VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

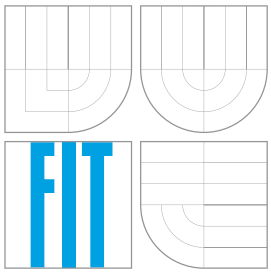
ODEMYKÁNÍ BRÁNY HLASEM

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE JAN BAUER
AUTHOR

BRNO 2015

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

ODEMYKÁNÍ BRÁNY HLASEM

GATE UNLOCKING BY VOICE

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

JAN BAUER

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. PETR SCHWARZ, Ph.D.

BRNO 2015

Abstrakt

Cílem této práce je vytvoření zařízení pro autentizaci řečníka podle hlasu. V řešení je použito knihovny BSAPI, která byla vyvinuta společností Phonexia. Knihovna je napsána v jazyce C++ a byla portována na zařízení Raspberry Pi B+. Správný chod je zajištěn skriptem napsaným v jazyce Python. Vytvořené řešení je určitě zajímavé a může se v budoucnu stát spolehlivým bezpečnostním systémem.

Abstract

The aim of this BSc. thesis is to create a device for authentication based on human voice. The solution is based on the BSAPI speech processing library developed by Phonexia. The library written in C++ was ported to the Raspberry Pi B+ device. The core functionality of the application was implemented in a Python script. The resulting solution is certainly interesting and may become a reliable security system in near future.

Klíčová slova

biometrie, hlas, BSAPI, Raspberry Pi, bezpečnost

Keywords

biometrics, voice, BSAPI, Raspberry Pi, security

Citace

Jan Bauer: Odemykání brány hlasem, bakalářská práce, Brno, FIT VUT v Brně, 2015

Odemykání brány hlasem

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Petra Schwarze, Ph.D.

.....
Jan Bauer
20. května 2015

Poděkování

Tímto bych chtěl poděkovat vedoucímu práce panu Ing. Petrovi Schwarzovi, Ph.D. za odborné vedení a cenné rady, které mi pomohly při tvorbě této práce. Dále bych chtěl poděkovat Ing. Tomášovi Ciprovi za pomoc a poskytnutí informací při portování knihovny BSAPI na zařízení Raspberry Pi. V poslední řadě děkuji všem lidem, kteří se zúčastnili testování.

Obsah

Kapitola 1

Úvod

Pro svou bakalářskou práci jsem si vybral téma „Odemykání brány hlasem“. Zaujala mě možnost rozšířit si obzory z oblasti bezpečnosti a zjistit, jak se biometrické systémy vyvíjejí. Biometrie je v dnešní době velmi žádaná a rychle se rozvíjející disciplína. Různé formy autentizace lze dnes najít všude kolem nás. Jedná se například o čtečky otisků prstů na mobilních telefonech nebo rozpoznávání obličejů. Biometrie skýtá obrovský potenciál. Již dnes velice spolehlivě slouží nejen v oboru kriminalistiky. Jednoznačně lze říci, že přispívá ke zvýšení bezpečnosti. Přináší s sebou taky určitou formu pohodlí. Biometrické vlastnosti je na rozdíl od různých přístupových karet těžší ztratit. Samozřejmě všechno má své pro i proti. I biometrie má svá úskalí. V dnešní době se dá zfalšovat i otisk prstu. Nehledě na to, že s každým poskytnutím svých jedinečných vlastností, přicházíme o značnou část svého soukromí.

Identifikace řečníka hlasem se v dnešní době uplatňuje převážně v oblasti bezpečnosti a obrany, v kriminalistických ústavech pro posouzení a prezentaci důkazního materiálu, v bankách pro zvýšení bezpečnosti přístupu k účtu nebo pro odhalovaní podvodníků během telefonických žádostí o úvěr, nebo pro vyhledávání ve videoarchivech. V těchto oblastech působí například firmy Phonexia, Agnitio, Nuance, SpeechPro nebo VoiceTrust. My chceme přinést výhody této technologie i do běžných domácností, kde hlasová biometrie může přinést větší bezpečnost nebo naopak komfort (například ztráta klíčů dětmi).

Cílem práce je navrhnout a implementovat jednoduchý dialogový systém pro odemykání a otevírání brány hlasem za použití knihovny BSAPI a minipočítače Raspberry Pi.

Kapitola 2

Biometrie

V této kapitole je popsána stručná historie a současné rozdělení biometrických metod. Malá část je také věnována jejich spolehlivosti. Dále jsou vysvětleny pojmy jako identita, identifikace a verifikace. V neposlední řadě je popsán biometrický systém a nakonec jakým způsobem jsou tyto systémy hodnoceny.

2.1 Histrorie

Slovo biometrie pochází původem z řečtiny ze slov *bios* = život a *metron* = měřítko. Dalo by se tedy říci, že biometrie je automatické rozpoznávání osob na základě jejich jedinečných biologických rysů. Mezi nejznámější metody autentizace patří např. otisky prstů, DNA nebo charakteristika písma. Lidé, aniž by si to uvědomovali, denně rozpoznávají jiné osoby právě na základě těchto anatomických vlastností: obličeje, hlasu, postavy, pohybu atd. [1]

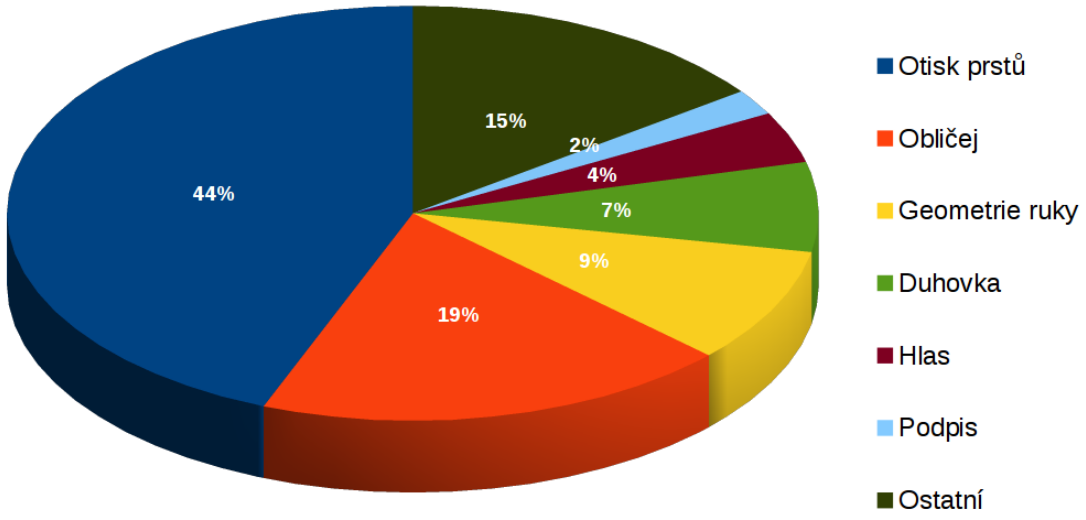
Biometrické systémy lidé používají již od nepaměti. Zmínky o ní můžeme najít již ve Starém zákoně. Je zde popsáno zavraždění tisíců osob právě na základě špatného vyslovení slova „shibboleth“, které používali Izraelité k rozpoznávání uprchlíků od ostatních. Otisky prstů používali už Babyloňané pro potvrzování obchodních smluv. O tento obor se velice zajímal český přírodovědec Jan Evangelista Purkyně. Ovšem jeho zájem byl čistě přírodovědecký, ačkoliv uznal možnost dělení obrazců papilárních linií podle geometrických vlastností. Velkým průkopníkem, co se daktyloskopie týče, byl William J. Herschel. Používal otisky prstů pro stvrzení o převzetí mzdy. Pracoval totiž jako koloniální úředník v Indii. Každý dělník musel při převzetí mzdy otisknout svůj palec na výplatní pásku. O zavedení daktyloskopie do praxe se zasloužil Juan Vucetich, který v roce 1891 v Argentině snímal otisky prstů obviněným osobám, které potom mohli využívat policejní úřady. V USA byla v roce 1924 zavedena identifikační divize, využívající právě otisků prstů, u FBI. Jejich sbírka obsahovala v roce 1946 100 milionů záznamů se všemi deseti otisky prstů [2].
Dnes používají systém IAFIS (*Integrated Automated Fingerprint Identification System*), který je v provozu 24 hodin denně a 365 dní v roce. Aktuálně tento systém obsahuje něco kolem 104 milionů otisků prstů. [3]

2.2 Současnost

Biometrie dnes už nenachází své uplatnění jen v oblasti kriminalistiky a soudnictví. Díky čím dále rychleji rozvíjejícím se technologiím, biometrická identifikace nachází své uplatnění i v běžném životě. Mezi její hlavní výhody určitě patří:

1. je velice obtížné ji ztratit
2. zvyšuje bezpečnost
3. není přenositelná
4. přináší větší komfort
5. je velice těžké ji zfalšovat

Jedny z nejznámější biometrické vlastností jsou například: otisky prstů, DNA, obličej, duhovka, dlaň, hlas atd. Jejich podíly využití na trhu můžeme najít na přiloženém obrázku 2.1 a detailnější srovnání jednotlivých metod v tabulkách 2.1 a 2.2:



Obrázek 2.1: Rozdělení biometrických aplikací na trhu [4].

Metoda	Univerzálnost	Jedinečnost	Stálost	Dostupnost	Přesnost
DNA	V	V	V	N	V
Obličej	V	N	S	V	N
Duhovka	V	V	V	S	V
Otisk prstu	S	V	N	V	S
Hlas	S	N	N	S	N
Sítnice	V	V	S	N	V
Podpis	N	N	N	V	N
Geometrie ruly	S	S	S	V	S

Tabulka 2.1: Tabulka srovnání biometrických vlastností [5].

Metoda	Přijatelnost	Odolnost
DNA	N	N
Obličej	V	V
Duhovka	N	N
Otisk prstu	S	S
Hlas	V	V
Sítnice	N	N
Podpis	V	V
Geometrie ruly	S	S

Tabulka 2.2: Tabulka srovnání biometrických vlastností [5].

Vysvětlivky k tabulkám: V - vysoká, S - střední, N - nízká

Identifikování osob pomocí biometrie dnes nabývá na významu a zajisté má velikou budoucnost. Lidé chtějí mít stále lepší systémy zabezpečení a přesně toto jim biometrie nabízí. Čím dál více roste počet online plateb. Společnost MasterCard, zabývající se platebními kartami, již spustila službu mobilních plateb na základě otisku prstu. Díky aplikaci *MasterPass*, která je jednou z prvních biometrických platebních aplikací na světě, budou mobilní platby zase o něco pohodlnější a bezpečnější. MasterCard zatím neeviduje žádné zneužití této technologie [6].

2.3 Identita, Identifikace, Verifikace

Tyto pojmy jsou úzce spojeny s biometrií. Slovo *Identita* vychází z latinského slova *identitas*, které bylo odvozeno ze slova *idem* - stejný. Každý z nás má svou identitu a jsme ji jednoznačně charakterizováni [2].
Identita je založena na těchto principech [1]:

1. **Něco co víme** (heslo, PIN)
   Tento princip je založen na získání informace, kterou je nutné si zapamatovat. Jedná se obvykle o bezpečnostní hesla. V tomto případě existuje nebezpečí získání takovýchto informací útočníkem. Nehledě na skutečnost, že danou informaci můžeme zapomenout.

2. **Něco co máme** (klíč, kartu)
   Základem je vlastnit něco, co nikdo jiný nemá. Zde existuje opět možnost odcizení dané věci, či ztráta. Není ani vyloučena možnost nelegálního okopírování.

3. **Něco co jsme** (chování,vzhled)
   U tohoto principu jsme sami sobě vstupním klíčem. Samozřejmě ztráta či zapomnění sebe sama asi nepřipadá v úvahu. V dnešní době ale není až tak velkým problémem vytvořit kopii otisku prstu. Nejlepší metodou zabezpečení je samozřejmě kombinace několika principů dohromady.

Pojem *Identifikace* znamená proces zjištění identity. Samotný proces spočívá v určení identity na základě biometrické vlastnosti, přičemž systém porovnává daný vzorek se všemi ostatními vzorky, které jsou uloženy v databázi. Proto se taky říká, že identifikace je porovnávání 1:N. Výsledkem takového porovnávání je buďto nalezená identita, nebo identita není nalezena. Porovnávání je samo o sobě dosti časově náročné, obzvlášť v případech rozsáhlých databází. V takových případech je databáze rozdělena do podkategorií a poté se vždy vyhledává jen v odpovídající podkategorii [1].

