

AKUSTICKÉ LISTY

České akustické společnosti
www.czakustika.cz

ročník 17, číslo 4

prosinec 2011

Obsah

Využití harmonicity při fonetické segmentaci řeči Using Harmonicity for the Phonetic Segmentation of Speech <i>Jana Heranová a Radek Skarnitzl</i>	3
Automatická segmentace hlásek při rychlém opakování slabik (/pa/-/ta/-/ka/) u hypokinetické dysartrie Automatic Segmentation of Phonemes during the Fast Repetition of (/pa/-/ta/-/ka/) Syllables in a Speech Affected by Hypokinetic Dysarthria <i>Michal Novotný, Jan Ruzs a Roman Čmejla</i>	10
Analýza nelineárních stojatých vln ve dvou vázaných akustických rezonátorech Analysis of Nonlinear Standing Waves in Two Coupled Acoustic Resonators <i>Michal Bednařík, Milan Červenka a Jaroslav Plocek</i>	17
Nízkofrekvenční zvukové svazky Low-frequency Sound Beams <i>Milan Červenka, Michal Bednařík, Jaroslav Plocek a Martin Šoltés</i>	21
Porovnání dvou spektrálních metod pro analýzu akustických signálů Comparison of Two Spectral Methods for Acoustic Signal Analysis <i>Václav Turoň, Ján Janík, Radim Špetík, Pavel Sovka a Miroslav Vlček</i>	26

Využití harmonicity při fonetické segmentaci řeči

Jana Heranová a Radek Skarnitzl

Fonetický ústav – FF UK, nám. J. Palacha 2, 116 38 Praha 1

e-mail: jana.heran@gmail.com, radek.skarnitzl@ff.cuni.cz

This paper examines harmonicity (HNR) as a possible indicator of segment boundaries. The aim is to assess whether harmonicity could be used to enhance the accuracy of automatic HMM segmentation. Since harmonicity may be regarded as a frequency-domain correlate of voicing, its greatest benefit was expected in vowel – voiceless obstruent and voiceless obstruent – vowel sequences: contexts with significant changes in voicing degree and also contexts where the performance of the automatic aligner is relatively low. We present a classification of harmonicity plots and examine the relationship between the harmonicity curve and manually located speechsound boundaries. It turned out possible to identify specific time intervals with respect to the maximum in the HNR curve in which most of the boundaries were situated. The dynamic properties of harmonicity thus proved to be a promising indicator of boundaries between vowels and voiceless obstruents and they are recommended for further research.

1. Úvod

Akustická analýza řeči vyžaduje rozčlenění spojitého řečového signálu na diskrétní jednotky. Pro většinu fonetických analýz, segmentálních i suprasegmentálních, těmto diskrétním jednotkám odpovídají hlásky. Hlásková segmentace může probíhat automaticky či manuálně, avšak automatické nástroje, většinou založené na skrytých Markovových modelech (HMM; např. [1–3]), nejsou pro potřeby fonetického výzkumu dostatečně přesné, a manuální opravy jsou tak nezbytné (detailní a konzistentní pravidla pro manuální segmentaci řeči, motivovaná foneticky významnými událostmi v akustickém signálu, jsou představena v [4]).

Protože se dnešní fonetický výzkum neobejde bez relativně velkých řečových vzorků, je automatický způsob segmentace – i přes zmíněné slabiny – při přípravě fonetických korpusů hojně využíván a hranice hlásek jsou postupně upravovány manuálně [5]. Z fonetického hlediska nacházíme ve výstupu automatického segmentátoru, *Prague Labeller* [2], několik typů „chyb“ [6]. Drobné odchylky v řádu milisekund bývají způsobeny časovým rozlišením metody, tedy délkou analyzačního okénka a časovým krokem. Problematické jsou dále konkrétní hlásky, především likvidy (l-ové či r-ové hlásky), jejichž akustické vlastnosti „prosakují“ do okolních hlásek. S nepřesným umístěním hláskové hranice se také často setkáváme při přechodech ze znělé hlásky na neznělou a obráceně, konkrétně ve spojeních vokálů s neznělými obstruenty; obstruenty jsou hlásky, při jejichž produkci vznikají šumové složky (explozivny jako [p d], frikativy jako [f z] a afrikáty [c č]), na rozdíl od sonor jako [m l j]). Je to právě tento poslední kontext, jímž se budeme v našem příspěvku zabývat.

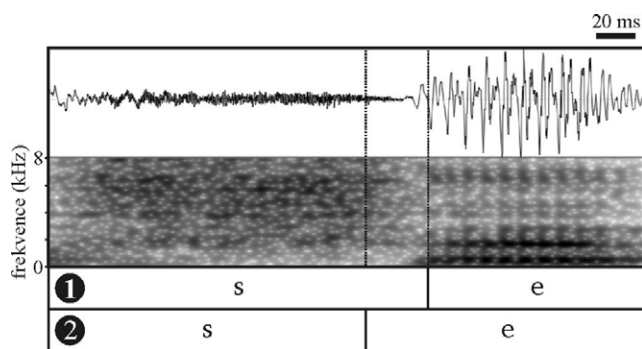
Podívejme se nejprve, co to vlastně znělost je a jakými způsoby ji můžeme měřit. Znělost je v mnoha jazycích světa využívána pro signalizaci fonologického kontrastu mezi hláskami jako [t]–[d] či [s]–[z]. Různé jazyky se znělostí zacházejí různými způsoby; čeština patří k jazykům, u nichž pozorujeme převážně shodu mezi znělostí fonolo-

gickou (v jazykovém systému definovanou) a fonetickou (v akustickém signálu skutečně přítomnou) [7, 8].

Z hlediska fyziologického je základem znělosti fonace neboli kvaziperiodické kmitání hlasivek, jehož akustickým odrazem je základní frekvence, F_0 . Ve zvukové vlně (tzn. v časové doméně) se znělost projevuje pravidelným opakováním složené vlny, periodičností, zatímco ve spektru a spektrogramu (ve frekvenční doméně) harmonickou strukturou, harmoničností (tzn. pravidelným rozestupem jednotlivých harmonických složek, celočíselných násobků základní frekvence).

Měření znělosti v časové oblasti je většinou záležitostí binární povahy: základní frekvence v daném časovém okamžiku buď detekována je, nebo není. Dynamické charakteristiky znělosti zachycuje tzv. metoda znělostního profilu [9], která vyjadřuje pravděpodobnost přítomnosti znělosti v průběhu dané hlásky. Použití této metody pro analýzu českých znělých obstruentů v intervokální pozici například ukázalo, že více než u poloviny položek /ř/ dochází během jeho průběhu k vytrácení znělosti. Oproti tomu k částečné ztrátě znělosti (k desonorizaci) dochází jen asi v 15 % položek /ž/, v 10 % položek /z/ a /v/ znělost prakticky neztrácí [8].

Ve frekvenční oblasti můžeme za určitý korelát znělosti považovat odstup harmonických složek od šumu (*harmonics-to-noise ratio*, HNR) neboli harmonicitu. Harmonicita, vyjádřená jako poměr harmonických a šumových složek v decibelech, je tradičně užívána k posuzování patologických rysů fonace, například chraplavosti [10–12]. Nedávný výzkum prokázal, že v míře harmoničnosti se odlišují i jednotlivé hláskové třídy, přičemž neznělé obstruenty a vokály se (dle očekávání) odlišují nejvíce [13]. Ačkoli zkoumání harmonicity vyžaduje poměrně dlouhé analyzační okénko [14], vyvstala otázka, zda je z fonetického hlediska užitečné zkoumat nejen průměrné hodnoty HNR, ale i průběh harmonicity, například v intervalu 5 či 10 milisekund. Harmonický profil, jakožto jakási paralela výše zmíněného zně-



Obrázek 1: Srovnání manuální a automatické segmentace spojení [se] ze slova *muset*. Ve vrstvě pod spektrogramem (označené 1) se nachází manuálně určená hranice umístěná na začátek plné formantové struktury vokálu. V následující vrstvě (označené 2) se nachází hranice automaticky rozpoznaná nástrojem *Prague Labeller*

lostního profilu, již byl použit ke kvantifikaci desonorizace [8].

V této studii nás zajímalo, zda průběh harmonicity ve výše uvedených hláskových přechodech – z neznělého obstruentu na vokál a z vokálu na neznělý obstruent – vykazuje konzistentní rysy. Pokud bychom takové opakující se rysy objevili a mohli je vztáhnout k manuálně stanovené hranici mezi cílovými hláskovými spojeními, bylo by možné formalizovat pravidla pro úpravu automaticky nalezených hranic. Cílem našeho příspěvku je tedy zjistit, zda je průběh harmonicity využitelný pro zpřesnění automatické hláskové segmentace v kontextu přechodů mezi vokály a neznělými obstruenty. Protože jsou taková hlásková spojení v řeči velmi častá, představoval by pozitivní výsledek pro fonetický výzkum významné zlepšení.

2. Metoda

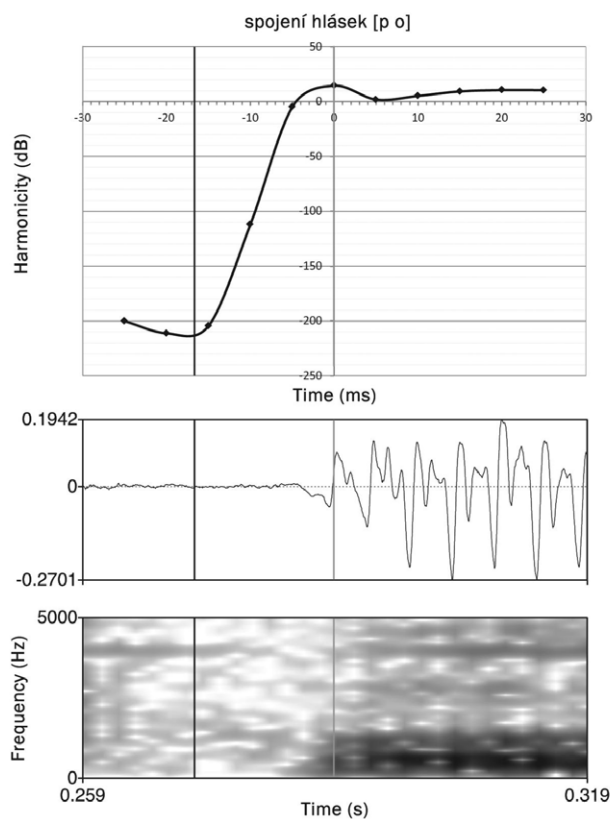
Pro tuto studii byly použity nahrávky 12 rodilých mluvčích češtiny (8 žen a 4 mužů) pořízené v nahrávací kabině Fonetického ústavu FF UK. Jednalo se o dospělé nekuřáky (věkové rozmezí se pohybovalo mezi 30 a 62 lety), kteří po krátké přípravě přečetli přibližně minutový text. Nahrávky byly zpracovány v softwaru Praat [15]; Praat je volně šiřitelný a průběžně rozvíjený program určený pro fonetické analýzy.

