

Česká zemědělská univerzita v Praze

Technická fakulta



Identifikace osob na základě hlasu a řeči

Bakalářská práce

Vedoucí bakalářské práce: Ing. Veronika Hartová, Ph.D.

Autor: Dominik Čáp

Praha, 2017

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Dominik Čáp

Obchod a podnikání s technikou

Název práce

Identifikace osob na základě hlasu a řeči

Název anglicky

Identification of persons on the basis of voice and speech

Cíle práce

Bakalářská práce je tematicky zaměřena na problematiku biometrických identifikačních systémů založených na porovnání hlasu a řeči. Hlavním cílem je provést zhodnocení této problematiky.

Dílčí cíle bakalářské práce jsou:

- vytvořit přehled řešené problematiky,
- charakterizovat základní principy systémů,
- provést ekonomické zhodnocení,
- zhodnotit možnosti bezpečnostních rizik u tohoto typu identifikace.

Metodika

Metodika řešené problematiky bakalářské práce je založena na studiu a analýzách odborných informačních zdrojů. Vlastní řešení je realizováno formou hodnocení základních principů. Na základě rozboru teoretických poznatků a výsledků hodnocení budou formulovány závěry bakalářské práce.

Doporučený rozsah práce

30 až 40 stran textu včetně obrázků, grafů a tabulek

Klíčová slova

biometrie, hlas, řeč, spolehlivost

Doporučené zdroje informací

HEŘMAN, J., et al.: Elektrotechnické a telekomunikační instalace. Praha: Verlag Dashöfer, 2008. ISSN 1803-0475.

JAIN, A.; BOLLE, R.; PANKANTI, S. „Biometrics. Personal Identity in a Networked Society.“ Norwell, Massachusetts, USA, Kluwer Academic Publisher, 1999, ISBN 0-7923-8345-1.

KŘEČEK, S., a spol.: Příručka zabezpečovací techniky. Blatná: Circetus, 2006. 313s. ISBN 80-902938-2-4. RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních

aplikacích.“ Praha, Nakladatelství Grada, 2012



Předběžný termín obhajoby

2016/17 LS – TF

Vedoucí práce

Ing. Veronika Hartová, Ph.D.

Garantující pracoviště

Katedra elektrotechniky a automatizace

Elektronicky schváleno dne 12. 1. 2016

prof. Ing. Jaromír Volf, DrSc.

Vedoucí katedry

Elektronicky schváleno dne 2. 3. 2016

prof. Ing. Vladimír Jurča, CSc.

Děkan

V Praze dne 30. 03. 2017

Prohlášení

„Prohlašuji, že jsem bakalářskou práci na téma: Identifikace osob na základě hlasu a řeči vypracoval samostatně a použil jen pramenů, které cituji a uvádím v seznamu použitých zdrojů.

Jsem si vědom, že odevzdáním bakalářské práce souhlasím s jejím zveřejněním dle zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů, ve znění pozdějších předpisů, a to i bez ohledu na výsledek její obhajoby.

Jsem si vědom, že moje bakalářská práce bude uložena v elektronické podobě v univerzitní databázi a bude veřejně přístupná k nahlédnutí.

Jsem si vědom, že na moji bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů, ve znění pozdějších předpisů, především ustanovení § 35 odst. 3 tohoto zákona, tj. o užití tohoto díla.“

V Praze dne 30. 3. 2017

.....

Dominik Čáp

Poděkování

Děkuji Ing. Veronice Hartové, Ph.D. za vedení této bakalářské práce, za její ochotu, odborné rady a cenné připomínky, které mi během psaní práce poskytovala. Dále bych chtěl poděkovat rodině za podporu během studia.

Abstrakt

Bakalářská práce se zabývá problematikou identifikace osob na základě hlasu a řeči. První část bakalářské práce se věnuje obecným pojmům biometrie jako identifikace, autentizace a verifikace, následně popisuje historii fonetických expertíz, fonetické aspekty zkoumání řečového signálu pro účely identifikace, obecné rozpoznání mluvěcího a jeho výhody a nevýhody nebo vysvětluje tvorbu lidského hlasu. Dále popisuje charakteristiku jednotlivých druhů hlasových systémů a výkonnost a bezpečnostní rizika verifikačních systémů. Další kapitola je věnována využití aplikací v praxi. Ve zbývajících teoretických částech se práce zaměřuje na algoritmy a metody kriminalistických audioexpertíz. V praktické části práce byly porovnány tři různé biometrické systémy. Konkrétně čtečka otisku prstů, sken obličeje a právě systém založený na principu identifikace osob dle hlasu a řeči.

Klíčová slova: biometrie, hlas, řeč

Abstract

This bachelor thesis deals with issues about identification of persons on the basis of voice and speech. First part of bachelor thesis deals with general concepts of biometrics such as identification, authentication and verification, then describes history of phonetic expertise, phonetic aspects of the research of speech signal for identification purposes, general recognition of speaker and its advantages and disadvantages and explains creation of human voice. One chapter is devoted to application usage in practice. Thesis also describes the characteristics of individual species voice systems and their efficiency. The remainder of the theoretical part focuses on algorithms and methods of forensic audio expertise. In the practical part of thesis were compared three different biometric systems. Specifically fingerprint reader, scan faces and system based on the principle of identifying persons by voice and speech.

Key words: biometry, voice, speech

Obsah

1	Úvod.....	1
2	Cíl práce	2
3	Metodika práce	3
4	Přehled řešené problematiky	4
4.1	Biometrie a základní pojmy	4
4.2	Historie fonetických expertíz	6
4.3	Fonetické aspekty zkoumání řečového signálu pro účely identifikace.....	7
4.4	Obecné rozpoznání mluvího	7
4.4.1	Výhody a nevýhody rozpoznání mluvího	9
4.5	Vytváření hlasu.....	10
4.6	Charakteristika základních principů systémů.....	12
4.6.1	Textově závislé systémy	12
4.6.2	Systémy s textovou výzvou.....	13
4.6.3	Textově nezávislé systémy.....	14
4.7	Výkonnost a bezpečnostní rizika verifikačních systémů	14
4.7.1	Ovlivňující faktory.....	16
4.8	Využití aplikací v praxi.....	17
4.8.1	Telefonní autorizace transakcí.....	17
4.8.2	Řízení přístupu.....	19
4.8.3	Správa řečových dat a personalizace	19
4.8.4	Forenzní rozpoznání mluvího	20
4.9	Algoritmy	20
4.9.1	Extrakce charakteristik.....	21
4.9.2	Ověření mluvího metodou GMM	22

4.9.3	Prahování.....	23
4.10	Metody kriminalistické audioexpertizy	25
4.10.1	Dlouhodobé spektrum řeči	29
4.10.2	Spektrální rozložení formantů	30
5	Praktická část práce	33
5.1	Přístupový snímač otisků prstů.....	33
5.2	Terminál na rozpoznání obličeje	34
5.3	Hlasový systém Multi-speech.....	35
5.4	Charakteristika metody	36
5.5	Zhodnocení systémů	36
5.5.1	Ekonomické zhodnocení	37
5.5.2	Uživatelské zhodnocení	38
6	Zhodnocení výsledků	39
7	Závěr	41
8	Seznam literatury	42
9	Seznam obrázků.....	44
10	Seznam tabulek.....	45

1 Úvod

Výraz identifikace se stal hojně používaným a moderním termínem. Za tím stojí zájem lidí, států a jejich institucí nebo různých organizací. V praxi je nutno čím dál častěji rozlišovat nebo ztotožňovat různé jevy a jejich důsledky, činnosti a akce, potřeba a zájmy, osoby, materiály, zvířata apod. a podle toho usměrňovat svoje cíle. ^{[1][2]}

Biometrické metody autentizace jsou založeny na předpokladu, že většina charakteristik jsou jedinečné pro každou živou bytost a zároveň v průběhu času nezaznamenávají velké změny. Při výběru jednotlivých vlastností, na základě kterých je prováděna identifikace, jsou určující i tyto aspekty: snadná získatelnost, ochotnost uživatele pro snímání dané charakteristiky a samozřejmě finanční náklady spojené se snímáním. ^[3]

V současné době je známo mnoho biometrických identifikačních metod, které využívají různé jedinečné charakteristiky. Nejznámější a nejpoužívanější metodou je rozpoznání otisků prstů, dalšími například rozpoznání obličejových rysů, sítnice oka nebo i mnohem komplexnější analýza DNA. ^[3]

Hlasová biometrie se zabývá ověřením a verifikací osob na základě jedinečných charakteristik jejich hlasu. V posledních letech již technologie velmi pokročila. Problémem menšího využívání je spíše přetrvávající nedůvěra odborné i laické veřejnosti v její spolehlivost. Využívání technologií zabývajících se hlasem lze nalézt například v navigačních systémech, ve virtuálních asistentech u mobilních telefonů nebo u firem v podobě virtuálních operátorů. V dohledné době se pravděpodobně rozšíří více i v bankách a podobných institucích. ^{[1][2]}

V současné době se poměrně rozvíjí odvětví s názvem audioexpertíza. Zkoumá se lidský hlas a řeč a úkolem je identifikovat neznámého mluvčího. Dalším odvětvím je potom forenzní fonetika, která se zabývá zkoumáním a porovnáváním některých akustických parametrů hlasu a řeči, také s cílem identifikovat mluvčího. ^[1]

2 Cíl práce

Hlavním cílem této práce bude provést zhodnocení problematiky identifikace osob na základě hlasu a řeči. Tématicky se bude zabírat problematikou biometrických identifikačních systémů. V praktické části se provede zhodnocení hlasového biometrického systému v porovnání s jinými biometrickými systémy používajícími se k identifikaci osob.

Dílčí cíle bakalářské práce jsou:

- vytvořit přehled řešené problematiky,
- charakterizovat základní principy systémů,
- provést ekonomické zhodnocení,
- zhodnotit možnosti bezpečnostních rizik u tohoto typu identifikace.

3 Metodika práce

Jako první bude proveden přehled řešené problematiky. Informace budou čerpány z odborné literatury a odborných článků. Tato část přiblíží čtenáři oblasti týkající se biometrie, identifikace, hlasu nebo například řeči.

První kapitola se bude zabírat úvodem do biometrie, kde se vysvětlí hlavní pojmy problematiky. V teoretické části se dále práce bude zaměřovat na historii fonetických expertíz, tvorbu hlasu z biologického hlediska, rozpoznání mluvčího, charakteristiku jednotlivých systémů, výkonnost a bezpečnost verifikačních systémů, algoritmy, využití v praxi nebo na metody kriminilistických audioexpertíz. Teoretická část bude obsahovat souhrn mnoha informací o problematice identifikace, autentizace a verifikace.

Praktická část práce bude obsahovat porovnání třech různých druhů biometrických identifikačních systémů. Přesněji porovnání čtečky otisku prstů, skenu obličeje a hlasového systému. Jednotlivé vybrané přístroje budou podrobněji přiblíženy a následně porovnány. Pro zdroje informací budou využity katalogy specializující se firmy. Informace ohledně cen a další specifikace produktů bude pocházet od prodejce biometrických systémů.

K vyhodnocení porovnání bude použita bodová multikriteriální analýza. Na základě určitých zvolených parametrů se stanoví vhodnost jednoho biometrického identifikačního systému vůči dalším dvěma ze dvou odlišných hledisek uživatele. Bude se hodnotit ekonomické hledisko a jak jsou systémy uživatelsky přívětivé. Po vyhodnocení obou bodových multikriteriálních analýz se porovnájí výsledky v grafu.

4 Přehled řešené problematiky

Audioexpertíza hlasu a řeči osob je poměrně mladé odvětví a v dnešní době se stále rozvíjí. Dříve se tato kriminalistická disciplína označovala jako fonoskopie. Cílem této metody je porovnat hlas a řeč cizí anonymní osoby s osobou uloženou v databázi a určit tak její identitu. Rozpoznávání osob podle hlasu a řeči se v současné době používá hlavně v oboru kriminalistiky nebo soudnictví. Odvětví se nazývá aplikovaná (forenzní) fonetika. ^[1]

Obor se začal rozšiřovat ve státech západní Evropy a USA přibližně od 60. let. Dnes již probíhá výzkum této problematiky takřka po celém světě. Přednost audioexpertízy tkví v tom, že díky řečovému signálu lze získat další informace jako přibližný věk, pohlaví osoby, lokální nebo národní příslušnost (přízvuk). ^[1]

Forenzní fonetika se řadí mezi obory poměrně kontroverzní, jelikož mnozí vědci nejsou stále přesvědčeni, jestli specialista z tohoto oboru disponuje dostatečnými informacemi, abymohl stanovit přesný závěr ohledně totožnosti člověka. Nejistota vychází zejména z toho, že fonetický výzkum se soustředí zejména na systémové jevy. U těchto jevů záleží na spoustě činitelů, jsou tedy velice variabilní a složitě se zobecňují. V okruhu odborníků na identifikace mluvčího se vyskytuje určité množství nedořešených problémů. Dosud totiž nelze najít a stanovit takový parametr hlasových a řečových charakteristik, který by byl dostačující pro jednoznačné stanovení totožnosti mluvčího. ^[1]

4.1 Biometrie a základní pojmy

Biometrie se zabývá zkoumáním živých organismů – člověka, u kterého měří anatomické a fyziologické vlastnosti. Název biometrie vychází z řečtiny. Slovo „bios“ znamená život, „metron“ potom měření. Jedná se tedy o měření a identifikaci jednotlivých charakteristik člověka. K rozpoznání člověka pomocí jeho unikátních vlastností či proporcí slouží různé metody. ^[2]

Důležité pojmy jsou autentizace, verifikace a identifikace.

