklé parametry tato kritéria zpravidla nesplňují, nebo alespoň ne beze zbytku. Kritéria 4 a 6, tj. dostupnost a měřitelnost, obvykle nepředstavují problém, ale u všech parametrů uvedených výše jsou patrné nedostatky s ohledem na kritéria ostatní. Jednou z možností, jak vyjádřit vzájemnou závislost kritérií 1 a 2, je spočítat věrohodnostní poměr (viz apendix B⁷⁸) – to však vyžaduje přístup k vhodným databázím, které jsou k mání pouze zřídka. Navzdory tomuto vážnému nedostatku probíhají pokusy na úrovni elementárního výzkumu, které ukazují možné směřování. Kinoshita [25] kupříkladu uplatnila věrohodnostní poměr v testu rozpoznávání dvojic mluvčích, v němž použila formantové informace získané z malé databáze řečových vzorků 13 mluvčích (mužů hovořících standardní japonštinou). Její experiment zahrnoval 90 dvojic *mluvčí A – mluvčí A* a 180 dvojic *mluvčí A – mluvčí B*. Pět ze 180 dvojic *mluvčí A – mluvčí B* bylo nesprávně klasifikováno jako nahrávky totožného mluvčího a devět z 90 dvojic jako nahrávky odlišných mluvčích. Podobný experiment provedl například i Alderman [1]⁷⁹, který vycházel z hodnot formantů F1, F2 a F3 extrahovaných z nahrávek 11 mluvčích (mužů hovořících australskou angličtinou). Vokály pocházely z prostředí /h_d/, aby bylo možné je porovnat s daty z referenční databáze (tzv. Bernardovými daty⁸⁰). Dvojice nahrávek totožného mluvčího se podařilo identifikovat v 72,2 % případů a dvojice nahrávek odlišných mluvčích v 99,6 % případů. Další experimenty obdobného charakteru jsou popsány v [52] a [54].

V oddílu 3.4 jsme viděli, že pokud se snažíme vyjádřit nějaký parametr na základě jediného ukazatele, např. základní frekvenci podle její průměrné hodnoty, nemůžeme při rozlišování dvou mluvčích očekávat dokonale přesné výsledky. Z jiných vědeckých oborů, např. genetiky, naproti tomu víme, že kombinací několika ukazatelů (které vyjadřují různé aspekty téhož parametru) se prediktivní schopnost zvyšuje, a to mnohdy výrazně. Pozitivní vliv těchto kombinací najdeme například v [26] či v [27]. V první jmenované studii zkoumala Kinoshita rozlišovací schopnost dlouhodobého průměru základní frekvence a směrodatné odchylky. Referenční data pocházela z korpusu spontánní řeči 90 rodilých mluvčích japonštiny (mužů) a testovaná data zahrnovala neformální promluvy 12 rodilých mluvčích japonštiny, kteří byli nahráváni pro účely výzkumu. Ukázalo se, že rozlišovací schopnost testovaných ukazatelů byla nízká. Kinoshita shrnuje výsledky takto: zjištěné věrohodnostní

⁷⁸ Není součástí práce; přibližuje problematiku věrohodnostního poměru, který následně porovnává s bayesovským přístupem.

⁷⁹ Opraveno; ve VT [5], což však odkazuje na jiný zdroj.

⁸⁰ Pojmenované podle fonetika Johna Bernarda, jehož akustický popis australských vokálů z let 1967 a 1970 se na dlouhou dobu stal standardní referenční databází. (Cox, 1998, s. 29).

poměry prakticky vykazují jednotu, tzn. důkaz získaný touto metodou nemá s ohledem na identitu daného mluvčího žádnou výpovědní hodnotu⁸¹. Druhá studie ([27]) již vedle dlouhodobého průměru F0 a směrodatné odchylky zahrnovala i špičatost, sešikmení, modus F0 a jeho četnost⁸¹. U všech ukazatelů bylo prokázáno, že jsou pro rozlišování mluvčích do jisté míry přínosné. Nejlepších výsledků však bylo dosaženo až tehdy, když se jednotlivé ukazatele uplatnily současně. Z uvedených dvou experimentů si můžeme odnést to, že ačkoli se na základě jediného deskriptoru určitého parametru může zdát, že daný parametr není dostatečně průkazný, nebo je dokonce pro srovnávání mluvčích nepoužitelný, zkombinováním několika jeho aspektů můžeme docílit daleko lepších výsledků.

Dalším faktorem, který bychom neměli opomenout, je robustnost při přenosu [42]. Patrně ji však můžeme zobecnit na „robustnost“ celkově. Přenos ve smyslu „kvalita kanálu“, kupříkladu v případě přenosu prostřednictvím sítí GSM, může být skutečně komplikací, ale kromě něj na robustnost negativně působí celá řada vlivů. Některým z nich, jako živosti projevu, mluvnímu úsilí či maskování jsme se již věnovali v souvislosti s analýzou rozličných parametrů. Zmínit můžeme také častou komplikaci ve formě hluku v pozadí či ozvěny. Nutno podotknout, že žádný z těchto problémů nelze řešit pomocí kombinace několika aspektů daného parametru ani s využitím věrohodnostního poměru. Pokud je totiž robustnost příslušných parametrů příliš nízká, zákonitě se to projeví i na kombinaci výsledků, a to bez ohledu na to, zda použijeme věrohodnostní poměr či nikoli. Je tudíž důležité hledat způsoby, jak vyjádřit co možná nejrobustnější komponenty. Jedním z výše uvedených ukazatelů (konkrétně základní frekvence) je i základní hladina [35], která je vůči některým z obvyklých forem zkreslení, jako jsou změny ve způsobu mluvy nebo špatná kvalita nahrávání, mnohem robustnější nežli běžně používaný aritmetický průměr či medián. Velice slibné jsou i temporální ukazatele, které jsme si již přiblížili výše a které by měly být zvláště robustní vůči snížené kvalitě kanálu. Měření délky trvání lze často provádět i v případě záznamů degradovaných tak výrazně, že jakékoli měření formantů, nebo dokonce i F0, více či méně postrádá smysl. V této oblasti je však zapotřebí dalšího výzkumu, než budeme moci vyvodit jednoznačné závěry.

