

METHODS OF NOISE SUPPRESSION FOR SPEECH RECOGNITION SYSTEMS

Zuzana Moldříková

Master Degree Programme (2), FEEC BUT

E-mail: xmoldr01@stud.feec.vutbr.cz

Supervised by: Jan Odstrčilík

E-mail: odstrcilik@feec.vutbr.cz

Abstract: This paper deals with noise suppression methods for speech recognition systems. Furthermore, the most common approach for noise suppression – method of spectral subtraction is introduced. The paper also describes environment where speech signals are produced and finally preliminary results are presented.

Keywords: speech signal, noise suppression, speech recognition system, spectral subtraction

1 ÚVOD

V současné době se v oblasti zpracování řečového signálu a konkrétně potlačení šumu a rušení u řečového signálu používá mnoho metod. Z jednodukánových metod patří mezi často užívané metoda spektrálního odečítání, Wienerova filtrace či metoda RASTA (Relative Spectral) a její modifikace. Každá z těchto metod má svá specifika, jinou časovou a výpočetní náročnost a také jiná metoda potlačuje jiný typ rušení. Žádná z metod není natolik univerzální, aby byla použitelná pro každý typ rušení. Proto je potřeba pro jiné prostředí a například jiný rozpoznávač vždy aplikovat nejvhodnější metodu pro potlačení šumu.

Na základě provedené literární rešerše byla vybrána jako vhodná metoda spektrálního odečítání pro její jednoduchost a zároveň úspěšnost. Metoda je často využívána pro vysokou pružnost v reálných podmínkách. V literatuře lze nalézt mnoho prací zabývajících se touto metodou s pozitivními výsledky, viz například významný zdroj. [2, 3]

2 DEFINICE PROBLÉMU

Pro tuto práci je zadán rozpoznávač řeči využívající technologie Sensory's Fluentsoft od společnosti Sensory Inc. Nahrávky řečových signálů musí být ve formátu .wav, 16-bit PCM signed mono a vzorkovány frekvencí 16 kHz. Zadaný rozpoznávač je součástí systému pro rozpoznání řeči v kokpitu dopravního letadla. Pilot letadla vysloví například pokyn ke zjištění stavu počasí v konkrétním městě, rozpoznávač identifikuje příkaz a další systém tento příkaz vyhodnotí a provede ve formě zobrazení konkrétní mapy s aktuálním počasím a doprovodným syntetickým hlasem o stavu počasí. Pro prostředí, ve kterém se systém rozpoznávání využívá je tedy kokpit dopravního letadla. Toto prostředí je typické hlukem, který vydávají motory letadla a další přístroje umístěné v kokpitu. Šum motoru se dá považovat za stacionární a zpravidla dosahuje hodnot 70 - 90 dB. Dalším zdrojem hluku mohou být jiné řečové signály vydávané řečníky jinými než je samotný pilot.

Cílem této práce je implementace metody pro potlačení šumu u řečových signálů vyslovených piloty v kokpitu dopravního letadla. U nahrávek je posuzována především úspěšnost rozpoznání řečového signálu zadaným rozpoznávačem včetně skóre rozpoznání, které je dáno vnitřním výpočtem rozpoznávače a udává míru úspěšnosti rozpoznání.

3 SPEKTRÁLNÍ ODEČÍTÁNÍ

Metoda spektrálního odečítání je vhodná ke zpracování signálu kontaminovaného aditivním šumem v časové nebo frekvenční oblasti. Metoda není výpočetně náročná a je velmi účinná i přes svou jednoduchost. Základním předpokladem pro úspěšnost této metody je nekorelovanost řečového signálu s aditivním šumem. Metoda také předpokládá stacionární charakter aditivního šumu a také přenosového kanálu ve smyslu neměnnosti spektrálních charakteristik v čase, případně dostatečně pomalých změn těchto charakteristik ve srovnání se změnami spektrálních charakteristik řeči. Při splnění těchto předpokladů lze provést odhad spektrálních charakteristik nezašuměného řečového signálu. [1, 2, 3]

Princip metody tedy spočívá v odečtení spektra šumu od spektra řečového signálu kontaminovaného aditivním šumem. Spektrum šumu je odhadováno a obnovováno během těch úseků signálu, kdy nezaznamenáváme řeč. Vstupní signál je segmentován pomocí okna, nejčastěji Hammingova. Na každý segment je nutno aplikovat rychlou Fourierovu transformaci a dále provést diskretní krátkodobou spektrální analýzu dle rovnice [3]

$$X_i(k) = \sum_{n=0}^{N-1} w(n)x_i(n)e^{-jk\frac{2\pi}{N}n}, \quad (1)$$

kde $k = 0, 1, \dots, N - 1$, parametr i udává počet segmentů vstupního signálu $x(n)$. V dalším kroku je uvažován pouze modul spektra i -tého segmentu, který je umocněn na parametr a . Při volbě parametru $a = 2$ je získán odhad výkonové spektrální hustoty. Metoda může využívat detektoru řečové aktivity pro odhad charakteristiky šumu. Je-li znám odhad výkonové spektrální hustoty řečového signálu a šumu, lze provést odečtení těchto odhadů dle rovnice [3]

$$|Y_i(k)|^2 = |X_i(k)|^2 - E\{|N_i(k)|^2\}, \quad (2)$$

kde $k = 0, 1, \dots, N - 1$ a $E\{|N_i(k)|^2\}$ značí střední hodnotu výkonové spektrální hustoty šumu. V dalším kroku je provedeno umocnění výkonové spektrální hustoty hodnotou $1/a$, čímž je získán modul spektra extrahované řeči. Po přidání nezměněného argumentu spektra původního signálu je aplikována inverzní krátkodobá diskretní Fourierova transformaci na každý segment. [1, 2, 3]

