

3 METODY PRO POTLAČENÍ ŠUMU U ŘEČOVÉHO SIGNÁLU

V současné době se pro potlačení šumu u řečového signálu používá mnoho různých metod. Jedná se například o metody spektrálního odečítání, Wienerovy filtrace, RASTA, mapování spektrogramu či algoritmy, které jsou založené na odhadu šumu.

3.1 Spektrální odečítání

Tato metoda se řadí do skupiny metod kompenzace šumu na signálové či příznakové úrovni. Princip těchto metod spočívá ve zpracování signálu postiženého aditivním šumem v časové nebo frekvenční oblasti. Tento signál je pak dále upravován v blocích – tzv. mikrosegmentech. Metoda spektrálního odečítání 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, případné změny jeho spektrálních charakteristik by měly být pomalé ve srovnání se spektrálními charakteristikami řeči [1], [2].

Princip metody spočívá v odečtení spektra šumu od spektra řečového signálu kontaminovaného aditivním šumem, přičemž se předpokládá, že signál šumu není korelovan s řečovým signálem. U signálu předpokládáme stacionaritu šumu nebo jeho pomalou proměnlivost v čase. Dále předpokládáme stacionaritu 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], [5].

V případě, že přenosový kanál bude ideální můžeme obraz řečového signálu $Y(\omega, k)$, který obsahuje aditivní šum, popsat rovnicí

$$Y(\omega, k) = S(\omega, k) + N(\omega, k), \quad (3.1)$$

kde k vyjadřuje čas spektra kterého segmentu, $S(\omega, k)$ je spektrum původního řečového signálu a $N(\omega, k)$ je spektrum aditivního šumu [1].

Odhad spektrálních charakteristik čisté řeči $|\hat{S}(\omega, k)|$ lze provést dle rovnice

$$|\hat{S}(\omega, k)| = |Y(\omega, k)| - \alpha \frac{1}{I} \sum_{i=0}^{I-1} |N(\omega, i)| = |Y(\omega, k)| - \alpha |\overline{N(\omega)}|, \quad (3.2)$$

kde $|Y(\omega, k)|$ je amplitudové spektrum řečového signálu kontaminovaného šumem, $|\overline{N}(\omega)|$ je amplitudové časově zprůměrované spektrum aditivního šumu a α je parametrem, který udává poměrnou část průměrného spektra šumu, kterou následně odečítáme od $|Y(\omega, k)|$ [1].

Odhad lze také provést ve výkonové oblasti dle rovnice

$$|\hat{S}(\omega, k)|^2 = |Y(\omega, k)|^2 - \alpha|\overline{N}(\omega)|^2, \quad (3.3)$$

kde $|\hat{S}(\omega, k)|^2$ je výkonové spektrum odhadovaného řečového signálu, $|Y(\omega, k)|^2$ je výkonové spektrum signálu obsahujícího aditivní šum a $|\overline{N}(\omega)|^2$ je časově zprůměrované výkonové spektrum aditivního šumu [1].

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 [2]. Další nevýhodou metody je velká citlivost na změny ve spektru šumu. U některých signálů je možné, že časově zprůměrovaná hodnota šumu odhadnutého z úseku, kdy není přítomna řeč, bude větší než spektrum řeči. V těchto případech má pak odhad amplitudového nebo výkonového spektra čisté řeči zápornou hodnotu. Této situaci se předchází zavedením mapovací funkce, která odhadu spektra řečového signálu přiřazuje nezápornou hodnotu. Nevýhodou zavedení mapovací funkce je možnost vzniku hudebního šumu. Tento šum pak může negativně ovlivnit rozpoznávání signálu [1].

Použití algoritmu bez další modifikace je ve velmi hlučném prostředí nevhodné z důvodu vysokého množství vznikajících hudebních tónů. V případě použití algoritmu v hlučném prostředí jej lze modifikovat například použitím Beroutiho algoritmu, nelineárního, pásmového či rozšířeného spektrálního odečítání [9].