- Autentizace (také autentifikace nebo legalizace) znamená rozpoznávání. Je to proces, při kterém se ověřuje identita objektu. Existuje mnoho metod např. heslo, použití předmětu nebo biometrického prvku. ^{[2][3]}

- Verifikace charakterizuje proces, kde díky biometrickému systému probíhá potvrzení totožnosti jedince. Již dříve uložený vzorek se porovnává se vzorkem právě sejmutým – jedná se tedy o porovnání 1:1. ^[2]
- Identifikace – po sejmutí biometrického údaje probíhá pokus o určení totožnosti neznámého jedince. Princip je znám jako one-to-many (1:N). ^[2]

Pojem *rozpoznávání mluvčího* je činnost s cílem přiřadit správně vzorek řeči k určité osobě na základě foneticko-akustických charakteristik. Během rozpoznávání osoby je možnost odlišit tzv. *naivní rozpoznávání*, kde se nepoužívá žádná zvláštní technika ani znalost a tzv. *technické rozpoznávání*, při kterém se speciální technika, metodologie a znalosti využívají. ^[1]

Při podrobném zpracování dat nahraných k identifikaci probíhá předběžné kvalitativní zhodnocení záznamu hlasu a řeči z jazykového i technického hlediska. S tím souvisí následující pojmy:

- Technická kvalita nahrávky – označuje míru šumů, hluku a dalších jevů, které vznikly při přenosu signálu nebo tvorbě nahrávky a ovlivňují kvalitu nahraných dat
- Jazyková kvalita záznamu – vypovídá o kvalitnosti nahrávky z hlediska délky textu a množství jazykových jevů, které lze zhodnotit a poté analyzovat podle běžných fonetických metod ^[1]

Při posouzení zpracovatelnosti nahrávky se hodnotí technická a jazyková kvalita. Výsledek se interpretuje na třístupňové škále: *dobře zpracovatelná*, *obtížně zpracovatelná*, *nezpracovatelná*. ^[1]

Akustika se zabývá fyzikálními jevy, které souvisí se vznikem zvukového vlnění, jeho šířením a vnímáním zvuku sluchem. V tomto oboru se vyskytuje další termín a to *zvukové spektrum*, což znamená rozklad zvukového signálu na jeho harmonické složky. ^[1]

Mezi důležité pojmy patří *sporná a srovnávací nahrávka*. Sporná nahrávka je nejčastěji běžný záznam telefonního hovoru, většinou tedy nemá dostatečnou jazykovou ani technickou kvalitu. Srovnávací nahrávkou se rozumí zvuková stopa nahraná pod odborným dohledem v laboratoři. Tyto termíny se využívají při identifikaci mluvčího, kde se porovnává hlas zaznamenaný na sporné nahrávce s hlasem na nahrávce odborně pořízené. ^[1]

4.2 Historie fonetických expertíz

První zmínky o fonetických forenzních expertízách pochází ze Spojených států amerických a Velké Británie, kde se používaly hlavně při řešení soudních sporů. Právě tyto dva státy potom řešily otázku, která ze dvou fonetických metod je vhodnější. Američané stáli za rozpoznáváním mluvcího díky spektrogramu (voiceprintu), které se jim zdálo spolehlivější. Fonetikové z Británie preferovali analýzu sluchovou, u které jako podpůrnou metodu používali akustickou analýzu. V dnešní době fonetici používají kombinovaně všechny dostupné metody. ^{[1][4]}

Zpočátku odborníci považovali individuální prokazatelnost hlasu a řeči za neprokazatelnou. Až v padesátých letech dvacátého století se objevují studie zaměřující se na tuto problematiku. Výzkum v této době nebyl ještě dostatečně rozvinutý, aby výsledky používání poslechoých analýz byly použity pro identifikace osob např. při policejních záležitostech. ^{[1][4]}

První významné studie vytvořili I. Pollack, J. M. Pickett nebo W. H. Sumby. Mezi další průkopníky se potom řadí J. N. Shearn, J. N. Holmes nebo P. L. Garvin. V článku časopisu *Phonetica* zmiňují identifikaci osob na základě instrumentální analýzy, ve které vidí budoucnost fonetiky. Také kladou důraz na určení validních kritérií pro rozpoznání mluvcího, která by byla vhodná pro metodu sluchovou i instrumentální. Soupeření mezi zastánci těchto dvou metod trvá doposud. ^[4]

Na začátku let šedesátých se na vývoji výrazně podílel i český fonetik P. Janota. Zaobíral se především oblastí dlouhodobého spektra řeči a zkoumáním individuálního hlasového tónu mluvcích (1961, 1967). ^[1]

Po Janotovi se v České republice fonetice a oblasti identifikace osob na základě hlasu nikdo téměř nevěnoval. Pražský Kriminologický ústav se na přelomu 70. a 80. let začal zabývat výzkumem forenzní problematiky mluvené řeči. Dále probíhal výzkum v oblasti poznatelnosti jazykových jednotek. ^[1]

V roce 1992 vznikla v Kriminologickém ústavu databáze mluvcích. Její princip spočívá ve zkoumání nejvýraznějších charakteristik hlasového projevu mluvcího, které jsou stanoveny v rámci sluchových a akustických analýz. ^[5]

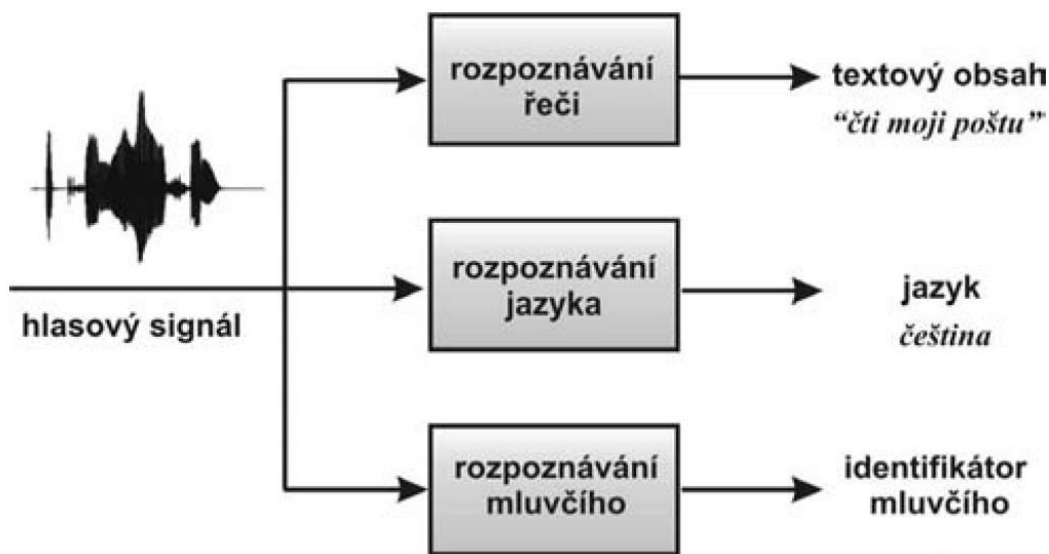
4.3 Fonetické aspekty zkoumání řečového signálu pro účely identifikace

Mezi mluvčími existují odlišnosti v řeči, které převážně spočívají v anatomických rozdílech mluvidel a změnou jejich tvaru při hovoření. Díky tomuto faktu může fonetik využít svých teoretických znalostí problematiky, tak i svou posluchačskou zkušenost a jazykový cit. Hodnoty získané přístrojem při instrumentální analýze tyto informace nezahrnují. ^[6]

4.4 Obecné rozpoznání mluvčího

Nejpřirozenějším způsobem komunikace mezi lidmi je mluvení. Vzhledem k očekávání progresu v oblasti interakce člověka s počítačem byly vynalezeny systémy pro extrakci různých informací obsažených v lidském hlase. Zpracování řečového signálu se dělí na tři základní úlohy viz obrázek 1. Při rozpoznání řeči se určuje slovosled tvořící finální zprávu, u rozpoznávání jazyka se identifikuje právě použitý jazyk řečového signálu a při rozpoznávání mluvčího je úkolem zjistit identitu mluvčího. ^{[1][7]}

Obrázek 1 - Zpracování řečového signálu



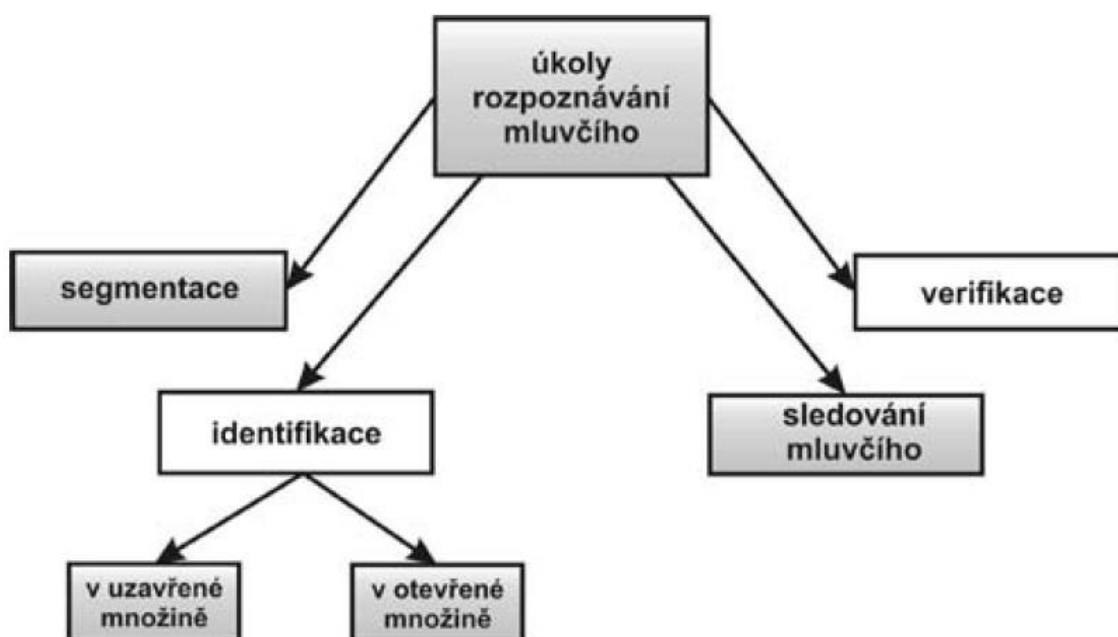
Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

V porovnání s dalšími biometrickými metodami identifikace (otisky prstů či snímání duhovek) je technologie rozpoznávání mluvčího často uváděna jako méně přesná. Má však dvě velká pozitiva. První poukazuje na velké rozšíření automatizovaných telefonních služeb, kde je

možné použít funkci rozpoznání mluvího. Za druhou důležitou skutečnost se považuje fakt, že mluvení je přirozené jednání člověka a jelikož nevyžaduje fyzický kontakt, uživateli je považováno za méně dotěrné. Vědecký pokrok ve spojení s těmito dvěma faktory činí biometriku založenou na řeči za významnou a poutavou technologií. [2]

Existují čtyři hlavní úkoly automatického rozpoznání mluvího. Jedná se o *identifikace mluvího*, *verifikaci mluvího*, *segmentaci mluvího* a *sledování mluvího* - viz obrázek 2.

Obrázek 2 - Úkoly rozpoznávání mluvího



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

Úkolem *identifikace mluvího* je zjistit o hlas jaké osoby se jedná. Porovnává se neznámý hlas s hlasem známých ze skupiny registrovaných mluvího. Jedná se o klasifikaci 1:N. Pokud je jistota, že neznámý hlas je obsažen ve skupině registrovaných mluvího, hovoří se o identifikaci v uzavřené množině. Jestliže naopak hlas obsažen není, mění se úloha na identifikaci. Identifikace mluvího se uplatňuje převážně u odposlechu. [1][2]

Úlohou *ověření mluvího* je určit identitu neznámé osoby. Mluvího označí svou identitu a systém tuto informaci buď potvrdí nebo vyvrátí. Jedná se o klasifikaci pouhých dvou skupin – skutečný mluvího a ostatní mluvího. Ověření mluvího se využívá například v bankách díky telefonnímu bankovníctví, kde zajišťuje bezpečnost. [1][2]

V oblasti problémů s více mluvčími jsou známé další dva úkoly. Prvním je *sledování mluvčího*, kde se zaměřuje na to, jak konkrétní mluvčí hovoří na záznamu několika dalších mluvčích. Při této úloze se také určují jednotlivé úseky, kdy konkrétní mluvčí mluví. Druhým úkolem je *segmentace*. Jejím úkolem rozdělit data několika mluvčích. Při segmentaci nejsou žádné další informace k dispozici, jedná se tedy o nejnáročnější z úkolů. ^{[1][2]}

4.4.1 Výhody a nevýhody rozpoznání mluvčího

Každá biometrická technologie má svoje výhody a nevýhody. Rozpoznání mluvčího tedy není vhodné pro všechna využití. Před použitím této technologie by se měla pozitiva a negativa vyhodnotit. ^{[1][8]}

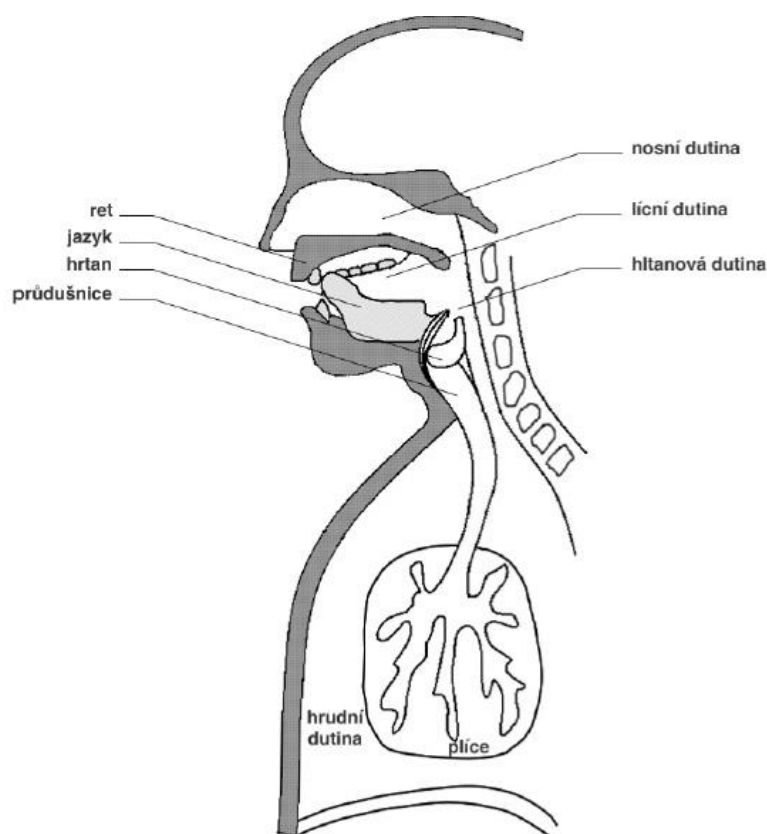
Aplikacím na počítači k rozpoznávání mluvčího stačí pouze zvukové karty a mikrofony, které nejsou drahé. U telefonních přístrojů nejsou zapotřebí žádné další zařízení. Výhodou jsou tedy nízké pořizovací náklady. Další výhodou je výborná bezpečnost proti záměrným útokům. Přístupové body k aplikacím dnes nejsou omezeny místem a lze se skrze mobilní telefonní síť připojit odkudkoliv. U rozpoznání osoby není vyžadován žádný fyzický kontakt a mluvení pro člověka znamená přirozené jednání. Jedná se o metodu příznivou pro koncového uživatele. ^{[1][8]}

