# SPEAKER RECOGNITION ON THE MOBILE PHONE

**Jan Pešán**

Master Degree Programme (2), FIT BUT

E-mail: xpesan00@stud.fit.vutbr.cz

Supervised by: Jan Černocký

E-mail: cernocky@fit.vutbr.cz

**Abstract**: This work aims to port Speaker Identication System (SID) from desktop computer to the mobile device. The basic principles, function and implementation of speaker identication system on Nokia N900 mobile phone are described.

**Keywords**: Speaker identification, iVector, Total Variability Space, Nokia N900, MOBIO, BS-CORE

## 1. ÚVOD

Když se podíváme na aktuální vývoj mobilních telefonů, je zřejmé že tato zařízení se stávají něčím více než jen nástroji pro komunikaci. Již nyní jsou mobilní telefony schopny v mnoha případech nahradit funkci plnohodnotného počítače. Tento fakt však s sebou nese určitá bezpečnostní a ergonomická omezení. Mnoho internetových služeb je založeno na autorizaci nebo autentizaci uživatele. Klasicky tato situace nastává u účtu na facebooku či webmailu. Zde je při každém připojení vyžadováno zadání uživatelského jména a hesla – což může být časem obtěžující. Pokud vezmeme aplikace z opačného spektra služeb, pak se dostaneme do oblasti vyžadující maximální míru zabezpečení. Sem patří například internetové bankovnictví.

K běžnému PC lze bez větších problémů připojit mnoho periferií, které se starají o autentizaci a autorizaci uživatele, jako jsou např. čtečky otisků prstů, SmartCard karty, nebo USB RSA tokeny. Použití stejného postupu u mobilních telefonů je však z mnoha důvodů nepraktické. Na druhou stranu má většina mobilních telefonů k dispozici frontální kameru a samozřejmě mikrofon.

V této práci je popsán systém pro ověření identity pro mobilní telefon za pomocí verfikace obličeje a hlasu. Ten je použit v demoaplikaci simulující správu a automatické přihlašování uživatele do webových aplikací pouze na základě jeho biometrických dat. Celý systém je součástí EU projektu MOBIO, ve kterém jsme měli za úkol návrh, implementaci a testy hlasové části.

## 2. TECHNOLOGIE ROZPOZNÁNÍ ŘEČNÍKA

Obecně se v rozpoznávání řečníka testují 2 hypotézy:

1. *H0* - mluvčí v testovací nahrávce **není** ten, kterého jsme nahráli při trénování.
2. *H1* - mluvčí v testovací nahrávce **je** ten, kterého jsme nahráli při trénování

Každá z hypotéz generuje *likelihood*, což je hodnota, která popisuje „věrohodnost“ dané hypotézy [3]. Podle jejich rozdílu lze odvodit, zda se jedná o řečníka kterého hledáme, nebo jiného řečníka.

Ve starších systémech pro SID byla hypotéza *H0* generována tzv. *UBM* (*Universal Backround Model*), který vznikne tak, že se natrénuje model na všech dostupných nahrávkách všech dostupných řečníků. Tím vznikne jakýsi model „univerzálního řečníka“. Hypotéza *H1* je naopak generována pouze modelem, který je adaptován na řeč konkrétního řečníka.

Tyto technologie fungují uspokojivě, pokud jsou zachovány stejné podmínky při trénování modelu i jeho testování. Bohužel takové podmínky v reálném nasazení neexistují a proto se využívá technika zvaná *Total variability space* (zkráceně *iVector systém*). Ten rozšiřuje výše popsaný koncept, tak, že UBM provádí pouze sběr statistik o řečníkovi a z těch je poté extrahován nízkodimenzionální vektor fixní délky („iVector"), který popisuje aktuálního řečníka. iVectory jsou skórovány pomocí pravděpodobnostní lineární diskriminační analýzy (PLDA), která je schopná přímo vyhodnotit skóre hypotéz *H0* a *H1*. Tento systém je velice robustní v měnících se podmínkách a dává velmi dobré výsledky i při krátkých testovacích nahrávkách. PLDA byla poprvé představena v [1] a po modifikacích pro rozpoznávání řečníka [2] se z ní stal de-facto standard v této oblasti. Tento systém jsme také použili v tomto projektu.

