

Detekce alkoholu v řečovém signálu

Diplomový projekt měl velice detailní zadání, což recenzentovi umožňovalo snadnou kontrolu vývoje a kvality prací a samozřejmě i jeho plnění.

Diplomant se dle zadání měl:

1. seznámit s problematikou zpracování řečových signálů a vlivu alkoholu na řečové parametry;
2. zpracovat rešerši významných publikací a patentů pojednávajících o vlivu alkoholu na parametry řečových signálů;
3. vybrané algoritmy na detekci alkoholu v kmitočtové a cepstrální oblasti naprogramovat a ověřit na reálném řečovém signálu z existující databáze;
4. určit nutnou minimální úroveň alkoholu pro správnou činnost detekčních algoritmů;
5. vytvořit databázi profesionálně předstírané a skutečné alkoholické řeči s referenčními údaji a poslechovými testy;
6. pomocí vytvořených algoritmů určit, které fonetické jednotky jsou nejvíce citlivé na alkohol;
7. vytvořené algoritmy optimalizovat z pohledu fonetické struktury, kvality a délky testovaných řečových signálů.
8. vytvořit ucelený program na detekci alkoholu v řečovém signálu a konečně
9. statisticky vyhodnotit spolehlivost detekce.

První dvě položky souvisí s přehledem stavu problematiky, příp. z něho nalezených problémů a postupů jejich řešení, podle nichž by se navazující práce měla řídit. Vlastní tvůrčí část projektu představují body 3 až 8, bod 9 reprezentuje dle zadání část závěrečnou, zabývající se výsledky práce a jejich hodnocením.

Stavu problematiky jsou věnovány první tři kapitoly práce. První z nich začíná statistikami vlivu alkoholu na dopravní nehody, které autor dále nerozvádí a při té příležitosti se naskýtá otázka, do jaké oblasti se předpokládá případné praktické využití výsledků práce. Uvádí jednotky měření koncentrace alkoholu v těle a pro laika nepochybně zajímavou tabulku s účinky koncentrace alkoholu v krvi. V kap.1.2 mám poněkud problém s terminologií, i když připouštím, že autor mohl použít pojmy z citované literatury. Soustava jednotek SI, kterou bychom se měli v oficiálních materiálech řídit, neumí pojem „váha“. I název „*promile*“ vyvolává jiné, i když příbuzné asociace. Na druhé straně je nepochybně na autorovi, jak označí veličiny, se kterými bude pracovat. Bohužel, ve vztazích (1.1) a (1.3) používá jeden pojem „*hmotnost*“ s různými významy – množství alkoholu obsažené v nápoji a celkové množství alkoholu, které se z těla odbourá za jednu hodinu. Druhému významu mi neseď ani fyzikální rozměr [g] uvedený ve vztahu (1.3).

Kap.2 se nejdříve velice stručně na dvou stranách zabývá generováním řeči člověkem a definicí základních veličin popisu řečového signálu, jak z hlediska akustiky, tak fonetiky českého jazyka. Problematice „zpracování řečového signálu“ (tj. metodám jeho zpracování a definicím a způsobům určení hodnot parametrů řeči), která by měla být jádrem této části práce, věnuje autor zbytek kap.2. a kap.3. Kap.2 se věnuje obecným způsobům popisu řečových signálů a metodám jejich analýzy, kap.3 pak pracím týkajícím se analýzy řeči ovlivněné požitím alkoholu. V kap.2 autor v podstatě uvádí všechny běžné základní parametry a metody a prokazuje tak dostatečné splnění prvního bodu zadání. Rešerše v kap.3 obsahuje relevantní prameny; recenzentovi nejsou známy žádné jiné podstatné publikační zdroje.

Co ale dle soudu recenzenta v tomto místě textu naprosto schází, je zhodnocení či rozbor publikovaných metod a výsledků a na něj navázaný návrh následujícího postupu. Tento nedostatek vidím i z pohledu např. třetího bodu zadání, který začíná: „vybrané algoritmy ...“. Tedy na základě čeho diplomant algoritmy vybíral či zda je vůbec vybíral? Dále, jaké vlastnosti by měla mít vytvářená databáze, aby bylo možné „určit minimální úroveň alkoholu pro správnou činnost detekčních algoritmů“ (4. bod zadání). Jak posléze stanovit, „které fonetické jednotky jsou nejvíce citlivé na alkohol“ (6. bod zadání)? Na základě kterých kritérií „vytvořené algoritmy optimalizovat z pohledu fonetické struktury, kvality a

délky testovaných řečových signálů“ (7. bod zadání)? Nepřítomnost tohoto rozboru pak způsobila, že není zcela jasné, zda vytvořená databáze i program dostatečně splňují očekávané.

Kap.4 „Implementace“ se pak věnuje stručnému popisu vlastností vzniklé databáze a jednotlivých segmentů experimentálního softwaru, ze kterého jsou v předložené práci uvedeny pouze dílčí skripty. Na to, že by tato část projektu měla tvořit podstatu práce, je kap.4 až nepřírozně střídmá. Celý program nebyl recenzentovi předložen (byť je zřejmě uložen v IS VUT) a tím pádem nebyl jeho text zkontrolován, příp. jeho funkce ověřena.