Za nevýhodu se považuje nutno opakovaných registračních sezení. Osoba musí poskytnout dostatečný vzorek řeči, aby šablona měla dostatečnou kvalitu. Naopak registrace pro uživatele by měla proběhnout v co nejkratším čase. Pro pokrytí všech proměnlivostí hlasu se sezení musí často opakovat a někdy také časově rozkládat. Mezi další nevýhodu se řadí střední přesnost rozpoznání oproti ostatním biometrikám. Jsou známy tři příčiny. První je inherentní variabilita, ve které se projevují emoce uživatele nebo jeho zdravotní stav a věk. Jako druhá příčina se označuje variabilita mezi mluvčími. Potenciální šance nalezení osoby se stejnými charakteristickými vlastnostmi nejsou tak nízké jako u ostatních druhů rozpoznávání člověka. Poslední důvod představují vnější zvukové vlivy v okolním prostředí jako hluk nebo šum. ^{[1][8]}

4.5 Vytváření hlasu

Hlasový signál je způsoben vytlačením vzduchu z plic. To způsobuje vibraci či klid hlasových chord a formuje hlasový trakt, skrze který proudí vzduch ven. Hlasové ústrojí obsahuje tři dutiny (viz. obrázek 3). Dutinu nosní, lícní a hltanovou. Hlasový trakt je tvořen dutinou hltanovou a ústní. Nosní trakt je tvořen dutinou nosní. Na zformovanosti hlasového traktu může mít vliv pozice rtů, jazyka, čelisti nebo měkkého patra. ^[9]

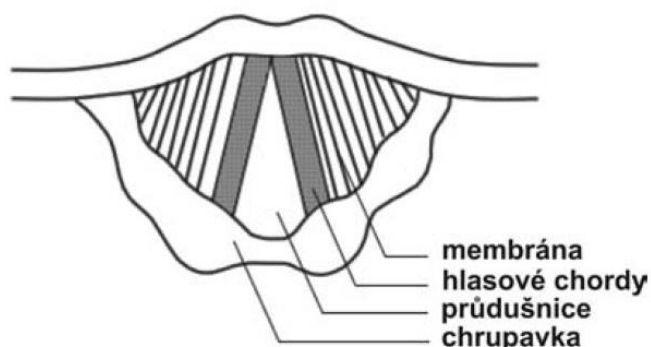
Obrázek 3 - Schématický pohled na lidské hlasové ústrojí



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

Vytlačený vzduch z plic putuje přes průdušnici, hrtan a hlasovým i nosním traktem. Významným orgánem při vzniku řečového signálu je hrtan. Nachází se v horní části průdušnice a je složen z pohyblivých chrupavek a membrán. Hrtan lze zcela uzavřít za pomoci hlasových chordů (viz. obrázek 4), které vytváří dva symetrické okraje. Jestliže jsou napnuty, hrtan se v určitých periodách otvírá nebo zavírá tak, aby vznikl zvuk hlasu. Vzduchové pulsy se označují také jako pulsy glotální. Vytvářeny jsou v hrtanu a rezonují v různých dutinách zformovaných pro vznik určitého zvuku. ^{[1][9]}

Obrázek 4 - Schématický pohled na hrtan

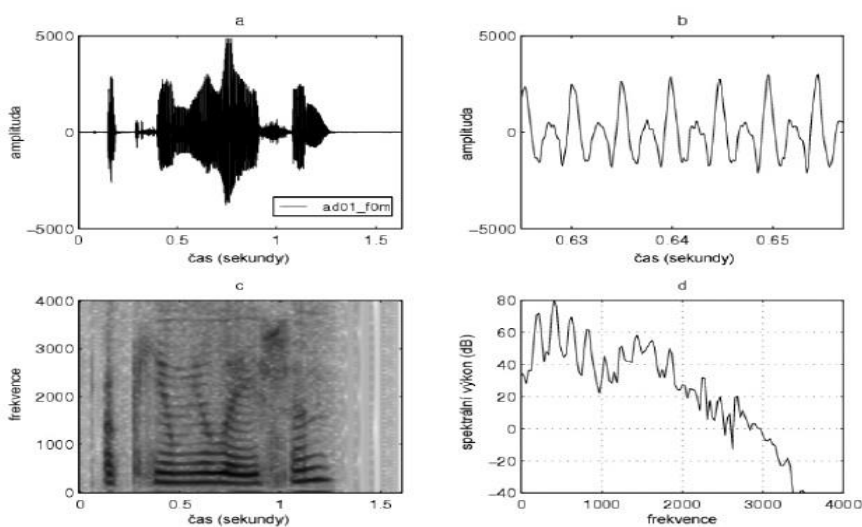


Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

Existují dva typy zvuků, které lze vytvořit. Prvním je neznělá hláska. Vzniká v případě, kdy v bodě zúžení hlasového traktu nastává turbulentní proudění vzduchu. Ve druhém případě jde o hlásku znělou. Ta vzniká při úplně uzavřeném hlasovém traktu a jeho náhlým uvolněním.

Na obrázku 5a můžeme vidět příklad hlasového vzorku. Pokud se hlas zkoumá v krátkých časových intervalech jako na obrázku 5b, nazývá se kvazistacionární nebo slabě časově závislý. Charakteristiky jsou v tomto případě výrazně stacionární (5-100 msec) na rozdíl od obrázku 5a, kde měření proběhlo v dlouho trvajících periodách a signál je nestacionární (>200 msec) a odráží různé vyslovené hlasové zvuky. [1][9]

Obrázek 5 - Příklady řečového signálu



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

Fonémy jsou lingvisticky lišící se hlasové zvuky v určitém jazyce. Jejich výčet lze zjistit ve standartních seznamech, které jsou k dispozici u lingvistických odborníků. Fonémy jsou tvořeny různou rychlostí a frekvencí. Rozdíly mezi rozpoležením mluvcího (zdravotní, emoční...) udávají variabilitu mluvcího (intra). Fyziologickými nebo behaviorálními rozdíly mezi mluvcími je potom dána variabilita mluvcího (inter). Posledním zdrojem variability je koartikulace. Ta vzniká při ochablosti hlasového traktu a způsobuje souvislost s tvořením fonémů vzhledem k okolnímu kontextu. Pro rozlišení jednotlivých mluvcího systémů rozpoznání mluvcího využívají variabilitu mezi mluvcími. ^{[1][10]}

Fyziologické vlastnosti a behaviorální rysy – tyto dva druhy informace obsahuje řečový signál. U fyziologických vlastností se rozeznávají rozměry jednotlivých dutin nebo hlasových chord. Na základě těchto rozměrů potom vznikají fonémy. Behaviorální rysy jsou rozdílné například vlivem typu osobnosti, ovlivněním rodiči, vzděláním nebo prostředím mluvcího. ^[10]

Systémy rozpoznávání mluvcího většinou spoléhají nízkou úroveň akustických vlastností spolu s fyziologickými profily. Částečně zachycují behaviorální rysy jako prozodii nebo délku jednotlivých fonémů. Rysy, které se obtížně získávají a modelují jako například preferovaný slovník nejsou systémy rozpoznání mluvcího obvykle vůbec modelovány. ^[1]

4.6 Charakteristika základních principů systémů

Úkolem rozpoznání mluvcího je ověření nebo přiřazení identity dané osoby. Většinou je k dispozici pouze omezené množství řečových dat určených k analýze hlasových charakteristik. Aby byl uživatel rozpoznán, musí vyslovit určitý typ textu. Systémy rozpoznání jsou klasifikovány právě na základě tohoto typu dat. Obecně jsou rozlišovány systémy textově závislé, s textovou výzvou nebo textově nezávislé. Tyto skupiny jsou aplikovány pro členění úkolů verifikace mluvcího. Pro úlohu identifikace mohou být použity pouze v menší míře. ^{[2][3]}

4.6.1 Textově závislé systémy

Textově závislé systémy disponují většinou dobrým rozpoznávacím výkonem. Pro registraci a následnou autentizaci je využívána stejná část textu. Charakteristiky získané z hlasových signálů jsou stabilnější, jelikož se u každého sezení používá shodná sekvence zvuků. Textová závislost dovoluje využít jemnější modelovací techniky způsobílé zaznamenat informaci ze sekvence zvuků. Největší nevýhodou textově závislých systémů představuje

možnost útoku přehráním, který lze jednoduše uskutečnit pomocí obvyklého zařízení neprodukcujícího řečový vzorek uživatele. ^[11]

Textově závislé systémy založené na hesle, kde se využívá krátký text, jenž by neměl být znám jiným uživatelům. Podmínkou použití tohoto typu je ovšem vytvoření systému pro obnovu hesel, která jsou ztracena nebo zapomenuta. Řízení takového systému představuje zátěž, i pokud se jedná o systém automatizovaný. Systémy založené na heslo a textově závislé systémy mohou být rozděleny do dvou skupin. ^[11]

První skupina popisuje situaci, kdy text nebo heslo je vybráno systémem a přiřazeno uživateli automaticky. Představuje pevnou frázi, která je tvořena slovy z omezeného slovníku. Příkladem jsou i systémy používající PIN kód (osobní číslo). Vzhledem k nutnosti pamatovat si daný text nebo číslo, nejedná se o systémy uživatelsky přívětivé. Jako výhodou tohoto systému lze považovat spojení verifikace do jednoho kroku. Vzorek řeči se nejdřív rozpozná a následně je verifikován. ^{[1][11]}

Případ, kdy text nebo heslo je vybrán uživatelem, představuje skupinu druhou. Způsob je pro uživatele příjemnější, může si totiž sám stanovit obsah textu. Systém je oproti prvnímu způsobu náročnější na vývoj. Disponuje slabším výkonem z důvodu nemožnosti využití znalosti textu modelovací techniky. Systém také musí umět dodatečně zkontrolovat, popř. zamítnout příliš dlouhé nebo naopak krátké heslo. Využití textově závislých systémů je především při potvrzení totožnosti jedince v aplikacích řízení přístupu. ^{[1][11]}

4.6.2 Systémy s textovou výzvou

Systémy s textovou výzvou po uživateli požadují vyslovit náhodně vybrané uskupení slov. Uživatel tedy předem nezná slovosled textu, který musí být vysloven. Systém zpracovává dva základní kroky. Při kontrole očekávané sekvence slov systém rozpozná obsah řeči uživatele. Pokud rozpoznání proběhne v pořádku, následuje verifikace. Tento způsob zvyšuje úroveň bezpečnosti při obraně proti útokům přehráním, nazývá se *výzva-odpověď*. ^{[1][11]}

Není zapotřebí si pamatovat žádné heslo, systém je tedy uživatelsky přívětivý. Ovšem nevýhodou je nutnost vést dialog s uživatelem, který je nucen poslouchat a výrok opakovat. Za další nevýhodu může být považováno rozpoznání řeči, jež může díky různým okolnostem selhat (přízvuk uživatele, zvuky v pozadí apod.) a zmařit snahu uživatele. Systémy s textovou

výzvou jsou velmi výkonné. Modelovací techniky jsou schopny využít znalosti textu. Nastavení verifikace umožňuje uživatele požádat o další informace, pakliže nebude úroveň shody úplná. [1][11]

Využití těchto systémů je hlavně v aplikacích řízení přístupu. Mezi vhodné možnosti patří zabezpečení přístupu k telefonním službám využívající hlas, jelikož automatizovaný rozhovor s uživatelem stejně musí proběhnout. [1]

4.6.3 Textově nezávislé systémy

U textově nezávislých systémů nezáleží na obsahu textu, který uživatel vyřkne. Není tedy známo předem, jaký text uživatel zvolí. Systémy v těchto případech modelují hlasový signál z globálního pohledu. Jelikož počet dat je omezen, některé druhy zvuků nejsou systémem zaznamenány. Díky tomu nejsou není většinou přesnost tak vysoká, jako je tomu u textově závislých systémů. [6][11]

Využití textově nezávislých systémů spočívá většinou v identifikaci mluvčího z telefonních hovorů, záznamů porad a podobně. Uplatnit tuto metodu lze v případech, kdy mluvčí nechce vyslovit určený text, používá se tedy ve forenzních aplikacích. Jsou také využívány v aplikacích řízení přístupu. V případě textově nezávislých systémů si uživatel nemusí pamatovat žádné heslo a systém může při ověřování požadovat extra data, dokud nebude úroveň shody úplná. Tyto dvě skutečnosti jsou hlavními výhodami. Za nevýhodu se považuje zranitelnost vůči útokům přehráním. [2][11]

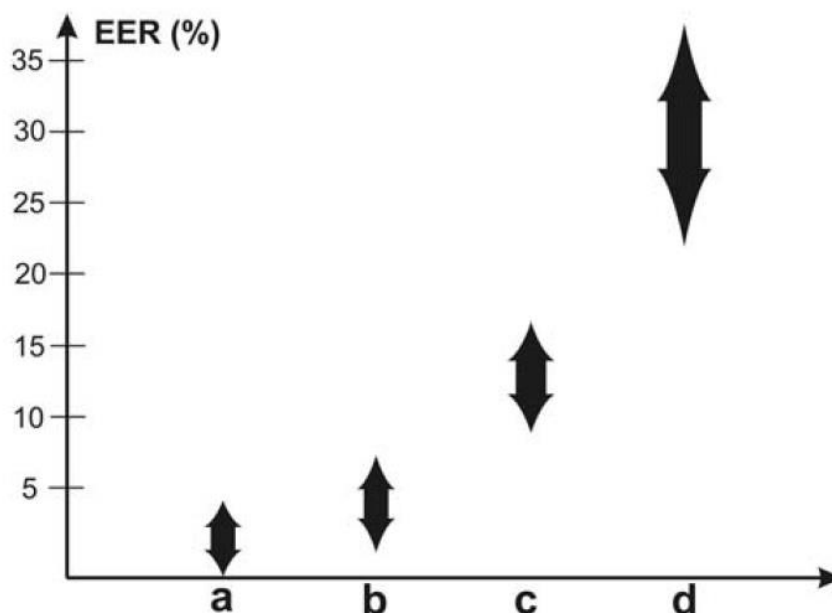
4.7 Výkonnost a bezpečnostní rizika verifikačních systémů