Abychom mohli analyzovat umístění hranic mezi segmenty, provedli jsme automatickou i manuální segmentaci materiálu. Soustředili jsme se na hlásková spojení obsahující neznělý obstruent [p t k f s š x c ě] a vokál [i i: e e: a a: o o: u u: ou̯ au̯ eu̯]. Automatická segmentace byla provedena pomocí HMM v nástroji *Prague Labeller* [2] s délkou rámce 16 ms a časovým krokem 8 ms. Manuální segmentace probíhala podle předem stanovených pravidel, podle nichž je hlavním kritériem pro segmentaci intervokálních obstruentů přítomnost plné formantové struktury vokálu [4]. Hranice jsme umísťovali na začátek nebo na konec plné formantové struktury vokálu, jež byla charakte-

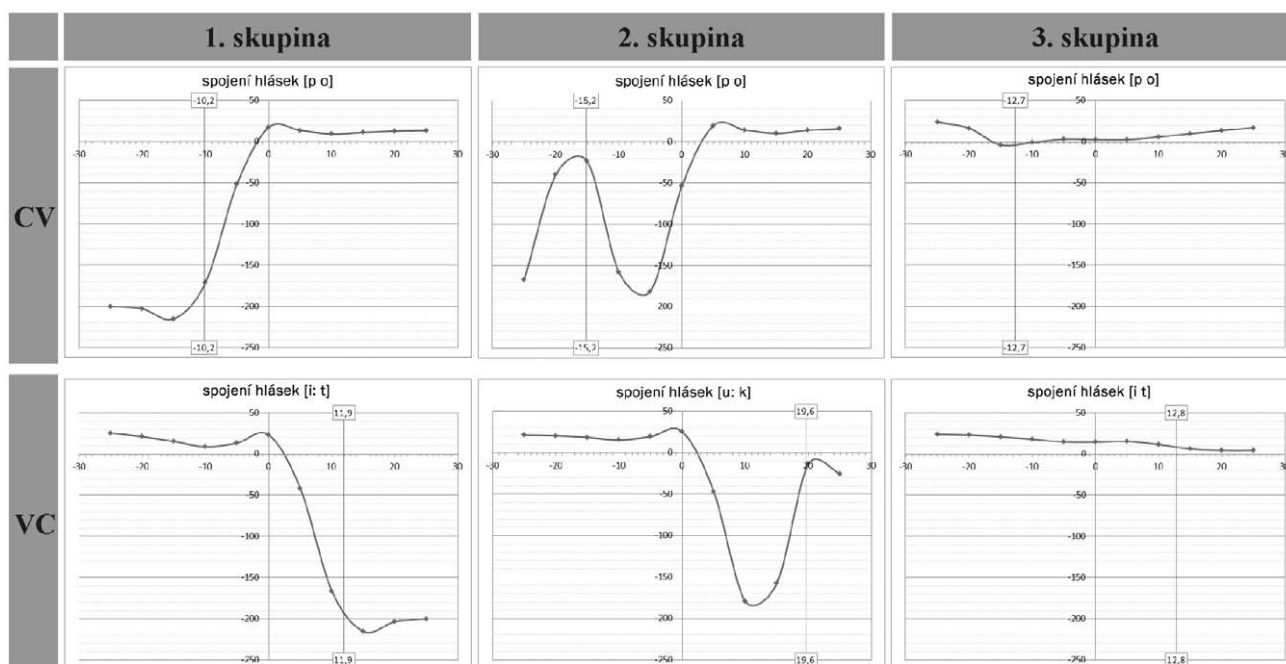
rizována sledem zřetelných formantových sloupků. Byla-li mezi segmenty přítomna přechodová oblast, umístili jsme hranici do temporálního středu této oblasti. Všechny hranice byly umísťovány do průchodu nulou (bod v oscilogramu, kde zvuková vlna protíná nulovou hodnotu amplitudy). Ukázkou srovnání automatické a manuální segmentace znázorňuje obrázek 1.

Zvukové nahrávky byly v programu Praat převedeny na objekt *Harmonicity* pomocí křížové korelace za použití standardního nastavení. Hodnoty HNR byly následně extrahovány pro výše zmíněná hlásková spojení s časovým krokem 5 ms, a to v intervalu ± 25 ms od manuálně určené hranice (viz příklad na obrázku 2).

Řečový materiál obsahoval celkem 1955 cílových hláskových spojení, z nichž jsme pro účely naší analýzy udělali užší výběr podle časového rozdílu mezi automaticky a manuálně umístěnou hranicí. Tento vzorek obsahoval 459 položek, u nichž se rozdíl v umístění obou hranic pohyboval v rozmezí 10–20 ms. Důvod pro volbu tohoto rozpětí je následující. Spodní hranice 10 ms odpovídá u průměrných mužských hlasů jedné hlasivkové periodě; odchylku v seg-



Obrázek 2: Příklad spojení neznělého obstruentu a vokálu. Nahoře graf průběhu HNR (hodnoty v dB). Manuální hranice se nachází v bodě nula a je označena šedou čarou. Automatická hranice je vyznačena černou čarou. Hodnoty HNR byly měřeny v intervalu ± 25 ms od umístění manuální hranice. Pod grafem HNR se nachází odpovídající oscilogram a spektrogram signálu



Obrázek 3: Klasifikace průběhů HNR (hodnoty v dB). Manuálně určená hranice se nachází v bodě nula. Automatická hranice je vyznačena druhou plnou čarou, spolu s odchylkou od manuální hranice v milisekundách

mentaci o jednu periodu považujeme za přijatelnou. Hodnota 10 ms zároveň leží těsně nad časovým rozlišením použitého automatického nástroje pro segmentaci, tzn. 8 ms. Naopak horní hranice 20 ms již představuje časový úsek, který může v některých případech odpovídat trvání jednoho celého segmentu. Nalezené události v průběhu harmonicity by tak již nemusely být relevantní pro cílové hláskové spojení. Tato studie se tedy soustředila právě na ty rozdíly v automatickém a manuálním umístění hranic, které se s největší pravděpodobností ještě nacházejí v rámci daného segmentu a zároveň jsou větší než délka jedné periody u hlubokých hlasů.

3. Klasifikace průběhů HNR

Naše klasifikace tedy vychází z analýzy 459 grafů, jež popisují průběh hodnot harmonicity na hranicích cílových hláskových spojení, konkrétně 25 ms před a 25 ms za manuálně určenou hranicí mezi neznělým obstruentem a vokálem nebo naopak. Výsledná klasifikace průběhů HNR byla východiskem pro další posouzení vztahu mezi umístěním segmentálních hranic a průběhem harmonicity.

Jak ukazuje obrázek 3, celkově jsme identifikovali tři skupiny průběhů: standardní průběhy (na obrázku označeny jako 1. skupina), nestandardní průběhy (2. skupina) a rozkolísané průběhy, které vykazují vlastnosti na pomezí obou předešlých skupin (3. skupina). Tyto tři skupiny jsme identifikovali jak u spojení konsonant – vokál (CV), tak u spojení vokál – konsonant (VC).

Považujeme za důležité podotknout, že uvedená hlásková spojení ([po], [i:t] apod.) představují pouze reprezen-

tativní zástupce skupin, ale daný průběh nezávisí na konkrétní (fonologické) identitě konsonantu ani vokálu (například rozdíl mezi dlouhým a krátkým vokálem tedy nehraje žádnou roli).

Standardní průběh křivky harmonicity (1. skupina) je charakterizován výraznou změnou v hodnotách HNR, která se objevuje v důsledku přechodu neznělého konsonantického segmentu na segment vokalický a naopak. Protipólem této skupiny jsou průběhy 3. skupiny, u nichž k takové zásadní změně v hodnotách HNR nedochází; jejich průběh lze popsat jako víceméně rovný. Skupina 2 byla vyčleněna jako kombinace standardního a nestandardního průběhu. U této skupiny můžeme pozorovat standardní průběh HNR ve vokalické části signálu. S nástupem konsonantického segmentu zde sice opět dochází k výraznému poklesu HNR jako u skupiny 1, ale další průběh HNR v konsonantické části signálu buď značně kolísá, nebo se vyvíjí konstantně v oblasti kladných (spíše než záporných) hodnot, což koreluje s průběhy skupiny 3.