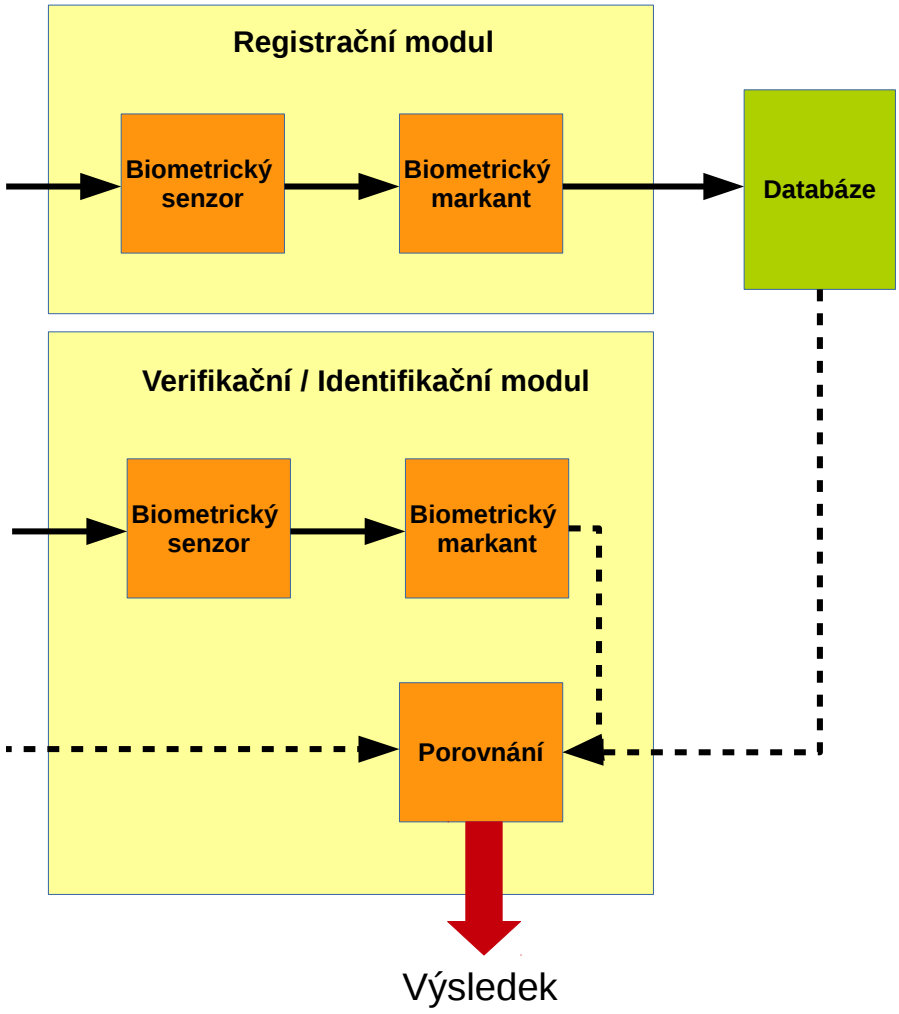
Naprosto rozdílným procesem je *verifikace*. Je to proces porovnávání jednoho vzorku s jiným vzorkem. Jedná se tedy o porovnání 1:1. Tento proces je samozřejmě mnohem rychlejší než proces identifikace. Nezáleží totiž na délce vstupu. Potřebný čas je stále stejný. Výsledkem může být vpuštění daného uživatele, nebo zamítnutí přístupu.

V oblasti biometrie ještě často narazíme na pojem *autentizace*. Autentizace určuje hodnověrnost osoby. Při pouhém porovnávání dvou hesel je autentizace jednoduchou záležitostí. Naopak u biometrického porovnávání se jedná o složitější úlohu [1].

2.4 Biometrický systém

Biometrický systém 2.2 se obvykle skládá ze dvou modulů [1]:

1. **Registrační modul**
2. **Verifikační/identifikační modul**



Obrázek 2.2: Biometrický systém [1].

Jak je již z obrázku patrné, oba moduly obsahují *biometrický senzor* a *biometrický markant*. Biometrický senzor slouží k získání a převedení vzorku do digitální podoby. Pod pojmem biometrický markant se ukrývají již vyextrahované rysy ze získaného vstupního vzorku. Registrační modul tento markant uloží do databáze. Verifikační modul dělá totéž co registrační modul, ale neukládá markant do databáze. Z databáze si načítá data k porovnání aktuálního markantu. Po porovnání dostaneme nějaký výsledek.

Výsledkem bývá skóre. Skóre udává podobnost mezi vzorkem v databázi a právě získaným vzorkem. Takovéto skóre je poté porovnáno s prahem, který jsme si předtím zvolili viz. obrázek 2.3.

Obrázek 2.3: Oblasti přijetí a odmítnutí na základě porovnání skóre s prahem [1].

Při porovnání může dojít k dvěma chybovým stavům:

1. Systém odmítne správného uživatele - chybné odmítnutí.
2. Systém pustí neoprávněného uživatele - chybné přijetí.

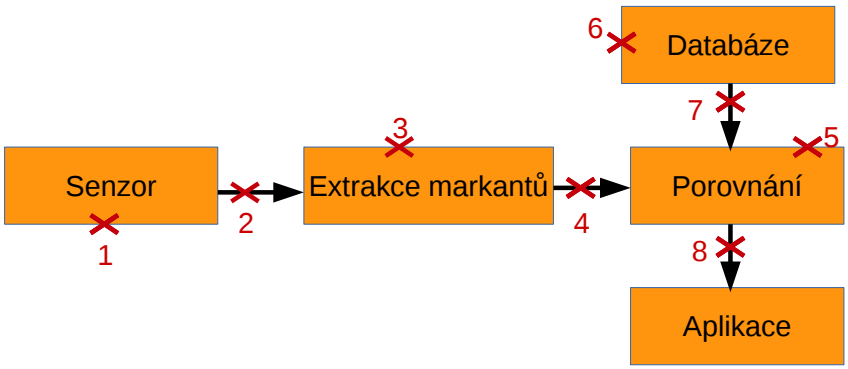
V souvislosti s těmito chybami bylo důležité zavést hodnotící metriky, které jsou popsány v podkapitole 2.5.

Nikdo není dokonalý a biometrický systém samozřejmě také ne. Na obrázku 2.4 jsou vyznačena místa, kde může dojít k problémům.

1. **Falešná biometrická vlastnost** - například kopie otisku prstu.
2. **Znovupoužití starých dat** - vyměňování informací mezi senzorem a extraktorem může být zachyceno a znovu použito.
3. **Úprava extraktoru** - extraktor je upraven tak, aby se útočník dostal do systému.
4. **Syntetický vektor rysů** - výsledek extraktoru je nahrazen jiným.
5. **Změna porovnání** - výsledek porovnání může být upraven.
6. **Modifikace šablony** - data uložená v databázi můžou být změněna
7. **Blokování kanálu** - kanál mezi databází a porovnáním může být zahlcen nesmyslnými dotazy.
8. **Změna výsledku** - je změněn výsledek celé operace.

Biometrický systém může být postaven na různém spektru biometrických vlastností. Ty se dělí do dvou skupin [1]:

1. **Statické vlastnosti**
   Statické vlastnosti jsou rozpoznatelné vždy, nehledě na stav člověka. Řadíme zde např. otisk prstu, obraz sítnice oka, DNA.
2. **Dynamické vlastnosti**
   Naopak dynamické vlastnosti jsou spojeny s určitým chováním osoby. Sem zcela jistě patří charakteristika písma, chůze a námi zkoumané rozpoznávání podle hlasu.



Obrázek 2.4: Možnosti napadení biometrického systému [1].

Taktéž biometrické systémy dělíme do dvou skupin [1]:

1. **Unimodální**
   Systémy patřící do této skupiny používají pro ověření jen jednu biometrickou vlastnost. V běžném životě se nejvíce setkáváme právě s těmito systémy.

2. **Multimodální**
   Jak již název sám napovídá, tyto systémy využívají buď více biometrických vlastností najednou, nebo více znaků jedné biometrické vlastnosti. Takové systémy jsou samozřejmě mnohem odolnější vůči jakýmkoliv útokům.

2.5 Hodnocení spolehlivosti biometrických systémů

1. **Míra chybného přijetí - FAR** [1]
   FAR (False Acceptance Rate) určuje pravděpodobnost, se kterou systém vyhodnotí dva odlišné vzorky jako shodné, a tedy vpustí neoprávněného uživatele.
   Její výpočet je jednoduchý:

$$FAR = \frac{\text{Počet chybně určených shod dvojic vzoru}}{\text{Celkový počet rozdílných dvojic vzoru}}$$

2. **Míra chybného odmítnutí - FRR** [1]
   FRR (False Rejection Rate) určuje pravděpodobnost, se kterou systém vyhodnotí dva vzorky od téže osoby jako různé, a tedy systém odmítne vpustit správného uživatele.
   Ke zjištění takové pravděpodobnosti použijeme vzorec:

$$FRR = \frac{\text{Počet chybně zamítnutých shod dvojic vzorku}}{\text{Celkový počet dvojic vzorku od stejné osoby}}$$

3. **Míra chybné shody - FMR** [1]
   FMR (False Match Rate) vyjadřuje ten stejný podíl jako FAR s jediným rozdílem, že se zde nezapočítávají pokusy, které selhaly ještě před porovnáním. Jedná se tedy o FAR zmenšené o počet FTA a FTE.

4. **Míra chybné neshody - FNMR** [1]
   FNMR (False Non-Match Rate) vyjadřuje ten stejný podíl jako FRR s tím rozdílem, že se zde nezapočítávají pokusy, které selhaly ještě před porovnáním. Jedná se tedy o FRR zmenšené o počet FTA a FTE.

5. **Míra neschopnosti nasnímat - FTA** [1]
   FTA (False to Acquire) nastává v případě odmítnutí biometrické charakteristiky, přestože je charakteristika přítomna. Její hodnota vyjadřuje, jak moc je daný senzor vhodný. Čím vyšší je míra neschopnosti nasnímat, tím méně vhodný je daný senzor.

6. **Míra neschopnosti zaregistrovat se - FTE**[1]
   FTE (Failure to Enroll) vyjadřuje podíl vzorků, které se systém není schopen naučit. Například v případě snímání hlasu nemusí být vždy hlas v nahrávce rozpoznám kvůli okolním zvukům.

7. **Míra vyrovnání chyb - EER** [2]
   EER (Equal Error Rate) vyjadřuje ideální hodnotu prahu. Na hodnotě prahu jsou závislé právě hodnoty FMR a FNMR. Čím vyšší stanovíme práh, tím menší bude pravděpodobnost, že systém vyhodnotí dva odlišné vzorky jako shodné, tedy sníží se FMR. Naopak v tomto případě vzroste pravděpodobnost FNMR. Při nízkém prahu nastane přesně opačná situace. Tedy zvýší se FMR a sníží se FNMR. EER vyjadřuje hodnotu prahu, při které dojde k rovnosti FMR a FNMR.

8. **ROC křivka** [2]
   ROC (Reciever Operating Curve) určuje kvalitu systému. Křivka obvykle vyjadřuje závislost FMR k FNMR (nebo FAR k FRR) a stala se standardem v popisu chování detekčních systémů.