Kapitola 5 sumarizuje dosažené výsledky zpracování signálů z vytvořené databáze jak v tabulkách, tak i přehledně v grafické podobě. Výsledky se autor snaží i slovně popisovat. Bohužel jsem v této kapitole, ani v následující závěrečné kapitole, nenašel explicitní odpovědi na některé otázky formulované v zadání (kromě stručné zmínky v kap.6 o minimální detekovatelné hladině alkoholu v krvi, ze které ovšem není zcela jasné, zda je to údaj literární či zjištěný v rámci tohoto diplomního projektu). Předpokládám, že diplomant dokáže na tyto otázky odpovědět při obhajobě.

Přírozně nelze očekávat, že práce tohoto typu bude psána jazykem Vančurovým. Nicméně, asi by bylo možné mít výhrady k některým řečnickým drobným jazykovým prohřeškům. Logická struktura textu v některých jeho částech také drobně pokulhává, ale obecně lze konstatovat, že se jazyková kultura práce významně nevyvíjí běžné úrovni prací tohoto typu.

Vlastní text práce je doplněn seznamem symbolů a zkratk (určitě užitečným pro orientaci v textu, i když některé zkratky asi nebylo zcela nezbytné v tomto seznamu uvádět, jako např. VUT) a v publikacích tohoto typu běžnými seznamy obrázků a tabulek, jejichž skutečná potřeba mi pro čtení a orientaci v práci dlouhodobě uniká.

Pro stanovení celkového počtu bodů byla použita následující dílčí kritéria:

- *úvodní přehledová kapitola (úplnost relevantních zdrojů a zpracování publikovaných poznatků)* – přehledová kapitola v podstatě obsahuje všechny relevantní informace, v některých případech pokulhává terminologie a logická návaznost některých informací, proto jen 18 bodů z 20;
- *vyvození závěrů pro navazující práci* – v práci zcela absentuje, ať samostatně či jako součást některé jiné kapitoly a z toho vyplývají mnohé problémy v navazujících aktivitách, proto 0 bodů z 10;
- *tvůrčí část práce (správnost použitých postupů z hlediska cílů i principů a podmínek jejich použití, srozumitelnost jejich popisu)* – z dokumentace vytvořeného programu, která je součástí textu diplomové práce, vyplývá, že nejsou explicitně řešeny některé body zadání – jak bylo zajištěno primární ověření činnosti programu (bod 3), optimalizace programu dle bodu 7 zadání; není rovněž jisté, zda je program schopen nalézt odpovědi podle bodů 4 a 6 zadání - proto 25 bodů z 30;
- *výsledky práce (prezentace experimentálních výsledků z hlediska jejich výpovědní hodnoty; úroveň diskuze nad výsledky; úroveň a správnost závěrů vzhledem k zadání)* – kvůli scházejícím odpovědím na některé otázky zadání pouze 15 bodů z 30;
- *jazyková úroveň, gramatická správnost* – 4 body z 5;
- *grafická úroveň práce* – 5 bodů z 5.

Předložený text diplomového projektu rozhodně představuje nezanedbatelný objem odborných prací, kterými diplomant dostatečně prokázal své inženýrské schopnosti. Proto jej doporučuji k obhajobě a navrhuji celkové bodové hodnocení určené podle výše uvedených dílčích kritérií

67 bodů.

prof. Ing. Jiří Holčík, CSc.

Otázky pro obhajobu

Otázky vyplývající ze zadání projektu, na které by měl diplomant rozhodně odpovědět:

- Jak byly vytvořené algoritmy optimalizovány z pohledu fonetické struktury, kvality a délky testovaných řečových signálů?
- Jaká je minimální úroveň alkoholu pro správnou činnost detekčních algoritmů?
- Které fonetické jednotky jsou nejvíce citlivé na alkohol?

Z dalších otázek lze vzhledem k jejich většímu počtu při obhajobě vybrat dle názoru zkušební komise jen některé:

- Může aplikační oblast ovlivnit způsob zpracování řeči při odhalování vlivu alkoholu na mluvčího?
- Jaký by měl být správný rozměr ve vztahu (1.3)?
- Proč mají řečové mikrosegmenty (kap.2.3, str.14) právě délku 10 ms.
- Prosím, vysvětlíte konstatování (str.14, dole), že fourierovské koeficienty „se nehodí jako vstupní příznaky pro řečové klasifikátory vzhledem k *nepodstatnému snížení informační redundancy ve srovnání s PCM.*“
- Existují české odborné ekvivalenty pro pojmy „jitter“ a „shimmer“?
- dB je jednotka poměrná. Jak to souvisí s definicí pomocí „středního **rozdílu** mezi amplitudami ...“ (definice shimmeru, str.19 nahoře)?