

# ANALÝZA POTLAČOVÁNÍ AKUSTICKÉHO ECHA A DTD DETEKCE V CHYTRÝCH TELEFONECH

*Jan Klapuch, Petr Pollák*

České vysoké učení technické v Praze, Fakulta elektrotechnická, K13131  
klapujan@fel.cvut.cz, pollak@fel.cvut.cz

## Abstrakt

Článek se zabývá testováním algoritmů pro potlačování akustického echa a detekci společné promluvy v prostředí chytrých mobilních telefonů. Akustické echo společně se šumem prostředí představuje obecně velmi nežádoucí jev v telekomunikacích a zejména v mobilních telefonech. Konstrukce současných mobilních telefonů a jejich použití způsobuje vysokou úroveň echa i zkreslení, čímž vzniká náročnější úloha než v případě klasických pevných telefonních systémů. Hlavní úlohou bylo především vytvořit vhodný detektor DTD, jenž je významným prvkem při potlačování echa. Byl analyzován detektor založený na průměrné koherenci signálů reproduktoru a mikrofону s využitím keprstrální detekce řeči v obou kanálech. K tomuto a dalším účelům byla vytvořena GUI aplikace v MATLABu, umožňující provádět analýzu a testování algoritmů. Simulace byly prováděny na nově vytvořené řečové databázi, jenž reprezentovala mobilní komunikaci. Vyhodnocení DTD bylo prováděno statisticky jako míra chybné detekce v celém signálu a zvláště jako míra chybné detekce v kritických segmentech signálu, která jsou důležitá pro správný běh systému AEC.

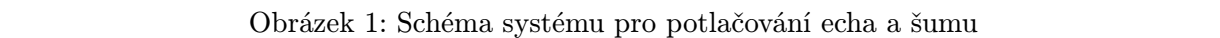
**Klíčová slova:** potlačování akustického echa, detekce řeči, koherence, AEC, DTD, VAD, LMS, ERLE

## 1 Úvod

Mobilní komunikace je čím dál více používaným komunikačním nástrojem. Její rozšíření zahrnuje nejen různé způsoby použití, jak v tichých místnostech, tak v dopravních prostředcích, na ulici, restauracích apod. Zejména zahrnuje i nepřeborné množství přístrojů s odlišnými způsoby zpracování zvuku a s různou kvalitou mikrofónů či reproduktorů. Vyšší variabilita prostředí a konstrukcí telefonních zařízení klade vyšší požadavky na systém pro potlačování nežádoucích jevů této komunikace, než jak tomu bylo v případě pevné veřejné telefonní sítě.

Významnými nežádoucími jevy jsou tedy akustické echo, vznikající akustickou cestou mezi reproduktorem a mikrofónem mobilního přístroje, a šum prostředí. Akustická cesta může být velice krátká díky přenosu zvuku po konstrukci telefonu nebo při velmi blízkém odrazu a to způsobuje, že echo může mít vyšší hladinu intenzity zvuku než užitečný signál blízkého mluvčího. Potlačování akustického echa a šumu prostředí v mobilní komunikaci je řešeno systémem jenž obsahuje více bloků zpracování signálu. Obrázek 1 znázorňuje uspořádání systému a vazbu mezi reproduktorem a mikrofónem. Signál vzdáleného mluvčího  $x[n]$  je převeden do analogové podoby a vyslán reproduktorem do okolního prostředí blízkého mluvčího. Na této straně je společně s jeho signálem  $s[n]$  přičten signál echa  $x'[n]$  a šum prostředí  $n[n]$ . Takto vzniklý signál  $d[n]$  reprezentuje výstup získaný z mikrofónu telefonu na straně blízkého mluvčího a je nutné v něm potlačit nežádoucí složky. V tomto případě pomocí potlačovače echa (AEC) a redukce šumu (NR).