Přesnou hodnotu výkonu lze definovat velmi těžko. Výkonnost ovlivňuje mnoho faktorů, které se liší pro registrační a testové fáze. Za výsledek systému rozpoznání mluvčího považujeme určitou hodnotu (značíme R_c) přímo úměrnou pravděpodobnosti, kdy je vstupní hlasový signál vložen rozpoznaným uživatelem. Systém akceptuje hodnotu R_c nad určitou rozhodovací hladinou, v opačném případě ji odmítne a verifikace není úspěšná. Mohou nastat dvě chyby systému. Nesprávné odmítnutí reálného klienta nebo přijmutí podvádějící osoby. Měření výsledků systémů rozpoznání mluvčího probíhá pomocí chybových hodnot nesprávného přijetí (*FA - false acceptance*) a nesprávného odmítnutí (*FR - false rejection*), které prezentují chybovost jako funkci prahové hodnoty T . V případě malých hodnot T systém

více přijímá uživatele, což zvyšuje hodnotu míry FA. Oproti tomu velké hodnoty T směřují k vyšším hodnotám míry FR. Testují se správné a podvodné pokusy o vstup do systému. Tyto hodnoty lze interpretovat v grafu s parametrem T. O míře se shodnou chybou (*EER* – *equal error rate*) se hovoří v případě, kdy vzniká speciální bod, kde se $FA = FR$. Mírou EER se indikuje výkon. ^{[1][4]}

Rozsahy výkonu EER pro čtyři kategorie systémů rozpoznání osob jsou extrémně rozdílné – od 0,1 až po 30% mezi systémy (viz. obrázek 6). Systémy textově závislé používající vysokokvalitní hlasové signály mohou mít velmi nízké EER, obvykle v rozsahu 0,1% až 2%. Pro výkony s nízkým EER je nutné provést důkladnou registraci, složenou z několika sezení o délce několika minut. Další podmínkou sezení je získání několikasekundových testovacích dat za identických podmínek jako při registraci. Výkon mezi 2% až 5% chybovosti vykazují textově založené aplikace na PIN kódech využívající telefonní linku. Nahrávky telefonní kvality získané díky různým přístrojům na kterých jsou založené textově nezávislé aplikace mají EER od 7% do 15%. Za předpokladu registrace složené z několika sezení o délce několika minut a testovacích dat o desítkách sekund. Textově nezávislé aplikace používající data za rádii s vysokou hodnotou šumu vykazují chybovost od 20% do 35%. ^{[1][4]}

Obrázek 6 - Výkony systémů verifikace mluvího



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „*Biometrie a identita člověka ve forenzních a komerčních aplikacích.*“ Praha, Nakladatelství Grada, 2012

- a) textově závislé aplikace založené na signálech vysoké kvality
- b) textově závislé aplikace založené na signálu telefonní kvality
- c) textově nezávislé aplikace založené na signálu telefonní kvality zaznamenaném během konverzace
- d) textově nezávislé aplikace založené na radiovém signálu s vysokým šumem

4.7.1 Ovlivňující faktory

Pokud se zvýší kvalita řeči, kontrola získávání záznamů, množství testovacích dat a omezí se slovník, zvýší se celkový výkon. Dále budou prezentovány detailněji faktory, které systémy působí výkonost systémů. ^[5]

Podmínky záznamu: Hlavním faktorem přesnosti systémů rozpoznání mluvího je kvalita záznamu. Kvalitu hlasového signálu snižuje okolní zvukové vlivy jako šum, rozhovor v pozadí nebo variabilita komunikačního kanálu. Snižování kvality hlasového signálu snižuje přesnost systému. U snímání záznamu závisí na prostředí, kde probíhá. Modelovací techniky mluvího částečně zaznamenávají charakteristiky prostředí. Odlišné předpoklady u registrace a testech většinou vedou ke snížení výkonosti. Pokud existuje možnost, využívají se normalizační techniky, které se pokouší zformovat model mluvího nezávislý na podmínkách nahrávání, ale kompletní dekoderace většinou není možná. Přejchod z pevné linky na mobilní telefon vede k rozdílným komunikačního kanálu nebo rozdílným reakčním křivkám. Typický problém ohledně telefonních aplikací je způsoben rozdíly mezi mikrofony aparátů. ^{[1][3]}

Spolupráce uživatele: Aplikace, ve kterých se usiluje o rozpoznání mluvího, bývají obvykle využívány v komerčních systémech. V těchto případech jsou upřednostňovány textově závislé systémy nebo systémy s textovou výzvou. Mluví ovšem musí mluvit srozumitelně a zřetelně, aby ho systém správně zaznamenal. Systémy tohoto typu většinou dosáhnou dobré úrovně výkonu. Nižší přesnost verifikačního systému zvyšuje variabilita hlasu, která je způsobena emocemi v hlase, které se objevují např. při konverzaci. ^{[1][3]}

Variabilita mluvího: Tato skupina se dělá na variabilitu dlouhodobou a krátkodobou. U dlouhodobé variability se zohledňuje vliv stárnutí, který je dostatečně pomalý, aby umožnil používání stejného modelu na dostatečnou dobu bez dalších nových registrací. Vyvíjení stárnutí lze sledovat díky inkrementálních registračních technik. Krátkodobá variabilita mluvího je ovlivněna různými vlivy. Výkonost systému může ovlivnit zdravotní stav uživatele - léky

nebo nemoci; potom například silné emoční události nebo únavy. Je obtížné porovnat opravdovou výkonnost systémů, jelikož existuje mnoho parametrů, které jí ovlivňují. Vědecká společnost se od roku 1997 orientuje na rozvoj technologie tohoto odvětví hodnocení. ^[1]

Množství registračních a testovacích dat: Významný vliv na přenost záznamu má délka registračních sezení. Kvalita se zvyšuje s narůstajícím počtem sezení a dobou jejich trvání. Registrační sezení v délce několika desítek sekund se používá pouze z důvodu jednoduššího používání aplikace. Registrační šablony ovšem mohou být zkreslené nebo nekompletní, a tak nedokážou pokrýt veškeré variability hlasu mluvčího. ^[1]

4.8 Využití aplikací v praxi

V posledních dvaceti letech technologie ověření mluvčího zažila velký pokrok a teď je využívána ve spoustě oblastech. Mezi nejdůležitější patří telefonní autentizace, řízení přístupu, správě řečových dat, personalizace a soudní vymáhání. ^{[1][5]}

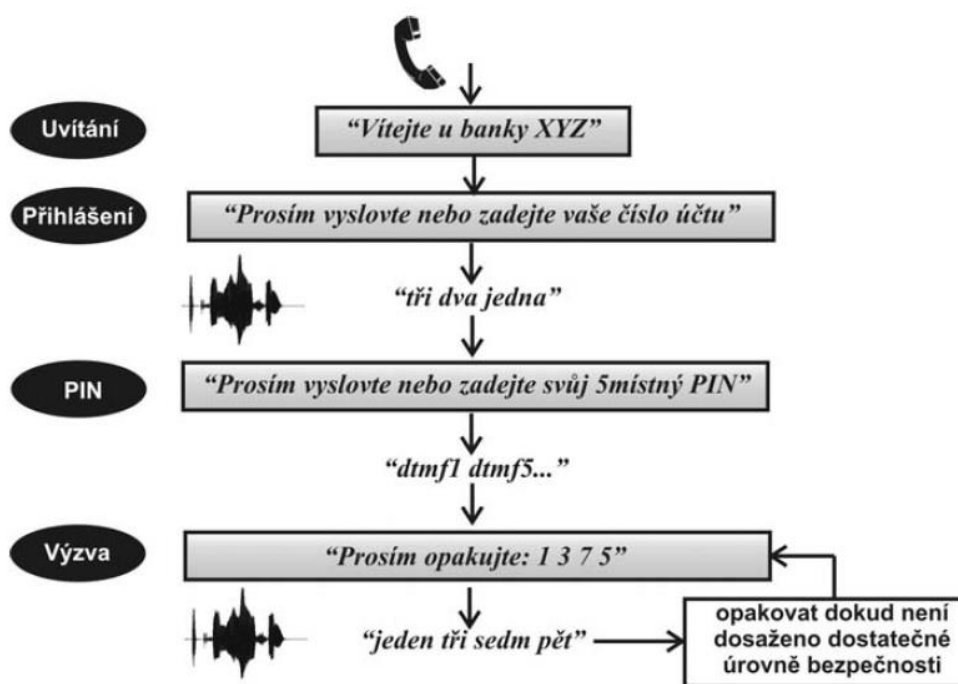
Rozpoznání mluvčího disponuje mnoha potencionálními aplikacemi. Mezi dvě nejvýznamější oblasti, kde se používá technologie rozpoznání mluvčího, se řadí bankovníctví a telekomunikace. Právě pro odvětví telekomunikace znamená rozpoznání mluvčího jedinou biometrickou disciplínu, kterou lze použít. Metoda se používá v aplikacích, kde se do určité míry ještě akceptují podvody. Přece jen rozpoznávání mluvčího není plně spolehlivá technologie, protože existuje velké množství faktorů, kterými je ovlivněno. Dále je možnost využít systém jako dodatečnou techniku k jiným autentizačním procesům. Mimo nízkých nákladů spočívá hlavní výhoda rozpoznání mluvčího v přívětivosti k uživateli. Mluvení je pokládáno za přirozené jednání a nevyžaduje žádný fyzický kontakt. ^{[1][5]}

4.8.1 Telefonní autorizace transakcí

Ať už banky nebo telefonní operátoři využívají již mnoho let automatizované telefonní systémy a umožňují tak zákazníkům provádět některé operace po telefonu. V těchto případech jsou využívány interaktivní hlasové odpovědní systémy (IVR – Interactive Voice Response) na telefonních platformách. Existují dvě skupiny těchto dialogových systémů. Prvním případem jsou dotykové klávesnicové systémy s využitím dvoutónových několika-frekvenčních signálů (DTMF – Dual Tone Multiple Frequency), druhý systém spočívá v aktivaci hlasem s pomocí rozpoznání řeči. Je možnost i využít oba systémy zároveň. Rozpoznání mluvčího u systémů

IVR je využíváno k autentizaci zákazníka u určité automatizované transakce nebo k zabezpečení přístupu k telefonním službám. Verifikační systém lze využít i během hovoru, kdy kontroluje případné změny hlasu mluvčího a tím snižovat riziko podvodu. Technologie verifikace uživatele může být dále používána jako náhrada nebo rozšíření typického autentizačního procesu spočívajícího ve využití PIN kódů. Jako obrana proti útokům přehráním se používají procedury s textovou výzvou. Také to znamená simplifikaci interakce s dialogovým systémem, kde je uživatel nucen vyslovit pouze to, co mu systém předkládá (viz obrázek 7). Úsek s textovou výzvou je zobrazen u části „výzva“, kde je zákazník vyzván k opakování 4 číslic. Tento krok často probíhá až do té doby, než systém získá dostatek dat pro určitý stupeň jistoty rozhodnutí. Pro dobrý výkon jsou většinou používány 3 výzvy o 4 číslicích. [1][7]

Obrázek 7 - Dialogový tok aplikace telefonního bankovníctví



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

Všechny hlasové signály mohou být využity jako materiál pro modul verifikace mluvčího pro rozhodnutí, jako například jeden ze stavů ořihlášení na obrázku 7. Pokud systém není schopen verifikaci dořešit, hovor se může přesměrovat na lidského operátora, který již problém vyřeší. Největší poskytovatelé automatizovaných telefonních služeb jsou telefonní operátoři.

Díky verifikaci mluvcího je možné zredukovat možné podvody, například u aplikací volacích karet. Pro provedení autentizace se používá číslo volací karty a PIN kód vyřknutý zákazníkem, jelikož transakce nejsou tak riskantní jako v případě telefonního bankovníctví. Není nutné používat protokol výzva-odpověď. Technologii ověření mluvcího mohou výhodně použít všechny telefonní služby pracující s kreditními kartami. Znáмым zdrojem neautorizovaného používání kreditních karet je například teleshopping. ^[3]

4.8.2 Řízení přístupu

Ověření mluvcího může být využito spolu s dalšími obvyklými mechanismy u systémů fyzického řízení přístupu. Například použití jmenovek či klíčů pak vede ke zvýšení bezpečnosti. V dnešní době jsou k dispozici na trhu například hlasem ovládané dveřní zámky nebo spínače u zapalování aut. Oblastí potencionálních aplikací jsou také autorizované aktivace mobilních telefonů. Textové aplikace obvykle spočívají ve využití tetově závislých provedůrách využívající jediné heslo. ^[3]

Pro přístup do počítačových sítí může být kombinovaně použito ověření mluvcího a obvyklý postup založený na hesle. V tomto případě je ovšem implementace méně rozpoznatelná. To je způsobeno vlivem rozdílné kvality mikrofonů, které ovlivňují signálové charakteristiky. Druhým důvodem je absence práv uživatelů k nastavení mikrofonů nebo jejich nezvyk je nastavovat. Posledním nejdůležitějším důvodem je fakt, že uživatelé nejsou zvyklí hovořit s počítači. ^{[1][3]}

Systém rozpoznání mluvcího je ve velkých firmách používán pro automatizaci resetování hesla. V případě zapomenutí hesla zaměstnancem, bezpečnostní postup vyžaduje nastavení hesla nového. Pokud je tento proces prováděn manuálně, může být poněkud nákladný. Pokud je jednoduše zautomatizován, může být považován za zdroj bezpečnostních problémů. Při použití plně automatizovaného systému ověření mluvcího je zachovaná dobrá bezpečnostní úroveň. ^[3]

4.8.3 Správa řečových dat a personalizace

Sledování mluvcího může být využito k organizaci informací v případě potřeby zjištění jaká osoba mluví a a kdy. Aplikace se využívají ve filmovém nebo mediálním průmyslu k indexování mluvcích nebo automatické detekci změny osoby, která mluví k automatickému

titulkování. Tuto technologii lze využít i pro záznamy z jednání. Na rozdíl od aplikací řízení přístupu mají tyto aplikace velmi odlišné charakteristiky. Pro vytvoření modelu může existovat velmi hodně tréninkového materiálu, doba zpracování pak není nijak omezena. Pro přesnější rozpoznání lze využít pokročilejší modelovací techniky. Z důvodu možnosti simultánní mluvy se považuje úloha sledování mluvčích za náročnější. Pokud není předem znám počet mluvčích na záznamu, sledování mluvčích se rozšiřuje o úkol segmentace. ^{[2][3]}