3.8 Předkládání důkazů u soudu

V tomto oddílu se zaměříme na to, jak lze u soudu prezentovat závěry analýz. Jak uvidíme, v této věci nepanuje shoda, a zdá se, že se ocitáme v určitém přechodném období – podle

⁸¹ Podle: Skarnitzl a Hývlová (2014, s. 57).

některých dokonce můžeme očekávat zásadní změnu v paradigmatu [54] – kde vedle sebe koexistuje několik možných způsobů jak postupovat. Dalo by se očekávat, že jediná momentálně existující profesionální organizace, která slučuje odborníky na fonoskopii, tj. již zmíněná IAFPA, má k dispozici směrnice, jimiž své členy instruuje, jakou formou výsledky provedených analýz u soudu prezentovat, ale v současnosti tomu tak není. Vzniklá situace v zásadě pouze reflektuje žhavou mezinárodní diskuzi o předkládání důkazního materiálu ve forenzním kontextu obecně.

3.8.1 Poslechově-akustická tradice prezentování důkazních materiálů

V současnosti nejpoužívanější metodou prezentování výsledků forenzněfonetického vyšetřování je za pomoci slovně hodnotící škály. 9bodovou stupnici uvedenou níže používáme ve většině případů my a doporučuje ji mimoto i Švédská policie [62].

+4 Výsledky potvrzují hypotézu téměř s jistotou

+3 Výsledky hovoří silně ve prospěch hypotézy

+2 Výsledky hovoří ve prospěch hypotézy

0 Výsledky nejsou průkazné

-1 Výsledky jsou do jisté míry v rozporu s hypotézou

-2 Výsledky jsou v rozporu s hypotézou

-3 Výsledky silně odporují hypotéze

-4 Výsledky odporují hypotéze téměř s jistotou

„Hypotéza“ obvykle zní, že analyzované vzorky řeči pronesl tentýž mluvčí. Slovní hodnocení tohoto či obdobného typu používá i policie ve Finsku, Francii a Německu (osobní komunikace) a s nejvyšší pravděpodobností i další země. Snad není třeba zdůrazňovat, že podobné stupnice mohou být částečně subjektivní, zejména v případě, když vyhodnocujeme a kombinujeme závěry analýz několika parametrů. To nemusí nutně znamenat, že jsou analýzy nespolehlivé; různí kvalifikovaní a zkušení odborníci by při zkoumání téhož materiálu pravděpodobně ve většině případů došli k totožným závěrům. Když jsme v minulosti prováděli analýzy ve spolupráci s dalšími stranami, k neshodám kvůli výsledkům docházelo jen zřídka.

3.8.2 Věrohodnostní poměr: možný posun v paradigmatu

Relativní subjektivita současné praxe se však stala terčem kritiky, v důsledku čehož můžeme pozorovat postupný příklon k metodě věrohodnostního poměru [38–39, 51]. Kritika obecnějšího rázu (forezních věd obecně) přichází jednak z USA [41], ale i ze Spojeného království [33]. Ačkoli se ani jedna ze zpráv fonoskopií explicitně nezabývá, je zřejmé, že se kritika vztahuje i na ni, přičemž vytykáno je následující: nesourodý postoj forezních odborníků, absence závazné standardizace, certifikace a akreditace, interpretace a hodnocení výkonu (s. 5–8, [41]) a dále, že „Skutečnost je taková, že interpretace důkazních materiálů ve forezním prostředí není vždy podložena vědeckými studiemi, jež by prokázaly její validitu“ (s. 9). Jak lze předpokládat, za vzor je považováno DNA.

S výjimkou analýzy jaderného DNA však u žádné forezní metody nebylo zjištěno, že by byla schopna konzistentně a s vysokou mírou jistoty prokázat souvislost mezi důkazním materiálem a konkrétním jedincem či zdrojem. (s. 7, [41])

Ve fonoskopii můžeme v reakci na výtku týkající se příliš subjektivního přístupu pozorovat zvýšenou tendenci využívat metodu věrohodnostního poměru. Nadto je zde patrná snaha povýšit status důkazů z oblasti fonoskopie pokud možno na roveň DNA, což je na jednu stranu žádoucí, ale zároveň bychom neměli zapomínat na některé zásadní rozdíly mezi nimi. Detailnější srovnání s DNA je zavádějící hned z několika důvodů. Sekvence DNA je diskrétní (po tom, co eliminujeme chyby z měření), zatímco proměnné ze sféry řeči zpravidla vyjadřujeme na určité stupnici. Řeč lze vyjádřit množstvím rozličných (a alespoň částečně na sobě nezávislých) proměnných, jako je barva hlasu, základní frekvence, struktura formantů, načasování apod. DNA zůstává v průběhu celého života jedince stabilní a neměnná, řeč nikoli. I když vezmeme v úvahu třeba i jediný parametr jako F_0 , viděli jsme, že se mění nejen během života jednotlivce, ale někdy také z minuty na minutu. Dále ačkoli můžeme aplikovat metodu věrohodnostního poměru na všechny parametry, které lze vyjádřit pomocí proměnných na určité stupnici, ne všechny aspekty řeči tímto způsobem vyjádřit lze, například způsob mluvy, dialekt, cizinecký přízvuk, přeréknutí a jiné. V zásadě by se dalo uvažovat o odhadovaném počtu mluvčích daného dialektu apod., ale to s sebou taktéž nese řadu komplikací; dialektičnost lze například rovněž vnímat jako proměnnou bez ostře vytyčených hranic. Daubertův standard (viz appendix A⁸²) od znaleckého posudku požaduje

⁸² Appendix A není součástí práce; popisuje vznik Daubertova standardu, který čítá 4 požadavky na důkazní materiál: 1) zda je výchozí teorie/hypotéza či metoda ověřitelná, 2) zda výchozí hypotéza či metoda prošla recenzním řízením, 3) je-li u konkrétní použité metody známa chybovost a jsou-li k dispozici.