Kapitola 3

Rozpoznávání řečníka

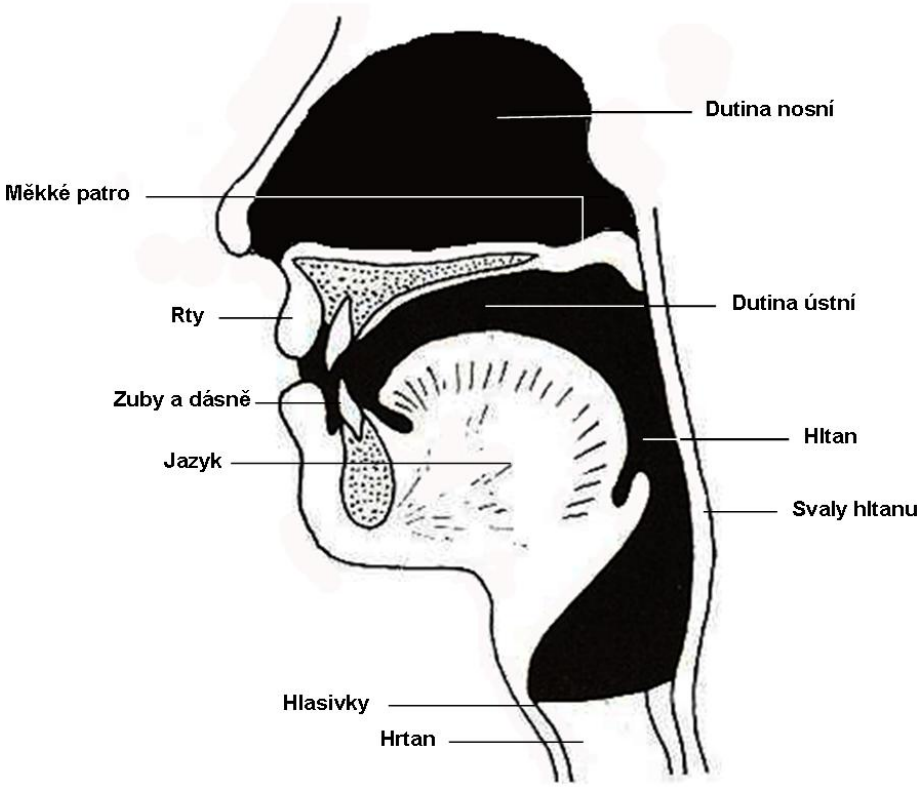
Rozpoznávání mluvčího je často zaměňováno s rozpoznáváním řeči. Přitom se jedná o dva naprosto odlišné pojmy. Rozpoznávání řeči se zabývá převedením řeči do textu. Touto oblastí se v práci nebudu zabývat. V této kapitole je popsáno něco málo z historie rozpoznávání řečníka, dále jak je vytvářen hlas a nakonec popíšu různé typy systémů k rozpoznávání řečníka.

3.1 Historie

Během druhé světové války byl vytvořen přístroj na vizualizaci spektrogramu [7]. Spektrogram je graf, který zobrazuje frekvence zvukového signálu v průběhu času. Po válce vědci tvrdili, že spektrogramy jsou jednoduchým nástrojem na identifikaci osob. Se spektrogramy se spojil název *voiceprint*. Termín voiceprint byl odvozen od otisku prstu - fingerprint [7]. Mezi roky 1960 až 1980, kdy se začaly objevovat první počítače, se tato oblast začala více rozvíjet. Přišly první přístupové systémy využívající tuto metodu. V osmdesátých letech přišla tato technologie i do telekomunikací. V této době byla vyvinuta jedna z klíčových technologií tzv. skryté Markovovy modely (HMM) [8]. V devadesátých letech byly díky Linguistic data consortium (LDC) odtajněny společné databáze hlasové verifikace [2]. Tento krok měl velký význam hlavně z pohledu budoucího výzkumu.

3.2 Tvorba hlasu

Hlas nám slouží především ke vzájemné komunikaci. Je tvořen proudem vzduchu vytlačovaného z plic pomocí sérií neuromuskulárních pokynů, které vedou k rozvybrování hlasivek. Následně je vzduch různě modulován tvářemi, čelistí, jazykem, patrem a rty [8]. Hlasové ústrojí lze vidět na obrázku 3.1.



Obrázek 3.1: Pohled na hlasové ústrojí člověka [9].

3.3 Systémy

Systém rozpoznání řečníka má za úkol identifikovat nebo verifikovat identitu uživatele na základě hlasových charakteristik. Tyto systémy rozdělujeme na textově závislé, s textovou výzvou a textově nezávislé [2].

3.3.1 Textově závislé

Textově závislý systém, jak již sám název napovídá, je postaven na zadání stejné fráze jak pro registraci, tak pro samotnou autentizaci. Při každé autentizaci se používá stejná sekvence zvuků, a tudíž jsou extrahované charakteristiky stabilnější. I proto výkon takovýchto systémů bývá na dobré úrovni. Do této kategorie řadíme i systémy, kde si uživatel nevolí vstupní frázi. V takovémto případě bývá heslo velmi krátké a bezpečnost je založena na nevědomosti neoprávněných uživatelů [2].

Mezi výhody textově závislých systémů patří již zmiňovaná větší přesnost, nevýhodou může být zapomenutí požadované fráze, či hesla.

3.3.2 S textovou výzvou

U systémů s textovou výzvou je požadovaná fráze vybrána systémem a poté sdělena uživateli. Vybraná sekvence slov je při každém použití jiná. Uživatel tedy nezná své heslo dopředu [10]. Tato metoda by se dala považovat za určitou formu kontrolu živosti osoby.

Jasnou výhodou systémů s textovou výzvou je, že uživatel nemůže zapomenout své heslo.

3.3.3 Textově nezávislé

Poslední možností jsou textově nezávislé systémy. Uživatel zde má naprosto volnou ruku. Může vyslovit jakoukoliv frázi chce. Odpadá tedy možnost zapomenutí potřebné fráze. V případě nejistoty, může systém vyžadovat další vstupní data, dokud nebude dosaženo chtěné hranice úspěšnosti. Nevýhodou textově nezávislých systémů bývá nižší výsledná přesnost [2].

Kapitola 4

Knihovna BSAPI

Pro účely této práce byla použita knihovna BSAPI, která je vyvíjená společností Phonexia s.r.o. Společnost Phonexia se zabývá zvukovou biometrií. Jejím cílem je zdokonalit analýzu řeči natolik, aby se dala používat v každodenní praxi. Tato metoda je založena na rozpoznávání konkrétních charakteristik zvukového signálu, který vydává zkoumaný objekt. V dnešní době se tato oblast velice rychle vyvíjí a má obrovský potenciál uplatnění. Běžně se používá v různých státních institucích či v bezpečnostních složkách. Důležitou roli zastává hlavně v oboru kriminalistiky. Společnost Phonexia ale chce tuto metodu přivést do běžného života, ať už pro zabezpečení domů či vstupu do chráněné oblasti.

Na zařízení Raspberry Pi byla portována optimalizovaná verze knihovny pro funkčnost i na výkonově slabších zařízeních, ke kterým Raspberry Pi bezesporu patří. Nakonfigurovali jsme systém s menším statistickým modelem a optimalizovaným výpočetním schématem pro rychlost.

Z pořízeného záznamu hlasu je vytvořen tzv. *voiceprint*. *Voiceprint* je soubor měřitelných vlastností lidského hlasu, který jednoznačně identifikuje jednotlivce. Tyto vlastnosti vycházejí z fyzických znaků úst, krku a hlasivek [11]. Velikost vytvořeného voiceprintu je pouhých 624 bajtů.

Čas potřebný pro vytvoření voiceprintu by se dal charakterizovat takto: 50 sekund zvukového záznamu je rovno 1 sekundě výpočetního výkonu strojového času CPU. Rychlost porovnávání je proti vytváření voiceprintu zanedbatelná. Společnost Phonexia s.r.o. uvádí, že milión porovnání se pohybuje v řádech několika sekund. Tyto údaje vyplynuly z testování na Intel(R) Core(TM) i5-2500 CPU @ 3.30GHz.

Kapitola 5

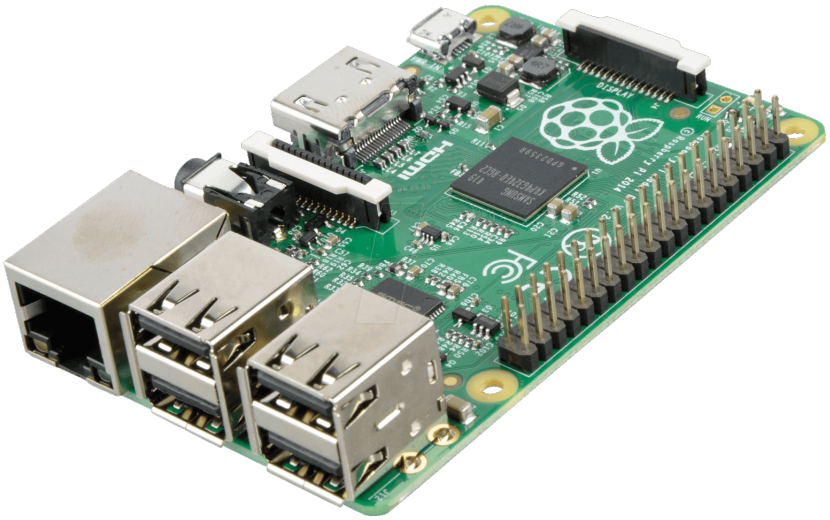
Raspberry Pi

Jak již bylo řečeno v úvodu, v této práci používáme minipočítač Raspberry Pi - konkrétně Raspberry Pi 1 model B+ viz. obrázek 5.1. V této kapitole trochu popíšu samotný minipočítač a jeho přídavný modul PiFace Digital 2.

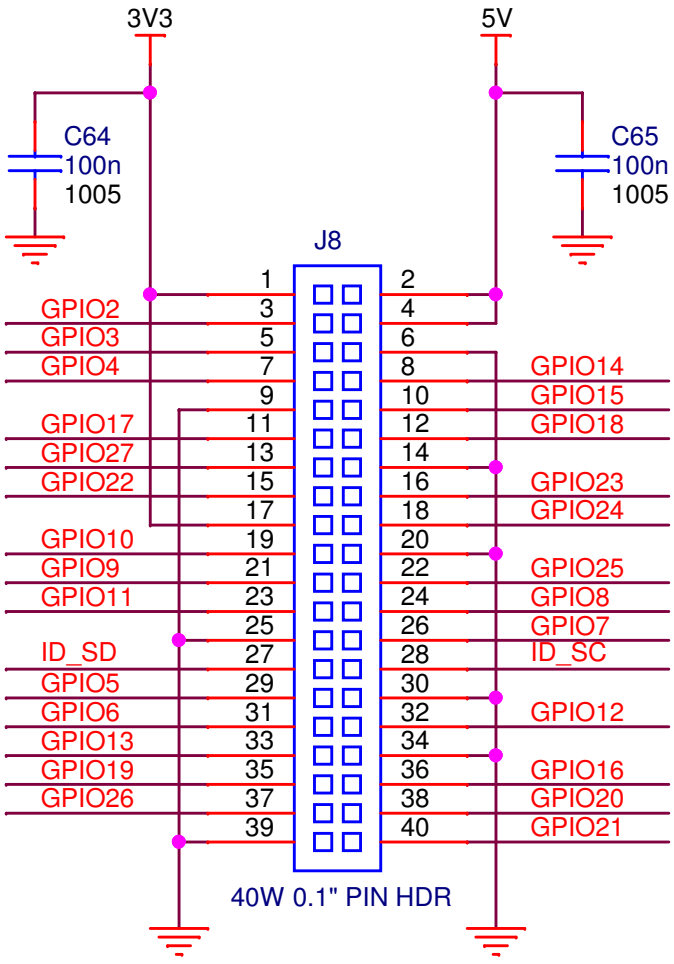
5.1 Popis zařízení

Tento minipočítač byl vybrán především z toho důvodu, že nám umožňuje jednoduché a praktické řešení, a to nejen díky svým malým rozměrům. Raspberry Pi B+ má rozměry 85 x 56 x 17 mm. Je vyvíjen společností The Raspberry Pi Foundation, aby podpořil celkovou výuku informatiky [12]. Zařízení za svou dobu prošlo určitým vývojem. Na trhu je k sehnání hned několik modelů [13]:

1. Co se týče společných vlastností. Raspberry Pi 1 obsahuje jednojádrový procesor BCM2835 700 MHz z rodiny ARM, grafický procesor Broadcom VideoCore IV s podporou OpenGL ES 2.0, MPEG-4 a audio výstup v podobě 3,5mm jacku. Audio vstup bohužel zařízení nemá. Jednotlivé modely mají tyto specifikace:
   - Model A - 256 MB RAM, slot pro SD nebo MMC kartu, 1 USB port.
   - Model A+ - 256 MB RAM, slot pro SD kartu, 1 USB port.
   - Model B - 512 MB RAM, slot pro SD nebo MMC kartu, 2 USB porty, ethernet 10/100 s konektorem RJ45.
   - Model B+ - 512 MB RAM, slot pro microSD kartu, 4 USB porty, ethernet 10/100 s konektorem RJ45.
2. Raspberry Pi 2 - zatím nejnovější typ minipočítače Raspberry Pi. Obsahuje čtyřjádrový proceor BCM2836 900 MHz, grafický procesor Broadcom VideoCore IV s podporou OpenGL ES 2.0, MPEG-4 a audio výstup v podobě 3,5mm jacku. Audio vstup bohužel stále chybí.
   - Model B - 1GB RAM, slot pro microSD kartu, 4 USB porty.



