

VÝZKUM FALŠOVÁNÍ HLASOVÉ IDENTITY NA FFUK V RÁMCI PROJEKTU START



Tomáš Nechanský – Alžběta Růžičková

Program START je programem, jenž si dal za cíl podpořit vědeckou činnost doktorandů studujících na Univerzitě Karlově. Patří do operačního programu Výzkum, vývoj a vzdělávání, projekty jsou realizovány od začátku dubna roku 2021 do konce března 2023. Z 274 podaných žádostí bylo podpořeno 97 týmů celkovou částkou téměř čtvrt miliardy korun. Upřednostňovány byly týmy skládající se z více řešitelů, respektive z jednoho hlavního řešitele a dalších vedlejších řešitelů. Zaměření projektu se nesmělo překrývat s tématem disertační práce žádného z členů projektového týmu. Aby informace o složení byly kompletní, je třeba dodat, že nad každým týmem stojí takzvaný mentor, jehož úlohou je poskytovat jednotlivým členům podporu nejen v oblasti výzkumu. Kromě řešení hlavního tématu celého projektu se studenti zavázali k plnění takzvaných vzdělávacích cílů, které se dělí na dva typy — formální a neformální. Mezi neformální řadíme zlepšení měkkých dovedností jako například jazykové či manažerské dovednosti. Splnění formálního cíle přísluší zejména hlavnímu řešiteli a spočívá v zahraniční stáži trvající alespoň tři měsíce a v absolvování kurzu zaměřeného například na rozvoj profesních dovedností.

Hlavním řešitelem projektu, o němž bude řeč níže, je Tomáš Nechanský. Vedlejšími řešiteli jsou Alžběta Růžičková a Vojtěch Skořepa. Mentorem celého týmu je ředitel Fonetického ústavu Filozofické fakulty Univerzity Karlovy, Radek Skarnitzl. Jelikož všichni řešitelé studují na zmíněném pracovišti a jejich hlavním zájmem je forenzní fonetika, v tématu projektu se jejich zaměření odrazilo. Název projektu zní *Falšování hlasové identity: fonetická analýza a identifikace mluvčích*.

Nicméně ještě než přistoupíme k detailnějšímu popisu zmíněného projektu, věnujeme pár odstavců tomu, co vůbec falšování hlasové identity a potažmo forenzní fonetika je. Forenzní fonetika je relativně mladá disciplína, jejíž oficiální počátek je datován na začátek 90. let minulého století, i když do jisté míry bychom úkony forenzního fonetika mohli spatřit v procesech důkazních šetření i o desetiletí dříve. Jedním z hlavních cílů fonetického experta je vyjádřit poměr pravděpodobností (z angl. *likelihood ratio*, LR) mezi následujícími hypotézami: „jak pravděpodobné je, že podobnosti a odlišnosti pozorované ve sporné nahrávce (získané například z telefonního odposlechu) a ve srovnávací nahrávce (např. z výslechu policie) mohou pocházet od téhož mluvčího“ a „jak pravděpodobné je, že bychom tyto podobnosti a odlišnosti pozorovali u jiného mluvčího ze stejné populace“. Při dnešních technických možnostech by se to mohlo jevit jako jednoduchý úkol, avšak analýza lidského hlasu neprobíhá stejně jako analýza otisku prstu nebo DNA, u kterých se jedná buď o shodu, či neshodu. Hlas nemůžeme považovat za biometrický údaj srovnatelný s papírárnými liniemi či strukturou DNA.

Pachatel také svou DNA nebo otisk prstu těžko zfalšuje. Ačkoliv falšování hlasové identity, jinak také maskování hlasu, u osob páchajících trestnou činností není obecně příliš častým jevem — v rámci všech případů zahrnujících identifikaci mluvčího jde



o jednotky až nižší desítky procent (Masthoff, 1996; Braun, 2006) — u určitých typů trestné činnosti se tento jev vyskytuje poměrně hojně, zejména v případech únosů, vydírání, sexuálního obtěžování a falešných telefonických oznámení (Künzel, 2000). V případech, kdy pachatel během páchaní trestného činu promlouvá a zároveň může předpokládat, že je jeho projev nahráván, je výskyt maskování hlasu velmi častý: mezi případy, jimiž se zabýval německý Spolkový kriminální úřad v letech 1988–1995, dosáhl jeho poměr 52 %, vůbec nejvyššího podílu pak dosáhl výskyt maskování hlasu v případech vydírání, a to 69 % (Masthoff, 1996). Jedná se tedy o jev, který není pouze okrajový a kterému je zapotřebí věnovat pozornost.