Akustické echo se nejčastěji potlačuje pomocí NLMS algoritmu [1]. Tento časový algoritmus je více citlivý na existenci společné promluvy v signále, oproti LMS algoritmu ve frekvenční



## 2 Koherenční detektor společné promluvy

V této práci byl navržen a analyzován kombinovaný koherenční detektor společné promluvy s pomocnými kepstrálními detektory řečové aktivity v obou kanálech mobilní komunikace. Práce je více zaměřena na detekci řeči v mobilní komunikaci a proto je teoretická část věnována této oblasti.

kde  $S_x[k]$  a  $S_d[k]$  jsou spektrální výkonové hustoty signálu  $x[n]$  respektive  $d[n]$  a  $S_{xd}[k]$  je jejich vzájemná spektrální výkonová hustota. Proměnná  $k$  reprezentuje index frekvenční složky  $f_k = \frac{f_s k}{N}$ , dle vzorkovací frekvence  $f_s$  a počtu vzorků segmentu  $N$ . Spektrální veličiny je nutné vyhlazovat průměrováním, a proto zde bylo použito exponenciální vyhlazování, viz vztahy (2),(3), (4), kde  $p$  je parametr vyhlazování. Zvolen byl vyšší stupeň vyhlazení, tedy  $p = 0.8 \div 0.9$ , odpovídající časové konstantě  $40 \div 80$  ms.

$$\overline{S_{x,i}[k]} = p \cdot \overline{S_{x,i-1}[k]} + (1 - p) \cdot |X_i[k]|^2 \quad (2)$$

$$\overline{S_{d,i}[k]} = p \cdot \overline{S_{d,i-1}[k]} + (1 - p) \cdot |D_i[k]|^2 \quad (3)$$

$$\overline{S_{xd,i}[k]} = p \cdot \overline{S_{xd,i-1}[k]} + (1 - p) \cdot X_i[k] D_i[k]^* \quad (4)$$

Koherence je pro odstranění odchylek dále vyhlazována exponenciálním průměrováním s parametrem  $p_{msc} = 0.5 \div 0.6$ , odpovídající 16  $\div$  20 ms. Koherenci signálu pak můžeme vypočítat dle vztahu:

$$\overline{MSC_i[k]} = p_{msc} \cdot \overline{MSC_{i-1}[k]} + (1 - p_{msc}) \cdot \frac{|\overline{S_{xd,i}[k]}|^2}{\overline{S_{x,i}[k]} \cdot \overline{S_{d,i}[k]}} \quad (5)$$

Segmentace signálu byla definována délkou okna pro AEC systém, zde bylo použito 256 vzorků. Pro výpočty detekcí byl zvolen 75% překryv, zpracovávalo bylo tedy 4x více segmentů než je tomu u samotného AEC. Pro délku okna 256 vzorků a vzorkovací frekvenci 8 kHz se jedná o 8 ms časový segment nepřekrývajících se dat. Při každém novém segmentu dat jsou veličiny aktualizovány pomocí vyhlazených odhadů spektrální výkonové hustoty a střední koherence. Segmenty byly váhovány Hammingovým oknem a transformovány do spektrální oblasti pomocí FFT transformace.

Následujícím krokem byla transformace frekvenčně závislé koherence na veličinu popisující daný segment signálu. Zvoleny byly 3 varianty: průměrná koherence z vybraného frekvenčního rozsahu; maximální koherence a minimální koherence. Průměrná koherence byla vypočtena dle vztahu (6), maximální koherence byla průměrována z 20 nejvyšších hodnot (7), kde  $l$  je index frekvenčních složek seřazených od maxima do minima, a minimální z 10 nejnižších hodnot (8), kde  $o$  je index frekvenčních složek seřazených od minima do maxima.