3.1.1 Beroutiho algoritmus

Tato úprava obecného algoritmu spektrálního odečítání byla navržena M. Beroutim a kolektivem v [19]. Postup spočívá v odečtení odhadu výkonového spektra šumu při zajištění neklesnutí hodnot výsledných spektrálních komponent pod předem stanovenou minimální hodnotu dle následujícího vztahu

$$|\hat{X}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha|\hat{D}(\omega)|^2 & \text{jestliže } |Y(\omega)|^2 > (\alpha + \beta)|\hat{D}(\omega)|^2 \\ \beta|\hat{D}(\omega)|^2 & \text{jinak} \end{cases} \quad (3.4)$$

kde α je odečítací faktor a platí $\alpha > 1$ a $\beta(0 < \beta \ll 1)$ je parametr spektrálního pozadí. Oba parametry poskytují algoritmu velkou flexibilitu. Parametr α ovlivňuje zkreslení zpracovaného signálu, parametr β ovlivňuje množství zbytkového šumu.

Pokud však bude hodnota parametru β příliš velká, hudební tóny budou potlačeny, ale ve výstupním signálu bude výrazně slyšet šum. Naopak pokud bude hodnota parametru β příliš malá, intenzita hudebních tónů ve výstupním signálu bude příliš velká. Odečtením odhadu spektra šumu je možné snížit amplitudu širokopásmových vrcholků a v některých případech je dokonce eliminovat [2], [9], [19].

Dále je nutné uvažovat hodnotu SNR zpracovávaného signálu. Pro signály s vysokou hodnotou SNR je nutné nastavit parametr α na nízkou hodnotu a naopak. Berouti a kolektiv [19] navrhli výpočet parametru α dle vzorce

$$\alpha = \alpha_0 - \frac{3}{20}SNR - 5dB \leq SNR \leq 20dB, \quad (3.5)$$

kde α_0 je požadovaná hodnota α při SNR = 0 dB. Použitím Beroutiho algoritmu je tedy možné částečně eliminovat nevýhodu metody spektrálního odečítání, a to vzniku hudebních tónů [2], [9].

3.1.2 Nelineární spektrální odečítání

Základním předpokladem metody nelineárního spektrálního odečítání je rovnoměrné postižení šumem všech spektrálních komponent. Přes celé spektrum je proto odečten odhad šumu za pomoci odečítacího faktoru α . Vzhledem k základnímu předpokladu této modifikace spektrálního odečítání není možné použít algoritmus na řečový signál kontaminovaný hlukem na ulici či v restauraci [2].

Nelinearita této metody spočívá v odečtení větších hodnot na frekvencích s nižším poměrem signál šum a menších hodnot na frekvencích s větším poměrem signálu k šumu. Tento postup lze popsat vztahem

$$|\hat{X}(\omega)| = \begin{cases} |\bar{Y}(\omega)| - a(\omega)N(\omega) & \text{jestliže } |\bar{Y}(\omega)| > a(\omega)N(\omega) + \beta \cdot |\bar{D}(\omega)| \\ \beta|\bar{Y}(\omega)| & \text{jinak} \end{cases} \quad (3.6)$$

kde β je proměnná udávající zaokrouhlení, $|\bar{Y}(\omega)|$ a $|\bar{D}(\omega)|$ jsou vyhlazené odhady zašuměného řečového signálu a šumu, $\alpha(\omega)$ je odečítací faktor závislý na frekvenci a $N(\omega)$ je nelineární funkce spektra šumu [2].

Odhady $|\bar{Y}(\omega)|$ a $|\bar{D}(\omega)|$ zašuměné řeči získáme následovně

$$|\bar{Y}_i(\omega)| = \mu_y |\bar{Y}_{y-1}(\omega)| + (1 - \mu_y) |\bar{Y}_i(\omega)| \quad (3.7)$$

$$|\bar{D}_i(\omega)| = \mu_d |\bar{D}_{d-1}(\omega)| + (1 - \mu_d) |\hat{D}_i(\omega)| \quad (3.8)$$

kde $|\bar{Y}_i(\omega)|$ je amplitudové spektrum zašuměného řečového signálu z i -tého segmentu a $|\hat{D}_i(\omega)|$ je odhad amplitudového spektra šumu z i -tého segmentu. Konstanty μ_y , μ_d nabývají hodnot v rozmezí $0, 1 \leq \mu_y \leq 0,5$ a $0,5 \leq \mu_d \leq 0,9$. Nelineární

funkce spektra šumu je získána výpočtem maxima z amplitudového spektra šumu z posledních 40 segmentů. [2]