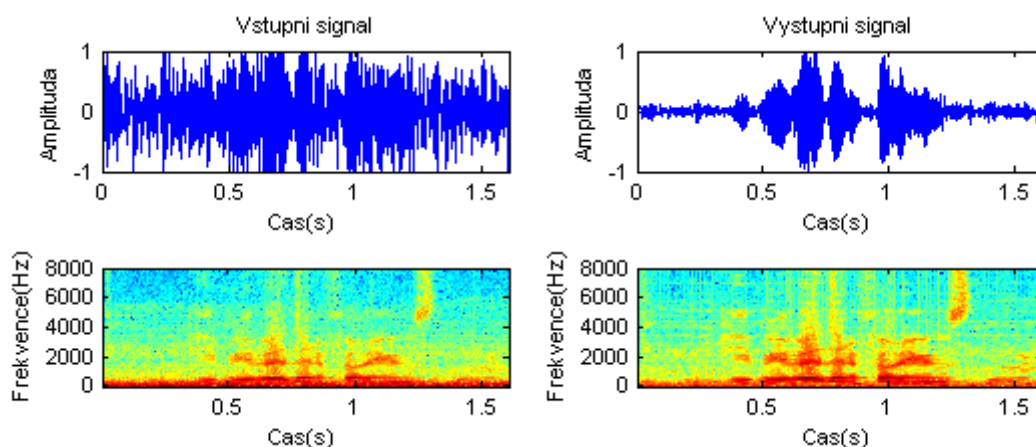
Nevýhody metody spektrálního odečítání spočívají v nutnosti sledování množství odečítané informace. Pokud je odečteno příliš mnoho informace, může dojít ke ztrátě části informace řečového signálu. Naopak pokud je odečteno málo informace, metoda nebude dostatečně účinná a řečový signál bude nadále obsahovat šum. Další nevýhodou metody je velká citlivost na změny ve spektru šumu. Použití algoritmu bez další modifikace ve velmi hlučném prostředí je nevhodné z důvodu vysokého množství vznikajících hudebních tónů. [1, 2, 3]

4 ŘEŠENÍ A MOŽNÉ VYUŽITÍ

V rámci řešení byla nahrána zkušební databáze v simulátoru kokpitu dopravního letadla, kterou tvoří dvacet řečových signálů. Pět vět typicky užívaných piloty bylo nahráno dvěma muži, přičemž jeden má vadu řeči, s použitím náhlavní soupravy a bez. Během prvního testování metody spektrálního odečítání bylo využito toolboxu Voicebox [4], který umožňuje zpracování řečových signálů v prostředí Matlab. Na Obrázku 1 lze pozorovat vliv použitého algoritmu. V případě použití náhlavní soupravy se po aplikaci tohoto algoritmu zvýšila úspěšnost rozpoznání řečových nahrávek z 80 % na 100 %. V případě, kdy se náhlavní souprava nepoužila, úspěšnost rozpoznání klesla ze 100 % na 70 %.

Dále je nahrávána obsáhlejší databáze řečových signálů typicky užívaných vět v simulátoru a reálném kokpitu dopravního letadla, opět ve variantách s použitím vhodně zvolené náhlavní soupravy a bez. Tuto databázi tvoří řečové signály pěti mužů a pěti žen při vyslovení deseti různých vět. Metoda

spektrálního odečítání je testována na všech variantách nahraných řečových signálů, a také na uměle zarušených nahrávkách.



Obrázek 1: Porovnání průběhu a spektrogramu řečové nahrávky věty "Give me weather for Houston" ve variantě bez použití náhlavní soupravy při použití algoritmu spektrálního odečítání

5 ZÁVĚR

Příspěvek popisuje vybranou metodu spektrálního odečítání pro potlačení aditivního šumu u řečových signálů vznikajících v kokpitu dopravního letadla a použití této metody na zkušební datazábí řečových nahrávek. Další kroky spočívají v modifikaci algoritmu spektrálního odečítání, a také v porovnání této metody s Winerovou filtrací. Nezbytnou součástí řešení je porovnání vlivu softwarového a hardwarového řešení daného problému. Náhlavní sady, které se v současnosti vyskytují na trhu se vyznačují schopností výrazně potlačit šum, a proto budou porovnány různé kombinace použití softwarového a hardwarového řešení. Pro efektivní testování a vyhodnocení metod je vytvářeno uživatelské rozhraní GUI v prostředí Matlab umožňující nejen odšumění nahrávek pomocí zvolené metody, ale také přehrání nahrávek a vykreslení průběhů signálů a jejich spektrogramů.

6 PODĚKOVÁNÍ

Tato práce vzniká ve spolupráci a na základě zadání firmy Honeywell, proto bych na tomto místě ráda poděkovala nejen vedoucímu práce, ale i konzultantovi Václavu Pfeiferovi za podnětné návrhy k řešení práce.

REFERENCE

- [1] PSUTKA, J., MULLER L., MATOUŠEK J., RADOVÁ V.: Mluvíme s počítačem česky, 1.vydání Academia Praha 2006, 752s., ISBN 80-200-1309-1.
- [2] BOLL, S.F.: Suppression of Acoustic Noise in Speech Using Spectral Subtraction, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-27, No.2., April 1979.
- [3] SMÉKAL, Z.: Číslicové zpracování řeči, Elektronická skripta pro magisterská studia, Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2009.
- [4] VOICEBOX: Speech Processing Toolbox for MATLAB, dostupný z internetu: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>.