$$MSC1_i = \frac{1}{N} \sum_{k=k_{min}}^{k_{max}} MSC_i(k) \quad (6)$$

$$MSC2_i = \frac{1}{20} \sum_{l=1}^{20} MSC_i(l) \quad (7)$$

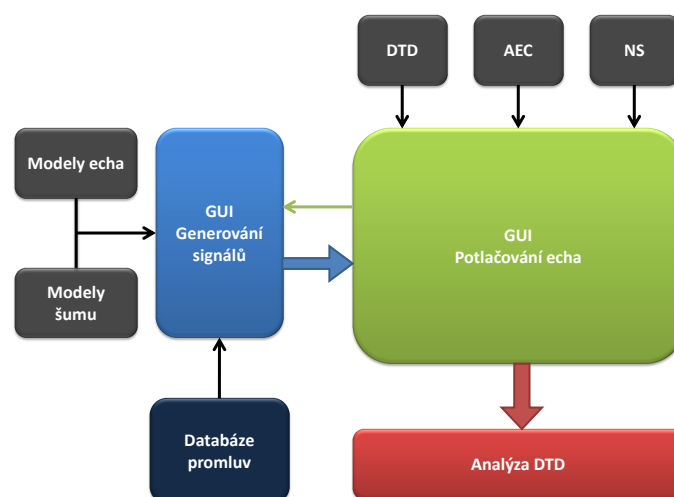
$$MSC3_i = \frac{1}{10} \sum_{o=1}^{10} MSC_i(o) \quad (8)$$

U koherence počítané jako průměrné hodnoty frekvenčního rozsahu, je tento rozsah omezen na frekvence od 400 Hz do 3800 Hz. Vztah (6) je tedy definován na rozsahu  $< k_{min}; k_{max} >$ , kde  $k_{min} = \frac{f_{min}}{f_s} \cdot N$  a  $k_{max} = \frac{f_{max}}{f_s} \cdot N$ . Tento rozsah je zvolen kvůli šumu prostředí, který maskuje dolní kmitočty signálu mikrofону a horní rozsah je přizpůsoben filtraci vyšších kmitočtů v telefonu. Následné prahování sledovaného typu koherence je opět vyhodnocováno ze změn dané veličiny a je aplikováno opět exponenciální vyhlazování pro rostoucí a klesající oblast odlišně. Parametry zapomínání byly zvoleny v rozsahu 0.8  $\div$  0.99999. Rozhodovací práh je zvolen pro průměrnou hodnotu koherence v 60 % rozsahu maxima a minima veličiny dle vyhlazování z předcházejících hodnot. Pro maximální a minimální koherenci je zvolen práh 80 % resp. 40 %.

Detektor společné promluvy je pro správné vyhodnocování doplněn detektory řeči, založené na mel-kepstrálních koeficientech [4], v kanále blízkého a vzdáleného mluvčího. To mu umožňuje zpřesnit detekci na segmenty, kde je obsažena řeč v mikrofону a kde je skutečně možné, aby se vyskytovalo echo vzdáleného mluvčího. Protože pokud v kanále vzdáleného mluvčího není obsažena řeč, není možné aby v kanále mikrofону bylo i echo, s uvažováním zpoždění echa oproti signálu reproduktoru. Toto zpoždění je nejvýraznější zejména pro VoIP komunikaci, kde dosahuje různým zpracováním i 150 ms a více. Pro takto vysokou hodnotu je nutné provádět detekci zpoždění, většinou pomocí korelační funkce a podle něj posunovat segmenty vzdáleného mluvčího.