„vědeckost“ a „spolehlivost důkazních materiálů“, ale Nejvyšší soud jej zároveň vnímá jako „flexibilní“. To lze jednak interpretovat jako oprávnění v co nejvyšší možné míře využívat metody blízké DNA, ale taktéž bychom mohli paletu obdobných zdrojů důkazů rozšířit o vědecký úsudek, který – ač obtížně kvantifikovatelný – určitou „spolehlivost důkazních materiálů“ prokazatelně poskytuje. Ačkoli bychom tedy opravdu měli usilovat o důslednější popisy, měli bychom zvážit, co je a není realistické. Právě to se stalo základem oficiálního stanoviska Spojeného království, které si přiblížíme v následujícím oddílu.

Výhodou slovního hodnocení, o němž jsme psali výše, je, že není náročné na pochopení a využívá se často; nevýhodou potom je, že je subjektivní. Na opačném pólu stojí věrohodnostní poměr, který je sice také částečně subjektivní (byť jen minimálně), ale především není ani zdaleka zřejmé, jak prezentovat výsledky z něj plynoucí tak, aby jim snadno porozuměl i neodborník. Pochopí porota či soudce, co představuje hodnota věrohodnostního poměru (rovna) 2374? Budeme-li se striktně držet bayesovského přístupu, tato hodnota znamená, že ať už byla naše hypotéza před uvážením důkazů jakákoli, nyní je 2374krát pravděpodobnější, že kombinace výsledků hovoří v neprospěch podezřelého. Přestože by bylo ideální, kdyby si soudce či porota do předpřipraveného vzorce Bayesovy věty pouze dosadili příslušné číslice, mnohem pravděpodobnější je, že o Bayesovi nikdy ani neslyšeli. Obdobné obavy vyjádřili i statistikové. Dobrý příklad založený na reálném případě nabízí [13].

Proběhly určité snahy pravděpodobnostní poměr vyjádřit pomocí slovního hodnocení ([9]; Gonzalez-Rodriguez et al. 2001).

1–10	Omezená podpora
10–100	Mírná podpora
100–1 000	Silná podpora
Více než 1 000	Velmi silná podpora

Tento návrh je poměrně opodstatněný, zvláště pokud zde vnímáme slovní hodnocení výhradně jako možnou interpretaci výsledného věrohodnostního poměru. V současnosti však neexistuje žádný konsenzus, jenž by stanovil, jakým způsobem je záhodno jej prezentovat u soudu.

3.8.3 Oficiální stanovisko Spojeného království

Důvodem k formulování oficiálního stanoviska Spojeného království [19] se stala skutečnost, že tradiční způsob, jak prezentovat důkazní materiály, může na straně obžaloby vést k dezinterpretaci. Jestliže totiž závěry hlasového srovnání prezentujeme jako „je vysoce pravděpodobné, že příslušné vzorky řeči pochází od téhož mluvčího“, aniž bychom uvážili možnost, že jiné vzorky, které jsme neanalyzovali, by mohly vykazovat stejnou podobnost, otevíráme tím prostor pro předpojatost vůči obžalovanému. Jak jsme mohli vidět, tento problém lze řešit metodou věrohodnostního poměru. Ale proč potom nepoužívat výhradně ji? Jako hlavní důvod zastánci zmíněného stanoviska uvádějí to, že je v současnosti patrný nedostatek vhodných referenčních databází, kvůli němuž u většiny parametrů není kvantitativní posouzení možné. Navrhovaným řešením je v zásadě verze věrohodnostního poměru vyjádřená slovy, přičemž klíčovými pojmy jsou zde *konzistence* a *specifičnost* (jež zhruba odpovídají *podobnosti* a *příznačnosti*). Konzistence se vztahuje k „rozhodnutí [...] zda jsou sporné a srovnávací vzorky kompatibilní či konzistentní s hypotézou, že je pronesl tentýž mluvčí“. Závěr může znít: konzistentní, nekonzistentní či bez rozhodnutí. Se specifičností souvisí posouzení, do jaké míry sdílí mluvčí v dané populaci určitý rys, jenž může být: mimořádně specifický, velmi specifický, specifický, mírně specifický či nespecifický.

Tento přístup v současnost používá většina odborníků ve Spojeném království, ale mám informaci, že i oni zvažují příklon k metodě věrohodnostního poměru – podobně jako policejní laboratoře ve Finsku, Německu a Francii (osobní komunikace), ale dosud nepadlo konečné rozhodnutí (pro ani proti).

3.9 Shrnutí a závěry

Na základě toho, co jsme si v této kapitole přiblížili, by mělo být zřejmé, že srovnávání mluvčích pro forenzní účely není tak jednoduché, jak si mnozí myslí. Řečové rysy, s nimiž se v této sféře můžeme setkat, se totiž uplatňují i v řečové komunikaci, a nejsou tedy neměnné. Jako příklad jsme uváděli frekvence formantů; ačkoli je pravda, že jsou částečně dány velikostí a tvarem vokálního traktu jedince, vlivem pozice a tvaru jazyka, pohybem dolní čelisti a rtů při tvoření řeči se neustále mění. To je v příkrém rozporu např. s metodou DNA či otisků prstů, jež se během života jedince nemění.

Úkol, před nímž nyní stojíme, je odhalit v lidské řeči takové rysy (pokud existují), které podléhají změnám co nejméně či vůbec, a zjistit, jestli vykazují dostatečnou interindividuální

variabilitu, abychom je mohli využít při srovnávání mluvčích. Z dosavadního bádání vyplývá, že žádný rys těmto požadavkům sám o sobě nevyhovuje, a že bychom se měli zaměřit spíše na kombinace faktorů daného parametru, jako je například u základní frekvence aritmetický průměr, směrodatná odchylka, špičatost, sešikmení, modus F0 a jeho četnost.

V naší praxi se osvědčilo rozlišovat mezi hlasem a řečí. Podobnost hlasů totiž není totéž co identita mluvčích. Je možné, že automatickým srovnáním dvou hlasů dojdeme k velmi vysokým hodnotám věrohodnostního poměru, zatímco analýzy jiných faktorů naznačují, že vzorky pochází od různých mluvčích. Poněkud extrémním, ale názorným příkladem může být situace, kdy nám srovnáním dvou vzorků vyjde věrohodnostní poměr rovný 1000, na základě čehož byste po vzoru starých televizních seriálů mohli být v pokušení zvolat „Našli jsme shodu“ – ale co když vyjde najevo, že srovnávání mluvčí hovoří dvěma různými jazyky? V reálné forenzní praxi bychom proto neměli zapomínat, že hlas a řeč nejsou totéž.