Výskyt jednotlivých skupin průběhů v rámci spojení CV a VC udává tabulka 1. Ve spojení CV je nejvíce zastoupen standardní průběh křivky HNR skupiny 1 (68,5 %), následuje skupina 2 (22 %) a nejméně výskytů jsme zaznamenali pro nestandardní průběh HNR skupiny 3 (9,5 %). Nenáhodnost těchto výskytů – oproti očekávaným rovnoměrným zastoupením průběhů v rámci skupin – jsme ověřili pomocí statistické metody chí-kvadrát: $\chi^2(2; n = 168) = 97,2; p < 0,001$. Také u spojení VC byly nejvíce zastoupeny průběhy 1. skupiny (59,8 %). V porovnání se spojením hlásek CV je zde

	CV	VC
1. skupina	115	174
	68,5 %	59,8 %
2. skupina	37	30
	22,0 %	10,3 %
3. skupina	16	87
	9,5 %	29,9 %
celkem	168	291

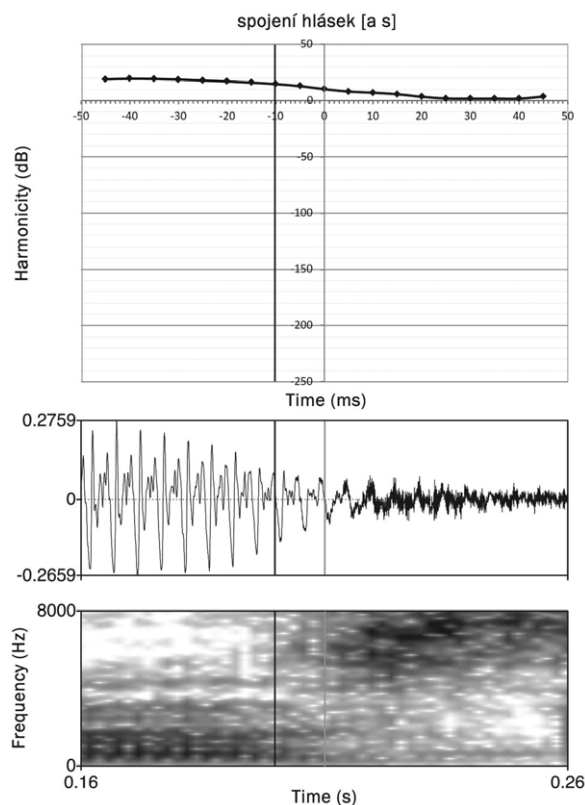
Tabulka 1: Procentuální zastoupení jednotlivých skupin průběhů HNR ve spojeních CV a VC

druhou nejpočetněji zastoupenou skupinou skupina ne-standardních průběhů, tj. skupina 3 (29,9 %). Skupina 2 byla ve spojení hlásek VC zastoupena nejméně (10,3 %). Četnost zastoupení průběhů jednotlivých skupin se i v tomto případě prokázala jako statisticky vysoce významná: $\chi^2(2; n = 291) = 108,4; p < 0,001$. Z výsledků vyplývá (opomineme-li nejčastěji zastoupené průběhy skupiny 1), že pro spojení CV existuje tendence k průběhům skupiny 2, zatímco ve spojeních VC shledáváme spíše tendenci k nestandardním průběhům skupiny 3.

Na základě kritéria plné formantové struktury, jež bylo zvoleno pro segmentaci cílových položek, jsme předpokládali, že formantová struktura vokálu bude korelovat s kladnými hodnotami HNR, zatímco neznělý obstruent, který by měl alespoň při kanonické realizaci harmonickou složku postrádat, bude vykazovat záporné hodnoty HNR. Kolem hranice ve spojení VC by tedy mělo docházet k přechodu z kladných hodnot HNR do hodnot záporných a ve spojení CV naopak k výraznému nárůstu hodnot harmonicity. Tento předpoklad se nám potvrdil u skupin 1 a 2, tedy celkem u 77,6 % položek zkoumaného vzorku (152 případy spojení CV a 204 případy spojení VC). Z těchto průběhů jsme dále vycházeli při analýze vztahu mezi průběhem HNR a umístěním hranice mezi segmenty. Nestandardní průběhy za tímto účelem použity být nemohly, jelikož u nich nedocházelo k žádné distinktivní změně hodnot HNR, kterou bychom mohli využít jako referenční bod pro posouzení umístění hranic mezi segmenty.

Než přejdeme k analýze vztahu mezi křivkou harmonicity a manuálně umístěnou hranicí, podívejme se krátce na relativně vysoký výskyt nestandardních průběhů ve spojeních VC a pokusme se tento neočekávaný výskyt vysvětlit. Zajímá nás především způsob artikulace konsonantů v těchto spojeních. Zatímco ve spojeních CV a stejně tak u exploziv a afrikát ve spojeních VC korelovala četnost zastoupení jednotlivých konsonantických skupin s četností zastoupení průběhů HNR v jednotlivých skupinách klasifikace, u frikativ ve spojeních VC je situace odlišná. Nejvíce

frikativ ve spojeních VC nacházíme právě ve 3. skupině ne-standardních průběhů (66), v 1. skupině pak 40 položek a ve 2. skupině 13 položek. Z 87 položek nestandardního průběhu tedy připadá 75,9 % na frikativy. Jedno z možných vysvětlení nabízí příklad na obrázku 4: ačkoli je patrné, že se skutečně jedná o neznělou frikativu, vysoká hodnota harmonicity (ačkoli jde o neznělou hlásku, nabývá HNR kladných hodnot) by mohla být způsobena neobvykle silným šumovým formantem přibližně kolem 5 000 Hz, který může vytvářet jakési zdání periodicity.



Obrázek 4: Nestandardní průběh HNR ve spojení [as] (hodnoty v dB). V realizaci konsonantu můžeme sledovat prominentní pásmo šumu, jež mohlo způsobit vysokou hodnotu harmonicity

4. Průběh HNR ve vztahu k umístění segmentálních hranic

Analýza umístění hranic mezi segmenty vycházela z realizací, u nichž došlo mezi vokálem a konsonantem k výrazné změně v hodnotách HNR, tedy z položek v 1. a 2. skupině. Za výchozí byl zvolen datový bod, který se vzhledem ke svému bezprostřednímu okolí vyznačoval maximální hodnotou HNR (srov. obr. 2, na němž se právě tento bod nachází nejbližší k manuálně umístěné hranici). Sledovali jsme tedy umístění manuálně nalezené hranice mezi hláskami vzhledem k lokálnímu maximu harmonicity. Pokud bychom objevili nenáhodný vztah, bylo by možné

výsledky využít ke zpřesnění automatické segmentace. Výsledky analýzy zobrazuje tabulka 2.

U spojení CV můžeme vidět, že 95,4 % umístění hranic se nachází v rozmezí 10 ms, tj. v časovém intervalu 5 ms před až 5 ms za hodnotou maxima HNR (tento interval je v tabulce vyznačen šedou barvou). U spojení VC již výsledky nejsou tak vyhraněné a variabilita umístění hranic segmentů je vzhledem k bodu maxima HNR mnohem vyšší. Nejvíce hranic se nachází v 10milisekundovém intervalu, který začíná maximem HNR a končí 10 ms po maximu, přičemž střed tohoto intervalu tvoří časový bod +5 ms. Toto pásmo zahrnuje 52,5 % umístění manuálních hranic v rámci kontextů VC. Pokud časové rozmezí rozšíříme na oblast -10 ms až +10 ms kolem hodnoty maxima HNR, zahrne tento časový interval 74,5 % všech analyzovaných hranic ve spojeních VC.

Větší rozptyl hodnot ve spojeních VC než ve spojeních CV si vysvětlujeme tím, že při přechodu z vokálního na konsonantický segment dochází k častějšímu výskytu přechodových oblastí, které jsou způsobené delší dobou přibližování artikulačních orgánů v porovnání s jejich oddálením, k němuž dochází u spojení CV (srov. [4]).

Výsledky této analýzy tedy potvrzují existenci určitých časových intervalů kolem lokálního maxima HNR, které zahrnují umístění většiny manuálně určených hranic mezi neznělými obstruenty a vokály. Umístění hranic vzhledem k průběhu HNR tedy není náhodné.

Další aspekt, na nějž jsme se při analýze manuálně umístěných hranic ve vztahu k harmonicitě zaměřili, jsou konkrétní hodnoty HNR, v nichž se většina hranic nachází. Na základě kritéria plné formantové struktury, podle něhož byl materiál segmentován, jsme očekávali umístění hranic do oblasti kladných hodnot HNR. Otázkou je, zda jsou hranice mezi segmenty umísťovány spíše do oblastí, kde se hodnoty HNR vyvíjejí konstantně, tedy na začátek či konec oblasti kladných hodnot HNR, nebo do oblastí přechodu mezi vokálním a konsonantickým segmentem, kde dochází k prudké změně HNR.

Ve spojeních CV se kladné hodnoty HNR vyskytují v časovém intervalu mezi maximem a +20 ms; tento interval zahrnuje 63,8 % všech analyzovaných hranic v rámci spojení CV. Vezmeme-li v potaz i hranice umístěné v časovém bodě -5 ms před maximem HNR, z nichž se polovina (26 z 52) pohybuje rovněž v kladných hodnotách, pracujeme až s 80,9 % hranic.

V rámci spojení VC se v oblasti kladných hodnot HNR pohybují hranice u 52 % analyzovaných realizací, což představuje hranice nacházející se v našem vzorku v časovém intervalu -20 ms až maximum HNR. Podobně jako v kontextu CV jsme i u spojení VC mohli pozorovat kladné hodnoty HNR u hranic nacházejících se jeden datový bod za hodnotou lokálního maxima HNR, tedy v čase +5 ms. To se týkalo 27 ze 40 hranic umístěných v tomto časovém bodě. Započteme-li i tyto hranice ke zmíněným 52 % hranic, zvýší se poměr hranic umístěných do oblasti kladných hodnot na 65,2 %.

	Umístění hranice (ms)	Počet a poměr položek	Celkem položek
CV	-10	3 2,0 %	152
	-5	52 34,2 %	
	MAX	67 44,1 %	
	+5	26 17,1 %	
	+10	3 2,0 %	
	+15	1 0,7 %	
VC	-20	9 4,4 %	204
	-15	14 6,9 %	
	-10	21 10,3 %	
	-5	24 11,8 %	
	MAX	33 16,2 %	
	+5	40 19,6 %	
	+10	34 16,7 %	
	+15	15 7,4 %	
	+20	10 4,9 %	
víc	4 2,0 %		

Tabulka 2: Umístění manuálně určených hranic vzhledem k lokálnímu maximu HNR (označeno MAX). Tento bod byl využit k usouvztažnění ostatních časových bodů. Šedě jsou vyznačené intervaly s největším počtem manuálně určených hranic

Podíl hranic umístěných do oblastí kladných hodnot HNR v rámci spojení VC (65,2 %) je zřejmě nižší než u spojení CV (80,9 %), což dáváme do souvislosti s již zmíněným častým výskytem přechodových oblastí ve spojeních VC, u nichž se hranice umísťují do temporálního středu přechodu [4]. V tomto místě již může být periodičita do značné míry potlačena, což se odrazí v záporných hodnotách HNR.