3.1.3 Vícepásmové spektrální odečítání

Tato metoda narozdíl od předchozí metody nelineárního spektrálního odečítání vychází z předpokladu, že šum nepostihuje spektrum řečového signálu rovnoměrně. Princip metody spočívá v segmentaci spektra řeči na N nepřekrývajících se pásem a metoda spektrálního odečítání je následně aplikována jednotlivě na každé pásmo. Segmentace spektra může být provedena v časové doméně za použití filtrů pásmových propustí nebo ve frekvenční doméně užitím vhodného okna. Způsob segmentace ve frekvenční doméně je dle autora [2] více využíváný. Odhad spektra čisté řeči v i -tém pásmu je získán dle rovnice

$$|\hat{X}_i(\omega)_k|^2 = |\bar{Y}_i(\omega)_k|^2 - \alpha_i \cdot \delta_i \cdot |\hat{X}_i(\omega)_k|^2 \quad (3.9)$$

kde $\beta_i \leq \omega_k \leq e_i$, $\omega_k = 2\pi k/N$ ($k = 0, 1, \dots, N - 1$) jsou diskrétní hodnoty frekvencí, $|\hat{D}_i(\omega)_k|^2$ je odhad výkonového spektra šumu (získáno a obnovováno během úseků signálu, kde se nevyskytuje řeč), b_i a e_i jsou počáteční a koncové frekvence v i -tém frekvenčním pásmu, α je odečítací faktor v i -tém pásmu a δ_i je přídatný pásmový odečítací faktor, který je možné individuálně nastavit pro každé frekvenční pásmo a přizpůsobit tím proces zpracování řečového signálu [2], [17].

Záporné hodnoty vznikající odečtením dle rovnice 3.9 jsou dále zpracovány dle vztahu

$$|\hat{X}_i(\omega)_k|^2 = \begin{cases} \hat{X}_i(\omega_k)^2 & \text{jestliže } |\hat{X}_i(\omega_k)|^2 > \beta |\bar{Y}_i(\omega_k)|^2 \\ \beta |\bar{Y}_i(\omega_k)|^2 & \text{jinak} \end{cases} \quad (3.10)$$

kde parametr β je nastaven na hodnotu 0,002 [2]. K dalšímu maskování zbývajících hudebních tónů je využito přidání části šumového spektra ke zpracovanému signálu dle vztahu

$$|\tilde{X}_i(\omega)_k|^2 = |\hat{X}_i(\omega)_k|^2 + 0,05 \cdot |\bar{Y}_i(\omega)_k|^2 \quad (3.11)$$

kde $|\tilde{X}_i(\omega)_k|^2$ vyjadřuje nové výkonové spektrum zpracovaného signálu [2].

Odečítací faktor v i -tém pásmu α je funkcí segmentálního SNR i -tého pásma a je vypočítán dle vztahu

$$\alpha_i = \begin{cases} 4,75 & SNR_i < -5 \\ 4 - \frac{3}{20}(SNR_i) & -5 \leq SNR_i \leq 20 \\ 1 & SNR_i > 20 \end{cases} \quad (3.12)$$

kde v i -tém pásmu je poměr SNR_i dán vztahem

$$SNR_i(dB) = 10 \log_{10} \left(\frac{\sum_{\omega_k=b_i}^{e_i} |\bar{Y}_i(\omega)_k|^2}{\sum_{\omega_k=b_i}^{e_i} |D_i(\omega)_k|^2} \right) \quad (3.13)$$

Přestože použitím odečítacího faktoru α_i je možné kontrolovat úroveň odečítaného šumu, je potřeba dále nastavit parametr δ_i , kterým je také možné kontrolovat množství odečítaného šumu v každém pásmu. Hodnoty tohoto parametru jsou stanovovány empiricky a dány následujícím vztahem

$$\delta_i = \begin{cases} 1 & f_i \leq 1kHz \\ 2,5 & 1kHz < f_i \leq \frac{Fs}{2} - 2kHz \\ 1,5 & f_i > \frac{Fs}{2} - 2kHz \end{cases} \quad (3.14)$$

kde f_i je vyšší frekvence v i -tém pásmu a Fs je vzorkovací frekvence v Hz. Většina energie řečového signálu se nachází v nižších frekvenčních pásmech, a proto je často užívána menší hodnota parametru δ_i pro pásma nižších frekvencí, aby nedocházelo k příliš velkému zkreslení zpracovávaného signálu [2], [17].