Systémy mohou být použity k automatickému rozřídění hlasové pošty a její další prohlížení či specifikaci. Inteligentní záznamníky jsou schopny rozeznat volajícího a umožnit personalizovanou odpověď. Využití aplikací je také v oboru personalizace reklam nebo služeb. Aplikace totiž rozpoznávají obecné charakteristiky mluvčích jako například věk nebo pohlaví. Další oblastí uplatnění jsou hry, kde rozpoznání mluvčích bývá použito k identifikaci vlastníka nebo lepší personalizaci interaktivity. ^{[2][3]}

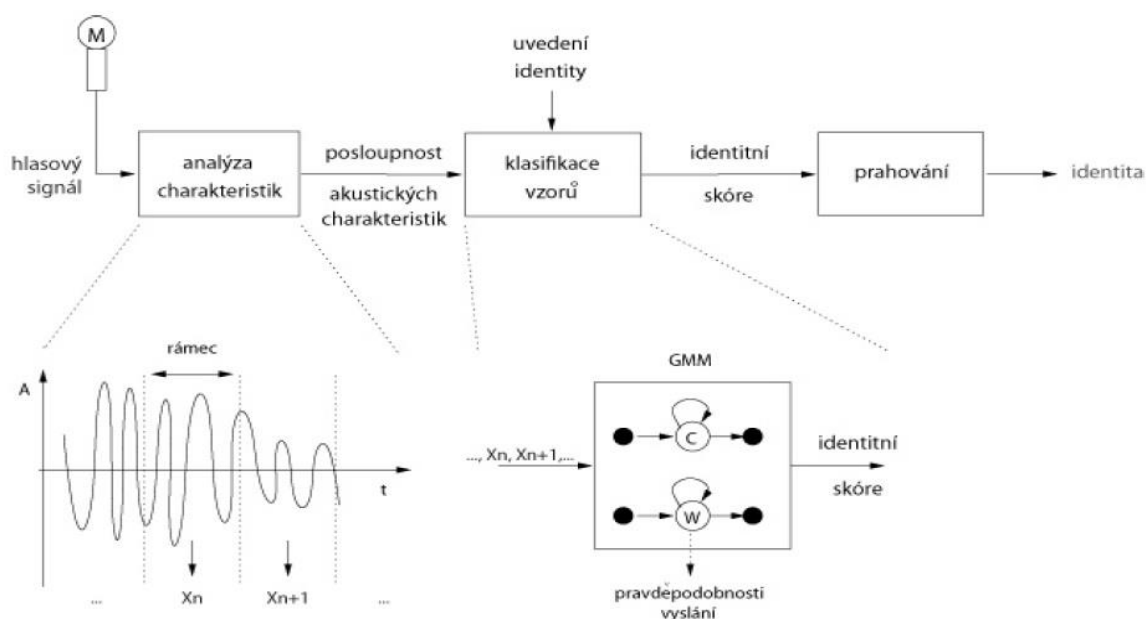
4.8.4 Forenzní rozpoznání mluvčích

U kriminálních případů jsou k dispozici záznamy hlasu pachatele trestného činu. Technologie ověření mluvčích pomáhá vyšetřovatelům určit viníka. Ve forenzní oblasti se využívá pojem otisk hlasu (voiceprint). Hlas ovšem není tak spolehlivý jako genetická data (DNA) nebo otisky prstů. V současné době se vědci shodují, že nelze použít verifikační shodu získanou automatizovaným systémem jako důkaz viny či nevinu. Při správném určení limitů technologie ji lze využít k vyšetřování. K forenzním aplikacím se řadí i automatizovaný odposlech například telefonních hovorů. ^[4]

4.9 Algoritmy

Přehled algoritmů se týká textově nezávislých systémů ověření mluvčích. Podobné algoritmy potenciálně umožňující podrobnější modelování určité vytyčené části vyřčeného textu využívají jak systémy textově závislé, tak systémy s textovou výzvou. Systémy ověření mluvčích lze rozdělit do třech modulů viz obrázek 8. Část charakteristik má za úkol vypočítat kompaktní a příhodnou reprezentaci hlasového signálu. Nejprve se uvede identita, ze které se nahraje klasifikátor vzorů modelu mluvčích a následně proběhne analýza vektorů charakteristik pro výpočet skóre. Skóre znázorňuje pravděpodobnost s jakou byl řečový signál vyřčen určeným mluvčím. V posledním modulu se skóre porovnává s prahovou hodnotou a rozhoduje se o přijetí či odmítnutí uvedené identity. ^{[1][7]}

Obrázek 8 - Systém ověření mluvčího



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

4.9.1 Extrakce charakteristik

Není dostatečně definovaná hranice mezi extrakcí charakteristik a klasifikací vzorů. Ideálně by měl extraktor charakteristik transformovat nezpracovaný vstup do lehce klasifikovatelného výstupu. Vzhledem k možnosti nesprávných předpokladů uskutečněných během extrakce charakteristik by klasifikátor měl být dostatečně dobrý. Extraktor charakteristik v případě hlasového signálu nejprve čelí dlouhé nestacionaritě. Kvůli tomu je hlasový signál většinou dělen na rámce o délce 10-30 msec a extrakce charakteristik se koná na každé části časového průběhu vlny. Algoritmus extrakce charakteristik také musí čelit krátké redundanci a získat zmenšenou a významnou akustickou informaci. ^{[1][7]}

Časový průběh vlny je z tohoto důvodu obvykle měněn z temporální domény na doménu frekvenční. Ve frekvenční doméně jsou krátkodobé temporální periodicity znázorňovány hodnotami větší energie ve frekvencích odpovídající periodě. O vyhlazení degradací signálu zapříčiněné komunikačním kanálem by se měla postarat extrakce charakteristik. U telefonního hovoru bude vymezena šířka pásma a variabilita kanálu bude potřebovat speciální manipulaci. Nakonec by se extrakce měla zabývat mapováním reprezentace řeči do formy, která bude kompatibilní s ostatními klasifikačními nástroji. ^[1]

Běžné techniky extrakce charakteristik jsou lineární prediktivní kódovací keprální analýzy a mel-frekvenční keprální analýzy. Takové algoritmy jsou v oblasti práce s řečí velmi rozšířeny. Výsledkem modulu extrakce charakteristik jsou temporální sekvence akustických vektorů $X = \{x_1, x_2, \dots, x_N\}$ délky N . Každý vektor x_n má potom neměnnou dimenzi D . Sekvence X znamená vstup do klasifikačního modulu. ^[1]

4.9.2 Ověření mluvího metodou GMM

Umělé neuronové sítě, podpůrné vektorové stroje, vektorová kvantizace, Gausovy mixovací modely, statistické metody druhého řádu, to všechno jsou příklady možností, jak vytvořit model mluvího. GMM jsou modely statistické a běžně se využívají v aplikacích ohledně rozpoznávání vzorů. GMM je schopen aproximovat libovolnou funkci hustoty rozdělení pravděpodobnosti, jestliže je k dispozici dostatečný počet mixů. Funkce hustoty pravděpodobnosti $p(x_n | M_{client})$, jinak zvaná taky pravděpodobnost D -dimenzionálního vektoru charakteristik x_n v modelu klienta M_{client} , je odhadována jako vážený součet mnohoproměnných gaussovských hustot jak ukazuje vzorec:

$$p(x_n | M_{client}) = \sum_{i=1}^I w_i N(x_n, \mu_i, \Sigma_i)$$

kde I znamená počet mixů, w_i váhu pro mix i a gausiánské hustoty N jsou parametrizovány střední hodnotou $D \times 1$ vektorem μ_i a $D \times D$ kovarianční maticí, Σ_i . Váhy mixů w_i splňují podmínku $\sum_{i=1}^M w_i = 1$. Mixy N mají tvar:

$$N(x_n, \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2} (x_n - \mu_i)' \Sigma_i^{-1} (x_n - \mu_i)\right)$$

Diagonální kovariační matice jsou někdy využívány pro tvoření hypotéz, které tvrdí, že charakteristiky nejsou korelované. Tento vzorec nám ilustruje, že GMM může být pochopen jako specifický případ skrytého markovovského modelu HMM, kde je pouze jediný stav automatu. ^{[1][11]}

Zjednodušeně řečeno, GMM modely utvářejí rozložení příznaků pro daného řečníka, kde se věrohodnost spočítá z emitovaných pravděpodobností. ^[1]

4.9.3 Prahování

Prahovací blok je poslední modul systému rozpoznání mluvčího, kde se provádí rozhodnutí, jestli daná identita bude schválena nebo odmítnuta. Rozhodování je prováděno porovnáním pravděpodobnosti R_c s prahovou hodnotou T :

$$R_c > T \rightarrow \text{přijetí}$$

$$R_c \leq T \rightarrow \text{odmítnutí}$$

V tomto případě mohou nastat dvě chyby. Systém může nesprávně odmítnout skutečného klienta. Tato událost je nazývána jako přehlédnutá detekce nebo nesprávné odmítnutí. Existuje velká množina pravých pokusů o rozpoznání uživatele z nichž je měřena pravděpodobnost odmítnutí klienta $P(\text{reject} | \text{client})$. Za druhou chybu se považuje, pokud systém akceptuje útočníka, také nazýváno jako falešný poplach nebo nesprávné přijetí. V tomto případě se pravděpodobnost $P(\text{accept} | \text{client})$ měří na velké množině podvodných pokusů. ^[11]

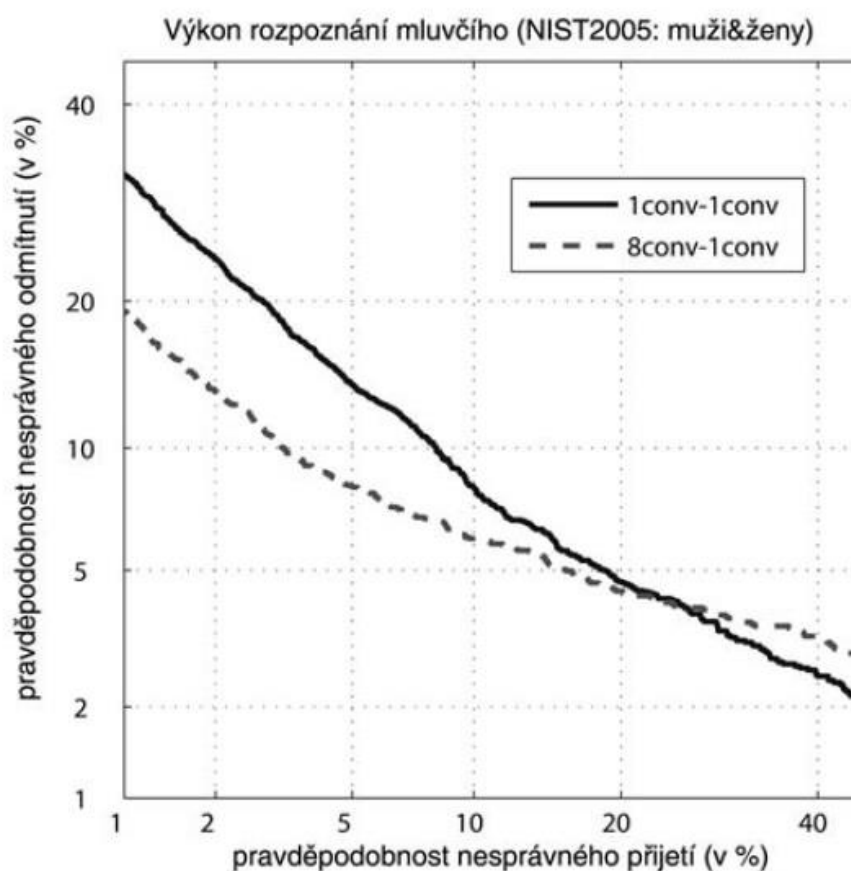
Pravděpodobnosti $P(\text{reject} | \text{client})$ a $P(\text{accept} | \text{client})$ nabývají různých hodnot podle konfigurace prahu T . Prahová hodnota T je definována tak, aby cenová funkce byla co nejmenší. Cenová funkce C_{det} znamená váženou hodnotu pravděpodobnosti nesprávného odmítnutí a nesprávného přijetí. ^{[11][11]}

$$C_{\text{det}} = C_{\text{fr}} P(\text{reject} | \text{client})P(\text{client}) + C_{\text{fa}} P(\text{accept} | \text{client}) P(\text{client})$$

Cena nesprávného odmítnutí se značí C_{fr} , cena nesprávného přijetí C_{fa} , $P(\text{client})$ je apriorní pravděpodobnost útočníka a klienta. Definování relativních hodnot C_{fr} a C_{fa} mají na starosti systémoví inženýři a probíhá většinou rozhovorem s projektovým manažerem a dle pocitu důležitosti obou typů chyb. V případě aplikací volacích karet chyby přijetí útočníka nebudou znamenat nikterak velkou katastrofu, ovšem odmítání reálných uživatelů by se mělo minimalizovat. Opačná situace by nastala v případě telefonního bankovníctví. Od projektového manažera jsou většinou k dostání odhady pravděpodobností pokusů podvodných i skutečných. Stává se, že je zapotřebí získat tyto hodnoty za pomoci testů v reálném prostředí nebo je odvodit z jiných zdrojů. ^[11]

Ve výše uvedené rovnici jsou pravděpodobnosti $P(\text{reject} | \text{client})$ i $P(\text{accept} | \text{client})$ funkcí prahové hodnoty T . Pro takovou testovací množinu lze graficky zakreslit křivku nesprávného přijetí (FA) a nesprávného odmítnutí (FR) určené dle hodnoty prahu T . Obecně se zaznamenává křivka FA na horizontální osu a pravděpodobnost FR na osu vertikální. Pakliže je úkolem porovnat podobné systémy, křivky mohou být nepraktické. Při porovnávání systémů sobě podobných se používá alternativní reprezentace přijímače, tzv. kompromis detekce chyb. Křivky budou lépe rozloženy, a proto je usnadněno pozorování rozdílů mezi systémy. Na obrázku 9 lze vidět příklad křivek, které byly získány systémem založeným na GMM. ^[1]