Můžeme tedy říci, že většině manuálních hranic odpovídají kladné hodnoty HNR: ve spojeních CV se jedná o 80,9 % případů, ve spojeních VC o 65,2 % případů. Hranice umístěné v souladu se segmentačními pravidly [4] se nacházejí spíše v těch oblastech – případně na jejich začátcích či koncích, kde se hodnoty HNR vyvíjejí konstantně v kladných hodnotách, nežli tam, kde dochází k přechodu z kladných hodnot HNR do záporných či opačně.

Další zajímavý výsledek nám poskytlo porovnání hranic určených automaticky pomocí HMM a hranic určených manuálně. U spojení CV se ze všech 152 realizací objevil pouze jeden případ, kdy byla automatická hranice umístěna až za hranici manuální, tj. později v čase. V naprosté většině případů jsme tak mohli sledovat tendenci v umísťování automatických hranic do oblastí, v níž byl přítomen aspirační šum nebo frikce neznělých obstruentů, případně tam zasahovala předznívající znělost (srov. obrázek 1). V těchto okamžicích se však ještě nevyskytovala formantová struktura vokálu.

V rámci spojení VC pozorujeme podobnou tendenci, která zde opět nebyla tak vyhraněná jako v případě spojení CV. Setkali jsme se celkem se 73 případy, kdy byla automatická hranice umístěna ještě před manuální, tj. časově dřív, do oblasti kladných hodnot HNR vokálu: tyto případy představují 35,8 % ze všech 204 analyzovaných spojení. Převládala ovšem spojení (131 položek, což odpovídá 64,2 % všech realizací VC), u nichž byla automatická hranice umístěna časově později v porovnání s hranicí manuální, tj. do oblasti podle manuální segmentace náležící již konsonantickému segmentu.

Na základě těchto výsledků bychom mohli říci, že automatický segmentátor *Prague Labeller* [2] má sklon vymezovat v signálu delší časové úseky odpovídající vokalickým segmentům než foneticky motivovaná segmentace manuální. Za příčinu této tendence můžeme považovat způsob trénování modelů HMM, který samozřejmě na takto explicitních fonetických pravidlech založen nebývá.

5. Diskuse a závěr

Cílem tohoto příspěvku bylo ověřit, zda existuje definovatelný vztah mezi umístěním manuálně stanovených hranic hlásek a hodnotami HNR, který bychom potenciálně mohli využít ke zpřesnění automatické segmentace tak, aby lépe odpovídala foneticky motivovaným pravidlům.

U spojení neznělého obstruentu s vokálem (CV) jsme identifikovali časový interval 10 ms (± 5 ms od lokálního maxima HNR), v rámci něhož se nachází 95,4 % manuálně určených hranic. Většina hranic nalezených pomocí automatického segmentátoru však leží v intervalu -25 až -5 ms vzhledem k maximu HNR, a vůči manuálně určené (a foneticky motivované) hranici jsou tedy značně posunuty.

U spojení vokálu a neznělého obstruentu (VC) byl rozptýlen dat větší. Interval zahrnující většinu (74,5 %) manuálně určených hranic je dlouhý 20 ms a nachází se v oblasti ± 10 ms od maxima HNR. Automaticky určená hranice se většinou pohybuje výrazně za maximem HNR (tedy více než $+10$ ms).

Považujeme za důležité připomenout, že naše analýza je založena na výběru položek, které splňovaly zadaná kritéria. Náš vzorek obsahoval 459 položek, u nichž se rozdíl v automaticky a manuálně určené hranici mezi cílovými hláskami pohyboval mezi 10 a 20 ms. (U většiny ostatních položek byl rozdíl mezi umístěním hranic nižší než 10 ms.) Na základě vybraných položek byla vytvořena typologie průběhů HNR; ze tří typů průběhů byl jeden pro následnou analýzu nevhodný, protože v křivce harmonicity nedocházelo k výrazným změnám, jichž by bylo možné ke stanovení následných pravidel pro segmentaci využít. Celkově jsme tedy pracovali s 356 položkami.

Shrneme-li výsledky našich analýz, můžeme konstatovat, že harmonicity, resp. její dynamické vlastnosti se projeví jako slibný indikátor hláskových hranic. Standardní průběh harmonicity (označovaný výše jako 1. skupina) převládá v četnosti svého zastoupení. Tento průběh ko-

responduje s očekávanými vlastnostmi spojení neznělých obstruentů s vokály a umístění maxima HNR v okolí hranice mezi cílovými hláskami není náhodné. Toto maximum tak může fungovat jako referenční bod využitelný pro přesnější stanovení hláskových hranic.

Na základě výsledků můžeme tedy formulovat následující pravidla, která jsou pomocí skriptu snadno implementovatelná v prostředí Praat:

- V případě hranice mezi neznělým obstruentem a následujícím vokálem se zdá nejvýhodnější posunout automaticky detekovanou hranici do bodu maxima HNR, které budeme hledat v intervalu 0 až 20 ms napravo od automaticky nalezené hranice.
- V případě hranice mezi vokálem a následujícím neznělým obstruentem budeme rovněž hledat maximum HNR, a to v intervalu -20 ms až $+20$ ms vzhledem k automaticky umístěné hranici. Za nejvýhodnější umístění nové hranice považujeme časový bod vzdálený $+5$ ms od nalezeného maxima.

V našem příspěvku jsme chtěli naznačit jedno z dalších možných využití harmonicity ve fonetickém výzkumu; harmonicity byla doposud nasazována především v oblasti poruch hlasu. Domníváme se, že harmonicity představuje vhodný parametr pro zpřesnění automatické hláskové segmentace tak, aby její výstup odpovídal foneticky motivovaným pravidlům pro umístění hláskových hranic.

Závěrem poznamenejme, že jsme v této studii pracovali s řečovým materiálem pořízeným v laboratorním prostředí. Jak jsme ukázali, hranice mezi cílovými hláskami se v takovém případě většinou nachází v kladných hodnotách HNR. Je otázkou pro budoucí výzkum, zda se harmonicity (HNR) – která vyjadřuje podobné vlastnosti jako odstup signálu od šumu (SNR) – ukáže jako užitečná i pro segmentaci méně kvalitních nahrávek, u nichž je hodnota SNR nižší.

Poděkování

Tento výzkum vznikl za podpory grantu VZ MŠM 0021620825.

Reference

- [1] Kominek, J., Bennett, C., Black, A. W.: Evaluating and Correcting Phoneme Segmentation for Unit Selection Synthesis, *Proc. of Eurospeech 2003*, p. 313–316, Geneva, 2003.
- [2] Pollák, P., Volín, J., Skarnitzl, R.: HMM-Based Phonetic Segmentation in Praat Environment, *Proc. of SPECOM 2007*, p. 537–541, Moscow, 2007.
- [3] van Niekerk, D. R., Barnard, E.: Important factors in HMM-based phonetic segmentation, *Proc. of the 18th PRASA*, p. 25–28, Pietermaritzburg, South Africa, 2007.

- [4] Machač, P., Skarnitzl, R.: *Fonetická segmentace hlásek*, Nakladatelství Epoque, Praha, 2009.
- [5] Skarnitzl, R.: Prague Phonetic Corpus: status report, *AUC Philologica 1/2009, Phonetica Pragensia*, XII, p. 65–67, 2010.
- [6] Volín, J., Skarnitzl, R., Pollák, P.: Confronting HMM-based Phone Labelling with Human Evaluation of Speech Production, *Proc. of Interspeech 2005*, p. 1541–1544, Lisbon, 2005.
- [7] Machač, P.: Desonorizace českých intervokálních frikativ. *AUC Philologica 2/2007, Phonetica Pragensia*, XI, p. 105–116, 2008.
- [8] Skarnitzl, R.: *Znělostní kontrast nejen v češtině*, Nakladatelství Epoque, Praha, 2011.
- [9] Möbius, B.: Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica*, 38, p. 5–26, 2004.
- [10] Yumoto, E., Gould, W. J., Baer, T.: Harmonics-to-Noise Ratio as an Index of the Degree of Hoarseness, *Journal of the Acoustical Society of America*, 71, p. 1544–1549, 1982.
- [11] Qi, Y., Hillman, R. E.: Temporal and Spectral Estimations of Harmonics-to-Noise Ratio in Human Voice Signals, *Journal of the Acoustical Society of America*, 102, p. 537–543, 1997.
- [12] Bauer, L., Ruzs, J., Čmejla, R.: Hodnocení vokálních parametrů u patologických hlasů, *Akustické listy*, 17(1–2), p. 13–18, 2011.
- [13] Heranová, J.: *Harmonicita českých segmentů*, nepublikovaná klauzurní práce, Fonetický ústav FF UK, 2009.
- [14] Boersma, P.: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, *IFA Proceedings*, 17, p. 97–110, 1993.
- [15] Boersma, P., Weenink, D.: Praat: doing phonetics by computer (Version 4.6.36). Staženo 2. listopadu 2007, <http://www.praat.org>

Automatická segmentace hlásek při rychlém opakování slabik (/pa/ – /ta/ – /ka/) u hypokinetické dysartrie

Michal Novotný^a, Jan Ruzs^{a,b} a Roman Čmejla^a

^aČeské vysoké učení technické v Praze, Fakulta elektrotechnická, Katedra teorie obvodů, Technická 2, 166 27 Praha 6

^bUniverzita Karlova v Praze, Neurologická klinika 1. LF UK a VFN, Kateřinská 30, 128 21 Praha 2

e-mail: [novotm26, ruzszjan, cmejla]@fel.cvut.cz

Hypokinetic dysarthria is a common manifestation of Parkinson's disease (PD). Articulation characteristics can provide useful information to distinguish dysarthric speakers from healthy subjects and monitor the severity of disease and treatment effects. The aim of this study was to design an algorithm for automatic segmentation of consonants and vowels based upon a rapid steady /pa/ – /ta/ – /ka/ syllable repetition. Data for subsequent analysis were used from 24 patients with early stages of PD and 22 healthy subjects. All syllables were manually labeled at three positions including explosion (E), vowel (V), and occlusion (O). In addition, the representative measurement of voice onset time (VOT) was included as difference between V and E position. The three thresholds including 5 ms, 10 ms, and 20 ms were set up to test the accuracy of algorithm in finding proposed positions and VOT. When compared to the manual labeled positions, the VOT is detected within the range 5 ms to 20 ms with a range of a success rate of 68.2–90.5 %, 44.1–75.2 %, and 57.2–83.5 % for normal, dysarthric, and all speakers. In conclusion, this study shows that algorithm based on the spectral, Bayesian, and polynomial approaches, that can be used to accurately detect the positions of consonant and vowels in normal and dysarthria-related utterances.