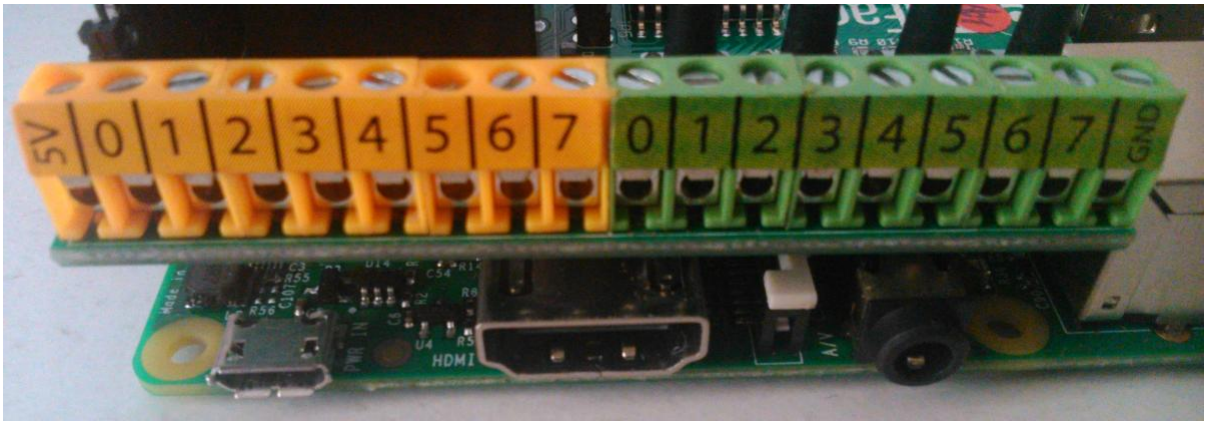
Obrázek 5.1: Zařízení Raspberry Pi B+ [14].



Obrázek 5.2: Schéma GPIO [15].

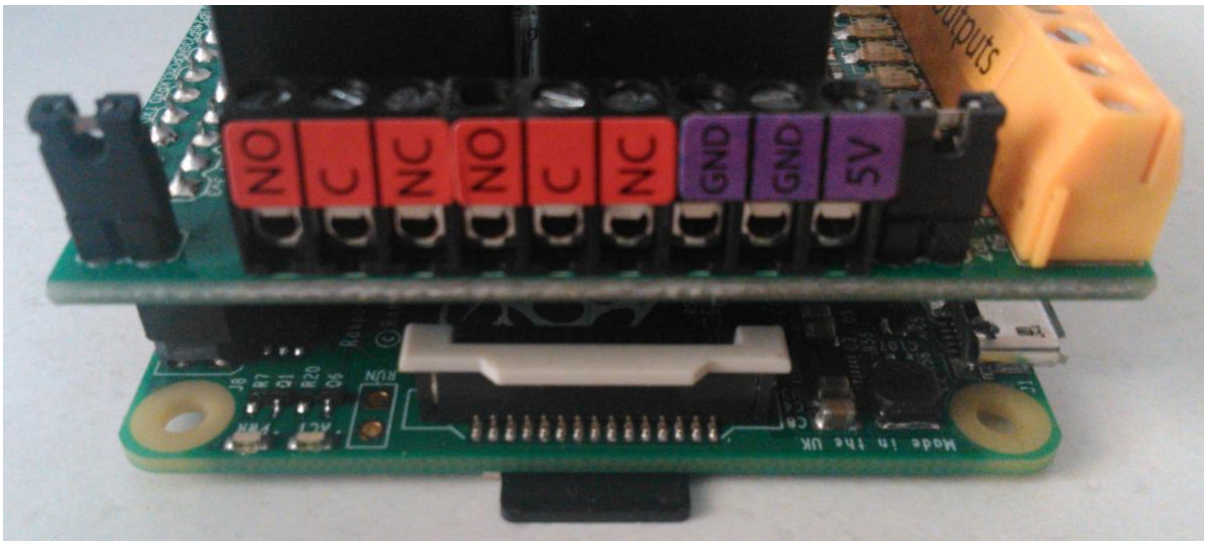
5.2 PiFace Digital 2

K zařízení Raspbbery Pi jsme připojili PiFace Digital 2. Jedná se o rozšiřující modul, který je s Raspberry spojen pomocí GPIO (General purpose input/output) viz. obrázek 5.2. Detailní pohled na vstupy a výstupy můžeme vidět na obrázku 5.3. Na obrázku 5.4 je pohled na výstupy relé. Celý modul pak lze vidět na obrázku 5.5



Obrázek 5.3: Detail na vstupy a výstupy.

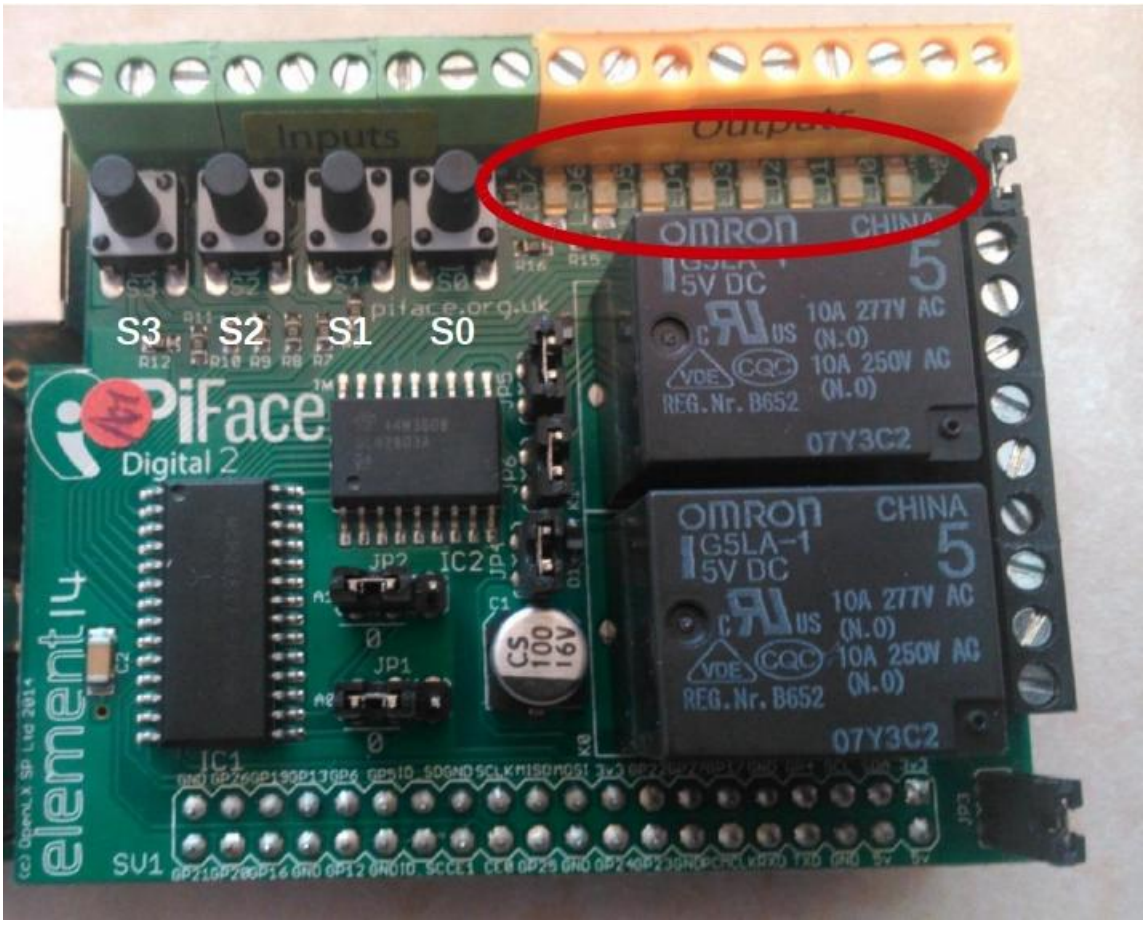
- 8 výstupů s otevřeným kolektorem (oranžový blok) - úplně vlevo je připojení na 5V, následuje 8 vstupů, které jsou očíslovány 0 - 7.
- 8 digitálních vstupů (zelený blok) - jako poslední mezi vstupy je uzemnění.



Obrázek 5.4: Detail na relé

- 2 přepínací relé - levé relé je ovládáno pomocí výstupu č. 0 a pravé relé je ovládáno pomocí výstupu č. 1.

- Obecně se elektromagnetické relé skládá z cívky, která je umístěna na jádře z magneticky měkkého materiálu. V blízkosti tohoto elektromagnetu je umístěna pohyblivá kotva, jejíž druhý konec se dotýká kontaktů, ke kterým je připojen ovládaný prvek. Jakmile začne cívkou protékat proud, kotva se k ní přitáhne a tím sepne kontakty. Proud potřebný k přitažení kotvy k cívce je menší, než proud procházející v obvodu ovládaného prvku.
- NC - Normally closed
- NO - Normall open
- C - Common



Obrázek 5.5: Zařízení PiFace Digital 2.

- Červeně je vyznačeno 8 LED indikátorů.
- Písmena S0, S1, S2, S3 označují 4 přepínače, které slouží k ovládání vstupů 0 - 3 v tomto pořadí.

PiFace Digital 2 se dá programovat pomocí Python modulu *pifacedigitalio*. Pifacedigitalio usnadňuje kontrolu celého zařízení.

Kapitola 6

Tvorba zařízení

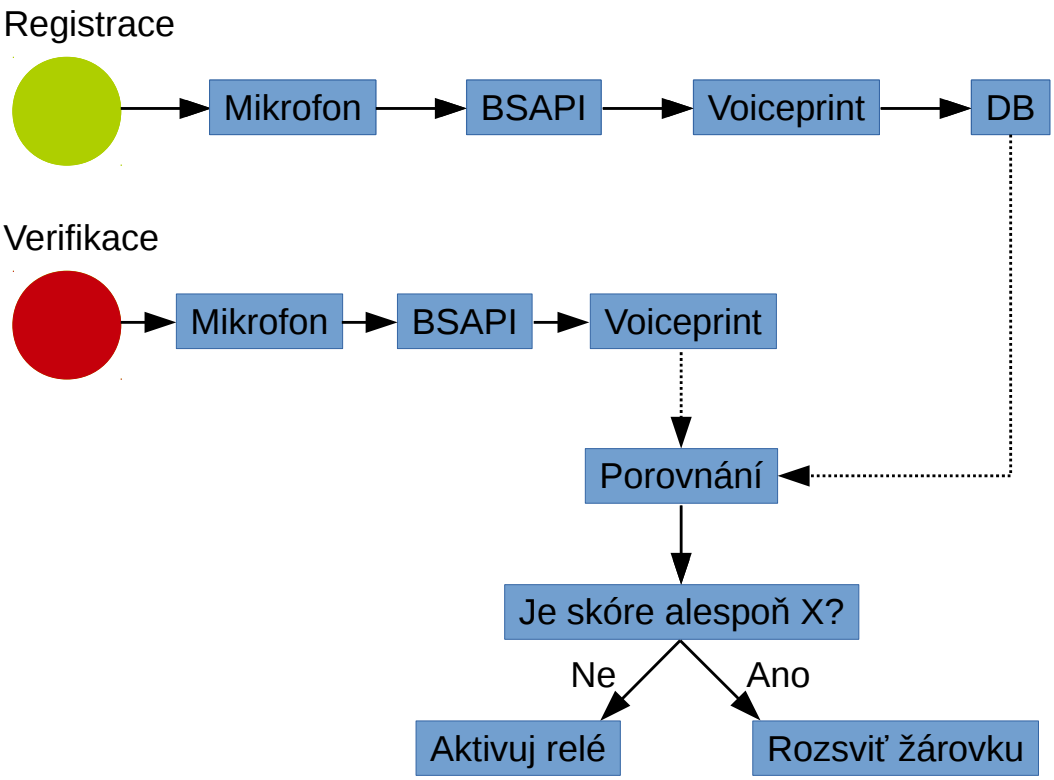
V této kapitole bude popsáno, jak vznikalo naše zařízení. Nejdříve si popíšeme jaké parametry by zařízení mělo mít a jaké součástky jsme k tomu použili. V další části se podíváme na výběr mikrofonu a zjištění optimální délky nahrávky. Dále bude popsáno, jak probíhalo uživatelské testování a jakých výsledků bylo dosaženo.