### 3 MATLAB implementace

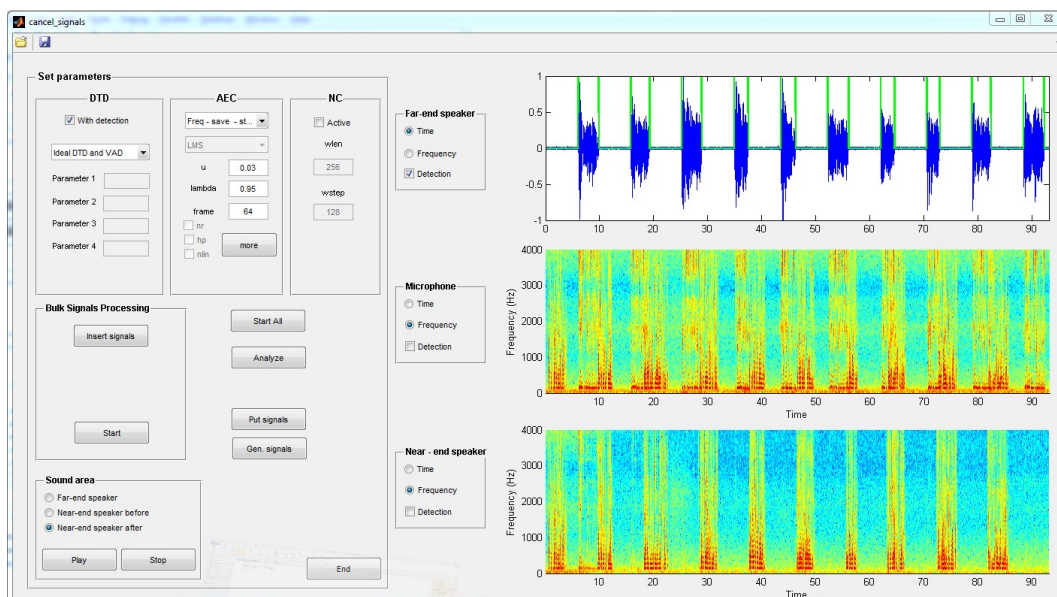
Pro testování algoritmů využívaných při potlačování akustického echa a šumu prostředí bylo vytvořeno grafické prostředí v programovém nástroji MATLAB. Jde o komplexní nástroj umožňující generování vlastních signálů simulujících hovor v mobilním zařízení v obou kanálech, potlačení nežádoucích jevů, testování detekce DTD a analýzu výsledků. Hlavní bloky programu jsou přehledně zobrazeny ve schématu na obrázku 2.



Obrázek 2: Blokové schéma vytvořeného GUI

Generování vlastních signálů umožňuje vytvoření dvojice signálů simulující vzdáleného a blízkého mluvčího společně s akustickým echem a šumem prostředí. Společně jsou vytvářeny i soubory obsahující ideální detekci, jenž je vhodná pro testování reálných detektorů řeči či společné promluvy. Signály jsou generovány z promluv české řečové databáze SPEECON a přidán je k nim šum a echo, dle vytvořených modelů vycházejících z reálných nahrávek mobilní komunikace. Více o modelování mobilní komunikace je možné nalézt v [5].

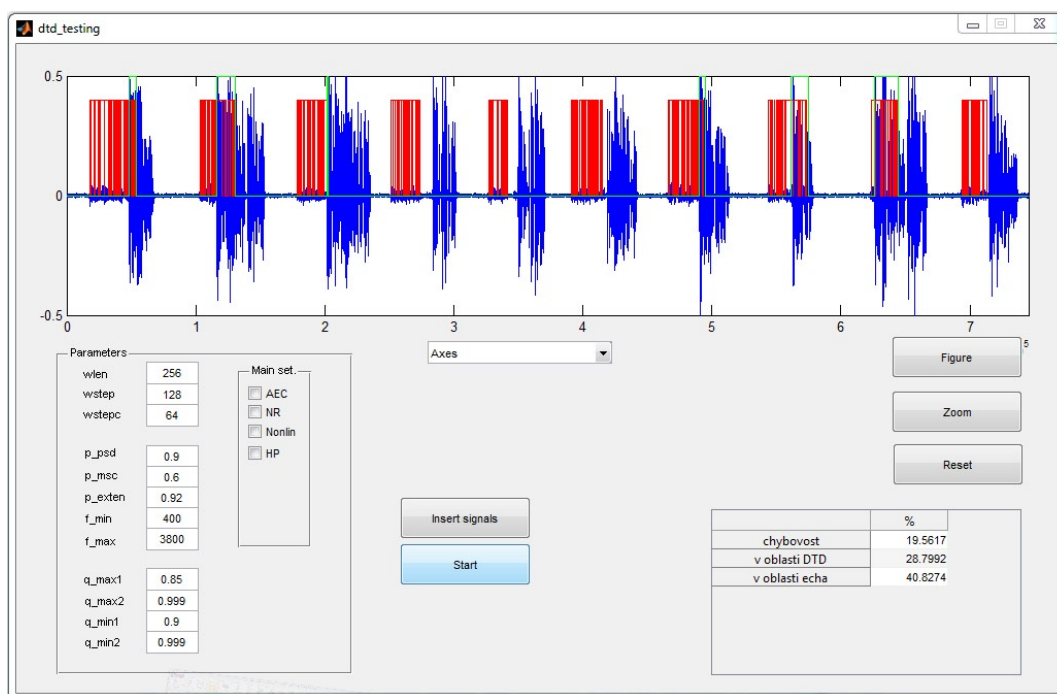
Hlavní prostředí programu je na obrázku 3, ve kterém je možné otevírat ostatní části programu. Toto hlavní prostředí slouží pro testování AEC v kombinaci s ideálním či reálným DTD a potlačením šumu. Pracovat lze s jednotlivými signály a testovat různé varianty nastavení s okamžitou zpětnou vazbou v podobě grafických průběhů v časové a frekvenční oblasti nebo v podobě přehrání zvuku. Avšak je zde i hromadné zpracování dat pro statistickou analýzu, při níž jsou výstupem tabulky s daty ve formátu pro vložení do textu v LaTeXu.