1. Úvod

Parkinsonova nemoc (PN) postihuje mediální část mozku, jmenovitě oblast substantia nigra [1]. Přes tuto oblast vedou neuronové okruhy spojující takzvaná bazální ganglia s dalšími částmi mozku. Bazální ganglia působí jako blokátor různých motorických úkonů, takže se pohyby nevykylují neplánovaně. V případě chtěného pohybu je pak inhibice potlačena a pohyb je proveden. Při Parkinsonově onemocnění dochází k odumírání neuronů v substantia nigra a to vede k zhoršené schopnosti potlačování inhibiční funkce bazálních ganglií a k hypokinesii [2]. Mezi čtyři základní motorické poruchy vyskytující se u parkinsoniků patří tremor (neovladatelný třes), bradykinesie (pomalost pohybů), ztuhlost (zvýšení svalového tonu), nestabilní držení těla [3]. Dalšími běžnými příznaky PN jsou poruchy nálad, chování, vnímání a různé změny řeči označované jako hypokinetická dysartrie [4].

Podle uveřejněných studií [5] a [6] se u lidí s PN vyskytují vady řeči s pravděpodobností 70–90 %. Tyto vady mohou být jedním z prvních příznaků PN [7]. Jsou definována tři stadia dysartrie: mírná (mild), střední (moderate) a těžká (severe) [4]. Existuje mnoho hlasových charakteristik spojených s hodnocením PN, tři základní oblasti tvoří *fonace*, *artikulace* a *prozódie* [8].

Fonace se vztahuje ke tvorbě hlasu v hlasívkách. *Artikulace* je modulování zvuku pomocí pohybu řečových orgánů, jako jsou jazyk, rty a čelisti. *Prozódie* je kolísání hlasitosti, intonace a načasování při běžné promluvě.

Mezi jednu z úloh pro hodnocení přesnosti artikulace u dysartrie patří diadochokinetická (DDK) úloha, která měří schopnost jedince provádět rychlé opakované pohyby. Pro účely analýzy se obvykle používá metoda opakování

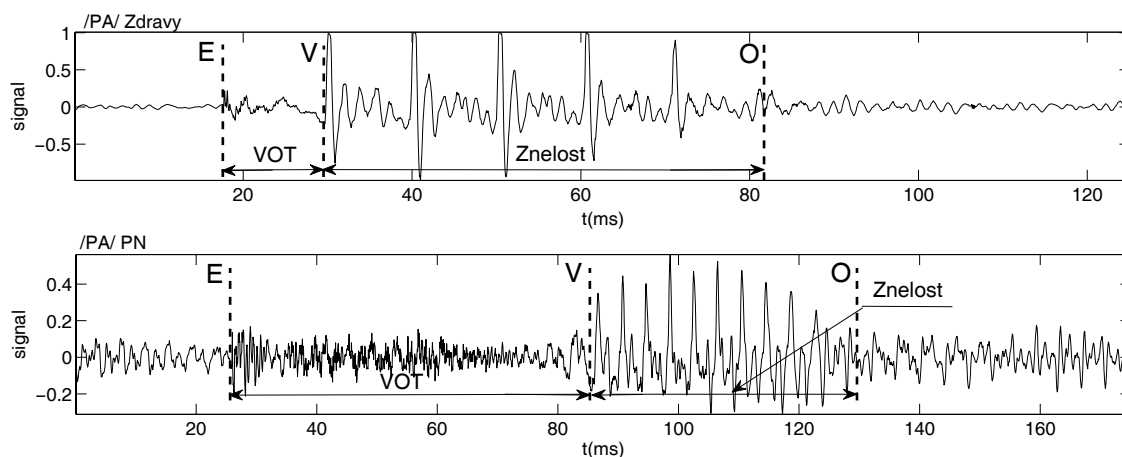
konsonantu a vokálu, kdy se střídá bilabiální, alveolární a velární artikulace. Pacient je požádán, aby co nejrychleji a nejdéle opakoval sekvenci /pa/ – /ta/ – /ka/ [9]. Sekvenci je mimo jiné možné dále použít pro hodnocení kvality konsonantů na základě měření VOT (VOT, Voice Onset Time), které může sloužit jako kritérium pro hodnocení dysartrie [10].

Tato práce se zabývá segmentací promluv v DDK úloze u zdravých jedinců a pacientů s výskytem mírné nebo poslechově nezaznamatelné hypokinetické dysartrie. Hlavním cílem práce je navržení robustního algoritmu pro automatické označení jednotlivých pozic vokálů a exploziv v signálu. Výsledný algoritmus je schopen zachytit již mírné změny řeči v rámci neurologických onemocnění a může sloužit jako nástroj pro diagnostiku, monitorování a terapii nejen u PN.

2. Metodika

2.1. Data

Data použitá v této práci jsou součástí předchozí studie [11], v rámci které byly pořízeny zvukové nahrávky 46 rodilých mluvčích. Z nich 24 (20 mužů a 4 ženy) bylo diagnostikováno s časným stadiem PN a jejich nahrávky byly pořízeny před zahájením symptomatické léčby. Jejich průměrný věk byl $60,9 \pm 11,2$ let, délka trvání PN symptomu $31,3 \pm 22,3$ měsíců a stadium nemoci dle stupnice Hoehnové & Yakra $2,2 \pm 0,5$ (stupnice Hoehnové & Yakra hodnotí stadium PN, kde 0 – bez příznaků a 5 – pacient odkázaný na vozík nebo upoután na lůžko). Data zdravé kontrolní skupiny byla pořízena od 22 účast-



Obrázek 1: Slabika /PA/ s vyznačenými pozicemi E, V a O a vyznačením VOT a znělosti, v horní části je signál zdravého člověka, v dolní je signál člověka s PN

níků (15 mužů a 7 žen) bez historie neurologických onemocnění s průměrným věkem $58,7 \pm 14,6$ let.

Úloha vyhodnocování DDK byla součástí skupiny řečových úkolů, se kterými byli účastníci dobře seznámeni terapeutem. Jednotlivé úlohy probíhaly vždy ve stejném pořadí a nebyly časově omezeny. V případě nespokojenosti účastníka mohla být úloha bez omezení opakována. Pokud byl nespokojen terapeut, který měření prováděl, požádal účastníka o opakování. Při úloze pro charakterizaci DDK byli účastníci požádáni o co možná nejdelší a nejrychlejší opakování /pa/–/ta/–/ka/, nejméně pětkrát za sebou. Pro celkové hodnocení byly použity dvě řečové nahrávky dané úlohy od každého účastníka studie.

Výsledná trénovací množina algoritmu byla složena z databáze o velikosti 80 záznamů (1644 slabik /pa/, /ta/ nebo /ka/). Z tohoto počtu bylo 40 záznamů lidí trpících počátečním stadiem PN (753 slabik) a 40 záznamů zdravých lidí (891). Signál byl pořizován externím kondenzátorovým mikrofonem ze vzdálenosti přibližně 15 cm od úst. Záznam byl prováděn se vzorkovací frekvencí $F_s = 48$ kHz a 16bitovým kvantováním.

2.2. Segmentace signálu

Pro účely automatického hodnocení jednotlivých kritérií je nutné segmentovat řečový záznam a detekovat tři základní pozice. První pozicí je exploze (E), počátek konsonantů (/p/, /t/, /k/), která je charakterizována uvolněním orálního závěru a nárůstem energie explozivního šumu. Druhou pozicí je počátek vokálu (V), který představuje počátek kmitání hlasivek. Poslední pozicí je okluze (O) nebo-li závěr, během které je v řečovém signálu minimální energie. Často se však setkáváme s přezníváním hlasivkového tónu do okluze. Čas mezi E a V představuje dobu od uvolnění orálního závěru do začátku kmitání hlasivek a označuje se jako VOT.

Problémem při hledání jednotlivých pozic v daném signálu byla neznalost jejich celkového počtu. Z tohoto dů-

vodu se ukázalo neefektivním vyhodnocovat signál jako celek, ale bylo nutné jej rozdělit na menší části. V těchto částech byla detekována vždy pouze jedna pozice E, V a O. Řešení algoritmu se tak rozpadlo na čtyři menší části (segmentování signálu na jednotlivé slabiky, detekce E, detekce V a detekce O).

2.3. Segmentace na jednotlivé slabiky

První částí zpracování signálu je segmentace na jednotlivé slabiky. Cílem tohoto kroku je rozdělit řečový signál na části obsahující vždy pouze jedno E, V a O. Jako výsledek jsou získány přibližné hranice slabik v signálu. Tento krok zjednodušuje následné detekce jednotlivých pozic.