Obrázek 9 - Výkon rozpoznání mluvího



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

Shodnou pravděpodobnost chyby (EER) definuje ta prahová hodnota, kde jsou shodné hodnoty nesprávného přijetí a nesprávného odmítnutí. Hodnota EER se využívá při porovnávání výkonností odlišných systémů. Z praktického hlediska nelze považovat hodnotu EER za příliš reálnou, protože představuje situaci, kdy je prahová hodnota vypočtena

až po provedení porovnání s hlasovými daty, které byly předklasifikovány. Co se týče praxe, hodnota by měla být předem nastavena a to není jednoduché. Hodnota EER neodpovídá prahové hodnotě, která se rovná jedné a to z toho důvodu, že modely mají zaujaté odhady díky nedostatku trénovacích dat. Za další problém vyskytující se v oblasti ověření mluvčího lze považovat fakt, že výsledné skóre je rozdělené nestejně pro rozdílné klienty systému. Pro tuto situaci existují dva důvody. Prvním je zaujaté modelování z důvodu nedostatku trénovacích dat, druhým rozdíl mezi podmínkami při trénování a testování (mikrofony, odlišné kanály, emoční stav klienta apod.). Model každého zákazníka bude generovat jiný pravděpodobnostní poměr a z toho důvodu není globální prahování příliš vhodné. Tento důvod lze však obejít normalizací pravděpodobnostních poměrů a tím se zbavit vlivu rozdílů při rozdělení. Z praktického pohledu na věc proces spočívá v normalizování log-pravděpodobnostního poměru s pokusy útočnicků. Postup, který se nazývá z-normování probíhá následovně. ^[1]

Je nutno použít velké množství útočných pokusů za předpokladu použití gausiánského rozložení, pro získání odhadu průměru log-pravděpodobnosti a dále standardní odchylky pro model každého klienta. U odhadů hraje roli pohlaví útočnicka, většinou se odhady vypočítávají právě ze skupiny stejného pohlaví. Během testování je dále log-pravděpodobnostní poměr nula-normalizován za pomoci hodnot průměru a standardní odchylky. ^{[1][11]}

Existuje ještě efektivnější způsob, jak globální prahování na normalizovaném log-pravděpodobnostním poměru uskutečnit. Další normovací techniky jsou používány dle podmínek využití rozpoznávacího systému. Například ohledně úkolů založených na telefonu se využívá bimodální rozdělení log-pravděpodobnostních skóre. V souvislosti s telefonními mikrofony se hovoří o dvou módech, které odpovídají dvou různým typům. Elektretové nebo uhlíkové. Typ telefonního zařízení se dá jednoduše určit, proto lze normalizaci provést v závislosti na druhu přístroje. Normalizační techniky dokážou i snížit vliv délky hlasových vzorků. ^{[1][11]}

4.10 Metody kriminalistické audioexpertizy

Forenzní identifikace mluvčího probíhá dle ustáleného schématu týkajícího se zpracování jazykového materiálu určitého typu. Provádí se instrumentální akustické analýzy i analýzy používající sluchové metody. Jakou metodu nebo její dílčí součást zvolit se určí dle jazykové a

technické kvalitě sporné nahrávky. Kvalita sporné nahrávky také určuje, jakou část a které aspekty lze použít ze srovnávací nahrávky. ^[5]

Sluchové zpracování zvukového záznamu hlasu má následující postup. Poslechne se celý záznam pro vytvoření názoru a lepší pozdější orientaci. Poté se vyhodnotí obtížnost s jakou lze záznam zpracovat a následně se stanoví parametry ohledně hodnocení. Naleznuté znaky se zaznamenávají pomocí standartního systému znaků (IPA – International Phonetic Alphabet), který je často experty doplňován o jejich další specifické zkratky a značky. Jednotlivé znaky jsou posuzovány odděleně. ^[5]

System hodnocení individuálních hlasových a řečových parametrů, který se používá při vlastní analýze, musí být běžný pro daný jazyk. Například pro češtinu byla při hodnocení využita kritéria vypracovaná pro obecné posuzování mluvního výkonu osob. Kritéria byla přizpůsobena určitým podmínkám a účelu na základě zkušeností se zpracováním hlasového materiálu a v závislosti na identifikační hodnotě jednotlivých parametrů byla sestavena výsledná stupnice. ^{[1][5]}

Hodnocena je například kvalita hlasu nebo způsob dýchání. Dále i segmentální rovina, což znamená způsob artikulace hlásek a jejich skupin, prozodie řeči, jako frázování, melodie, dynamika, tempo, přízvuk, pauzy a nakonec celková úroveň projevu – výslovnost a stavba jazyka. ^[1]

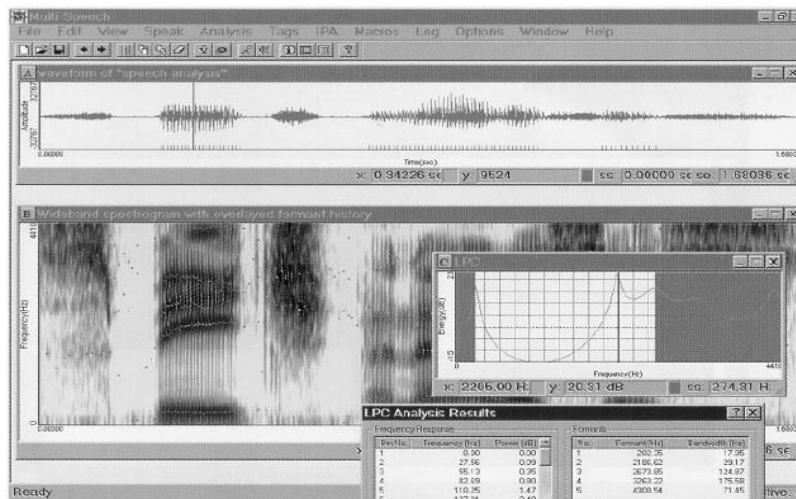
Pro finální posouzení výsledků analýzy jsou základem obecné tendence v jazyce a řeči, na jejichž základě se vyhledávají jednotlivé znaky a zvláštnosti. Za signifikantní v procesech identifikace mluvího můžou být považovány právě ty znaky a zvláštnosti, u kterých je výskyt ustálený a pravidelný v mezích individuální variability řeči. ^[5]

Základním předpokladem a předností instrumentálních měření je skutečnost, že akustické metody dokáží zaznamenávat v číselných hodnotách rozdíly v řeči mluvího, které jsou vnímány sluchem. Mohou však rozlišit i rozdíly, které nejsou pro lidskou bytost dost dobře rozpoznatelné. Jako test se uvádí metody, kdy dva mluvíci vysloví stejnou slabiku nebo samohlásku. Člověk vnímá oba hlasy jako jeden určitý foném systému. Mluvíci mohou mít i podobný hlasový tónbr. Hodnoty jednotlivých formantů jsou ovšem rozdílné a může se odlišovat i jejich průběh. ^[5]

Akustické metody ovšem nejsou bezchybné. Základní problém může být nedostatečná znalost o rozpětí akustického spektra, ve kterém se může vyskytovat řeč a hlas jednoho mluvčího. Doposud chybí přesvědčivé potvrzení o spolehlivosti technologií akustických analýz a stabilitě měření na jazykovém materiálu rozdílné kvality. ^{[1][5]}

Mezi přístroje využívající se k akustické analýze se řadí například sonagraf nebo frekvenční analyzátor. Analyzovat zvuk jde také se správným softwarem ve výbavě počítače. Takřka všechny jsou založeny na principu zjišťování spektra zvukového signálu. Zvukový signál v grafické podobě obsahuje časový rozměr (t) a okamžitou vlnovou amplitudu. Pokud se uvažuje takové zobrazení, nelze přímo odečíst počet ani vlastnosti jednotlivých spektrálních složek. Dle účelu zaměření analýzy se dále provádí podrobné výpočty. Existují především dva způsoby, jak dostat spektrum složeného řečového signálu ze signálu původního. Buď za pomoci analyzátoru nebo díky početní analýze. Analyzátor zpracovává zvuky například pomocí filtrů. Mezi metody analýzy patří například Fourierova analýza. Podstata této metody spočívá v možnosti jakoukoli periodickou časovou funkci (t) buď v celku nebo po částech rozdělit na jednoduché časové funkce se sinusovými průběhy a u periodicky nebo quasiperiodicky proměnného signálu za složky harmonické. V dnešní době se využívá analýz za pomoci počítače, a proto se různé metody početních analýz složeného signálu vracejí. Na obrázku 10 je ukázka výsledné analýzy provedené počítačem, kde v okně A je zobrazena zvuková vlna a v okně B formantové spektrum řečového signálu. Při využití analyzátoru se zjišťuje intenzita zvuku v jednotlivých frekvenčních pásmech, u kterých lze nastavit jejich šířku. Získá se dvourozměrné spektrum (intenzita a frekvence) bez informací ohledně časové následnosti jednotlivých složek, což pro některé druhy analýz není potřeba. Pokud je potřeba získat trojrozměrné spektrum, vhodným přístrojem je sonagraf. Takové spektrum ukazuje frekvenci i intenzitu jednotlivých složek v čase. ^{[1][5]}

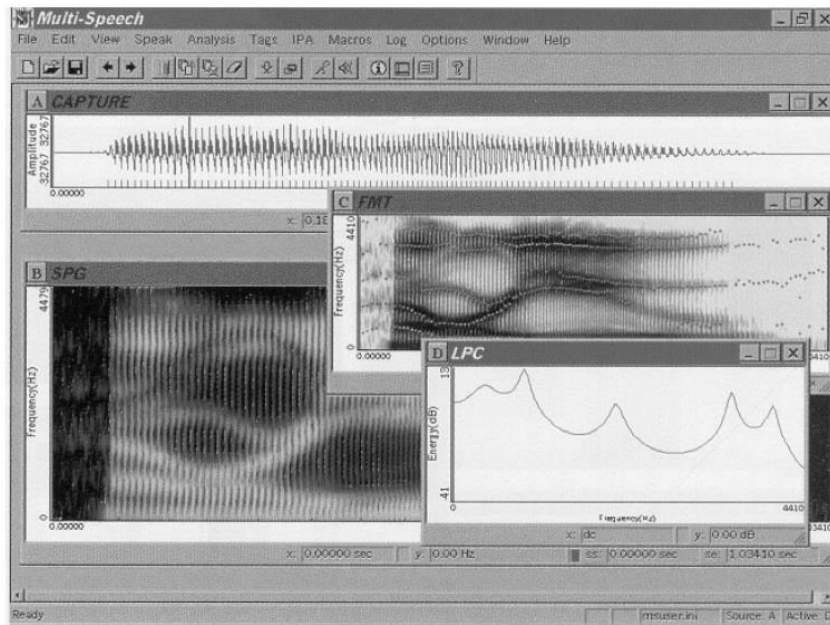
Obrázek 10 - Grafická a číselná analýza



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

Dnes je bezesporu nejvyužívanější metodou počítačové zpracování řečového signálu. Používají se nelineární modely analýzy hlasu. Využívají se rozdílné časové vektory a je nutno získat i další doplňující informace. Časové vektory obsahují údaje o spektrálním, keprstrálním, nebo lienárně prediktivním charakteru řečového signálu, popřípadě o dalších parametrech. K dispozici jsou nejrůznější počítačové systémy, z nichž k nejrozšířenějším patří Multi-Speech od americké firmy Kay-Elementrics. Na obrázku 11 jsou zobrazeny hlavní parametry analyzované počítačovou analýzou pomocí zmíněného systému. V okně A lze vidět zvukovou vlnu, v okně B formantové spektrum, v okně C formantové spektrum s grafickým zachycením průměrné hodnoty formantů a okno D obsahuje záznam dlouhodobého spektra zvukového signálu. [1][12][13]

Obrázek 11 - Ukázka z programu Multi-Speech



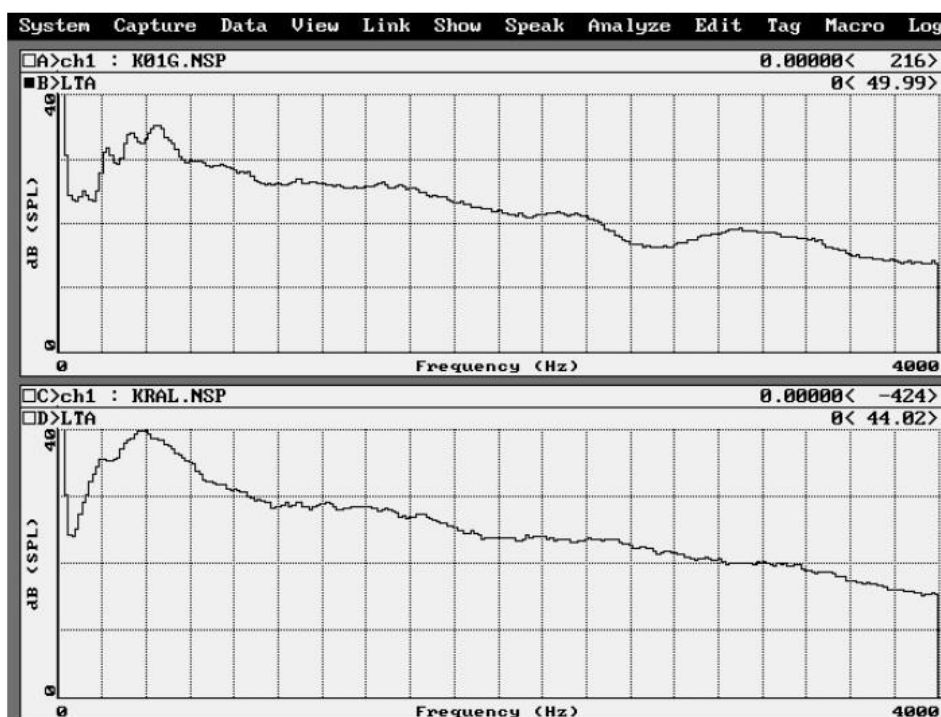
Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

4.10.1 Dlouhodobé spektrum řeči

Průměr okamžitých spekter akustického signálu za dostatečný časový úsek je z technického hlediska dlouhodobé řečové spektrum. U mluvčího se průběh spektra stabilizuje přibližně za časový úsek asi 30 sekund a vyhodnoceno je zhruba 220 okamžitých spekter. Hladší křivka zobrazená počítačem znamená větší počet zachycených spekter. Pakliže se nepodaří zachytit v určitém časovém intervalu potřebný počet okamžitých spekter, křivka má hranatější tvar. V případě, že signál je kvalitní a souvislý, výsledná křivka je hladká a dvojrozměrném spektru jsou výrazně vidět pouze vrcholy průměrných formantů. [1][8]