Obrázek 3: Hlavní okno aplikace

Hlavní částí programu pro analýzu a testování detektoru společné promluvy je prostředí Analýza DTD, viz obrázek 4. Zde probíhá experimentální část výzkumu hledáním vhodných nastavení systémů. Data je zde možné vkládat jednotlivě či hromadně a výsledky jsou zobrazovány přímo v grafu či ukládány do souboru jako výsledky chybovosti algoritmu.

Kromě základního nastavení délky zpracovávaných segmentů lze měnit parametry vyhlazených odhadů veličin a rozsah vyhodnocovaných frekvencí signálů. Pod základním grafem v rozbalovacím menu lze snadno měnit zobrazované průběhy, které jsou, kromě signálu v časové či frekvenční doméně, průběhy koherencí, detekce VAD v jednotlivých kanálech či kepstrální vzdálenosti v nich.



Obrázek 4: GUI - Analýza DTD

## 4 Dosažené výsledky

Základem pro testování algoritmu je mimo testovací prostředí nutné zejména dostatečné množství dat, na kterých simulace probíhá. V našem případě jde o řečovou databázi simulující mobilní komunikaci v obou kanálech. Za tímto účelem byla vytvořena nová databáze pomocí aplikace GUI - Generování signálů, obrázek 4. Z této databáze byla použita kategorie č. II (tabulka č.1), jenž obsahuje překrývající komunikaci blízkého a vzdáleného mluvčího, tedy variantu vhodnou pro testování DTD. Signály z databáze jsou sestaveny pomocí promluv české řečové databáze SPEECON a vytvořenými modely echa a šumu z analyzovaných nahrávek mobilních telefonů.

Tabulka 1: Parametry databáze - Kategorie II

| Kategorie II, F10 - N10, s náhodným překryvem mluvčích |          |               |            |            |
|--------------------------------------------------------|----------|---------------|------------|------------|
| Skupina                                                | ERL (dB) | zpoždění (ms) | šum F (dB) | šum N (dB) |
| 1                                                      | 8        | 35            | 30         | 30         |
| 2                                                      | 10       | 35            | 30         | 30         |
| 3                                                      | 15       | 35            | 30         | 30         |
| 4                                                      | 15       | 35            | 15         | 15         |

Nejčastěji používanou veličinou pro analýzu AEC je ERLE (Echo Return Loss Enhancement) [6], tedy míra potlačení echa v dB, porovnávající signál echa před a po průchodu systémem.

$$ERLE = 10 \log \left( \frac{E\{d^2[n]\}}{E\{e^2[n]\}} \right) \cdot VAD[n] \quad (9)$$

Avšak pro vyhodnocení detekce u většího množství dat je nutné porovnávat reálnou detekci s ideální. V nové databázi tato ideální detekce je obsažena a lze ji pro toto vyhodnocení lehce využít. Výstupem simulací je tedy chybovost reálné detekce. Chybovost byla vypočtena pro celkovou dobu trvání signálu (10) a pro 2 významné úseky, pro úsek skutečné společné promluvy (11) a pro úsek samostatného echa (12).