6.1 Návrh zařízení

Hlavní myšlenkou bylo vytvořit takové zařízení, které umožní osobám přístup do vymezeného prostoru na základě rozpoznání jejich hlasu. Chtěli jsme, aby byly možnosti využití v co největším rozsahu. Ideální zařízení by mělo být spolehlivé, nenáročné na provoz a připojitelné k nejrůznějším přístrojům. Na obrázku 6.1 je znázorněn základní návrh zařízení. Zelená a červená barva značí barvy tlačítek na zařízení.

Naše zařízení můžeme zařadit mezi systémy textově nezávislé. Tento způsob jsme zvolili především kvůli pohodlnosti pro uživatele. Úspěšnost rozpoznávání podle hlasu stoupá se získanou délkou hlasu. Pro naše zařízení jsme tedy zvolili takový způsob, že je stanovena pevná délka, po kterou se bude nahrávat. Uživatel si tedy sám nemůže zvolit, jak dlouhý záznam bude. Ke stanovení této délky jsme došli na základě experimentů viz. sekce 6.5. Celé zařízení se skládá ze dvou modulů: registračního a verifikačního. K získání záznamu hlasu je použit mikrofon Trust Starzz, který vyšel z našeho testování s nejlepšími výsledky. Po získání záznamu hlasu se využije třídy *SVoicePrintExtractorI* z knihovny BSAPI, která extrahuje voiceprint z dané nahrávky. Poněvadž registrovaných osob může být více, musíme zamezit tomu, aby se v databázi všech registrovaných uživatelů někdo ocitl dvakrát. Po dokončení registrace je tedy vytvořený voiceprint porovnán se všemi ostatními, které se již nachází v databázi. Pokud dojde ke shodě, je nový voiceprint odstraněn. K samotnému porovnávání se využívá třída *SVoicePrintComparatorI* z knihovny BSAPI.

Na základě porovnání dvou rozdílných voiceprintů se rozhodne, zda se rozsvítí žárovka, která simuluje samotnou bránu. Jak již bylo uvedeno výše, zařízení PiFace Digital 2 se dá programovat pomocí jazyka Python. Celý systém tedy zastřešuje skript napsaný v jazyce Python. Zároveň využíváme Python modul *pifacedigitalio*, který slouží především ke snímání vstupů a sepínání výstupů.



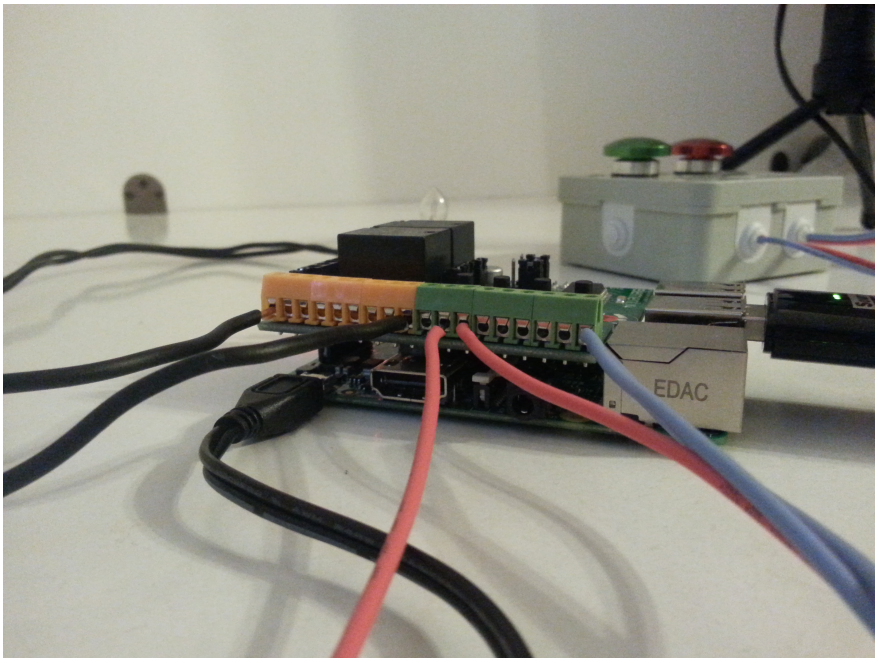
Obrázek 6.1: Základní logika zařízení

6.2 Konstrukce zařízení

Jak již bylo zmíněno výše, srdcem celého zařízení je minipočítač Raspberry Pi B+ společně s PiFace Digital 2. Jelikož Raspberry Pi nemá audio vstup, museli jsme přes USB připojit externí zvukovou kartu. Vybrali jsme Creative Sound Blaster SB1140. Jako paměť pro ukládání Raspberry Pi slouží microSD karta. Zvolili jsme 16 GB kartu značky Kingston, která je pro naše účely naprosto dostačující. Na ni jsme nainstalovali systém Raspbian. Raspbian je svobodný operační systém. Jak již název napovídá je založen na Debianu a je optimalizován pro samotné zařízení Raspberry Pi.

Dalším důležitým prvkem je mikrofon. Ten je připojen ke zvukové kartě pomocí 3.5mm konektoru. Vyzkoušeli jsme několik mikrofonů viz. sekce 6.4. Z testování vyšel jen jeden vítěz a to mikrofon Trust Starzz. Poskytuje dostatečnou kvalitu záznamu i z rozumné vzdálenosti. Jelikož zařízení musí rozeznávat registraci a verifikaci osoby, bylo nutné nějak tyto funkce oddělit. Nejlepší variantou se jevilo použití dvou tlačítek. K PiFace Digital 2 jsou tedy připojeny dvě tlačítka v zelené a červené barvě. Obě tlačítka jsou z důvodu lepší manipulace zasazeny do krabičky. Jak již bylo zmíněno výše, místo brány nám jako identifikátor slouží žárovka, která je taktéž jako tlačítka zapojena k PiFace Digital 2. Jedná se o 4.8V žárovku zasazenou do obyčejné objímky.

Tlačítka jsou zapojeny do vstupů č. 1 a 2. Žárovka je připojena do výstupu č. 7 a na 5V zdroj, který je vyveden hned na okraji výstupů viz. obrázek 5.3. Celé zařízení lze vidět na obrázku 6.2 a 6.3.



Obrázek 6.2: Detailní pohled na zapojení tlačítek a žárovky

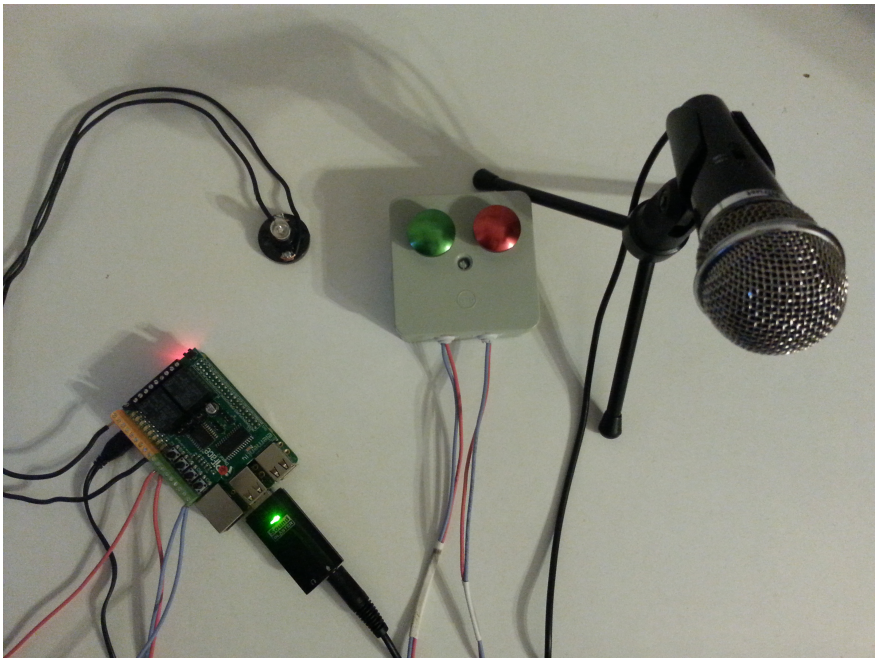
6.3 Obsluha zařízení

Samotné použití zařízení je velice snadné. Zelené tlačítko slouží k registraci osoby a jeho stisknutím začne po dobu 8 vteřin nahrávání. Se začátkem nahrávání se rozsvítí červená LED dioda na PiFace Digital 2. Po uplynutí 8 vteřin dioda zhasne a dojde k uložení nahrávky. Může se stát, že nahrávka neobsahuje dostatečnou délku hlasového záznamu. Např. v případě, že uživatel po většinu doby nahrávání mlčí. V takovémto případě se sice vytvoří voiceprint, ale úspěšnost porovnání bude hodně malá. Minimální délka získané řeči by měla být 3 vteřiny. Takovouto délku totiž vyžaduje knihovna BSAPI k spolehlivému porovnávání. Červené tlačítko slouží k samotné verifikaci osoby. Po jeho stisknutí se aktivuje mikrofon a uživatel má opět 8 vteřin na nahrání hlasu. Celých 8 vteřin znovu signalizuje červená LED dioda. V případě kladeného vyhodnocení se rozsvítí žárovka. Pokud verifikace nebyla úspěšná sepne výstup č. 0, což má za následek rozsvícení LED diody doplněné o zvuk vydávaný relé.

Po dobu nahrávání hlasu je dobré dodržet rozumnou vzdálenost od mikrofonu. Jinak záznam nebude dostatečně kvalitní a nebude obsahovat požadované 3 vteřiny řeči.

Obsluha našeho zařízení tedy obnáší:

1. Pořízení nahrávky o délce 8 vteřin ve formátu WAV.
2. Vytvoření voiceprintu ze získané nahrávky.
3. Porovnání získaného voiceprintu s již uloženými voiceprinty.
4. Rozsvícení žárovky v případě kladného vyhodnocení, nebo sepnutí výstupu č. 0.



Obrázek 6.3: Náhled na celé zařízení pro otevírání brány hlasem

6.4 Výběr mikrofonu

Mikrofon nám slouží k získání nahrávky hlasu a plní tedy roli biometrického senzoru. Je jedním z nejdůležitějších součástí celého zařízení, a proto byl jeho výběr velice důležitý. Při jeho volbě jsme se soustředili jak na porovnání skutečné délky nahrávky s vyextrahovanou délkou hlasu, tak na porovnání získané délky řečové nahrávky vzhledem ke vzdálenosti řečníka od mikrofonu.

Testování se zúčastnili pouze dva lidé, poněvadž nám zde nešlo o rozpoznávání řečníka, nýbrž jen o kvalitu mikrofonu. Vždy se nahrávalo ve stejné místnosti se stejnými okolními podmínkami. Intenzita okolního hluku nepřesáhla 35dB. Můžeme tedy říct, že vliv okolí byl při porovnávání mikrofonů zanedbatelný.