Dalším důležitým poznatkem, který bychom si z uvedeného výkladu měli odnést, je irelevance pojmu *identifikace*, na nějž bychom měli rezignovat. Spolehlivá rozpoznatelnost ve smyslu identifikace konkrétního jedince na základě vzorků řeči není jen obtížná, jako spíše prakticky nemožná. Někteří jsou názoru, že to ovšem neplatí pouze pro důkazy vyplývající z řeči, ale pro všechny typy důkazních materiálů včetně DNA (vynikající přehled nabízí [54!]).

U většiny typů analýz se dnes setkáme s komplikací ve formě nedostatku relevantních referenčních databází. Mnoho pracovišť pracuje na nápravě vytvořením nových databází, ale prozatím po dobu ještě několika let budeme muset odhadovat specifčnost na základě nepříliš spolehlivých dat.

Závěrem se pozastavme nad možným rizikem, jež dnes vyplývá z činnosti soudního znalce. S dostupnějšími automatickými metodami, které na uživatele nekladou přílišné nároky, roste riziko, že přístroje používané v této oblasti budou obsluhovat nekvalifikovaní pracovníci, kteří se bez nich neobejdou a kteří zároveň nebudou mít potuchu, na jakých principech ony přístroje fungují. Nemáme důkaz, že se tak v současnosti opravdu děje, ale měli bychom mít určité povědomí o vývoji tímto směrem a potenciální hrozbě pro vědeckost forenzní praxe.

Reference

1. Alderman TB (2005) Forensic speaker identification—a likelihood ratio-based approach using vowel formants. LINCOS Europa, München
2. Atal BS (1972) Automatic speaker recognition based on pitch contours. *J Acoust Soc Am* 52(6):1687–1697
3. Baldwin J, French P (1990) Forensic phonetics. Pinter, London
4. Bonastre J-F, Scheffer N, Matrouf D, Fredouille C, Larcher A, Preti A et al (2008). ALIZE/SpkDet: a state-of-the-art open source software for speaker recognition. In: Paper presented at the Odyssey 2008: the speaker and language recognition workshop, Stellenbosch, South Africa, 21–24 January 2008
5. Bonastre J-F, Wils F, Meignier S (2005) ALIZE, a free toolkit for speaker recognition. In: Paper presented at the ICASP 2005, Philadelphia, PA, USA
6. Bouchrika I, Nixon MS (2008) Gait recognition by dynamic cues. In: Paper presented at the 19th IEEE international conference on pattern recognition, Tampa, FL, USA
7. Braun A (1995). Fundamental frequency – how speaker specific is it? In: Braun A, Köster J-P (eds) *Studies in forensic phonetics*. WVT Wissenschaftlicher, Trier, pp 9–23
8. Byrne C, Foulkes P (2004) The ‘mobile phone effect’ on vowel formants. *Int J Speech Lang Law* 11(1): 83–102
9. Champod C, Evett IW (2000) Commentary on A. P. A. Broeders (1999) ‘Some observations on the use of probability scales in forensic identification’, *Forensic Linguistics*, 6(2): 228–41. *Int J Speech Lang Law* 7(2): 238–243
10. Dellwo V (2010) The automatic extraction of time-domain based speaker idiosyncratic features. In: Paper presented at the IAFPA 2010, Trier, Germany, 18–21 July 2010
11. Dellwo V, Koreman J (2008) How speaker idiosyncratic is measurable speech rhythm? In: Paper presented at the IAFPA 2008, Lausanne, Switzerland, 20–23 July 2008
12. Dellwo V, Ramyeed S, Dankovicova J (2009) The influence of voice disguise on temporal characteristics of speech. In: Paper presented at the IAFPA 2009, Cambridge, UK, 3–5 August 2009
13. Donnelly P (2005) Appealing statistics. *Significance* 2(1):46–48
14. Elliott J (2000) Comparing the acoustic properties of normal and shouted speech: a study in forensic phonetics. In: Paper presented at the 8th Australian International Conference on Speech Science and Technology, Canberra, Australia
15. Endres W, Bambach W, Flösser G (1971) Voice spectrograms as a function of age, voice disguise, and voice imitation. *J Acoust Soc Am* 49: 1842–1848
16. Eriksson A (2010) The disguised voice: imitating accents or speech styles and impersonating individuals. In: Llamas C, Watt D (eds) *Language and identities*. Edinburgh University Press, Edinburgh, pp 86–96

17. Eriksson A, Wretling P (1997) How flexible is the human voice?—A case study of mimicry. In: Paper presented at the EuroSpeech '97, Rhodes, Greece
18. Farrús M, Hernando J (2009) Using jitter and shimmer in speaker verification. *IET Signal Process Special Issue Biometr Recogn* 3(4): 247–257
19. French JP, Harrison P (2007) Position statement concerning use of impressionistic likelihood terms in forensic speaker comparison cases. *Int J Speech Lang Law* 14(1):137–144
20. Grabe E, Low EL (2002) Durational variability in speech and the rhythm class hypothesis. In: Gussenhoven C, Warner N (eds) *Papers in laboratory phonology*, vol 7. Mouton de Gruyter, Berlin, pp 515–546
21. Guillemin BJ, Watson C (2008) Impact of the GSM mobile phone network on the speech signal: some preliminary findings. *Int J Speech Lang Law* 15(2): 193–218
22. Hollien H (1990) *The acoustics of crime*. Plenum Press, New York
23. Hollien H (2002) *Forensic voice identification*. Academic Press, San Diego
24. Horii Y (1975) Some statistical characteristics of voice fundamental frequency. *J Speech Hear Res* 18(1):192–201
25. Kinoshita Y (2001) Testing realistic forensic speaker identification in Japanese: a likelihood ratio-based approach using formants. Unpublished Ph.D thesis, The Australian National University, Canberra
26. Kinoshita Y (2005) Does Lindley's LR estimation formula work for speech data? Investigation using long-term f0. *Int J Speech Lang Law* 12(2): 235–254
27. Kinoshita Y, Ishihara S, Rose P (2009) Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. *Int J Speech Lang Law* 16(1)
28. Kirkland J (2003) Forensic speaker identification using Australian English fucken: a Bayesian likelihood ratio-based auditory and acoustic phonetic investigation. Unpublished Honours Thesis, Australian National University
29. Künzel HJ (1987) *Sprechererkennung: Grundzüge forensicher Sprachverarbeitung*. Kriminalistik, Heidelberg
30. Künzel HJ (1997) Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguist* 4(1), 48–83
31. Künzel HJ (2000) Effects of voice disguise on speaking fundamental frequency. *Forensic Linguist* 7:149–179
32. Künzel HJ (2001) Beware of the 'telephone effect': the influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguist* 8(1): 80–99
33. Law Commission (2009) *The admissibility of expert evidence in criminal proceedings in England and Wales: a new approach to the determination of evidentiary reliability*
34. Lawrence S, Nolan F, McDougall K (2008) Acoustic and perceptual effects of telephone transmission on vowel quality. *Int J Speech Lang Law* 15(2):161–192