Za nejefektivnější způsob falšování hlasové identity lze považovat automatické maskování hlasu, které může identifikaci mluvčího i zcela znemožnit (Künzel, 2000; Clark & Foulkes, 2007) a jehož četnost v poslední době narůstá společně se zvyšující se kvalitou i dostupností nástrojů, které k němu lze využít (Amino et al., 2017). Z celkového množství případů maskování hlasu ve forenzní praxi však elektronické maskování představuje menšinu (Gfroerer, 1994, citováno v Eriksson, 2010; Masthoff, 1996; novější údaje bohužel nejsou k dispozici).

Ačkoliv se výjimečně můžou objevit i poměrně sofistikované maskovací strategie, většina mluvčích přistupuje ke změně pouze jednoho, případně dvou parametrů (Masthoff, 1996). Nejčastěji se při maskování hlasu mluvčí uchylují ke změnám základní frekvence, prostřednictvím kterých lze poměrně jednoduše dosáhnout percepčně velmi výrazné modifikace hlasu. Výrazné změny základní frekvence mohou být spojeny i se změnou typu fonace: při jejím zvyšování lze přejít až do falzetu, naopak při jejím snižování lze dosáhnout třepené fonace (Künzel, 2000). Mimo modifikace hlasivkového nastavení se mezi maskovacími strategiemi objevují také změny rezonančních charakteristik vokálního traktu, a to jednak prostřednictvím změn artikulačního nastavení, např. palatalizace, labializace, faryngalizace či změny nazality (Laver, 1980; Skarnitzl, 2016), a jednak pomocí umístění cizího předmětu do úst nebo před ústa (Figueiredo & Britto, 1996; Künzel, 2000).

Vedle modifikací jednotlivých parametrů se mezi strategie maskování hlasu řadí také imitace regionálního dialektu (Markham, 1999) či cizineckého přízvuku (Neuhauser & Simpson, 2007; Neuhauser, 2008), případně i řečových charakteristik určitého mluvčího (Růžičková & Skarnitzl, 2017). Tyto strategie obvykle zahrnují změny více parametrů zároveň, např. základní frekvence, artikulačního nastavení, temporálních charakteristik atd. Napodobení dialektu zpravidla působí přirozenějším dojmem než např. maskování hlasu pomocí cizího předmětu v ústech, proto může v takových případech být obtížné rozpoznat, že se jedná o maskovaný projev (Markham, 1999) — i rozeznání, že daný mluvčí nehovoří svým přirozeným hlasem, je přitom zásadní pro jeho fonetickou identifikaci. I imitace cizineckého přízvuku zní mnohdy velmi autenticky a identifikaci mluvčího tedy může výrazně zkomplikovat (Neuhauser & Simpson, 2007).

Mapování strategií maskování hlasu se věnovala také studie provedená u českých mluvčích (Růžičková & Skarnitzl, 2017). 100 mluvčích obecně češtiny mužského pohlaví bylo instruováno, aby libovolným způsobem maskovali svůj řečový projev tak, aby znemožnili svou identifikaci podle hlasu. Následná podrobná poslechová analýza ukázala, že většina z nich uplatnila změnu výšky hlasu, a to nejčastěji zvýšení (41 %),



u nižšího počtu mluvčích se pak objevilo její snížení (29 %), 4 mluvčí zvolili šepot, při němž je základní hlasová frekvence nepřítomná a výšku hlasu proto nelze posuzovat. Pouze u 26 mluvčích ze 100 při maskování hlasu nedošlo ke slyšitelným změnám výšky hlasu. Druhým nejčastěji měněným parametrem byla změna typu fonace, jednalo se zejména o třepenou a tlačenu fonaci (obojí u 13 % mluvčích), v jednom případě se dokonce objevila ingresivní tvorba hlasu. Vyskytly se také změny nazalality: 11 mluvčích hovořilo hypernazalizovaně, zatímco 2 si ucpali nos a jejich hlas tak byl naopak denazalizovaný. V 10 případech se objevil posun artikulace vokálů. 6 mluvčích napodobovalo řečové vady a 3 mluvčí imitovali cizinecký přízvuk (ve zkoumaném materiálu se naopak neobjevila imitace regionálního dialektu), jeden mluvčí přikročil k imitaci řečového projevu jiné osoby, konkrétně Miloše Zemana. V poměrně málo případech došlo ke změně tempa řeči; u 11 mluvčích se objevilo jeho snížení, u 3 naopak zvýšení. Ve většině případů docházelo ke kombinacím modifikací více parametrů, 31 mluvčích však zvolilo modifikaci pouze jednoho z nich (většinou výšky hlasu) a 3 mluvčí dokonce nedosáhli žádné slyšitelné změny svého řečového projevu.