Pro nahrávání jsme použily utilitu *arecord*, která je součástí balíku *alsa-utils* [16]. Spouští se přímo z příkazového řádku a podporuje několik formátů. Pro naše účely jsme vybrali formát WAV. Jedná se o nekomprimovaný formát a je jedním z nejstarších formátů zvuku. Už Microsoft a IBM ho pokládali za standard pro ukládání bitstreamového zvuku na počítač [17]. Použitá verze knihovny BSAPI umí pracovat pouze s nahrávkami o vzorkovací frekvenci 8kHz. Tato frekvence se dnes používá především pro telefonování. Příklad spuštění arecord:

```
arecord -D plughw:0,0 --format=S16 --rate=8000 --duration=15 test1.wav
```

Význam jednotlivých parametrů:

- `-D` zařízení, přes které se bude nahrávat
- `--format` způsob kódování
- `--rate` vzorkovací frekvence
- `--duration` délka nahrávky v sekundách

- poslední parametr značí název souboru

Do testu vstoupily tyto mikrofony:

1. Trust Microphone MC-1200
2. Sluchátka s mikrofonem Genius HS-04V
3. Trust Starzz
4. Canyon CRN-MIC1
5. Defender Mic-142

Délka každé nahrávky byla stanovena na 15 vteřin a byl použit vždy stejný text, aby rozdíly byly opravdu co nejmenší. Text pro nahrávání byl náhodně vybrán z knihy a uživatel vždy tedy četl stejný text. Tento postup byl zvolen především z důvodu vyhnutí se dlouhých pauz, způsobených přemýšlením, co vlastně říct. S každým mikrofonem bylo uskutečněno pět nahrávek z každé vzdálenosti. To nám tedy dává celkem 15 nahrávek pro každý mikrofon. První vzdálenost byla co možná nejmenší a pohybovala se od 2 do 4 centimetrů. Druhá měřená vzdálenost byla v rozmezí 8 až 12 centimetrů. Poslední vzdálenost, která vstoupila do experimentování se pohybovala od 15 do 20 centimetrů. Ke zjištění vyextrahované délky hlasu z 15 vteřinové nahrávky byla opět použita třída *SVoicePrintExtractorI* z knihovny BSAPI.

6.4.1 Výsledky měření

Provedené měření ukázalo na nedostatky jednotlivých mikrofonů. V tabulce 6.1 můžeme vidět konečné výsledky jednotlivých mikrofonů. Tabulka je rozdělena do tří částí podle vzdáleností od mikrofonu. Výsledné hodnoty u jednotlivých mikrofonů určují délku vyextrahovaného hlasu v sekundách. Samozřejmě čím delší je výsledná hodnota, tím je mikrofon vhodnější pro naše zařízení. Z měření vyplývá, že nejhorších výsledků dosáhly sluchátka Genius HS-04V. Naopak nejlepších výsledků se podařilo dosáhnout mikrofonu Trust Starzz, který proto byl použit jako nejvhodnější mikrofon pro naše zařízení.

Podrobné naměřené hodnoty můžeme najít v příloze v tabulkách B.1, B.2 a B.3.

Mikrofon	**Vzdálenost od mikrofonu v cm**		
	2-4	**8-12**	**15-20**
Trust Microphone MC-1200	10s	8s	6s
Sluchátka s mikrofonem Genius HS-04V	8s	6s	4.5s
Trust Starzz	**12s**	**10.5s**	**9.5s**
Canyon CRN-MIC1	10s	8.5s	6s
Defender Mic-142	11s	9.5s	8.5s

Tabulka 6.1: Tabulka srovnání jednotlivých mikrofonů.

6.5 Délka nahrávky a její vliv na práh

V úvodu této kapitoly jsme si uvedli, že uživatel nemá volnou ruku ohledně stanovení délky vstupní nahrávky. Museli jsme tedy zjistit, jakou délku zvolit. Především bylo potřeba brát ohled i na uživatele, aby čas určený pro záznam hlasu nebyl příliš dlouhý, ale zároveň aby nedocházelo k neúspěchu při registraci z důvodu nedodržení minimální délky hlasu. Experimentování s délkou nahrávání probíhalo podobně jako testování mikrofonů. Opět se jednalo o klidové prostředí. Celkově bylo osloveno 10 lidí z toho 5 mužů a 5 žen. Věkové rozmezí se pohybovalo od 23 do 44 let. S každým učinkujícím byly pořízeny dvě nahrávky o délkách 15, 12, 8 a 6 vteřin. Celkem jsme tedy získali osm nahrávek od jednoho učinkujícího. Každou délku jsme požadovali od jedné osoby dvakrát, aby mohla být porovnávána sama se sebou. Vždy mezi sebou byly porovnávány záznamy se stejnou délkou.

Po vytvoření a porovnání voiceprintů obdržíme od třídy *SVoicePrintComparatorI* z knihovny BSAPI skóre v rozmezí 0 až 100. Výchozí nastavení prahu knihovny BSAPI bylo staveno na hodnotu 50. Tedy pokud výsledné skóre po porovnání dvou voiceprintů bylo větší než 50, mělo by se jednat o téhož jedince. Na základě prováděného měření bylo zjištěno, že výchozí hodnota prahu není pro naše zařízení optimální.

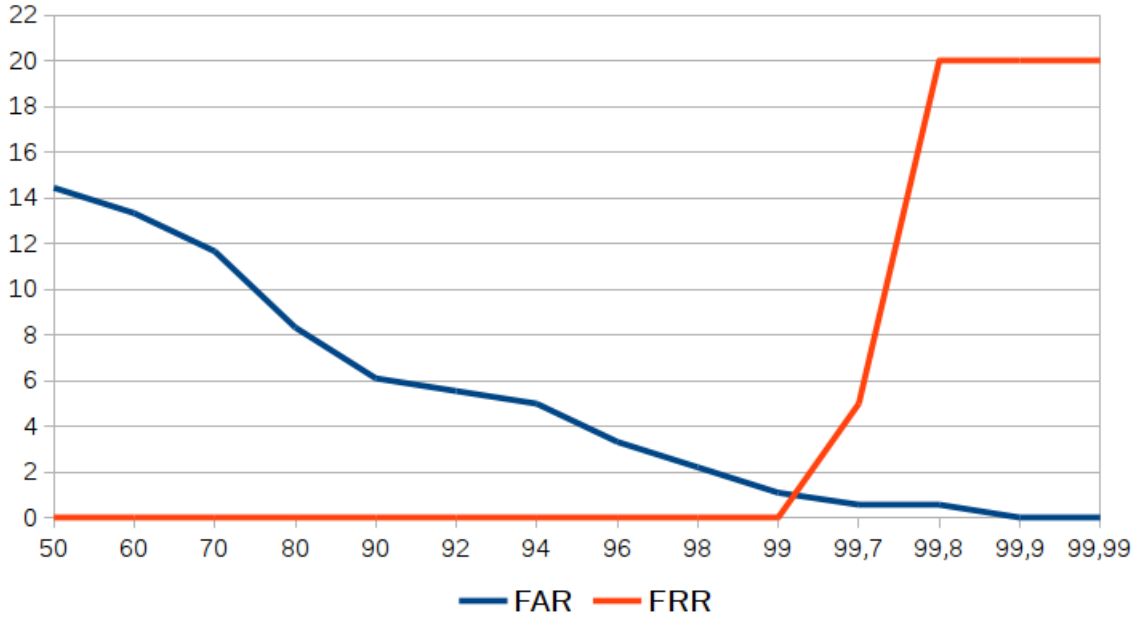
6.5.1 Vyhodnocení výsledků

Z tohoto testování vyplynulo, že 6 vteřinová délka nahrávky není vhodná. V jejím případě došlo k příliš velkému počtu neúspěchů při registraci. Z celkových 20 nahrávek se ve čtyřech případech stalo, že vyextrahovaná hlasová délka nebyla ani požadované 3 vteřiny. Lze tedy konstatovat, že míra neschopnosti zaregistrovat se v tomto případě rovná

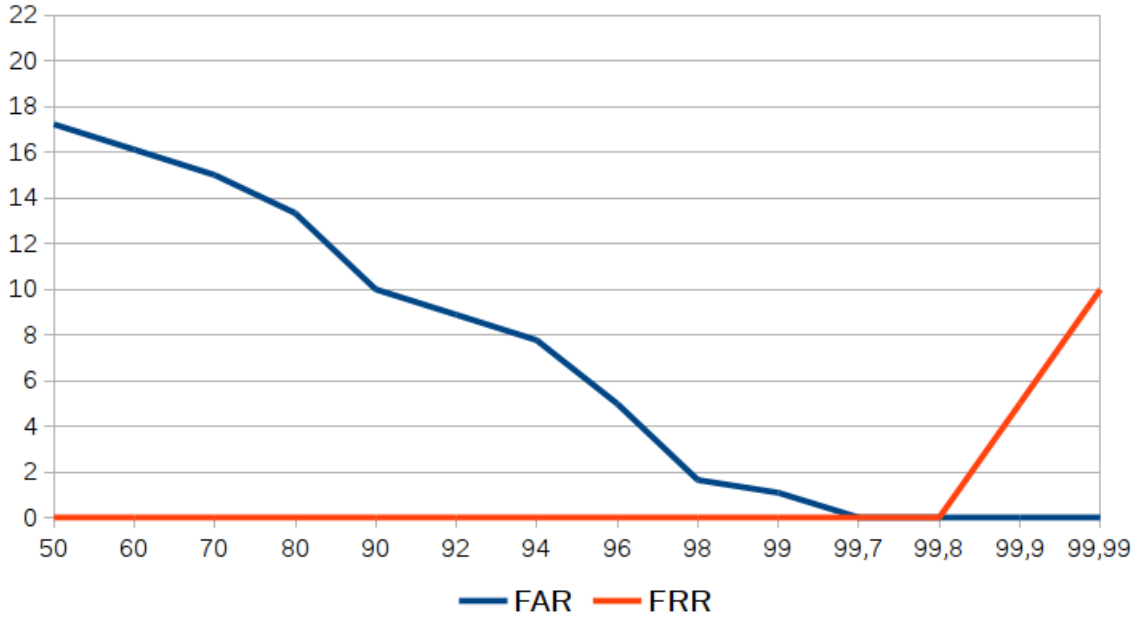
$$FTE = \frac{4}{20} = 20\%$$

V ostatních případech k takovéto situaci nedošlo ani jednou. V následujících grafech 6.4, 6.5 a 6.6 lze vidět hodnoty FAR a FRR pro jednotlivé délky nahrávek. Na ose X se vždy nachází práh a na ose Y je uvedena míra chyby v procentech.

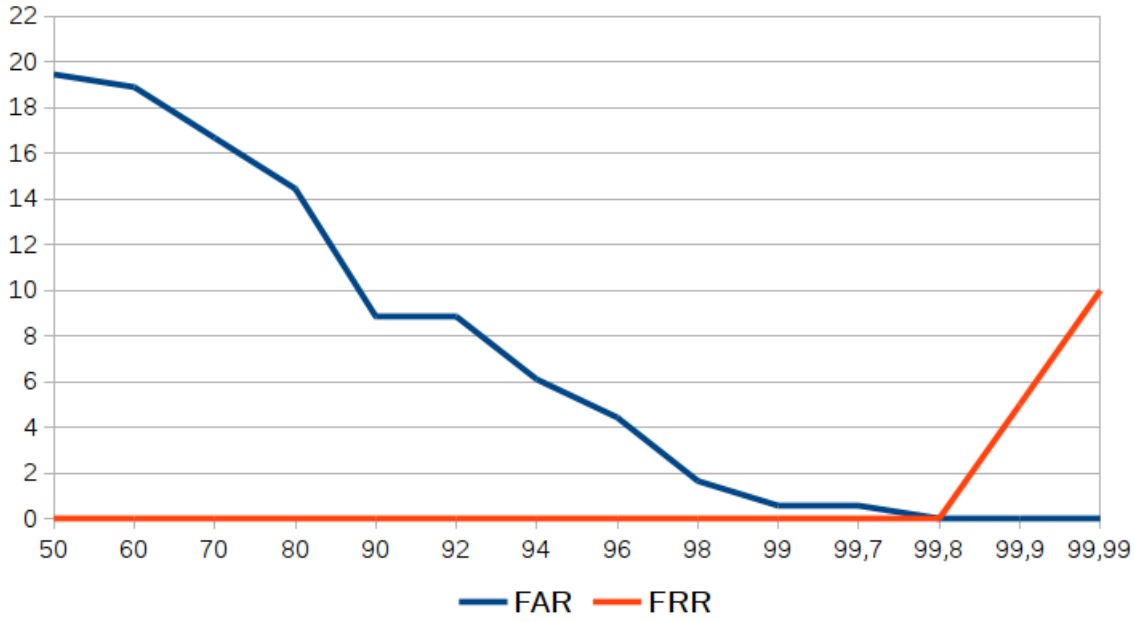
Pro účely naší práce jsme nakonec zvolili délku nahrávky na 8 vteřin. Při této délce nedošlo k problémům s nedostatečnou délkou. Delší varianty se nám pro naše použití jevily až příliš dlouhé. Nastavení délky lze však jednoduše kdykoliv změnit podle potřeby užití. Se změnou délky by ale přicházela do úvahy i změna prahu.