35. Lindh J, Eriksson A (2007) Robustness of long time measures of fundamental frequency. In: Proceedings of Interspeech 2007, Antwerp, Belgium, pp 2025–2028
36. McDougall K (2004) Speaker-specific formant dynamics: an experiment on Australian English /aI/. *Int J Speech Lang Law: Forensic Linguist* 11(1):103–130
37. McDougall K (2006) Dynamic features of speech and the characterization of speakers. *Int J Speech Lang Law* 13:89–126
38. Morrison GS (2009a) Forensic voice comparison and the paradigm shift. *Sci Justice* 49(4):298–308
39. Morrison GS (2009b) The place of forensic voice comparison in the ongoing paradigm shift. In: Proceedings of the 2nd international conference on evidence law and forensic science conference, vol 1, pp 20–34
40. Morrison GS (2010) Forensic voice comparison. In: Freckelton I, Selby H (eds) *Expert Evidence*. Thomson Reuters, Sydney (Ch 99)
41. National Research Council (2009) *Strengthening forensic science in the United States: a path forward*. The National Academies Press, Washington D.C.
42. Nolan F (1983) *The phonetic bases of speaker recognition*. Cambridge University Press, Cambridge
43. Öhman L, Eriksson A, Granhag PA (2010) Mobile phone quality vs. direct quality: how the presentation format affects earwitness identification accuracy. *Eur J Psychol Appl Legal Context* 2(2):161–182
44. Papcun G (1988) What do mimics do when they imitate a voice? *J Acoust Soc Am* 84(S114)
45. Perrot P, Aversano G, Chollet G (2007) Voice disguise and automatic detection: review and perspectives. In: Stylianou Y, Faundez-Zanuy M, Esposito A (eds) *Progress in nonlinear speech processing*. Springer, Berlin, pp 101–117
46. Perrot P, Preteux C, Vasseur S, Chollet G (2007) Detection and recognition of voice disguise. In: Paper presented at the IAFPA 2007, Plymouth, UK
47. Project on Scientific Knowledge and Public Policy (2003) Daubert—the most influential Supreme Court ruling you’ve never heard of
48. Ramus F, Nespor M, Mehler J (1999) Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3):265–292
49. Rose P (2002) *Forensic speaker identification*. Taylor & Francis, New York
50. Rose P (2006) Technical forensic speaker recognition: evaluation, types and testing of evidence. *Comput Speech Lang* 20(2–3):159–191
51. Rose P, Morrison GS (2009) A response to the UK position statement on forensic speaker comparison. *Int J Speech Lang Law* 16(1):139–163

52. Rose P, Osanai T, Kinoshita Y (2003) Strength of forensic speaker identification evidence: multispeaker formant- and cepstrum-based segmental discrimination with a Bayesian likelihood ratio as threshold. *Forensic Linguist* 10:179–202
53. Rothman HB (1979) Further analysis of talkers with similar sounding voices. In: Hollien H, Hollien PA (eds) *Current issues in the phonetic sciences*, vol 9. John Benjamins, Amsterdam, pp 837–846
54. Saks MJ, Koehler JJ (2008) The individualization fallacy in forensic science evidence. *Vanderbilt Law Rev* 61(1):199–219
55. Solzhenitsyn AI (1968) *The first circle* (trans: Whitney TP). Northwestern University Press, Evanston
56. Steinberg JC (1934) Application of sound measuring instruments to the study of phonetic problems. *J Acoust Soc Am* 6(1):16–24
57. Terzuolo CA, Viviani P (1980) Determinants and characteristics of motor patterns used for typing. *Neuroscience* 5(6):1085–1103
58. Traunmüller H (1994) Conventional, biological and environmental factors in speech communication: a modulation theory. *Phonetica* 51:170–183
59. Traunmüller H (2000) Evidence for demodulation in speech perception. In: *Proceedings of ICSLP 2000, Beijing, China*, pp 790–793
60. Traunmüller H, Eriksson A (1995) The perceptual evaluation of F0-excursions in speech as evidenced in liveliness estimations. *J Acoust Soc Am* 97(3):1905–1915
61. Traunmüller H, Eriksson A (2000) Acoustic effects of variation in vocal effort by men, women and children. *J Acoust Soc Am* 107(6):3438–3451
62. Utlåtandeskalan (2008)
<http://www.skl.polisen.se/Global/www%20och%20Intrapolis/Informationsmaterial/SKL/Utlatandeskalan.pdf>
63. Wagner I (1995) A new jitter-algorithm to quantify hoarseness: an exploratory study. *Forensic Linguist* 2(1):18–27
64. Wiget L, White L, Schuppler B, Grenon I, Rauch O, Mattys SL (2010) How stable are acoustic metrics of contrastive speech rhythm? *J Acoust Soc Am* 127(3):1559–1569
65. Yoo J-H, Hwang D, Moon K-Y, Nixon MS (2008) Automated human recognition by gait using neural network. In: *Paper presented at the IPTA 2008, first workshops on image processing theory, tools and applications*
66. Yoon T-J (2010) Capturing inter-speaker invariance using statistical measures of rhythm. In: *Paper presented at the proceedings of speech prosody 2010, Chicago, IL, USA, 10–14 May 2010*