## 3. IMPLEMENTACE

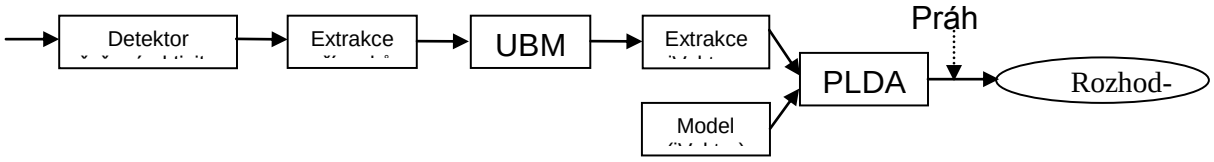
| | Paměť (MB) | Emulátor (s) | **N900 (s)** | PC (s) |
|---|---|---|---|---|
| 512G | 47 | X | **600** | 3 |
| 256G | 23 | X | **312** | 1 |
| 128G | 11 | X | **160** | <1 |
| 512G_precomputed | 359 | 20 | **24** | 2 |
| 256G_precomputed | 180 | 16 | **13** | <1 |
| 128G_precomupted | 90 | 8 | **7** | <1 |
| 512G_algo_precomputed | 48 | 9 | **10** | 1 |
| 256G_algo_precomputed | 24 | 9 | **8** | <1 |
| **128G_algo_precomputed** | **12** | **6** | **5** | **<1** |

**Tabulka 1:** Vliv velikosti systému a optimalizací na rychlost provádění

Jako nejvhodnější zařízení na trhu, dostupné v době rozhodování, byl vybrán mobilní telefon Nokia N900. Tento telefon byl zvolen z několika důvodů: (1) operační systém Linux na bázi OS Debian, (2) dostatečná výpočetní síla (CPU ARM Cortex A-8 600 MHz s FPU koprocesorem), (3) dostatek vnitřní paměti (32GB) a (4) existující kompilátor a vývojové prostředí kompatibilní s GCC.

Celá MOBIO aplikace je implementována v programovacím jazyce C++. Klíčové části rozpoznávače řečníka byly použity z knihovny BS-CORE produkované společností Phonexia s.r.o. ve spolupráci s VUT. Některé části bylo nutno doimplementovat ručně a došlo také k masivní optimalizaci celého systému.

Systém byl modulárně navrhován tak, aby bylo možno vyvíjet jednotlivé subsystémy efektivně, nezávisle na ostatních.




**Obrázek 1:** Struktura řečového subsystému

Jelikož implementace měla k dispozici pouze omezené zdroje vyplývající z povahy mobilního zařízení, bylo nutno standardní SID PC implementaci výrazně upravit aby byla schopna fungovat v reálném čase. Optimalizace, které jsme během testů provedli, zahrnují:

1. Optimalizace celkové velikosti systému – zmenšení systému z 512 Gaussovek na 128
2. Optimalizace zatížení CPU – použití předpočítaných výpočetních matic z paměti telefonu (řádky v tabulce **_precomputed*)
3. Optimalizace algoritmů – použití výpočetní metody z [4]. (řádky **_algo_precomputed*)
4. FPU utilizace – přepis maticového násobení do FPU nativních instrukcí, které přímo vytěžují pouze koprocesor pracující v plovoucí desetinné čárce.
5. Přetaktování CPU - softwarové přetaktování procesoru pomocí upraveného kernelu nahraného do mobilního telefonu ze základních 600MHz na stabilních 1GHz.

## 4. DEMO APLIKACE

Videa z provozu demoaplikace jsou ke shlédnutí zde: http://www.mobioproject.org/demonstrations


**Obrázek 2:** Vlevo: Potvrzená identita uživatele Vpravo: Odmítnutá identita uživatele

## 5. ZÁVĚR

Výsledkem práce je funkční systém integrovaný do mobilního telefonu, který byl testován ve dvou výše popsaných scénářích. Systém byl úspěšně prezentován na výstavě Biometrics 2010 v Londýně, kde také došlo k finální obhajobě projektu MOBIO. O aplikace, které byly jeho výstupem, projevilo zájem mnoho akademických, státních i soukromých subjektů.

Projekt jako celek je možno hodnotit jako úspěšný „proof-of-concept“ celého řešení a je připraven k reimplementaci do klientských řešení.

## PODĚKOVÁNÍ

Práce byla podporována EU-FP7 projektem MOBIO. Částečnou podporu poskytly také projekty GAČR č. 102/08/0707, a Výzkumný záměr MŠMT č. MSM0021630528

## REFERENCE

[1] Prince, S. J. D. and Elder, J. H.:Probabilistic Linear Discriminant Analysis for inferences about identity, 11th International Conference on Computer Vision, 2007

[2] N. Brummer, L. Burget, P. Kenny, P. Matejka, E. Villiers de, M. Karafiat, M. Kockmann, O. Glembek, O. Plchot, D. Baum and M. Senoussauoi, “Abc system description for NIST SRE 2010,” in Proc. NIST 2010 Speaker Recognition Evaluation. 2010, pp. 1–20, Brno University of Technology

[3] C. M. Bishop, Pattern Recognition and Machine Learning, chapter 4.2, Springer, 2006.

[4] Ondřej Glembek, Lukáš Burget, Pavel Matějka, Martin Karafiát and Patrick Kenny, „Simplification and optimization of i-vector extraction“ in Proc. ICASSP 2011, Brno University of Technology