Obrázek 6.4: Graf pro 8 vteřinové nahrávky.



Obrázek 6.5: Graf pro 12 vteřinové nahrávky.



Obrázek 6.6: Graf pro 15 vteřinové nahrávky.

6.6 Uživatelské testování

Celé zařízení bylo potřeba otestovat na opravdových uživatelích. I při tomto testování byl zvolen podobný postup jako při testování mikrofonů. Jednotlivý uživatelé si volili libovolný text dle vlastního výběru. Dá se tedy říct, že každý člověk měl rozdílný text. Chtěli jsme se tímto co nejvíce přiblížit skutečnému provozu našeho zařízení, které je, jak již jsme si uvedli výše, textově nezávislé.

Od každé osoby byly vždy pořízeny dvě nahrávky, aby bylo možné porovnat daného uživatele se sebou samým. Cílem bylo oslovit co možná největší věkové spektrum lidí, což se nakonec povedlo. Jedinou věkovou skupinou, kterou se nepovedlo zařadit do testování jsou osoby mladší 20ti let. Nicméně tyto osoby netvoří hlavní skupinu uživatelů, pro které je naše zařízení určeno.

Nejmladšímu členovi bylo v době testování 21 let. Naopak nejstarší osobě, která nám poskytla svůj hlas bylo 70 let. Věkový průměr učinkujích je 36 let. Podrobnější informace o staří uživatelů lze najít v příloze v tabulkách B.4 a B.5. Celkově bylo osloveno 30 lidí. Mezi nimi bylo 16 mužů a 14 žen. Nahrávání jednotlivých hlasů již neprobíhalo tak jako u testování mikrofonů ve stejné místnosti, nicméně se vždy jednalo o podobné prostředí, kde míra okolního hluku nepřesáhla 35dB. Hodnota 35dB by se dala přirovnat k relativnímu tichu v obsazeném hledišti kina, šepotu, velmi tichému bytu nebo velmi tiché ulici [18]. Dohromady jsme tedy pořídili 60 nahrávek. Délka nahrávky byla opět stanovena na 8 vteřin.

Z každé nahrávky byl opět vytvořen voiceprint pomocí třídy *SVoicePrintExtractorI* z knihovny BSAPI. Následně byly všechny tyto voiceprinty porovnány mezi sebou. Celkově tedy proběhlo 3600 porovnání. Zvolili jsme tento způsob testování, poněvadž dvakrát oslovit 30 lidi, aby se nejprve zaregistrovali, a poté se vyzkoušeli přihlásit na 30 dalších účtů je hodně komplikované. Výsledek porovnání by ovšem neměl být odlišný.

6.6.1 Vyhodnocení uživatelského testování

V tabulkách 6.2 a 6.3 můžeme vidět naměřené hodnoty FAR a FAR v procentech. Tytéž hodnoty zanesené do grafu lze vidět na obrázku 6.7. Na ose X se vždy nachází práh a na ose Y je uvedena míra chyby v procentech. Výpočet hodnota FAR a FRR byl v tomto případě následovný:

$$FAR = \frac{\text{Počet hodnot rozdílných vzorů nad prahem}}{\text{Celkový počet porovnání rozdílných vzorů}}$$

tedy

$$FAR = \frac{\text{Počet hodnot rozdílných vzorů nad prahem}}{3480}$$

$$FRR = \frac{\text{Počet hodnot shodných vzorů pod prahem}}{\text{Celkový počet možných shod}}$$

tedy

$$FRR = \frac{\text{Počet hodnot shodných vzorů pod prahem}}{120}$$

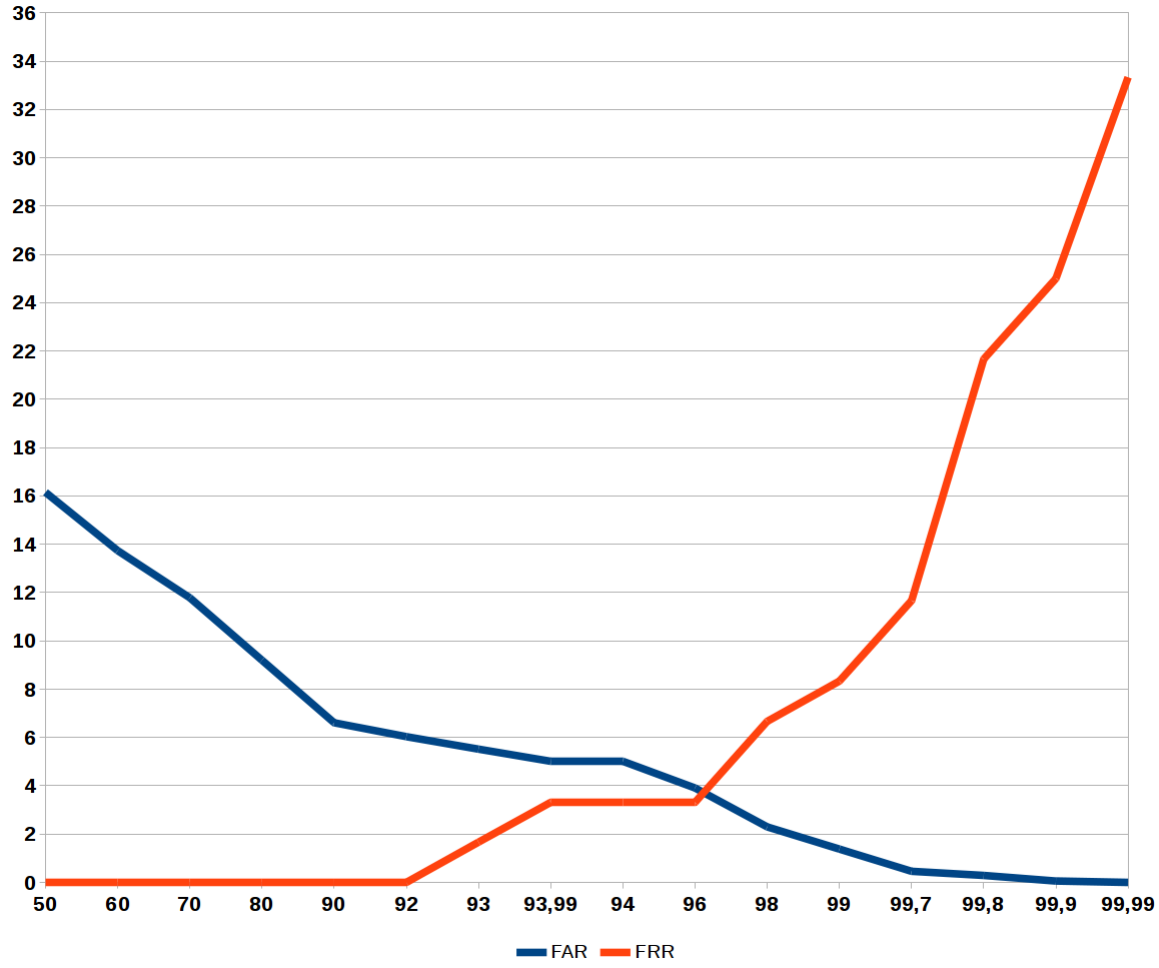
Od jednoho člověka máme celkem dvě nahrávky. To nám dává celkem 4 skóre při porovnávání téže osoby. Celkem jsme oslovili 30 osob. Dohromady je tedy počet celkových počet možných shod 120. Hodnotu 3480 získáme odečtením 120 od 3600.

Práh	50	60	70	80	90	92	94
FAR	16,1494%	13,7356%	11,7816%	9,1954%	6,6091%	6,03450%	5,000%
FRR	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%	3,3333%

Tabulka 6.2: Tabulka hodnot pro FAR a FRR.

Práh	96	98	99	99,7000	99,8000	99,9000	99,9900
FAR	3,9080%	2,2988%	1,3793%	0,4597%	0,2873%	0,0574%	0,0000%
FRR	3,3333%	6,6667%	8,3333%	11,6667%	21,6667%	25,0000%	33,3333%

Tabulka 6.3: Tabulka hodnot pro FAR a FRR.



Obrázek 6.7: Graf pro všech 30 nahrávek o délce 8 vteřin.

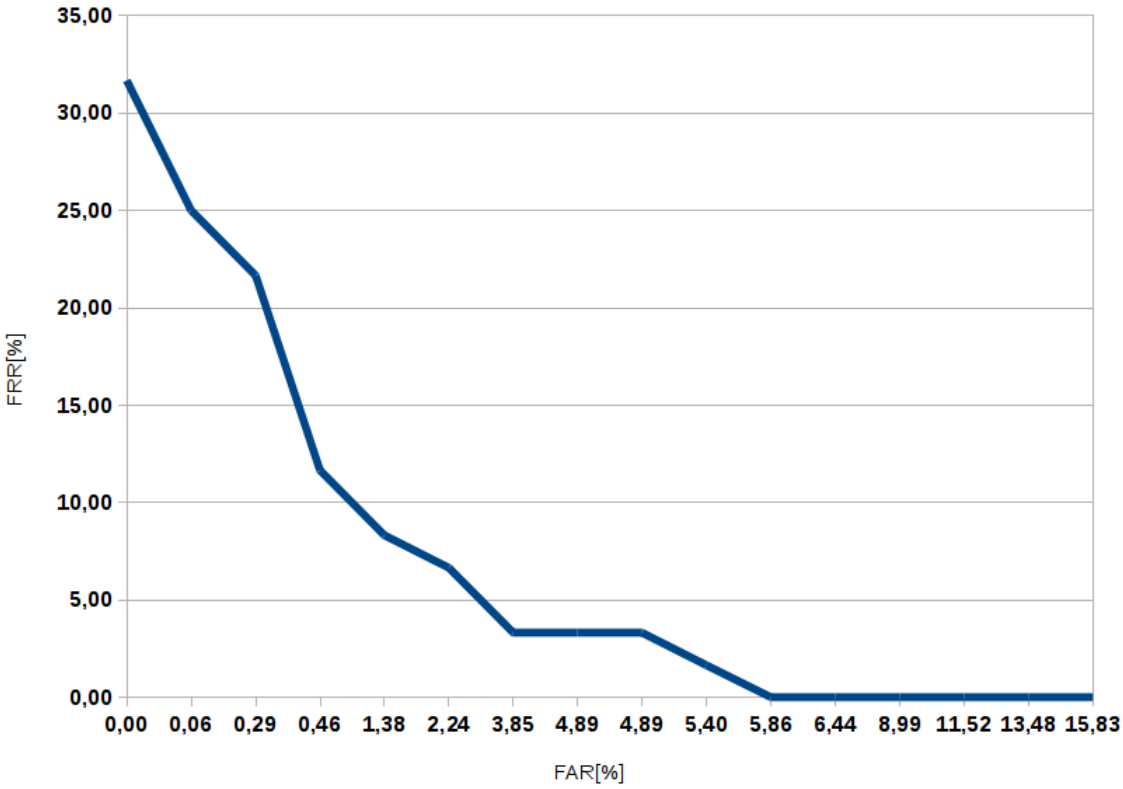
V následujících tabulkách 6.4 a 6.5 můžeme vidět číselné vyjádření počtu chybně přijatých a chybně odmítnutých osob. Jelikož porovnávání je symetrické, mohli bychom tyto hodnoty vydělit dvěma. Museli bychom ale taktéž změnit celkové počty.

Práh	50	60	70	80	90	92	94	96	Z celkového počtu
Chybně přijato	562	478	410	320	230	210	174	136	3480
Chybně odmítnuto	0	0	0	0	0	0	4	4	120

Tabulka 6.4: Tabulka srovnání jednotlivých prahů.

Práh	98	99	99,7	99,8	99,9	99,99	Z celkového počtu
Chybně přijato	80	48	16	10	2	0	3480
Chybně odmítnuto	8	10	14	26	30	56	120

Tabulka 6.5: Tabulka srovnání jednotlivých prahů.