GLOSÁŘ

accused	obžalovaný/obviněný
acoustic phantom picture	hlasový portrét
advanced filtering	pokročilé metody filtrace
adverse conditions	nepříznivé podmínky
air stream	proud vzduchu
arithmetic mean	aritmetický průměr
articulation rate	artikulační tempo
articulatory targets	cílové artikulační pozice
audio authentication	autentizace zvukové nahrávky
audio enhancement	„čištění“ audiozáznamu
aural voice recognition	poslechová identifikace hlasu
automatic formant tracking	automatická detekce formantů
automatic speaker recognition	automatické rozpoznávání mluvčího
automatic speaker verification system	systém automatické verifikace mluvčího
automatic voice comparison	automatické porovnávání hlasů
average fundamental frequency	průměrná základní frekvence
band-pass filters	filtry pásmové propustnosti
baseline	základní hladina
Bayes' formula	vzorec Bayesovy věty
bugged mobile phone	odposlech mobilního telefonu
carrier	nosná pozice
carrier wave	nosná vlna
conclusive	průkazný
creaky voice	třepeň fonace
Daubert ruling	Daubertový standard
defendant	obžalovaný/obviněný
degree of dialect	míra dialektu
dialect	dialekt
dialect mismatch	dialektová neshoda
diphthong	diplofonie, dvojhlasí

discrimination power	rozlišovací schopnost
distance	velká vzdálenost a/nebo časový odstup
distribution bias	zkreslení distribuce
eigenvector	vlastní vektor
engineers and computer scientists	techničtí a IT pracovníci
falsetto	falzet
familiar-speaker identification	identifikace známého mluvčího
fillers	hezitační zvuky (event. i výplňková slova)
foreign accent	cizinecký přízvuk
forensic speech analyst	forezní fonetik
forensic speech expert	znalec se specializací na fonoskopii
forensic speech science	fonoskopie, forezní analýza řeči
fundamental frequency mean	průměrná základní frekvence
fundamental frequency (F0)	základní frekvence (F0)
fundamental frequency level	hladina základní frekvence
global speaking rate	globální tempo řeči
harsh voice	tlačená/drsná fonace
hesitations	hezitace
high-pitched voice	vysoko položený hlas
hoarse voice	chrapot/dysfonie
human speaker perception	rozpoznatelnost mluvčího člověkem
channel bandwidth	šířka pásma kanálu
idiosyncrasies	idiosynkrazie
idiosyncratic choice of words	zvláštní výběr slov
informant	konzultant/informátor
inter-individual variation	interindividuální variabilita
International Association for Forensic Phonetics and Acoustics	Mezinárodní asociace forezní fonetiky a akustiky (IAFPA)
interrogation	výslech
impersonator	imitátor
intra-speaker variation	intraindividuální variabilita

jaw (upper/lower)	čelist (horní/dolní)
jitter	třas
kidnapping	únos
kurtosis	špičatost
Likelihood Ratio	věrohodnostní poměr
linguistic communities	jazyková společenství
lip rounding	zaokrouhlení rtů
Lombard condition	Lombardův efekt
long-term formant distribution (LTF)	dlouhodobá distribuce formantů (LTF)
low-pitched voice	hluboký hlas
manner of speaking	způsob mluvy
measurement of vowel centre formant frequencies	měření formantových frekvencí ve středu vokálu
Mel-Frequency Cepstral Coefficients	mel-kepstrální frekvenční koeficienty
mimic	imitátor
mismatch conditions	nesoulad v okolnostech vzniku
mispronunciation	chybná/nesprávná výslovnost
Modulation Theory of Speech	modulační teorie řeči
narrowband	úzkopásmové pípání
native speaker identification	identifikace rodilým mluvčím
neutral speech conditions	neutrální mluva / neutrální způsob mluvy
noise	rušivý signál
onset	nástup
open and closed nasality	nazálnost/nenazálnost
overlapping speech	překrývající se promluvy
Pairwise Variability Index (PVI)	index párové variability (PVI)
pathological speech problems	patologické řečové problémy / vady řeči
patterns of intonation	intonační kontury
perpetrator	pachatel
pertinent creaky voice	třepená fonace
phantom pictures / photofit pictures	portrétování (podle výpovědi svědků)

positive skewing	kladné sešikmení
recording quality	kvalita záznamu
register	registr
repetitions	repetice
reverbation	ozvěna
robbery	loupež
sentenced	odsouzen
shimmer	tremolo
skew	sešikmení
slur	nedbalá výslovnost
sound files	zvukové soubory
speaker classification	klasifikace mluvčích
speaker comparison	srovnání mluvčích
speaker profiling	profilování mluvčího
speaker recognition	rozpoznávání mluvčích
speaker-specific	individuální rysy mluvčího
speaker-specific characteristics	individuální rysy mluvčího
speaking condition	komunikační situace
speaking rate	tempo řeči
speaking style	způsob mluvy
spectogram	spektogram
speech	řeč
speech engineers	odborníci na řečové technologie
speech errors	přeřeknutí
speech pathology	patologie řeči
speech rate	mluvní tempo
speech recognition	rozpoznávání řeči
speech sample	vzorek řeči
speech sounds	hlásky
speech synthesis	řečová syntéza
standard deviation	směrodatná odchylka

stereotypical expressions	výplňková slova
suspect	podezřelý
syllable rate	slabičné tempo
syll/s	slab/s (slabik za sekundu)
technical speaker identification	technická identifikace mluvčího
tested variables	testované proměnné
timbre	témbr
timing	načasování (artikulačních fází) / frázování (vyšších artikulačních celků) / rytmizace
transmission channel	přenosový kanál
UK position statement	oficiální stanovisko Spojeného království
unfamiliar-speaker identification	identifikace neznámého mluvčího
user-friendliness	snadné ovládání
variation	variabilita
verbal scale	slovní hodnotící stupnice
vocal effort	(zvýšené) mluvní úsilí
voice analysis	analýza hlasu
voice breathiness	(hlasová) dyšnost
voice comparison	srovnávání/srovnání hlasů
voice conversion	hlasová konverze
voice creak	třepeň fonace
voice disguise	maskování/deformace hlasu
voice line-up	hlasová konfrontace
voice profiling	hlasové profilování
voice quality	barva hlasu, hlasové zabarvení, témbr, timbre
voice transformation	hlasová transformace
voiceprinting	metoda otisku hlasu
vowel space	vokalický prostor
whisper	šepot
white noise	bílý šum
witness statement	svědecká výpověď
wpm	s/m (slov za minutu)