Signál je nejprve převzorkován na vzorkovací frekvenci $f_s = 16$ kHz a filtrován dolnoprostopustným filtrem s šířkou propustného pásma 800 Hz pro zvýraznění harmonické složky signálu. Poté je využit špičkový detektor

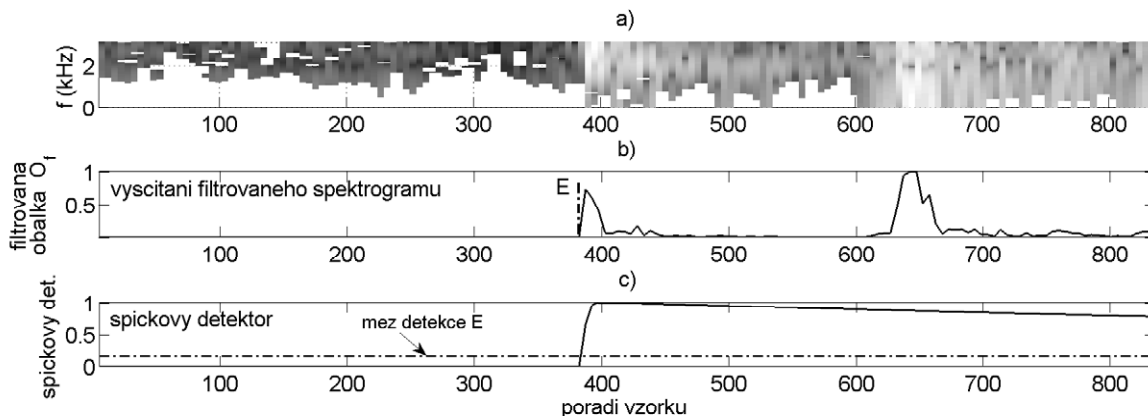
$$s(n) = (1 - k(n)) |y(n)| + k(n)s(n-1), \quad (1)$$

kde $|y(n)|$ značí absolutní hodnotu n -tého vzorku signálu a $k(n)$ je definováno jako

$$k(n) = \begin{cases} 0,9 & |y(n)| < s(n-1), \\ 0,997 & |y(n)| \geq s(n-1). \end{cases} \quad (2)$$

Výstup špičkového detektoru je vyhlazen metodou klouzavého průměrování s šířkou okna 800 vzorků. Vyhlazený průběh je dále normován. V normovaném průběhu jsou pak hledána lokální maxima větší než určitá empiricky stanovená minimální hodnota a navzájem od sebe vzdálená minimálně 800 vzorků. Tímto postupem je předcházeno falešným detekcím.

Po nalezení všech lokálních maxim je vypočítána největší vzdálenost mezi dvěma sousedními maximy. Tato vzdálenost je zvětšena o určitý rezervní počet vzorků, čímž dostaneme počet vzorků potřebný k obsazení i toho nejdelšího segmentu. Tento počet spolu s pozicí lokálních ma-



Obrázek 2: Postup detekce E

xim poté slouží k rozdělení signálu na kratší úseky obsahující pouze jednu slabiku. Pozice lokálních maxim se pohybuje v oblasti vokálu a bohužel ne přesně ve středu slabiky, proto je nezbytné použít i nesymetrické části okna před tímto maximem a za ním. Před maximem je použita pouze $\frac{1}{4}$ počtu vzorků a část za maximem zahrnuje $\frac{3}{4}$ počtu vzorků. Předpoklad pohybu pozice lokálního maxima pouze v rozmezí vokálu umožnil další zjednodušení, kdy byly části před lokálním maximem a za ním vyhodnocovány zvlášť. V levé části byly hledány pozice E a V a v pravé byla detekována O. Příklad segmentu zdravého a nemocného člověka je možné vidět na obrázku 1.

2.4. Detekce exploze

Detekce exploze je realizována pomocí filtrace spektrogramu, k němuž je přístupováno jako k matici \mathbf{P} s m řádky a n sloupci. Pro zvýraznění jednotlivých špiček je stanovena filtrační mez. Všechny hodnoty v \mathbf{P} , které jsou menší než stanovená mez, jsou nastaveny na nulu. Stanovená mez se mění s frekvencí a lze ji zapsat jako

$$h_{mez}(i) = w_e \frac{1}{n} \sum_{j=1}^n P(i, j), \quad (3)$$

kde w_e značí empiricky zjištěnou váhu, která poskytuje nejlepší výsledek při odstranění šumu. Výsledná úprava je pak provedena jako

$$P_f(i, j) = \begin{cases} P_f(i, j) & P(i, j) \geq h_{mez} \\ 0 & P(i, j) < h_{mez} \end{cases} \quad (4)$$

Hledanou pozicí je exploze, jejíž spektrum je tvořeno převážně šumem, proto se E projevuje spíše na vyšších frekvencích. Další úpravou matice \mathbf{P} je váhování podél frekvenční osy. V tomto kroku nejsou části spektrogramu nulovány jako v předcházející úpravě, ale je zde pouze upřednostňována informace nesená vyššími frekvencemi. Tato konstrukce může být vyjádřena jako

$$P_{fwf}(i, j) = w_f(i) P_f(i, j). \quad (5)$$

Filtrovaný spektrogram je zobrazen na obrázku 2a.

Další známou informací je přibližná poloha ve středu segmentu. Proto je zahrnuto i váhování podél časové osy. Pro sestavení váhovacího vektoru \mathbf{w}_t je použito Hammingovo okno, které dává větší důraz na střed časové osy a vhodně potlačuje výskyt předchozího vokálu na počátku a výskyt vokálu náležejícího k hledané explozi na konci segmentu. Tento postup je možno zapsat jako

$$P_{fwfw}(i, j) = w_t(j) P_{fwf}(i, j). \quad (6)$$

Energetická obálka je vytvořena zpětným vysčítáním v jednotlivých časech

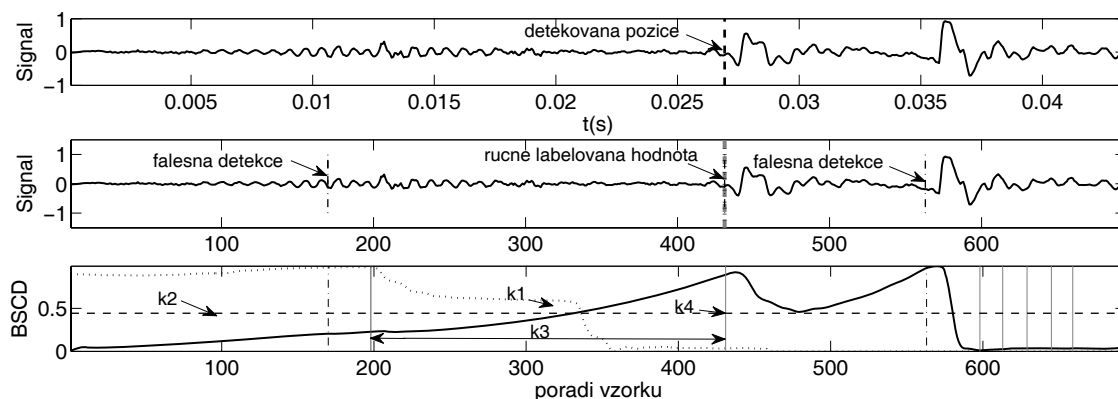
$$O_f(j) = \sum_{i=1}^n P_f(i, j). \quad (7)$$

Energetická obálka (viz obrázek 2b) slouží jako vstup pro špičkový detektor (viz rovnice (1)) a poté, na základě empiricky stanovené meze založené na střední hodnotě (viz obrázek 2c), je určena pozice exploze (viz obrázek 2b).

2.5. Detekce počátku vokálu

Detekce počátku vokálu je založena na prudkém nárůstu energie signálu, detekované pomocí Bayesovského skokového detektoru (BSCD, Bayesian step changepoint detector) [12], pomocí kterého je nalezena skupina možných pozic V.

V některých případech nedokonalého segmentování dochází k posunu E a V oproti předpokládaným pozicím, takže se může E nacházet příliš vpravo nebo V příliš vlevo. To se pak na výstupu BSCD projevuje existencí jedné extrémně vysoké špičky na pozici E nebo V. Pro rozlišení, zda jde o E nebo V, je využíváno periodicity signálu. Tento přístup je založen na předpokladu periodického vokálu a neperiodického konsonantu. V malém okolí špičky je v signálu určeno spektrum. Ve spektru je poté nalezena výška největší špičky různé od stejnosměrné hodnoty. V případě překročení empiricky stanovené meze je špička dostatečně



Obrázek 3: Postup detekce V

vysoká a okolí je označeno jako periodické a pozice jako V, v opačném případě je pozice označena jako E.

V případě normálního výstupu z BSCD je výstupem skupina více možných kandidátů na pozici V. Z tohoto důvodu byla vytvořena skupina kritérií, podle kterých je dále rozhodováno, která špička je reálnou pozicí V. Mezi tato kritéria patří *kritérium obráceného výkonu*, *kritérium střední hodnoty*, *kritérium vzdáleností mezi pozicemi* a *kritérium pořadí jednotlivých pozic*.

Kritéria obráceného výkonu, střední hodnoty a vzdáleností mezi pozicemi s výstupem z BSCD jsou zobrazena v dolní části obrázku 3. V horní části je pak zobrazena detekovaná poloha V a ve střední části jsou znázorněny možné pozice určené podle jednotlivých kritérií. Je vidět, že v místě reálné pozice se několik kritérií překrývá.

Kritérium obráceného výkonu ($k1$) vychází z předpokladu vysokého výkonu v oblasti vokálu. Této vlastnosti je využíváno pro návrh pohyblivé meze. Ze signálové obálky je vyhlazením pomocí klouzavého průměrování (okno dlouhé jednu pětinu délky výstupu BSCD) získána křivka, jejímž zrcadlovým převrácením podle horizontální osy je vytvořena pohyblivá mez. Po získání meze je hledáno místo, kde výstup BSCD přeroste přes tuto mez. První lokální maximum za tímto bodem je poté označeno za možnou pozici V.

Kritérium střední hodnoty ($k2$) je založeno na stejném principu jako mez aplikovaná pro detekci E ve výstupu špičkového detektoru. Vypočítáme váhovanou střední hodnotu výstupu z BSCD. Poté porovnáme výstup BSCD a takto získanou mez. Jako V je pak označena první pozice lokálního maxima následující za průnikem výstupu BSCD a jeho váhovanou střední hodnotou.