3.1.4 Spektrální odečítání využívající adaptivního průměrování zisku

Princip metody spektrálního odečítání, která využívá adaptivního průměrování zisku, spočívá v segmentaci vstupního signálu na L segmentů a další rozdělení těchto segmentů na M ($M < L$) vzorků. Dále je vypočítáno spektrum každého vzorku a toto spektrum je průměrováno s cílem zisku neměnného odhadu amplitudového spektra. Zisková funkce je vypočtena dle vztahu

$$G_i^{(M)}(\omega) = 1 - k \frac{|\hat{D}_i^{(M)}(\omega)|}{|\bar{Y}_i^{(M)}(\omega)|} \quad (3.15)$$

kde $|\hat{D}_i^{(M)}(\omega)|$ je odhad amplitudového spektra signálu, $|\hat{D}_i^{(M)}(\omega)|$ je odhad amplitudového spektra šumu, které je obnovováno během úseků bez řečové aktivity a k je odečítací faktor, dle [21] nastavený na hodnotu 0,7 [2], [21].

Zisková funkce $G_i^{(M)}(\omega)$ je dále průměrována v čase za účelem snížení její variability dle vzorce

$$\bar{G}_i^{(M)}(\omega) = a_i \bar{G}_{i-1}^{(M)}(\omega) + (1 - a_i) G_i^{(M)}(\omega) \quad (3.16)$$

kde $\bar{G}_i^{(M)}$ vyjadřuje ziskovou funkci v i -tém segmentu a a_i je adaptivní vyhlazovací parametr, který je stanoven dle vztahu

$$a_i = \begin{cases} \gamma a_{i-1} + (1 - \gamma)(1 - \beta_i) & \text{jestliže } a_{i-1} < 1 - \beta_i \\ 1 - \beta_i & \text{jinak} \end{cases} \quad (3.17)$$

kde γ je vyhlazovací parametr, v [21] stanoven na hodnotu 0,8 a parametr β vyjadřuje spektrální odchylku. Hodnota parametru α_i může rychle klesat a tímto umožňuje ziskové funkci rychlou adaptaci na nový vstupní signál. Tato modifikace metody spektrálního odečítání standardně využívá diskrétní přímé a inverzní Fourierovy transformace ke zpracování řečového signálu a také metody OLA (Overlap and add) k převedení zpracovaných segmentů zpět do jednorozměrného signálu [2], [21].

3.2 Wienerův filtr

Wienerův filtr klade požadavek stacionarity jak na šum, tak řečový signál. Návrh Wienerova filtru vychází ze střední kvadratické chyby mezi originálním řečovým signálem a odhadem tohoto signálu. Odhad původního řečového signálu je výstupem Wienerova filtru [1], [4].

Nejobecnější tvar Wienerova filtru popisuje rovnice

$$M(\omega) = \frac{S_{xy}(\omega)}{S_{yy}(\omega)}, \quad (3.18)$$

kde $M(\omega)$ je frekvenční charakteristika restauračního filtru, $S_{xy}(\omega)$ vyjadřuje vzájemné spektrum mezi originálním a zkresleným signálem a $S_{yy}(\omega)$ je výkonové spektrum zkresleného signálu. Návrh dle rovnice 3.18 představuje optimální lineární filtr pro lineární i nelineární zkreslení [4].

Rovnice 3.19 je součinem prostého inverzního filtru a Wienerova korekčního faktoru, který nabývá pouze hodnot v intervalu $< 0, 1 >$. Cílem použití Wienerova korekčního faktoru je snížit amplitudový přenos v některých frekvenčních oblastech.

$$M(\omega) = \frac{1}{G(\omega)} \frac{|G(\omega)|^2}{|G(\omega)|^2 + \frac{S_{vv}(\omega)}{S_{xx}(\omega)}}, \quad (3.19)$$

kde $G(\omega)$ je frekvenční charakteristika odpovídající zkreslení a její převrácená hodnota tedy odpovídá prostému inverznímu filtru. $S_{vv}(\omega)$ je dále výkonové spektrum šumu a $S_{xx}(\omega)$ je výkonové spektrum původního signálu [4].