Závěr

Tato práce nabídla anotovaný překlad dvou textů z forenzněfonetického prostředí, konkrétně *The forensic phonetician: Forensic speaker identification by experts* Michaela Jessena a *Aural/Acoustic vs. Automatic Methods in Forensic Phonetic Case Work* Anderse Erikssona. Obě kapitoly byly vybrány na základě konzultace s ředitelem Fonetického ústavu pod Filozofickou fakultou Univerzity Karlovy, doc. Mgr. Radkem Skarnitzlem, Ph.D., který se danou problematikou odborně zabývá již několik let. Při výběru bylo bráno v potaz mimo jiné stáří textů, jejich relevance pro současné bádání a erudice jejich autorů.

V úvodu jsme si ve stručnosti zmapovali současný stav české forenzní fonetiky a poukázali na její interdisciplinární charakter, z něhož pro překladatele plyne nejedna komplikace, neboť je na něj kladen požadavek obeznámenosti hned v několika různých oborech. Jak jsme mohli vidět, české protějšky některých termínů nadto zatím nejsou pevnou součástí úzu a v některých případech se adekvátní ekvivalent, který by nebylo nutné řešit opisem, vůbec nepodařilo dohledat (jako v případě *earwitness*).

Překlad obou kapitol provázel poznámkový aparát, který se vztahoval k relevantním pasážím a výrazům. Na základě komentářů terminologické i neterminologické povahy z něj extrahovaných byla následně konstruována syntetizující kapitola, která zjištěné poznatky kategorizovala dle jednotlivých jazykových rovin, a to foneticko-fonologické, morfologické, lexikální a syntakticko-textové. V závěru kapitoly byl věnován prostor i pro chyby ve výchozím textu coby specifickou a nepříliš diskutovanou součástí translátologické činnosti. Ze zvláště relevantních termínů byly nakonec vytvořeny anglicko-český glosář forenzněfonetické terminologie.

Tato práce vznikla s cílem alespoň částečně zpřístupnit českým čtenářům problematiku forenzní fonetiky, neboť nejde o disciplínu v českém prostředí dostatečně etablovanou a známou; česky psané oborové literatury je zde nadto pouhé minimum. S rozmachem mobilních telefonů a technologií obecně lze očekávat rostoucí potřebu analyzovat zvukové nahrávky pro forenzní účely. Často problematická průkaznost těchto analýz u soudu a skutečnost, že jsou v současnosti Ministerstvem spravedlnosti ČR v tomto oboru evidováni pouze dva soudní znalci, je signálem, že si tato disciplína zaslouží mnohem více pozornosti.

Anotace

Autor diplomové práce: Mánek Jiří

Název katedry a fakulty: Katedra bohemistiky, Filozofická fakulta Univerzity Palackého v Olomouci

Název diplomové práce: Vybrané kapitoly z forenzní fonetiky (komentovaný překlad)

Vedoucí diplomové práce: doc. PhDr. Božena Bednaříková, Ph.D.

Počet znaků: 156 477 (včetně mezer)

Počet příloh: 0

Počet titulů použité literatury: 25

Klíčová slova: forenzní fonetika, fonoskopie, audioexpertiza, srovnávání mluvcích, profilování mluvcích, rozpoznávání mluvcích, modulační teorie řeči, maskování

Anotace diplomové práce:

Diplomová práce nabízí překlad dvou vybraných kapitol zabývajících se problematikou forenzní fonetiky (*The forensic phonetician: Forensic speaker identification by experts* a *Aural/Acoustic vs. Automatic Methods in Forensic Phonetic Case Work*), a to z anglicky psaných odborných publikací. Překládané texty jsou doplněny o poznámkový aparát vztahující se k relevantním pasážím a výrazům. Součástí práce je také teoretická kapitola pojednávající o vybraných problematických jevech s ohledem na jednotlivé jazykové roviny a anglicko-český glosář forenzněfonetické terminologie.

Resumé

This diploma thesis is concerned with an annotated English to Czech translation of two selected chapters on forensic phonetics, namely *The forensic phonetician: Forensic speaker identification by experts* by a German forensic phonetician named Michael Jessen, and *Aural/Acoustic vs. Automatic Methods in Forensic Phonetic Case Work* by a Swedish forensic phonetician named Anders Eriksson. These texts have been chosen after an initial assessment of the date of their origin, their relevance and the level of erudition of the authors. The commentary deals with problematic areas of the translation (mainly concerning terminology and possible lack of adequate Czech equivalents for certain phenomena). The theoretical part of the thesis concerns selected relevant features encountered during the course of translation, which are then categorized according to their affiliation to a specific area of linguistics, namely phonetics/phonology, morphology, lexicology and syntax / text linguistics. Furthermore, a glossary of English–Czech forensic phonetic terminology has been made.

Seznam literatury

Překládané texty:

JESSEN, Michael (2010). *The forensic phonetician*. In: M. Coulthard a A. Johnson (Eds.), *The Routledge Handbook of Forensic Linguistics*. London: Routledge, s. 378–394.

ERIKSSON, Anders (2012). *Aural/Acoustic vs. Automatic Methods in Forensic Phonetic Case Work*. In: A. Neustein a H. A. Patil (Eds.), *Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism*. New York: Springer, s. 41–69.

BEDNAŘÍKOVÁ, Božena (2009). *Slovo a jeho konverze*. Olomouc: Univerzita Palackého v Olomouci. ISBN 978-80-244-2220-6.

ČERMÁK, František (2011). *Jazyk a jazykověda: přehled a slovníky*. Vyd. 4., V Karolinu 2., dopl. Praha: Karolinum. ISBN 978-80-246-1946-0.

