

Software pro mobilního klienta a pro operátorskou konzoli systému PocketEAR

František Pártl¹

1 Úvod

V současné době probíhá na Katedře informatiky a výpočetní techniky Fakulty aplikovaných věd ZČU v Plzni vývoj rozpoznávače, který klasifikuje akustické signály na základě analýzy mel-frekvenčních keprstrálních koeficientů (MFCC). Pro kompletní využití této inteligence je třeba vytvořit infrastrukturu, kterou bude možné vyměňovat data mezi rozpoznávačem a mobilní aplikací koncového uživatele. Další plánovanou částí této infrastruktury je operátorské stanoviště, které přijímá audio záznamy, jejichž obsah nebylo možné rozpoznávačem stanovit. Posádkou tohoto stanoviště jsou operátoři, kteří určí obsah zmíněného nerozpoznaného zvuku, čímž se rozpoznávač bude dále učit. Tento celistvý systém byl označen jako *PocketEAR* (česky „kapesní ucho“).

Funkční systém, přesněji mobilní aplikace, by pak mohl napomáhat například sluchově postiženým jedincům při orientaci v blízkém okolí. Dále by mohl varovat před hrozícím nebezpečím, tj. například projíždějícím automobilem, sirénou, houkáním, pískáním a podobně.

2 Architektura systému PocketEar

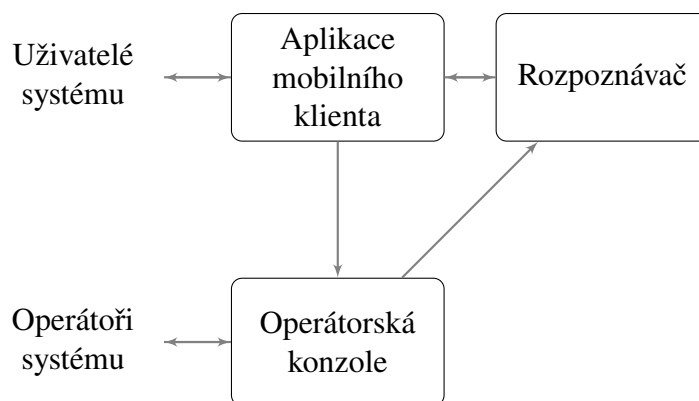
Původní architektura systému PocketEAR byla postavena na myšlence rekonstrukce akustického signálu z daných vektorů mel-frekvenčních keprstrálních koeficientů. Během práce se ukázalo, že rekonstrukce akustického signálu z MFCC je možná, ale pro potřeby rozpoznávání obecného zvuku je dle subjektivních testů velmi nepřesná. Na druhou stranu je dosažená kvalita rekonstrukce podle těchto testů použitelná minimálně pro rozpoznávání obsahu mluvené řeči.

Kvůli nedostatečné kvalitě rekonstruovaného akustického signálu nebyla tato myšlenka využita a výsledný systém má podobu zobrazenou na obrázku 1.

3 Knihovna libpe

V rámci implementace systému PocketEAR byla vytvořena knihovna `libpe`, která obsahuje veškeré potřebné prostředky jak pro parametrizaci akustického signálu metodou MFCC, tak pro jeho následnou rekonstrukci. Implementovaná knihovna je napsána v jazyce C++ s využitím frameworku `Qt`. Pro výpočet rychlé Fourierovy transformace, která je součástí metody MFCC, byla využita knihovna s otevřeným zdrojovým kódem – `KissFFT`. V současném systému má uplatnění minimálně při parametrizaci akustických signálů v mobilních aplikacích a dále v podobě modulu operátorské konzole pro parametrizaci klasifikovaných úseků akustického signálu pro potřeby dalšího trénování rozpoznávače na pozadí.

¹ student bakalářského studijního programu Inženýrská informatika, obor Informační systémy, e-mail: fpartl@students.zcu.cz



Obrázek 1: Schéma architektury systému PocketEAR.

4 Mobilní aplikace

Jednou z hlavních částí celého systému PocketEAR je mobilní aplikace, skrze kterou koncoví uživatelé využívají služeb vzdáleného rozpoznávače. V případě neúspěchu při rozpoznávání obsahu akustického signálu mobilní aplikace zabezpečí jeho doručení operátorské konzoli.

Kvůli potřebám multiplatformního řešení a předchozí implementaci potřebné knihovny `libpe` byl pro tvorbu aplikace užit jazyk C++ a framework Qt. Pro převod vzorků akustického signálu do souboru formátu *MP3* byla použita knihovna *LAME*.

5 Operátorská konzole

Úseky akustického signálu, jejichž rozpoznání nebylo úspěšné, jsou skrze mobilní aplikaci ve formě *MP3* souborů odesílány operátorské konzoli. Zde jsou tyto záznamy přidělovány registrovaným operátorům, kteří pomocí webového rozhraní přiřadí obsaženým zvukům příslušnou kategorii a typ. Výsledná data operátorské konzole, tj. klasifikované úseky akustického signálu, jsou připravena pro další trénování rozpoznávače.

Pro implementaci operátorské konzole byla vybrána rodina svobodného softwaru LAMP (*Linux, Apache, MySQL a PHP*). Pro vizualizaci a práci s akustickým signálem pak byla využita knihovna *Wavesurfer.js*.

6 Dosažené výsledky

Byla vytvořena knihovna `libpe`, která je využitelná v oblasti parametrizace akustických signálů metodou MFCC a jejich následné rekonstrukci při kvalitě dostačující pro určení obsahu mluvené řeči. Této knihovny využívá plně stabilní multiplatformní mobilní aplikace. Další úspěšně implementovanou částí systému je operátorská konzole, jež je připravena přijímat nerozpoznané záznamy ke klasifikaci skupinou operátorů. Na základě těchto klasifikací je možné další trénování rozpoznávače.

Literatura

Min, G. et al. (2015) *Speech Reconstruction from Mel-frequency Cepstral Coefficients via L1-norm Minimization*. Dostupné z: https://www.researchgate.net/publication/308201459_Speech_reconstruction_from_mel-frequency_cepstral_coefficients_via_l1-norm_minimization.