Rovnice 3.18 a 3.19 předpokládají přístup k originálnímu signálu. Ne vždy je možné této podmínce vyhovět, a proto je potřeba rovnici modifikovat tak, aby obsahovala členy, jejichž hodnoty je reálně získat měřením. Wienerův filtr lze tedy popsat i rovnicí

$$M(\omega) = \frac{1}{G(\omega)} \frac{S_{yy}(\omega) - S_{vv}(\omega)}{S_{yy}(\omega)}, \quad (3.20)$$

přičemž všechny hodnoty této rovnice lze reálně získat [4].

3.3 RASTA

Metoda RASTA, celým názvem Relative Spectral, byla navržena na základě studia vlastností lidského sluchu. Metoda potlačuje spektrální složky s vyšší či nižší rychlostí změn než jsou změny řeči. Metoda je rozšířením metody PLP (Perceptual Linear Predictive), jejímž účelem je odstranit zkreslení vznikající přenosovým kanálem či aditivním šumem [1], [5], [11].

Princip metody RASTA spočívá v následujících krocích. Stejně jako u metody spektrálního odečítání je potřeba řečový signál segmentovat a počítat krátkou Fourierovu transformaci jednotlivých segmentů. Po segmentaci následuje číslicová filtrace časového průběhu modulu jednotlivých spektrálních složek. Dále je potřeba určit poměr signál šum, jehož hodnota slouží jako podklad k volbě kmitočtového číslicového filtru typu pásmová propust. Po filtraci časových relativních změn modulu jednotlivých harmonických změn pro všechny frekvence je na signál aplikována inverzní kmitočtová číslicová filtrace a nakonec metoda přičtení přesahu. Metoda je neúčinná v případě shody časové změny řeči a šumového signálu [5].

3.4 Mapování spektrogramu

Metoda mapování spektrogramu se zásadně liší od metod pro potlačení šumu, které byly v práci dosud uvedeny. Metoda využívá vlastnosti spektrogramu k práci s řečovým signálem v časově-frekvenční oblasti. Pomocí této metody je možné zachovat oblast řečové aktivity (oblast s vyšší energií užitečného signálu než je energie pozadí) při současném odstranění šumu ze vstupního signálu využitím binární masky. Algoritmus této metody zahajuje zpracování vstupního řečového signálu segmentací a váhováním Hammingovým oknem. V dalším kroce je proveden výpočet komplexního spektrogramu z několika po sobě následujících segmentů vstupního signálu. Metoda využívá detektor řečové aktivity pro stanovení počáteční hodnoty prahu z úseku bez řečové aktivity. Dále dochází k vytvoření binární masky a zpracování obrazu filtrací. Poté je vypočten součin masky a původního spektrogramu a dále je aplikována zpětná Fourierova transformace komplexního spektrogramu. Posledním krokem algoritmu je převedení zpracovaných segmentů do jednorozměrného signálu pomocí metody OLA (Overlap and Add). Mezi nevýhody této metody patří například výpočetní a paměťová náročnost, dále také vznik hudebních tónů [5], [14].

3.5 Algoritmy založené na odhadu šumu

Tyto algoritmy se od ostatních liší především v přístupu k šumu. Narozdíl od ostatních metod šum není odhadován pomocí detektoru řečového signálu, tzn. z úseků,

kde není přítomna řeč, ale v průběhu celého signálu. Proto jsou algoritmy založené na odhadu šumu vhodné především pro vysoce nestacionární prostředí. Tyto algoritmy lze v zásadě rozdělit do tří skupin. První skupina je založena na sledování minima výkonového spektra řečového signálu obsahujícího šum, a to buď s použitím okna nebo v průběhu celého signálu. Druhá skupina využívá pro odhad šumu časově rekurzivní průměrování. Šumová charakteristika je tedy odhadována jako vážený průměr předešlých odhadů šumu a aktuálního spektra šumu. Třetí skupina využívá technik vycházejících z vlastností histogramu. Odhad šumu vychází z histogramu hodnot výkonového spektra - jako odhad šumu je brána hodnota odpovídající maximální hodnotě v histogramu [2].