U vrcholů formantů je podstatnější jejich umístění než jejich tvar či výška. Nacházejí se ve frekvenčních oblastech a právě dlouhodobé spektrum řeči přibližně ukazuje, ve kterých oblastech se vyskytují. Finální vyhodnocení spočívá v porovnání lokálních maxim grafu. Celkový tvar spektra totiž ovlivňují různé faktory mimo mluvčího, jako hluk, šum v pozadí nebo technická kvalita nahrávky. Při hodnocení spekter jednotlivých záznamů tudíž nelze čekat absolutní shodu jejich tvaru. Na obrázku 12 je znázorněno dlouhodobé spektrum dvou rozdílných mluvčích. V horní části se nachází záznam sporné nahrávky, ve spodní je záznam nahrávky srovnávací. [1][12][14]

Obrázek 12 - Dlouhodobé spektrum dvou různých mluvčích



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

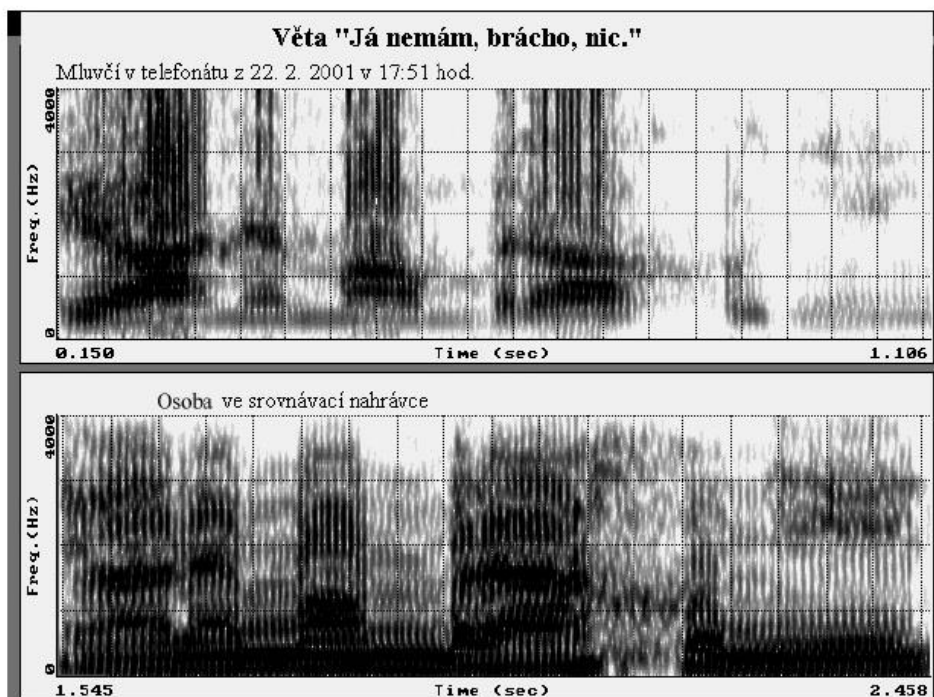
Vzhledem k výsledkům výzkumů ohledně využití forenzní identifikace zvukového materiálu češtiny, porovnání spekter nelze považovat za dostatečně přesné ani v rámci identifikace ani co se týče eliminace mluvčího. U nahrávky jednoho mluvčího pořízené za dobrých podmínek v laboratoři se v dostatečné míře nepodobá ani umístění lokálních maxim. Nelze přesně doložit z jakého důvodu, pravděpodobně za tím stojí nepřesnost měření zaviněné limitovanými možnostmi techniky. Expertíza je poměrně často využívána, ale většinou jen jako orientační analýza. Hlavní přednosti této analýzy jsou rychlost a nenáročnost na předchozí úpravu materiálu. ^{[1][12][13]}

4.10.2 Spektrální rozložení formantů

Zobrazení rozložení formantů se nazývá spektrogram, respektive sonagram. Příklad sonogramu lze vidět na obrázku 13. Z technického pohledu je to zobrazení tří složek spektra určitého časového úseku hlasového signálu. Na svislé ose se zobrazuje kmitočet (Hz), na vodorovné ose čas (s) a zčernání grafu znázorňuje hladinu zvuku (dB). Jednotlivým hlasivkovým kmitům v řečovém signálu odpovídají svislé čáry, které lze pozorovat

na sonagramu. Tyto čáry, přesněji jejich vzdálenosti a časové měřítko, pomáhají ke stanovení výšky hlasu. Jestliže v některých úsecích chybí, většinou je to z důvodu potlačení hlasivkového tónu a tedy nerozpoznání počítačovým programem. Zejména v rušeném signálu čáry znázorňující hlasivkové kmity často chybí. ^{[1][13][15]}

Obrázek 13 - Ukázka sonagramu



Zdroj: RAK, R.; MATYÁŠ, V.; ŘÍHA, Z. a kolektiv. „Biometrie a identita člověka ve forenzních a komerčních aplikacích.“ Praha, Nakladatelství Grada, 2012

Změna šířky pásma pásmové propustě při vytváření spektra způsobuje změnu tvaru. V případě malé šířky (např. 40 Hz) se zobrazují rozdrobené formantové oblasti na řadu pásem, které odpovídají jednotlivým harmonickým tónům. Jednotlivé formanty a formantové tranzienty ovšem nejsou dobře zřetelné. Proto se používá šířka vyšší, kolem 200-300 Hz. Při této hodnotě je na sonagramu rozložení jednotlivých formantů, jejich průběh i vzájemné rozestupy dobře čitelné. ^{[13][16]}

Zrakové posuzování spektrogramů je označováno za sporné. Musí jej vykonávat odborník s potřebnou kvalifikací a ideálně dlouho praxí v této činnosti. Prvotní experimenty s hodnocením spektrogramů byly zejména subjektivní záležitostí. Za použití matematických analýz byla vypracována přesnější kritéria hodnocení. ^{[1][13]}

V dnešní době lze určit přesné absolutní hodnoty formantů, jejich odchylky, rozpětí atd., díky moderním formám počítačových analýz. Opět ale závisí na kvalitě nahrávky. Vyskytují se i zkreslené hodnoty způsobené například zpracováním zašuměného signálu. Tento fakt je pravděpodobně jedním z důvodů, proč zatím neexistuje věrohodný způsob vyhodnocování spektrogramů.^{[13][17][18]}

5 Praktická část práce

V praktické části práce bylo provedeno porovnání z uživatelského a finančního hlediska třech vybraných biometrických systémů. Byly porovnány systémy na principu otisku prstu, skenu obličeje a identifikace hlasu.

5.1 Přístupový snímač otisků prstů

Pro porovnání se zvolila pokroková čtečka otisku prstu „F007EM-II“ od společnosti Sebury – viz. obrázek 14. Jedná se o biometrický snímač, který umožňuje využít technologii otisku prstu, bezkontaktních karet nebo jejich kombinaci. ^{[19][20]}

Obrázek 14 - Sebury F007EM-II



Zdroj: <http://katalog.abbas.cz/f007em-s26209/>

Čtečka vyniká v rychlé a přesné identifikace otisku prstu. Mezi další výhody se řadí ekonomická příznivost nebo možnost venkovního použití. K provozu není potřeba žádný software. Obsahuje optickou i akustickou signalizaci všech svých režimů. ^{[19][20]}

Vyniká elegantním a tenkým designem. Programovat čtečku lze třemi způsoby – master kartami, master otisky prstů nebo dálkovým ovladačem. Přístup pomocí dálkového ovladače je chráněn heslem. Čtečku lze využít i v režimu vzájemné blokace dvou dveří. ^{[19][20]}

Tabulka 1 - Specifikace čtečky Sebury F007EM-II

Napájecí napětí	12 V DC
Proudový adaptér	100 mA
Typ čtečky	Emarine
Vstupy	odchodové tlačítko, Wiegand 26-bit
Výstupy	spínaný napěťový, zatížitelnost 3A, Wiegand 26-bit
FAR (False Acceptance Rate)	< 0,0001%
FRR (False Reject Rate)	< 0,1%
Provozní teplota	-20 až 60 °C
Max. počet uživatelů	2 500
Senzor otisku prstu	Optoelektronický
Software	zcela bez potřeby SW
Max. počet otisků prstů	1 000
Další vlastnosti	rozlišení senzoru otisku prstu 450 DPI
Šířka	70 mm
Výška	155 mm
Hloubka	35 mm
Hmotnost	0,5 kg

Zdroj: <http://katalog.abbas.cz/f007em-s26209/>

5.2 Terminál na rozpoznání obličeje

Pro porovnání se zvolil speciální terminál na rozpoznání obličeje Broadway 3D B – viz obrázek 15, který umožňuje úžasné spojení pohodlí při identifikaci s vysokou bezpečností. 3D biometrie vylučuje svou podstatou možnost zapůjčení přístupové karty někomu jinému nebo použití fotografie. ^{[20][21]}

Obrázek 15 - Broadway 3D B



Zdroj: <http://katalog.abbas.cz/broadway-3d-b-s19566/>

Identifikace systémem Broadway 3D trvá pouze zlomek vteřiny, stejně jako rozpoznání lidmi. Je schopen rozeznat obličej uživatele i při jeho běhu. Na rozdíl od lidského vnímání je schopen rozlišit i nepatrný nesoulad v řádech milimetrů. Zásadou této funkce se řadí mezi nejpřesnější a nejbezpečnější identifikační biometrické systémy. ^{[20][21]}

Tabulka 2 - Specifikace čtečky Broadway 3D B

Doba rozpoznání	1 s
Doba zadání	3 s
Rychlost chůze	8 km/h
Velikost biometrické předlohy	4 kB
Operační systém	Linux
FAR (False Acceptance Rate)	0,0001%
FRR (False Reject Rate)	0,10%
Rozhraní pro administraci	webserver
Úhel záběru	32 °
Zdroj světla	blesková žárovka
Pracovní vzdálenost	0,8 až 1,6 m
Napájecí napětí	230 V AC
Provozní teplota	15 až 30 °C
Snímkování	15 sn./s
Doba expozice	0,2 ms
Další vlastnosti	rozhraní Wiegand I/O, 2x relé
Šířka	250 mm
Výška	1 623 mm
Hloubka	250 mm
Hmotnost	13 kg

Zdroj: <http://katalog.abbas.cz/broadway-3d-b-s195>

5.3 Hlasový systém Multi-speech

Software Multi-speech se řadí mezi programy běžící na operačním systému Windows. Jedná se o program určený k analýze hlasu a řeči. K pořízení nahrávky používá standardní multimediální hardware. ^[22]

Program Multi-speech poskytuje komplexní záznam řeči, zpětnou vazbu a následné vyhodnocování záznamu. Dokáže zpracovat stejnou funkční analýzu jako CSL software, což je jeden z nejvyužívanějších hlasových programů. Je limitován pouze pouze specifikacemi a funkcemi audio systému používaného počítače. ^[22]

Multi-speech software může být využíván pro výuku, výzkum, hlasové měření, artikulaci v dalším jazyce, klinickou zpětnou vazbu nebo pro forenzní vědu. Program obsahuje různá předepsaná makra, další si uživatelé mohou vytvořit sami. Například podporuje makro, které načítá nahrávku a filtruje jí tak, aby zněla jako telefonní hovor a tím ukazuje spektrum změn typických pro telefonní spojení. Všechny reporty a výsledky analýz jsou graficky zpracovány a automaticky je provedena statistická analýza. [22]

5.4 Charakteristika metody

K ekonomickému i uživatelskému zhodnocení byla zvolena multikriteriální analýza. Postup byl následovný. Sestavila se tabulka s pěti nejdůležitějšími parametry z pohledu hodnotitele. Zvolila se cena, šance na chybné přijetí, šance na chybné odmítnutí, doba snímání a zohlednila se i skutečnost, jestli musí při snímání dojít k přímému kontaktu s uživatelem.

Následně se každé kritérium obodovalo od 1 do 10, kdy 10 bodů znamená nejlepší možné hodnocení. Dále se určí váha kritéria tak, aby jejich součet dával číslo 1. Opět platí, že čím vyšší je váha, tím větší je důležitost parametru. Poté se už vypočítá výsledné bodové hodnocení, kdy se sečtou body každého parametru pro daný systém a násobí se s váhou příslušného kritéria. Seřazením bodových zisků od nejvyššího po nejmenší vyjde výsledné pořadí.

5.5 Zhodnocení systémů

Čtečka otisků prstů a terminál pro sken obličeje byly porovnány s hlasovým systémem pomocí bodové multikriteriální analýzy. Bylo zvoleno 5 nejdůležitějších parametrů, podle kterých se systémy hodnotily – viz tabulka 3. V každé analýze byly zvoleny jiné váhy kritérií podle daného hlediska k dosažení co nejpřesnějších výsledků.

Tabulka 3 - Vstupní hodnoty analýzy

	cena	FAR	FRR	kontaktní snímání	doba snímání
čtečka otisku prstů	4 465 Kč	< 0,0001 %	< 0,1 %	ano	> 1 s
terminál na rozpoznání obličeje	477 950 Kč	< 0,0001 %	< 0,1 %	ne	< 1 s
hlasový systém	135 000 Kč	< 0,0001 %	< 0,1 %	ne	> 1 s

Zdroj: vlastní zpracování autora

5.5.1 Ekonomické zhodnocení

Nejdůležitějším parametrem s nejvyšší váhou kritéria byla zvolena cena. Hodnoty FAR (False acceptance rate – chybné přijetí) a FRR (False recognition rate – chybné odmítnutí) byly do analýzy zvoleny jako výkonnostní prvky, které jsou podstatné při výběru přístupového systému. FAR a FRR mají pro všechny systémy stejnou hodnotu, avšak musí být jako nejdůležitější parametry bezpečnosti při výběru systému v analýze zmíněny. Dalšími kritérii v hodnocení byla doba snímání a zohledněna byla i skutečnost, jestli k identifikaci je třeba přímý kontakt uživatele s přístrojem.

V tabulce 4 lze vidět body přidělené hodnotitelem. Maximální počet bodů je 10, minimální počet je 1 bod. Váha jednotlivých kritérií byla zvolena podle důležitosti parametru. V tomto případě platí, že čím vyšší je hodnota váhy kritéria, tím důležitějším parametrem je podle hodnotitele.