Ať už se daný jedinec rozhodne pro jakoukoliv strategii, ukazuje se, že maskování hlasu má tendenci k postupné deklinaci, a to zejména v případě, že mluvčí skutečně více maskovaných projevů v průběhu času. V některých případech může nakonec maskování hlasu i zcela vymizet (Künzel, 2000).

Hlavním cílem projektu *Falšování hlasové identity: fonetická analýza a identifikace mluvčích* je poskytnout komplexní pohled na možné způsoby, jak mluvčí záměrně modifikují svůj hlas, a následně vyhodnotit, jaké to má dopady na fonetickou identifikaci mluvčího. Zkoumat budeme tři hlavní oblasti — cizinecký přízvuk, individuální strategie maskování hlasu a profesionální imitaci mluvčího. Pro tyto účely nám poslouží pět různorodých databází. První z nich je databáze obecné češtiny, ve které 100 mluvčích nejdříve produkovalo spontánní řeč, a poté četlo text „svým“ hlasem a hlasem modifikovaným dle svého uvážení (srov. Růžičková & Skarnitzl, 2017); druhou je česko-anglická databáze, v níž opět 100 mluvčích četlo český a anglický text; třetí z nich je česko-ruská databáze, jejímž cílem bylo nahrát české mluvčí imitující ruský přízvuk a porovnat jejich imitaci s ruskými mluvčími češtiny; čtvrtá databáze obsahuje profesionální hlasové imitátory (v tuto chvíli je zaznamenaný hlas Petra Jablonského, který imituje například Miloše Zemana); a konečně poslední databáze sestává z nahrávek českých fonetiků, kteří opakovaně čtou jeden text s různým hlasovým nastavením (například palatalizovaně či se zavřenou čelistí).

Jelikož naším cílem je poskytnout komplexní pohled na falšování hlasové identity, je potřeba nasbíraný materiál taktéž komplexně zkoumat. Při fonetické identifikaci mluvčího expert využívá dvou hlavních přístupů k analýze řečových vzorků — poslechový (kdy percepčně, tedy za pomoci svého sluchu, analyticky porovnává určité složky hlasu, jakými jsou například výška a barva hlasu či řečové vady, a zaměřuje se na celkový, holistický dojem) a akustický (kdy v počítačových fonetických programech analyzuje řečové parametry, tedy akustické koreláty zmíněných percepčních charakteristik, jako například základní frekvence, formanty či trvání). Trochu stranou stojí automatické rozpoznávání mluvčího (angl. *automatic speaker recognition*, ASR), což je ve své podstatě taktéž akustický přístup, který využívá složitých algoritmů a ještě donedávna navozoval dojem „magické skříňky“. Nicméně



s rozvíjejícími se technologiemi a přibývajícím studiem je zřejmé, že právě ASR je budoucností v identifikaci mluvčího. Na vstupu stojí sporné a srovnávací nahrávky, na výstupu (pokud je k dispozici populace pro srovnání) poměr pravděpodobností (LR, viz výše).

Tyto tři hlavní přístupy odráží metody, které jsme zvolili pro analýzu našeho materiálu, a tak jednotlivé nahrávky budeme porovnávat akusticky, percepčně i automaticky. Co se týče manuálních akustických metod, analyzovat budeme zejména základní frekvenci a dlouhodobé formantové distribuce (průměr formantových hodnot z celé nahrávky nehledě na segment, ze kterého pocházejí). Automatický přístup zastane program VOCALISE (Oxford Wave Research, 2019), který využívá takzvaných *i*-vektorů a faktorových analýz. Konečně percepční přístup bude spočívat ve vytvoření pěti percepčních testů (každý bude náplní odpovídat jedné z databází), jež budou distribuovány vždy mezi 50 respondentů. Abychom imitovali poslechový přístup jak z hlediska expertů (například soudních znalců), tak z hlediska laiků (například obětí či svědků), posluchače plánujeme najímat z řad fonetiků i mimo ně. Celkem tedy cílíme na 250 administrovaných testů. Jen pro úplnost dodáme, že respondentům bude jejich účast kompenzována finanční odměnou. Jak vyplývá z výše zmíněného, porovnáme, jak úspěšné v identifikaci mluvčího budou jednotlivé přístupy. Předčí percepční holistický přístup ten automatický či manuálně-akustický? Jsou určité strategie maskování hlasu efektivnější než jiné? Bude pro posluchače obtížné identifikovat ty hlasové vzorky, které byly problematické i pro akustické metody?