DUŠKOVÁ, Libuše a kol (2009). *Mluvnice současné angličtiny na pozadí češtiny*. [online] Praha: Ústav anglického jazyka a didaktiky FF UK. [cit. 5. 8. 2019]. Dostupné z: <http://emsaff.cuni.cz/12.21.4>.

HÁLA, Bohuslav (1975). *Fonetika v teorii a v praxi: vysokošk. učebnice*. Praha: SPN. Učebnice pro vysoké školy (Státní pedagogické nakladatelství).

HUDDLESTON, Rodney a Geoffrey K. PULLUM (2005). *A Student's Introduction to English Grammar*. Cambridge: Cambridge University Press, s. 245–265. ISBN 978-05-218-4837-4.

HUDDLESTON, Rodney a Geoffrey K. PULLUM (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. ISBN 978-0-521-43146-0.

KRČMOVÁ, Marie (2008). *Fonetika a fonologie*. 2. vyd. Brno: Masarykova univerzita. Dostupné z: http://is.muni.cz/elportal/estud/ff/ps09/fonetika/tisk_2009/Fonetika_a_fonologie_logo.pdf.

LADEFOGED, Peter a Keith JOHNSTONE (2011). *A Course in Phonetics*. 6th ed. Boston, MA: Wadsworth/Cengage Learning.

- MATHESIUS, Vilém (1913). *O nominálních tendencích v slovesné predikaci novoanglické*. In: Sborník filologický. 4. ročník. Praha: III. třída České akademie věd a umění, s. 325–339.
- PALKOVÁ, Zdena (1997). *Fonetika a fonologie češtiny s úvodem do obecné problematiky oboru*. Praha: Karolinum. ISBN 8070668431.
- PETR, Jan, et al. (1986). *Mluvnice češtiny I*. Praha: Academia.
- RENDÁR, Ľubomír (2015). *Fonácia a hlasové začiatky v slovenčine*. Trnava: Pedagogická fakulta Trnavskej univerzity v Trnave. ISBN 978-80-8082-877-6.
- ROMPORTL, Milan (1981). *Základy fonetiky*. 2. vyd. Praha: Státní pedagogické nakladatelství.
- ŘEŘIČHA, Václav (1998). *Anglicko-český glosář fonetických a fonologických termínů*. 2., přeprac. vyd. Olomouc: Univerzita Palackého, Filozofická fakulta. ISBN 80-7067-935-2.
- SGALL, Petr (2011). *Jazyk, mluvení, psaní*. Praha: Karolinum, s. 103.
- SKALIČKOVÁ, Alena (1982). *Fonetika současné angličtiny: učebnice pro vysoké školy*. Praha: Státní pedagogické nakladatelství. Učebnice pro vysoké školy (Státní pedagogické nakladatelství).
- SKARNITZL, Radek, ed. (2014). *Fonetická identifikace mluvího*. Praha: Filozofická fakulta Univerzity Karlovy. ISBN 978-80-7308-548-3.
- SKARNITZL, Radek (2014). *Forenzní fonetika*. In: Fonetická identifikace mluvího. Praha: Filozofická fakulta Univerzity Karlovy, s. 11–20.
- SKARNITZL, Radek, et al. (2014). *Vokálníkové formanty*. In: Fonetická identifikace mluvího. Praha: Filozofická fakulta Univerzity Karlovy, s. 21–48.
- SKARNITZL, Radek a Dita HÝVLOVÁ. (2014). *Statistický popis hodnot základní frekvence*. In: Fonetická identifikace mluvího. Praha: Filozofická fakulta Univerzity Karlovy, s. 49–64.
- VOLÍN, Jan a Tomáš BOŘIL (2014). *Základní frekvence v konturách a průbězích*. In: Fonetická identifikace mluvího. Praha: Filozofická fakulta Univerzity Karlovy, s. 65–76.
- WEINGARTOVÁ, Lenka, BOŘIL Tomáš a Jitka VAŇKOVÁ (2014). *Spektrální sklon*. In: Fonetická identifikace mluvího. Praha: Filozofická fakulta Univerzity Karlovy, s. 77–94.

WEINGARTOVÁ, Lenka a Jan VOLÍN (2014). *Temporální charakteristiky*. In: Fonetická identifikace mluvího. Praha: Filozofická fakulta Univerzity Karlovy, s. 95–103.

BOŘIL, Tomáš a Jitka VAŇKOVÁ (2014). *Telefonní přenos*. In: Fonetická identifikace mluvího. Praha: Filozofická fakulta Univerzity Karlovy, s. 104–115.

BOŘIL, Tomáš a Lenka WEINGARTOVÁ (2014). *Rozhodování a statistika*. In: Fonetická identifikace mluvího. Praha: Filozofická fakulta Univerzity Karlovy, s. 116–135.

SVOBODOVÁ, Marie a Luboš VOŘÍŠEK (2014). *Identifikace mluvího z pohledu autentické kriminalistické praxe v České republice*. In: Fonetická identifikace mluvího. Praha: Filozofická fakulta Univerzity Karlovy, s. 136–144.

SKARNITZL, Radek, Pavel ŠTURM a Jan VOLÍN (2016). *Zvuková báze řečové komunikace: fonetický a fonologický popis řeči*. Praha: Univerzita Karlova v Praze, nakladatelství Karolinum. ISBN 978-80-246-3272-8.

SPENCER, Andrew (1988). *Bracketing paradoxes and the English lexicon*. In: *Language* 64, s. 663–682. Dostupné z: <http://www.jstor.org/stable/414563>.

SVOBODOVÁ, Marie (1997). *Forenzní lingvistika: obsah a možnosti*. In: *Slovo a slovesnost* 58, s. 124–129. Dostupné z: <http://sas.ujc.cas.cz/archiv.php?lang=en&art=3726>.

WELLS, John. (1982a). *1: An introduction*. Cambridge [etc.]: Cambridge university press.

WELLS, John. (1982b). *2: The British isles*. Cambridge [etc.]: Cambridge university press.

WELLS, John. (1982c). *3: Beyond the British isles*. Cambridge [etc.]: Cambridge university press.