Tabulka 4 – Ekonomické zhodnocení

	cena	FAR	FRR	kontaktní snímání	doba snímání	získané body	pořadí
čtečka otisku prstů	10	10	10	1	1	8,20	1.
terminál na rozpoznání obličeje	1	10	10	10	10	5,50	3.
hlasový systém	5	10	10	10	1	6,60	2.
váha kritéria	0,50	0,15	0,15	0,10	0,10		

Zdroj: vlastní zpracování autora

V analýze ekonomického zhodnocení se dospělo k jasným závěrům. Vzhledem k nejvyšší váze kritéria u ceny hrála právě ekonomická stránka nejdůležitější roli. Terminál na rozpoznání obličeje je ceněn na bezmála půl milionu korun, a proto se umístil nejhůře. Hlasový systém je ekonomicky dostupnější volba a se svou cenou přes sto tisíc korun obsadil druhou pozici. Nejvíce bodů získala a tedy nejlepší umístění z této analýzy ekonomického hlediska obsadila čtečka otisku prstů, vzhledem ke svým nízkým pořizovacím nákladům. Pro ekonomické zhodnocení platí tedy následující pořadí:

1. Čtečka otisků prstů
2. Hlasový systém
3. Terminál pro snímání obličeje

5.5.2 Uživatelské zhodnocení

V uživatelském zhodnocení byly váhy kritérií změněny. Nejvíce byla zohledněna pohodlnost zákazníka při používání jednotlivých systémů. Nejvyšší váhy tedy byly přiděleny kontaktnímu snímání a době snímání. Nejnižší váha byla přidělena ceně jednotlivých systémů. V tabulce 5 lze vidět přidělené body jednotlivým parametrům a výsledky analýzy.

Tabulka 5 - Uživatelské zhodnocení

	Cena	FAR	FRR	kontaktní snímání	doba snímání	získané body	pořadí
čtečka otisku prstů	10	10	10	1	1	4,60	3.
terminál na rozpoznání obličeje	1	10	10	10	10	9,10	1.
hlasový systém	5	10	10	10	1	6,80	2.
váha kritéria	0,10	0,15	0,15	0,30	0,30		

Zdroj: vlastní zpracování autora

V uživatelském hodnocení získal nejvyšší počet bodů terminál na rozpoznání obličeje a v analýze tedy zvítězil. Na druhém místě se umístil hlasový systém a poslední skončila čtečka otisků prstů. Platí tedy následující pořadí:

1. Terminál na rozpoznání obličeje
2. Hlasový systém
3. Čtečka otisků prstů

6 Zhodnocení výsledků

V praktické části byly blíže přiblíženy tři různé systémy. Každý ze systémů využíval jinou biometrickou oblast k identifikace uživatele. Zvolena byla čtečka otisků prstů, terminál na sken obličeje a hlasový systém. Tyto systémy figurovaly v obou analýzách.

Byla provedena dvě různá zhodnocení bodovou multikriteriální analýzou. Tato analýza umožňuje objektivně posoudit vhodnost jednotlivých systémů. Všechny tři systémy byly v obou zhodnoceních porovnávány se stejnými parametry. Jako nejdůležitější byly vybrány pravděpodobnosti FAR a FRR, cena systému, doba snímání a zohledněn byl i fakt, jestli při snímání je třeba přímého kontaktu uživatele s přístrojem. V ekonomickém zhodnocení byla nejdůležitějším kritériem cena. Nejvyšší počet bodů získala čtečka otisků prstů vzhledem ke své nízké hodnotě. Druhou příčku obsadil hlasový systém a poslední skončil sken obličeje. Pořadí ekonomického zhodnocení:

1. Čtečka otisků prstů
2. Hlasový systém
3. Terminál na rozpoznání obličeje

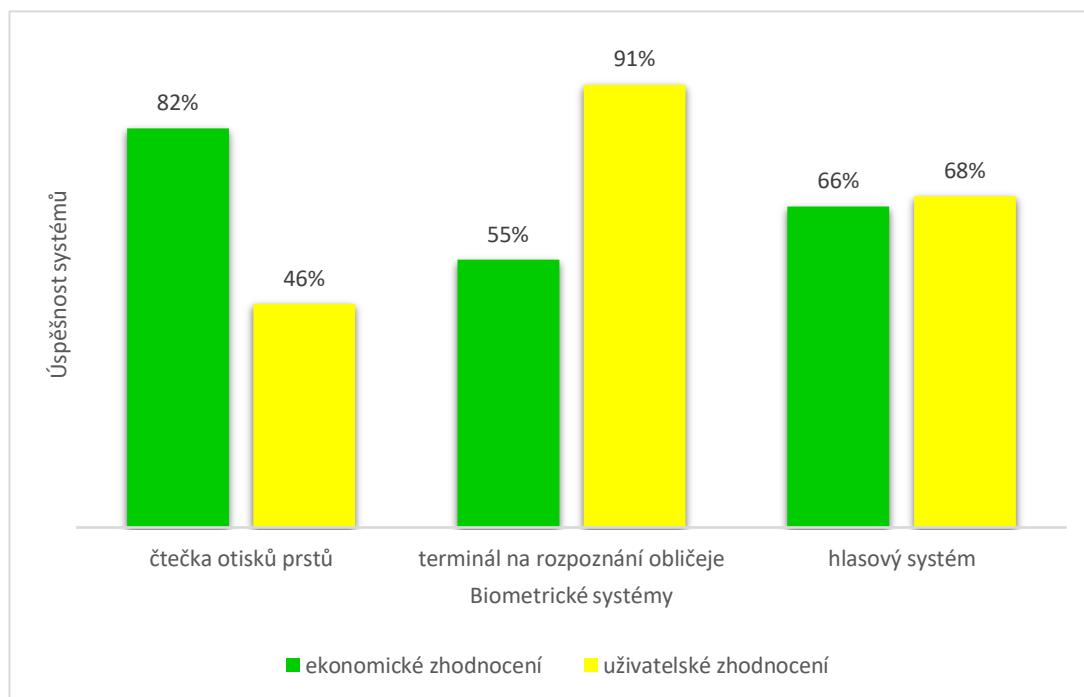
V hodnocení z uživatelského hlediska se dosáhly výsledky rozdílné. V této analýze byly za nejdůležitější parametry zvoleny doba snímání a kontaktní snímání. Nejlépe z analýzy vyšel sken obličeje, který dokáže snímat uživatele i za chůze a tak je velmi uživatelsky přívětivý. Druhé místo obsadil hlasový systém, jelikož není třeba přímý kontakt uživatele s přístrojem při snímání. Nejméně bodů získala čtečka otisku prstů. Pořadí uživatelského zhodnocení:

1. Terminál na rozpoznání obličeje
2. Hlasový systém
3. Čtečka otisků prstů

Na obrázku 16 je graf porovnání výsledků ekonomického a uživatelského zhodnocení pro jednotlivé systémy. Přesněji na něm lze vidět, z kolika procent byl který systém úspěšný při rozdílných vahách kritérií. Čtečka otisků prstů dosáhla v ekonomickém zhodnocení 82 %, ale z uživatelského hlediska analýzou prošla pouze se 46% úspěšností. Podobný rozdíl vyšel i u terminálu na rozpoznání obličeje, kde se u ekonomického zhodnocení získala hodnota 55 % a u uživatelského vysokých 91 %. Pokud by při výběru nehrála roli cena, systém na rozpoznání

uživatelé by byl podle této analýzy nejvhodnější volbou. Nejvyrovnanější výsledky přinesl hlasový systém. Rozdíl mezi uživatelským a ekonomickým zhodnocením jsou u něj pouze 2 % a dá se tedy považovat za jakýsi kompromis mezi testovanými systémy.

Obrázek 16 - Graf úspěšnosti systémů v multikriteriálních analýzách



Zdroj: vlastní zpracování autora

7 Závěr

Bakalářská práce byla zaměřena na problematiku identifikace osob na základě hlasu a řeči. V teoretické části byl blíže přiblížen pojem biometrie a základní pojmy tohoto odvětví. Byla zmíněna i historie fonetických expertíz a vysvětleno obecné rozpoznání mluvcího. Následně byla objasněna tvorba hlasu z biologického hlediska. Byly charakterizovány a vysvětleny principy jednotlivých systémů rozpoznání mluvcího. Byla zhodnocena výkonnost a bezpečnostní rizika verifikačních systémů a jedna kapitola se věnuje i využití aplikací v praxi. V závěru teoretické části byly přiblíženy algoritmy hlasových systémů a metody kriminalistické audioexpertízy.

V praktické části práce byly vybrány systémy k ekonomickému i uživatelskému zhodnocení. Byla aplikována multikriteriální bodová analýza. Zvolena byla čtečka otisků prstů, hlasový systém a sken obličeje. Za nejdůležitější parametry z pohledu hodnotitele byly zvoleny: hodnoty FAR a FRR, cena systému, doba snímání a zohledněna byla i nutnost přímého kontaktu uživatele s přístrojem. Ekonomicky nejdostupnějším systémem vyšla čtečka otisku prstů, avšak z uživatelského hlediska se umístila nejhůře, vzhledem k nutnosti přímého kontaktu uživatele s přístrojem při snímání. Ekonomicky nejméně dostupným systémem byl analyzován sken obličeje, ovšem uživatelsky je to nejprívětivější možnost, jelikož uživatel ani nemusí při snímání zastavovat chůzi. Hlasový systém představuje kompromis mezi předtím zmíněnými možnostmi.

Hlasové systémy se dle mého budou používat postupem času ve větší míře. Na základě zjištěných informací jsou uživatelsky podobně přívětivé jako skeny obličeje nebo skeny oční duhovky. Pravděpodobnosti chybného přijetí či odmítnutí mají již všechny systémy velmi podobné a dají se považovat za poměrně bezpečné.

8 Seznam literatury

- [1] RAK, R., MATYÁŠ, V. a ŘÍHA, Z. *Biometrie a identita člověka ve forenzních a komerčních aplikacích*. Praha: Grada, 2012. Profesionál. ISBN 978-80-247-2365-5.
- [2] ŠČUREK R. [Online]. červen 2008, [Citace 14. 02. 2017] URL: <http://www.biometricypodpis.cz/PDF/biometricke_metody.pdf>
- [3] BENEŠ, R. *Autentizační metody založené na biometrických informacích*. Access server. [Online] 18. 11 2010. [Citace: 14. 02. 2017.] URL: <<http://access.feld.cvut.cz/view.php?cisloclanku=2010110002>>
- [4] BALDWIN, J. R., a FRENCH, P. *Forensic phonetics*. London: Pinter, 1990.
- [5] SVOBODOVÁ, M. *Fonetika v kriminalistice*. In: Spisovná čeština a jazyková kultura 1993 : sborník z olomoucké konference 23.-27.8.1993. Sv. 2 / usp. Jančáková J. a kol. (eds.) - Praha: Filozofická fakulta Univerzity Karlovy, 1995.
- [6] BALDWIN, J., *Phonetics and speaker identification, Medicine, Science and the Law*. London, 1979.
- [7] ATA, B. S., *Automatic recognition of speakers from their voices*. 1976
- [8] NOLAN, F. J. D., *The phonetic bases of speaker recognition*. Cambridge: University of Cambridge, 1980
- [9] REICHL, J., *Vznik lidského hlasu*. [Online][Citace: 14. 03. 2017.] URL: <<http://fyzika.jreichl.com/main.article/view/188-vznik-lidskeho-hlasu>>
- [10] JANOTA, P. *Personal characteristics of speech*. Praha: Academia, 1967.
- [11] POLLÁK, P. *Biometrické metody založené na rozpoznávání hlasu*. [Online] 16. 11. 2016, [Citace 27. 02. 2017] URL: <https://cw.fel.cvut.cz/wiki/media/courses/a6m33bio/a6m33bio_spkverif_ii_handout.pdf>
- [12] SVOBODOVÁ, M. *Identifikace mluvčího v češtině na základě akustického spektra hlásek*, Acta Universitatis Nicolai Copernici: Studia Slavice (Toruń), 2000.

- [13] FURUI, S. *50 years of progress in speech and speaker recognition*. Patras, Řecko, 2005
- [14] KOENIG, B. E. *Enhancement of forensic audio recordings*. JAES, 1988
- [15] FURUI, S. *An overview of speaker recognition technology*. Martigny, Švýcarsko, 1994
- [16] REYNOLDS, D., QUATIERI, T. a DUNN, R. *Speaker verification using adapted Gaussian mixture models*. 2000
- [17] MÁLEK, J. *Využití úzkopásmového analyzátoru zvukového spektra pro identifikaci mluvího*. Československá kriminalistika, 1990
- [18] STEINER, D. *Akustika* [Online] [Citace 09. 03. 2017] URL: <<http://www.steiner.cz/david/akustika/>>
- [19] *Biometrická čtečka Sebury F007EM-II*, [Online] [Citace 20. 03. 2017] URL: <http://www.biometricke-ctecy.cz/produkty/ctecy-otisku-prstu/sebury-f007em-ii/>
- [20] *Abbas katalog*, [Online] [Citace 20. 03. 2017] URL: <<http://katalog.abbas.cz/>>
- [21] *3D čtečky obličejů*, [Online] [Citace 20. 03. 2017] URL:< <http://www.biometricke-ctecy.cz/produkty/3d-ctecy-obliceju/>>
- [22] *SwordMedical katalog*, [Online] [Citace 28. 03. 2017] URL: <<http://www.swordmedical.ie/>>

9 Seznam obrázků

Obrázek 1 - Zpracování řečového signálu.....	7
Obrázek 2 - Úkoly rozpoznávání mluvčího	8
Obrázek 3 - Schématický pohled na lidské hlasové ústrojí.....	10
Obrázek 4 - Schématický pohled na hrtan	11
Obrázek 5 - Příklady řečového signálu	11
Obrázek 6 - Výkony systémů verifikace mluvčího	15
Obrázek 7 - Dialogový tok aplikace telefonního bankovníctví.....	18
Obrázek 8 - Systém ověření mluvčího	21
Obrázek 9 - Výkon rozpoznání mluvčího	24
Obrázek 10 - Grafická a číselná analýza.....	28
Obrázek 11 - Ukázka z programu Multi-Speech	29
Obrázek 12 - Dlouhodobé spektrum dvou různých mluvčích.....	30
Obrázek 13 - Ukázka sonogramu.....	31
Obrázek 14 - Sebury F007EM-II.....	33
Obrázek 15 - Broadway 3D B	34
Obrázek 16 - Graf úspěšnosti systémů v multikriteriálních analýzách.....	40

10 Seznam tabulek

Tabulka 1 - Specifikace čtečky Sebury F007EM-II	34
Tabulka 2 - Specifikace čtečky Broadway 3D B.....	35
Tabulka 3 - Vstupní hodnoty analýzy	36
Tabulka 4 – Ekonomické zhodnocení.....	37
Tabulka 5 - Uživatelské zhodnocení.....	38