Obrázek 6.8: Závislost FAR na FRR.

Jako výchozí práh jsme nakonec stanovili hodnotu 98. Zvolili jsme raději striktnější hodnotu, při které nedojde k tak velkému počtu chybných přijetí. Práh lze samozřejmě měnit dle optřeby použití.

Naměřené hodnoty porovnání pro jednotlivé uživatelé nebyly zahrnuty ani do přílohy této práce z důvodu velkého obsahu. Na přiloženém CD se však nachází soubor *score.txt* s podrobnými výsledky.

6.7 Vliv okolního prostředí

Doposud jsme se jen dozvěděli, jakých výsledků naše zařízení dosahuje v relativně klidných místech, kde hodnota okolního hluku nepřesahuje 35dB. Dá se však předpokládat, že možnosti využití se nachází i místech, kde okolí nebude úplně potichu, nebo zde budou působit okolní zvuky narušující kvalitu nahrávání. Bylo tedy nutné vyzkoušet i jiné reálnější prostory, ve kterých by mohlo být zařízení používáno. Zvolili jsme tři rozdílná místa, u kterých se dala předpokládat různá intenzita okolních ruchů. Cílem experimentování bylo, zjistit v jakém prostředí je zařízení ještě schopné dosáhnout uspokojivých výsledků a v jakém již nemá smysl.

Testování se ve všech případech zúčastnilo 15 lidí a probíhalo stejně jako uživatelské testování v klidovém prostředí. Opět tedy byly pořízeny vždy dvě nahrávky od jedné osoby. Délka každé z nich byla stanovena na 8 vteřin a práh byl dán hodnotou 98. Vždy byly mezi sebou porovnávány voiceprinty ze stejného místa experimentování.

6.7.1 Kancelář

První oblastí, kde jsme zkoušeli zařízení, byla kancelář. Jedná se určitě o potenciální místo, kde by se mohlo zařízení uplatnit např. pro odemykání dveří. V tabulce 6.6 můžeme vidět výsledky.

		Z celkového počtu
Chybně přijato	15	840
Chybně odmítnuto	4	60

Tabulka 6.6: Tabulka výsledků pro kancelář.

6.7.2 Venkovní příjezd do firmy

Tento prostor byl vybrán především z důvodu, že se jednalo o místo, kde skutečně dochází k otevírání brány. Přesně na takovéto místa se zařízení zaměřuje, a proto nemohlo v testování chybět. V následující tabulce 6.7 nalezneme naměřené výsledky.

		Z celkového počtu
Chybně přijato	25	840
Chybně odmítnuto	7	60

Tabulka 6.7: Tabulka výsledků pro venkovní příjezd do firmy.

6.7.3 Rušná ulice před domem

Poslední zkouškou, kterou zařízení prošlo, byl vchod do domu přímo u hlavní cesty. Jednalo se o opravdu rušnou ulici, po které jezdí mnoho aut včetně tramvají. V tomto prostředí bylo naše zařízení až příliš háklivé na okolní hluk, což dokazuje velký počet chybných odmítnutí v tabulce 6.8.

		Z celkového počtu
Chybně přijato	1	840
Chybně odmítnuto	30	60

Tabulka 6.8: Tabulka výsledků pro rušnou ulici

6.8 Spuštění skriptu

Ke správnému běhu skriptu je zapotřebí veškerého výše popsaného zařízení a především knihovny BSAPI. Mnou napsaný skript v jazyce Python se spouští následujícím způsobem `./gate.py`, pokud se uživatel nachází ve složce `./home/pi/BSAPI/gate/sid2`. Je nutné dodržet spouštění právě z této složky, protože skript využívá i jiné programy, které se nachází ve stejné složce. Vytvořené voiceprinty jsou uloženy do složky `vp`. Po spuštění skriptu je zařízení připraveno k ověřování osob podle hlasu. Ukončení skriptu se dá provést stisknutím přepínače *S0* na PiFace Digital 2.

6.8.1 Doba zpracování nahrávky

V kapitole 4 jsme si uvedli, že vytvoření voiceprintu by se dalo charakterizovat tak, že 50 sekund záznamu je rovno 1 sekundě strojového času CPU. Bohužel Raspberry Pi nedisponuje dostatečným výkonem k realizaci takovéto rychlosti. Nicméně rychlost porovnávání je stále oproti vytvoření voiceprintu zanedbatelná. Měřením bylo zjištěno, že zpracování 8 vteřinové nahrávky (vytvoření voiceprintu a porovnání) trvá v průměru 4 vteřiny. Rychlost zpracování by se dala snížit např. použitím nejnovějšího zařízení Raspberry Pi 2, který již disponuje lepším výkonem viz. kapitola 5 nebo portovat na Raspberry Pi nejnovější verzi knihovny BSAPI, která je lépe optimalizována. Ta již bohužel nebyla použita z důvodu náročnosti kompilace.

Kapitola 7

Závěr

V rámci práce jsem měl možnost seznámit se s technologií identifikace řečníka podle hlasu. Část práce se také věnovala zařízení Raspberry Pi, na kterém je celý systém postaven. Největší část práce se ale věnovala samotnému zařízení a také jeho testování. Z výsledků testování je patrné, že vytvořené zařízení lze využít v klidnějších prostředích. Jak je totiž patrné ze sekce 6.7, použití zařízení ve velice rušném prostředí nedosahuje uspokojivých výsledků.

Celé zařízení by se dalo vylepšit jak po designové, tak po technické stránce. Můžeme dále zlepšovat kvalitu vlastního algoritmu rozpoznáváni řečníka, nebo můžeme kombinovat technologii identifikace řečníka například s požadováním zadání hesla. Tento postup je nejspolehlivější, protože se opírá o biometrické měření a zároveň skrytou znalost, kterou zná pouze oprávněná osoba. Společnost Phonexia již dnes úspěšně zdokonaluje algoritmus pro rozpoznávání řečníka. Naše zařízení do budoucna slibuje širokou možnost uplatnění. Jako nejpravděpodobnější možnost se zatím jeví zabezpečení domu, auta či lodě.

Literatura

[1] Ph.D. Ing. Martin Drahanský. Biometrické systémy bio. studijní opora, 2006. https://www.fit.vutbr.cz/study/courses/BIO/private/BIO_Studijni_opora.pdf.

[2] Zdeněk Říha Roman Rak, Václav Matyáš. *Biometrie a identita člověka ve forenzních a komerčních aplikacích*. Grada Publishing,a.s., 2008. ISBN 978-80-247-2365-5.

[3] FBI.gov. Integrated automated fingerprint identification system, 2014. http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis.

[4] Michal Černý. Jak vyzrát na biometrii. *CHIP*, (8), Srpen 2010.

[5] Salil Prabhakar Anil K. Jain, Arun Ross. An introduction to biometric recognition. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, 14(1), Leden 2004.

[6] STEPHANIE MLOT. Zwipe, mastercard show off credit card with a fingerprint sensor, Říjen 2014. http://www.pcmag.com/article2/0,2817,2470568,00.asp.

[7] B. Hazen. Effects of differing phonetic context on spectrographic speaker recognition. *Journal of the Acoustical Society of America*, (54), 1973.

[8] Biing-Hwang Juang Lawrence rabiner. *Fundamentals of speech recognition*. PTR Prentice Hall, 1993. ISBN 0-13-285826-6.

[9] lungovav. Stavba a funkce hlasového ústrojí, 2012. http://pfyziollfup.upol.cz/castwiki/?p=2661.

[10] Ran Gazit Yaakov Metzger. Text-prompted without text: a language-independent, 2004. http://www.isca-speech.org/archive_open/archive_papers/odyssey_04/ody4_149.pdf.

[11] Margaret Rouse. voiceprint, Září 2005. http://searchsecurity.techtarget.com/definition/voiceprint.

[12] The Raspberry Pi Foundation. What is a raspberry pi?, 2015. https://www.raspberrypi.org/help/what-is-a-raspberry-pi/.

[13] Rpi hardware, 2015. http://elinux.org/RPi_Hardware.

[14] Raspberry pi b+, 2015. https://www.reichelt.de/Einplatinen-Computer/RASPBERRY-PI-B-/3/index.html?ACTION=3&GROUPID=6666&ARTICLE=146194&OFFSET=16&.

[15] Schematics, 2015. https://www.raspberrypi.org/documentation/hardware/raspberrypi/schematics/README.md.

[16] launchpad. Utilities for configuring and using alsa, 2014. https://launchpad.net/ubuntu/trusty/+package/alsa-utils.

[17] WAV.name. Wave / wav, 2011. http://wav.name/.

[18] Jiří Mikulčák. *Matematické, fyzikální a chemické tabulky a vzorce pro střední školy*. Prometheus, 2003. ISBN 80-7196-264-3.

Příloha A

Obsah CD

- Skript psaný v jazyce Python.
- Manuál k obsluze ve formátu .txt.
- 60 nahrávek ve formátu .wav
- Soubor se skóry všech porovnání ve formátu .txt.
- Bakalářská práce ve formátu .pdf a .tex soubory

Příloha B

Výsledky měření

Mikrofon	**Vzdálenost od mikrofonu**				
	2-4cm				
Trust Microphone MC-1200	9.9s	9.8s	10.1s	10.1s	10.1s
Sluchátka s mikrofonem Genius HS-04V	8.1s	7.9s	7.9s	7.8s	8.2s
Trust Starzz	12.1s	12.1s	11.8s	12.0s	12.0s
Canyon CRN-MIC1	10.3s	10.1s	9.8s	9.8s	10.0s
Defender Mic-142	10.8s	11.0s	11.1s	11.0s	11.1s

Tabulka B.1: Tabulka naměřených sekund pro délku 2 - 4cm.

Mikrofon	**Vzdálenost od mikrofonu**				
	8-12cm				
Trust Microphone MC-1200	7.8s	8.0s	8.2s	8.0s	8.0s
Sluchátka s mikrofonem Genius HS-04V	6.1s	5.8s	5.9s	6.0s	5.9s
Trust Starzz	10.4s	10.6s	10.5s	10.4s	10.6s
Canyon CRN-MIC1	8.5s	8.6s	8.3s	8.6s	8.5s
Defender Mic-142	9.4s	9.5s	9.7s	9.4s	9.5s

Tabulka B.2: Tabulka naměřených sekund pro délku 8 - 12cm.

Mikrofon	**Vzdálenost od mikrofonu**				
	15-20cm				
Trust Microphone MC-1200	6.5s	5.8s	5.8s	5.9s	6.0s
Sluchátka s mikrofonem Genius HS-04V	4.8s	4.6s	4.5s	4.4s	4.4s
Trust Starzz	9.4s	9.3s	9.6s	9.6s	9.5s
Canyon CRN-MIC1	6.1s	5.8s	5.9s	6.1s	6.1s
Defender Mic-142	8.3s	8.5s	8.6s	8.6s	8.5s

Tabulka B.3: Tabulka naměřených sekund pro délku 15 - 20cm.

Název	**Věk**
1.wav	27
1_2.wav	27
2.wav	23
2_2.wav	23
3.wav	63
3_2.wav	63
4.wav	40
4_2.wav	40
5.wav	23
5_2.wav	23
6.wav	21
6_2.wav	21
7.wav	23
7_2.wav	23
8.wav	23
8_2.wav	23
9.wav	66
9_2.wav	66
10.wav	54
10_2.wav	54
11.wav	21
11_2.wva	21
12.wav	23
12_2.wav	23
13.wav	28
13_2.wav	28
14.wav	27
14_2.wav	27
15.wav	23
15_2.wav	23

Tabulka B.4: Tabulka stáří jednotlivých uživatelů.

Název	**Věk**
16.wav	51
16_2.wav	51
17.wav	45
17_2.wav	45
18.wav	44
18_2.wav	44
19.wav	33
19_2.wav	33
20.wav	20
20_2.wav	20
21.wav	28
21_2.wav	28
22.wav	67
22_2.wav	67
23.wav	70
23_2.wav	70
24.wav	25
24_2.wav	25
25.wav	27
25_2.wav	27
26.wav	52
26_2.wav	52
27.wav	46
27_2.wav	46
28.wav	27
28_2.wav	27
29.wav	23
29_2.wav	23
30.wav	23
30_2.wav	23

Tabulka B.5: Tabulka stáří jednotlivých uživatelů.