

Posudek vedoucího diplomové práce

Jana Wallacha

Aplikace Fourierovy analýzy na rozpoznávání kvality samohlásek podle jejich formantů

Během mé spolupráce s kolegy z Lékařské fakulty vznikl nápad na automatizovaný systém, který by uměl nejen rozpoznávat jednotlivé samohlásky, ale i nějak hodnotit jejich kvalitu. Taková aplikace se může hodit, například když se pacienti po mrtvici znovu učí mluvit. Počítač s podobným softwarem by jim mohl pomoci naučit se znova správně artikulovat.

Vývoj takové aplikace rozhodně není téma na bakalářskou práci, spíše na několikaletý grant, ale přesto stojí za to učinit jakýsi průzkum terénu. Téma je velmi multi-disciplinární, spojuje v sobě základy fyziky (akustika), biologie (princip tvorby hlasy a princip sluchu), matematickou teorii zpracování digitálních signálů (zejména Fourierovu transformaci), jisté programátorské dovednosti (práce s MatLabem) a trochu statistiky (zejména Linear Predictive Coding). Téměř nic z výše uvedeného není obsahem studijních plánů bakalářského stupně, takže je třeba ocenit odvahu kolegy Wallacha se do tématu pustit. Práci jsem mu navíc zkomplikoval svým dlouhodobým pobytem v zahraničí, o to více si cením výsledku.

Autor nejprve popisuje fyzikální podstatu zvuku, princip sluchu a model tvorby zvuku v lidském hlasovém ústrojí. Zde představuje klíčový pojem formantu. V další kapitole je představena Fourierova transformace, vzorkování, aliasing, diskrétní Fourierova transformace a její rychlá verze (FFT). Následuje představení metody lineární predikce (LPC), což je klíčová metoda pro extrakci formantů ze záznamů samohlásek. Jsou uvedeny i dva široce používané algoritmy na výpočet koeficientů LPC. Velmi dobrý vhled do problému poskytuje autor pomocí srovnání LPC a FFT metod analýzy spekter. Následuje samostatná kapitola o technických detailech předzpracování digitálních signálů. Poslední kapitola podává zprávu o praktické části výzkumu. Autor sesbíral několik nahrávek českých samohlásek, sestavil z nich databázi a navrhuje a testuje několik metod na rozpoznání a hodnocení kvality nových nahrávek, pomocí srovnání s předzpracovanou databází. Tato metoda samozřejmě výrazně závisí na velikosti a kvalitě sesbírané databáze. Cílem práce nebylo vytvořit co nejlepší databázi, ale spíše prozkoumat různé možnosti hodnocení kvality samohlásek. Autor uvádí celkem šest přístupů, jak tento problém řešit.

Dle mého názoru je práce napsaná velmi dobře. Autor prokázal schopnost rychle se zorientovat v několika různých oblastech a velmi smysluplně je propojit. Velice se mi líbí, že autor uvádí jen tu teorii, kterou bude potřebovat, a to navíc ve velmi čtivě a pěkně napsané formě. Práce tak velmi dobře drží pohromadě a jasně směřuje k finální aplikaci. Praktická část práce vyžadovala mnoho úsilí, jak při sběru dat, tak při tvorbě programových kódů, a dle mého názoru snese srovnání s mnoha magisterskými pracemi.

Vždy je potřeba někdy prohlásit práci za hotovou, ačkoliv každý nápad a každé řešení vzbuzuje další otázky a další nápady. Jistě by stálo za to otestovat výše zmíněných šest metod poněkud systematictější způsobem nebo učesat programové kódy do standardnější podoby. Zmíněných šest metod (zejména poslední tři) mohlo rovněž být představeno o mnoho podrobněji a pečlivěji, včetně příkladů výpočtu. Rozhodně ovšem autorovi nechci vyčítat, že to neudělal, i tak je objem práce

úctyhodný. Téma lze jistě rozvíjet dál v rámci dalších diplomových prací, dokonce se domnívám, že by mohlo být předmětem výzkumného projektu.

Pokud bude u obhajoby čas, rád bych se zeptal na následující:

1. Proč se formant hlásky "U" na straně 34 rozpadá do tří oblastí? Je to tím, že když se první a druhý formant slejí v jeden, vertikální složka souřadnice středu elipsy se posune do třetího případně čtvrtého formantu? Nebo se děje ještě něco jiného?
2. Jak si vysvětlujete nesoulad mezi literaturou a vašimi výsledky na obrázku 16 a 17?
3. Jak se projeví, zda jde o mužský či ženský hlas?

Celkově znovu konstatuju, že práce je velmi přínosná, dobře provedená a obtížná, práci **doporučuju** k obhajobě a navrhuju hodnocení **výborně**.

Evanston, 12. 5. 2013

Tomáš Füst