Kritérium vzdáleností mezi pozicemi ($k3_{1,2,3}$) vychází z tvaru výstupu BSCD a poměrně dlouhé, hladké, stoupající hrany, která je dobře vidět na spodní části obrázku 3. Na konci této hrany se často nachází hledaná pozice, proto bereme první tři největší vzdálenosti mezi špičkami. Ty jsou porovnány s ostatními vzdálenostmi a v případě, že jsou dostatečně dlouhé oproti ostatním, jsou poté jako

možné pozice V označeny hranice pravých stran těchto vzdáleností.

Kritérium pořadí jednotlivých pozic ($k4$) – zde bylo zohledněno pozorování, kdy se reálná hodnota většinou vyskytovala v určité části nalezených pozic. Z toho důvodu byla jako nejslabší kritérium vybrána i pevně nastavená pozice. První lokální maximum po této pozici bylo zaznamenáno jako možná poloha V.

Tato čtyři kritéria mohou poskytnout až 6 možných pozic V (kritérium vzdáleností poskytuje až 3 polohy). Pro výběr nejpravděpodobnější polohy byl určen postup, kdy jsou nejprve do jednoho vektoru zapsány všechny možné pozice tak, že $\mathbf{v} = (k1, k2, k3_1, k3_2, k3_3, k4)$. Dalším krokem je nalezení počtu shod. V případě, že se na jedné pozici shodne nejvíce kritérií, je tato pozice označena jako V. Pokud se ovšem vyskytuje více pozic se stejným počtem detekování, určuje výběr pořadí kritéria ve vektoru \mathbf{v} . Je vybráno $k1$ a porovnáno s nejvíce zastoupenými pozicemi. V případě, že na některou ukazuje, je tato pozice označena jako V. V případě, že zastoupeno není, postupujeme ke $k2$ a celý proces opakujeme. Takto postupujeme vektorem \mathbf{v} ve směru od $k1$ ke $k4$.

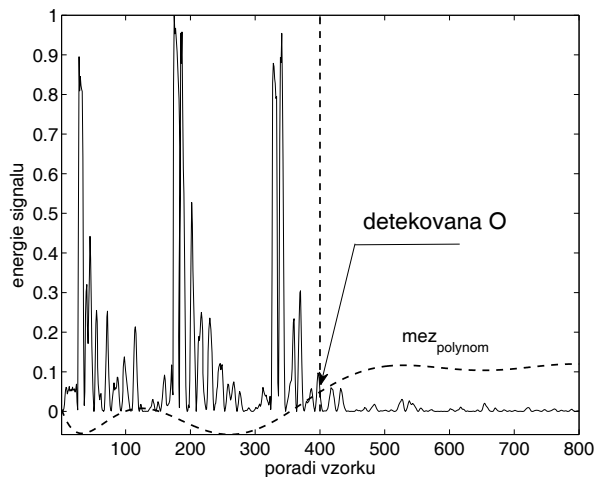
2.6. Detekce okluze

Principem detekce okluze je nalezení proměnlivé meze, která se optimalizuje vzhledem ke změnám signálu. Signál je nejprve filtrován dolnoproústným filtrem se šířkou pásma 500 Hz. Následně je z filtrovaného signálu vypočítán kvadrát. Mez je poté tvořena invertovanou polynomiální aproximací devátého řádu, která je zároveň posunuta o offset tvořený dvojnásobkem střední hodnoty. Tato mez může být vyjádřena jako

$$\text{mez}_{\text{polynom}} = \prod_{i=1}^9 (a_i \mathbf{x} + b(i)) + 2\bar{x}, \quad (8)$$

kde \mathbf{x} je vektor hodnot osy x s první hodnotou rovnou jedné a s délkou rovnou délce zkoumaného segmentu n .

Na obrázku 4 je vidět určení pozice pomocí polynomiální meze a obálky signálu. Koeficienty a_i a b_i jsou koeficienty i -tého řádu polynomu a \bar{x} je střední hodnota energie signálu.



Obrázek 4: Postup detekce O

2.7. Vyhodnocení úspěšnosti

Vyhodnocování funkčnosti detekcí jednotlivých pozic bylo prováděno vzhledem k celkovým počtům jednotlivých slabik, nikoliv k úspěšnosti (v tabulce 1 uvedená jako úspěšnost) detekcí v jednotlivých signálech.

Pro posouzení úspěšnosti byly využity ručně označené pozice jednotlivých E, V a O. Pro porovnání byl vypočítán rozdíl automaticky detekovaných pozic a těchto refe-

renčních hodnot. Absolutní hodnota rozdílu pak byla porovnána se třemi mezemi, a to 5 ms, 10 ms a 20 ms, jako úspěšná byla hodnocena každá detekce, která byla menší, nebo se rovnala stanovené mezi

$$|x_{\text{změřené}} - x_{\text{labelované}}| \leq \text{mez}_{5,10,20}, \quad (9)$$

kde $x_{\text{změřené}}$ a $x_{\text{labelované}}$ jsou detekované a labelované pozice. Procentuální úspěšnost byla stanovena jako podíl součtu všech úspěšných detekcí slabik ku celkovému počtu.

Pro hodnocení úspěšnosti detekce VOT byly stanoveny stejné meze jako pro hodnocení detekce E, V a O. Porovnáván byl rozdíl mezi detekovanou a labelovanou délkou VOT, vypočítanou jako rozdíl $\text{VOT} = V - E$.

Meze byly zvoleny právě takto, z důvodu hodnocení použitelnosti algoritmu při charakterizaci řečového signálu. Další motivací byla možnost porovnávat dosaženou úspěšnost s dříve publikovanými pracemi [13] a [14].

3. Výsledky

V tabulce 1 jsou znázorněny počty správných detekcí a procenta úspěšnosti pro tři skupiny (zdraví, PN, celkově). Celkový výsledek úspěšnosti činil při 5 ms 64,0% pro E, 71,2% pro V a 32,1% pro O. Nicméně pro O je relevantnější mez 10 ms, jejíž úspěšnost činila 64,5%. Pro všechny meze poskytovala nejlepší výsledky skupina zdravých, která se lišila o 6,2–10,4%. Mezi jednotlivými detekcemi pak nejlepších výsledků dosahovala detekce V, jejíž úspěšnost se lišila o 0,3–9,9% od detekce E a o 39,1–43,3% od detekce O.

V poslední části tabulky 1 jsou uvedeny úspěšnosti pro výpočty délek VOT.

Detekce	Zdraví		PN		Celkově		
	počet správných detekcí/celkový počet	úspěšnost (%)	počet správných detekcí/celkový počet	úspěšnost (%)	počet správných detekcí/celkový počet	úspěšnost (%)	
E	mez						
	≤ 5 ms	644 / 891	71,7	428 / 753	56,8	1072 / 1644	64,0
	≤ 10 ms	694 / 891	77,3	488 / 753	64,8	1182 / 1644	70,6
	≤ 20 ms	787 / 891	87,6	558 / 753	74,1	1345 / 1644	80,3
V	mez						
	≤ 5 ms	733 / 891	81,6	459 / 753	61,0	1072 / 1644	71,2
	≤ 10 ms	810 / 891	90,2	551 / 753	73,2	1361 / 1644	81,3
	≤ 20 ms	855 / 891	95,2	618 / 753	82,1	1473 / 1644	88,0
O	mez						
	≤ 5 ms	344 / 891	38,3	193 / 753	25,6	537 / 1644	32,1
	≤ 10 ms	688 / 891	76,6	393 / 753	52,2	1081 / 1644	64,6
	≤ 20 ms	834 / 891	92,9	500 / 753	66,4	1334 / 1644	79,7
VOT	mez						
	≤ 5 ms	608 / 891	68,2	332 / 753	44,1	940 / 1644	57,2
	≤ 10 ms	681 / 891	76,4	439 / 753	58,3	1120 / 1644	68,1
	≤ 20 ms	807 / 891	90,6	566 / 753	75,2	1373 / 1644	83,5

Tabulka 1: Shrnutí výsledků

4. Diskuze

Z uvedených statistik vyplývá, že nejvyšší úspěšnosti bylo dosaženo u detekce V. Detekce E zaostává zhruba o 10 %. Výsledky detekce O se s výsledky detekce E a V těžko srovnávají, a to z důvodu mnohem strmějšího nárůstu úspěšnosti se snižováním nároků na přesnost. Zatímco u detekcí E a V odpovídá každému snížení nároků zvýšení úspěšnosti o 6–10 %, u detekce O se nárůst úspěšnosti pohybuje mezi 15–30 %. To je částečně způsobeno pozvolným odezněním vokálu, které ztěžuje ruční labelování přesné pozice O. Z tohoto důvodu může být deterministický přístup, který se drží pevně nastavených pravidel výhodnější. Zároveň předpokládaný účel využití neklade na přesnou detekci O tak velký důraz jako na detekci E a V. Toto se dá demonstrovat například na charakteristikách VOT a periodicitě. Zatímco VOT je získán jako rozdíl E a V, a je tedy nutné znát co možná nejpřesnější polohy E a V, je periodicita dále dopočítávána z úseku znělosti ohraničeného V a O a není tak náchylná na zkrácení signálu. Větším problémem je prodloužení, kdy do výpočtu periodicity zanášíme i neperiodickou šumovou část po odeznění vokálu. Požadavky na detekci O tedy byly spíše posunuty směrem doleva z důvodu omezení pravděpodobnosti zanešení neperiodické části signálu do výpočtu periodicity.

Úspěšnost detekce VOT je možné rámcově porovnat s pracemi [13] a [14]. Studie [13] se zabývá hodnocením délky VOT pro účely rozlišování akcentů. Pro porovnání byly vybrány výsledky hodnocení subjektů majících jako mateřský jazyk americkou angličtinu, jelikož se nejvíce podobaly signálům v naší databázi. Úspěšnost nalezení VOT pro mez rovnou 10 % délky VOT je 74,9 % a průměrná chyba ve správných detekcích je 0,735 ms [13]. Náš algoritmus dosahuje v celkových výsledcích horšího skóre 57,2 % a 1,273 ms. Výsledky uvedené v práci [14], která se zabývá návrhem algoritmu pro měření VOT znělých i neznělých konsonantů (/b/, /p/, /d/, /t/, /g/, a /k/), jsou pro mez 10 ms 72,6 % a pro mez 20 ms 87,8 %. Naše výsledky jsou na téměř srovnatelné úrovni, 68,1 % pro 10 ms mez a pro 20 ms mez jsou 83,5 %.