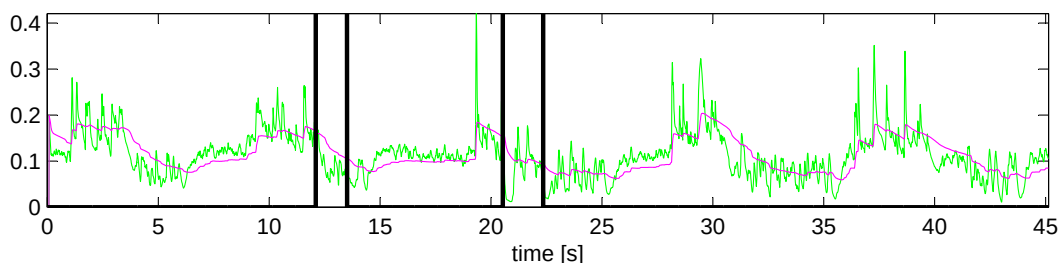
$$ERR_1 = \sum_{i=0}^{M/64} (DTD_{ideal}[i] \otimes DTD_{real}[i]) / M * 100, \quad (10)$$

kde  $M$  je délka signálu, jenž se dělí délkou překryvu segmentů. Chybovost se v tomto případě počítá pomocí funkce  $XOR$  z ideální a reálné detekce jenž je vztažena k délce celého signálu. Oproti tomu vztah (11) ukazuje způsob výpočtu chybovosti jen pro úsek skutečné společné promluvy a vztah (12) jen pro úsek samostatného echa.

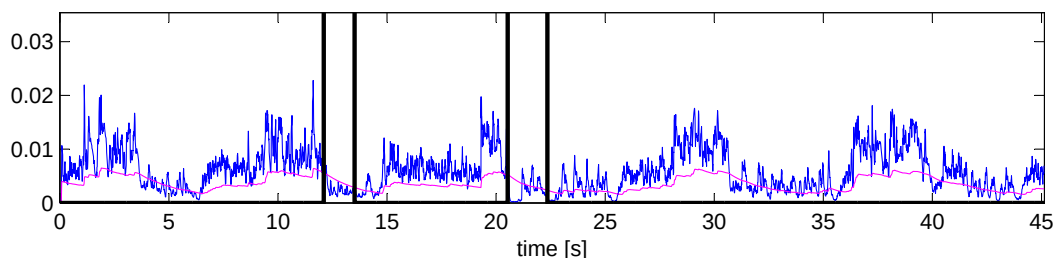
$$ERR_2 = \sum_{i=0}^{M/64} (DTD_{ideal}[i] \cdot \overline{DTD_{real}}) / \sum_{i=0}^{M/64} (DTD_{ideal}[i]) * 100 \quad (11)$$

$$ERR_3 = \sum_{i=0}^{M/64} (VAD_{FAR}[i] \cdot \overline{VAD_{NEAR}} \cdot DTD_{real}[i]) / \sum_{i=0}^{M/64} (VAD_{FAR}[i] \otimes \overline{VAD_{NEAR}}) * 100 \quad (12)$$

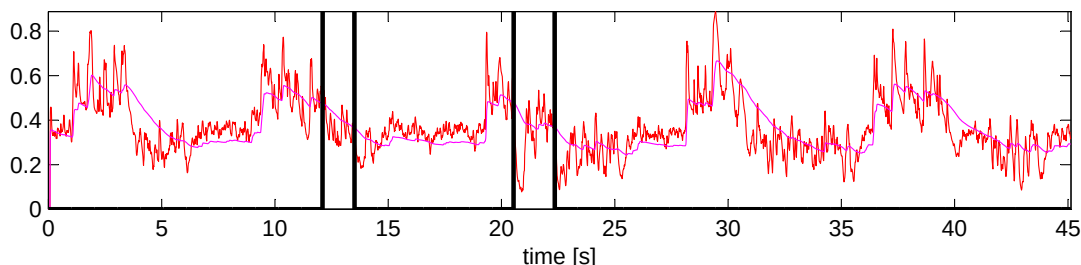
Průběhy jednotlivých koherencí jsou na následující obrázcích, které jsou doplněny ideální detekcí pomocí černých svislých čar.