Závěrem zmíníme čtyři plánované projektové výstupy. Prvním výstupem, ke kterému jsme se zavázali, je příspěvek na forenzně-fonetické konferenci IAFPA (International Association for Forensic Phonetics and Acoustics), která proběhne v Praze v létě roku 2022. Vzhledem k časové posloupnosti plnění jednotlivých projektových aktivit předpokládáme, že na této konferenci budeme prezentovat pouze předběžné výsledky. Hlavními výstupy jsou potom tři články, které budou nabídnuty odborným, převážně impaktovaným periodikům v tuzemsku i zahraničí. Jednotlivé články budou odrážet oblasti zmíněné výše. První bude informovat, zdali jsou vybrané metody schopny identifikovat mluvčího, pokud imituje či produkuje cizinecký přízvuk; druhý se zaměří na to, zdali jsou akustické a poslechové metody schopny od sebe odlišit záměrně maskované hlasy od nemaskovaných, či zdali se vůbec o falšování hlasu jedná (a do jaké míry je přirozené); a třetí se zaměří na to, jestli jsou naše metody schopné od sebe odlišit skutečné hlasy známých osobností od imitací nahraných profesionálními hlasovými imitátory.

Na závěr chceme podotknout, že vzhledem k šíři pojímaného tématu bychom rádi naše poznatky využili i po skončení projektového období. I přestože se nejedná o oficiální výstup projektu, je naším záměrem vydat monografii shrnující celou problematiku do jednoho celku a v návaznosti na náš výzkum podporovat studenty bakalářských a magisterských programů v dalším výzkumu falšování hlasové identity.

BIBLIOGRAFIE

- Amino, K., Makinae, H., & Kamada, T. (2018). Auditory discrimination of natural speech and synthetic speech used as voice disguise. *Acoustical Science and Technology*, 39(1), 48–50. <https://doi.org/10.1250/ast.39.48>
- Braun, A. (2006). Stimmverstellung und Stimmenimitation in der forensischen Sprechererkennung. In T. Kopfermann (Ed.), *Das Phänomen Stimme: Imitation und Identität* (pp. 177–181). St. Ingbert.
- Clark, J., & Foulkes, P. (2007). Identification of voices in electronically disguised speech. *Speech, Language and the Law*, 14, 195–221. <https://doi.org/10.1558/ijssl.v14i2.195>
- Eriksson, A. (2010). The disguised voice: Imitating accents or speech styles and impersonating individuals. In C. Llamas & D. Watt (Eds.), *Language and Identities* (pp. 86–96). Edinburgh University Press.
- Figueiredo, R. M., & Britto, H. S. (1996). A report on the acoustic effects of one type of disguise. *International Journal of Speech, Language and the Law*, 3(1), 168–175. <https://doi.org/10.1558/ijssl.v3i1.168>
- Künzel, H. J. (2000). Effects of voice disguise on speaking fundamental frequency. *International Journal of Speech, Language and the Law*, 7(2), 149–179. <https://doi.org/10.1558/sll.2000.7.2.149>
- Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge University Press.
- Masthoff, H. (1996). A report on a voice disguise experiment. *International Journal of Speech Language and the Law*, 3(1), 160–167. <https://doi.org/10.1558/ijssl.v3i1.160>
- Markham, D. (1999). Listeners and disguised voices: The imitation and perception of dialectal accent. *International Journal of Speech, Language and the Law*, 6(2), 289–299. <https://doi.org/10.1558/sll.1999.6.2.289>
- Neuhauser, S. (2008). Voice disguise using a foreign accent: Phonetic and linguistic variation. *The International Journal of Speech, Language and the Law*, 15(2), 131–159. <https://doi.org/10.1558/ijssl.v15i2.131>
- Neuhauser, S., & Simpson, A. P. (2007). Phonetic correlates of accent authenticity in voice disguise. *Proceedings of IAFPA 2007*.
- Oxford Wave Research (2019). iVOCALISE 2019A. Available from <https://oxfordwaveresearch.com/products/vocalise/>.
- Růžičková, A., & Skarnitzl, R. (2017). Voice disguise strategies in Czech male speakers. *AUC PHILOLOGICA*, 2017(3), 19–34. <https://doi.org/10.14712/24646830.2017.30>
- Skarnitzl, R. (2016). Co dokáže náš hlas? Fonetický pohled na variabilitu řečové produkce. *Slovo a smysl*, 26, 95–113.

Tomáš Nechanský | Fonetický ústav FF UK
<tomas.nechansky@seznam.cz>

Alžběta Růžičková | Fonetický ústav FF UK
<ruzickovaalzbeta@seznam.cz>