Obrázek 5: Průměrná hodnota koherence



Obrázek 6: Průměrná hodnota z minimální koherence



Obrázek 7: Průměrná hodnota z maximální koherence

Tabulka 2: Vypočtené hodnoty pro DTD

| Skupina | ERR1 [%] | ERR2 [%] | ERR3 [%] |
|---------|----------|----------|----------|
| 1       | 22,75    | 32,89    | 39,44    |
| 2       | 19,39    | 16,05    | 42,89    |
| 3       | 20,40    | 18,67    | 35,17    |
| 4       | 17,12    | 21,51    | 19,35    |

## 5 Závěr

Analýzou závislosti DTD na celkovém systému AEC bylo zjištěno, že při použití DTD je zkreslení užitečného signálu blízkého mluvčího minimální. Při snaze AEC adaptovat se na signál společné promluvy byly koeficienty filtru tak ovlivněny, že způsobovaly degradaci užitečného signálu a adaptační algoritmus echo téměř neodhadoval.

Použitím ideální detekce DTD v systémech AEC mělo za následek zlepšení potlačení o více než 6 dB pro algoritmus NLMS v časové oblasti a více jak 3 dB pro frekvenční algoritmus na

bázi LMS s metodou přesahů overlap - save. Koherenční detektor dosahoval nejlepších výsledků při sledování průměrné či maximální koherence signálů a nejnižší hodnota chybovosti v rozsahu celého signálu byla menší než 18 %. Největší pozornost byla věnována sledování přesnosti detekce v oblasti společné promluvy, zde byla dosažena hodnota chybovosti kolem 16 %.

Důležitým výsledkem práce bylo vytvoření GUI rozhraní aplikace v prostředí MATLABu, jenž umožnilo zmíněné testování a analýzu použitých algoritmů. Jednoduché a přehledné grafické zobrazení různých částí zpracování systémů sloužilo k analýze algoritmu na různých úrovních a hromadné zpracování signálů z databáze umožnilo snadné testování na větším množství signálů.

## Poděkování

Výzkum popisovaný v tomto článku vznikl za podpory ČVUT grantu SGS12/143/OHK3/2T/13 "Algoritmy a hardwarové realizace digitálního zpracování signálů". Signály pro testování a vytvoření nové řečové databáze byly získány ve spolupráci se společností Acrobits s.r.o. (<http://www.acrobits.cz>).

## Reference

- [1] DINIZ, Paulo Sergio Ramirez. *Adaptive filtering: algorithms and practical implementation*. 3rd ed. New York: Springer, 2008, 625 s. ISBN 978-0-387-31274-3.
- [2] SHYNK, J.J. *Frequency-domain and multirate adaptive filtering*. IEEE Signal Processing Magazine. 1992, roc. 9, c. 1, s. 14-37. ISSN 1053-5888. DOI: 10.1109/79.109205.
- [3] TASHEV, I.J. *Coherence based double talk detector with soft decision*. Proceedings of the ... IEEE International Conference on Acoustics, Speech, and Signal Processing. 25-30 March 2012, Kyoto, [cit. 2012-10-20]. ISSN 1520-6149.
- [4] UHLÍŘ, J. et al. *Technologie hlasových komunikací*. ed. 1. Praha: CTU, 2007, 276 p. ISBN 978-80-01-03888-8
- [5] KLAPUCH, J.: Modelling of Real Acoustic Echo and Environmental Noise Background in Smartphone Devices. In POSTER 2012 - 16th Int.Student Conference. Praha: CTU in Prague, 2012, ISBN 978-80-01-05043-9.
- [6] HANSLER, E., SCHMIDT, G. *Acoustic echo and noise control: a practical approach*. Hoboken: John Wiley, 2004, 444 s. ISBN 04-714-5346-3

---

Ing. Jan Klapuch  
klapujan@fel.cvut.cz

Doc. Ing. Petr Pollák, CSc.  
pollak@fel.cvut.cz