Při porovnávání je však nutné uvážit zaměření jednotlivých prací. Data použitá v článku [13] byla poskytnuta lidmi s různým akcentem (americká angličtina, hindština, čínština), ve článku [14] byla využita databáze založená na (americké) angličtině, dalším limitujícím faktorem je zastoupení PN lidí, které detekci VOT výrazně ztěžuje. V případě práce [13] navíc přesně nesouhlasí meze. Zatímco práce [13] bere meze jako relativní odchylku 10 %, v naší práci byla využita odchylka absolutní, z důvodu chování VOT u PN. V případě zanedbání akcentů a vynechání skupiny PN jsou výsledky s uvedenými pracemi na srovnatelné úrovni. Úspěšnost detekce VOT u zdravých pro mez 5 ms vychází 68,2 % a průměrná chyba činí 0,961 ms oproti 74,9 % a 0,735 ms uvedeným v [13]. Úspěšnost pro 10 ms je 76,4 % a úspěšnost pro 20 ms činí 90,57 %. Úspěšnosti uvedené v [14] jsou 72,6 % pro 10 ms a 87,8 % pro 20 ms.

V současné době pracují všechny tři detekce nezávisle na sobě, proto je zde prostor pro zlepšení výsledků zpětnou kontrolou. Jedním z přístupů může být výpočet VOT nebo délky znělosti a v případě zjištění hodnot, které nejsou fyziologicky pravděpodobné (předpokladem je, že minimální délku VOT je možné experimentálně odhadnout), opětovné přeměření a detekce v mezích založených na předchozích měřeních.

Robustnost algoritmu by také mohlo zvýšit zlepšení úspěšnosti algoritmu pro rozdělení signálu na jednotlivé podsignály. I do této části by mohl být zařazen kód porovnávající délku VOT. Případné špatné detekce totiž ve většině případů vedou k záměně E a V. Tato chyba vede na délku VOT rovnu nule a ta může být snadno označena fyziologicky nepravděpodobnou.

Zlepšení detekce O by mohlo pomoci plánované zavedení algoritmu pro výpočet ideálního stupně polynomu pro určení tvaru meze. V současné době je stupeň polynomu nastaven na pevnou hodnotu a ne ve všech případech je tato hodnota tou nejlepší.

5. Závěr

Algoritmus prezentovaný v článku pracoval při uplatnění meze rovné 5 ms, s úspěšnostmi rovnými 64,0 % pro detekci exploze, 71,2 % pro detekci vokálu a 57,2 % pro detekci voice onset time. Úspěšnost detekce okluze byla 64,6 % při uplatnění 10 ms meze. Tyto výsledky jsou uspokojivé a srovnatelné s ostatními pracemi, ovšem další zlepšení robustnosti algoritmu jsou nezbytná z důvodu plánovaného použití u poruch artikulace způsobených dysartrií.

Poděkování

Autoři děkují Mgr. Haně Růžičkové, MUDr. Janě Picmausové, MUDr. Veronice Majerové, MUDr. Janu Klempiřovi, Ph.D., doc. MUDr. Janu Rothovi, CSc. a prof. MUDr. Evženu Růžičkovi, DrSc., za poskytnutí souboru klinických dat. Tato práce je podporována z grantů SGS10/180/OHK3/2T/13, GAČR 102/08/H008 a NT 12288-5/2011 a z výzkumných záměrů MSM 0021620849 a MSM 6840770012.

Reference

- [1] Hornykiewicz, O.: Biochemical aspects of Parkinson's disease, *Neurol. Suppl.*, 51(2), S2–S9, 1998.
- [2] Obeso, J. A., Rodríguez-Oroz, M. C., Benitez-Temino, B., et al.: Functional organization of the basal ganglia: Therapeutic implications for Parkinson's disease, *Movement Disorders*, 23, S548–S559, 2008.
- [3] Jankovic, J.: Parkinson's disease: clinical features and diagnosis, *J. Neurol. Neurosurg. Psychiatr.*, 79(4), 368–376, 2008.

- [4] Darley, F. L., Aronson, A. E., Brown, J. R.: Differential diagnostics patterns of dysarthria, *J. Speech. Hear. Res.*, 12, 426–496, 1969.
- [5] Logemann, A. J., Fisher, H. B., Boshes, B., Blonsky, E. R.: Frequency and cocurrence of vocal tract disfunction in the speech of a large sample of Parkinson patients, *J. Speech Hear. Disord.*, 43, 47–57, 1978.
- [6] Duffy, J. R.: Motor Speech Disorders: Substrates, Differential Diagnosis and Management, 2nd ed., *Mosby, New York, NY, 2005*, p. 1–592.
- [7] Harel, B., Cannizzaro, M., Snyder, P. J.: Variability in fundamental frequency during speech in prodromal and incipient Parkinson’s disease: A longitudinal case study, *Brain Cogn.*, 56, 24–29, 2004.
- [8] Ruzs, J., Čmejla, R., Růžicková, H., Růžička, E.: Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson’s disease, *J. Acoust. Soc. Am.*, 129(1), 350–367, 2011.
- [9] Fletcher, S.: Time – by – count measurement of diadochokinetic syllable rate, *J. Speech. Hear. Disord.*, 15, 757–762, 1972.
- [10] Fisher, E., Goberman, A. M.: Voice onset time in Parkinson’s disease, *J. Commun. Disord.*, 43, 21–34, 2010.
- [11] Ruzs, J., Čmejla, R., Růžicková, H., Klempíř, J., Majerová, V., Picmausová, J., Roth, J., Růžička, E.: Acoustic assesment of voice and speech disorders in Parkinson’s disease through quick vocal test, *Mov. Disord.*, 26(10), 1951–1952, 2011.
- [12] Čmejla, R., Sovka, P.: Recursive Bayesian Autoregressive Changepoint Detector for Sequential Signal Segmentation, *EUSipco Proceedings, Wien(2004)*, 245–248.
- [13] Hansen, J. H. L., Gray, S. S., Kim, W.: Automatic voice onset time detection for unvoiced stops (/p/,/t/,/k/) with application to accent classification, *Speech Communication*, 52(10), 777–789, 2010.
- [14] Stouten, V., Van Hame, H.: Automatic voice onset time estimation from reassignment spectra, *Speech Communication*, 51(12), 1194–1205, 2009.

Analýza nelineárních stojatých vln ve dvou vázaných akustických rezonátorech

Michal Bednařík, Milan Červenka a Jaroslav Plocek

České vysoké učení technické v Praze, Fakulta elektrotechnická,
Technická 2, 166 27 Praha 6
e-mail: bednarik@fel.cvut.cz

The paper deals with description of nonlinear standing waves in acoustic resonators that are coupled mechanically by means of an elastically mounted wall which is implemented between the resonators. The coupling represents a linear oscillators. For the purpose of the behavior description of the nonlinear acoustic fields, the system of three model equations were derived. Two of them are the modified inhomogeneous Burgers equations and the third model equation is the oscillator's equation of motion. The investigated resonant system is excited by the harmonically vibrating pistons. The system of model equations was solved numerically in the frequency domain. The whole system obtains many parameters which can be changed. With help of these parameters we can adjust various configurations of the resonant system. The configurations, which offer interesting results, were studied. One of the configurations ensures that the resonant system behaves as a frequency convertor. Other selected configuration causes suppression of higher harmonic components in the one of the resonators.

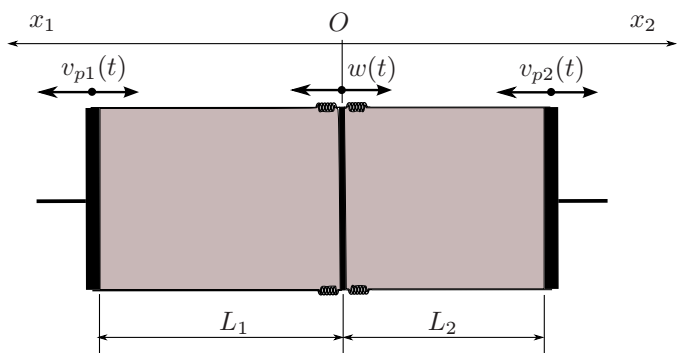
1. Úvod

Součástí řady zařízení je akustická soustava sestávající se z nelineárních rezonátorů, které jsou vzájemně spojeny vazbou. K těmto zařízením patří např. termoakustické stroje, akustické kompresory či některé hudební nástroje. Vázané rezonátory, resp. nelineární stojaté vlny v těchto rezonátorech, vzájemně interagují, čímž mohou ovlivnit funkci celého zařízení, jehož jsou součástí. Pro takové případy je nutné umět popsat a ovlivnit chování výsledných stojatých vln na základě volby příslušných parametrů rezonanční soustavy. Z tohoto důvodu se tato práce věnuje právě zkoumání interakce mezi nelineárními stojatými vlnami prostřednictvím lineární mechanické vazby. Tato vazba je realizována pomocí stěny, kterou považujeme za dokonale tuhou, ale pružně upevněnou mezi dva rezonátory. Předpokládáme, že výchylky stěny jsou natolik malé, že lze na tuto kmitající stěnu pohlížet jako na lineární buzený mechanický oscilátor. Dále předpokládáme, že rezonátory kruhového průřezu jsou tvořeny dokonale tuhými stěnami a jsou vyplněny stejnou tekutinou, která se v obou rezonátorech nachází v shodném termodynamickém stavu. Nelineární stojaté vlny je možné budít dvěma harmonicky kmitajícími písty, viz obr. 1.

Aby bylo možné modelovat chování nelineárních stojatých vln v rezonátorech, byla nalezena soustava tří rovnic. Tuto soustavu tvoří dvě nehomogenní Burgersovy rovnice popisující nelineární stojaté vlny v rezonátorech a pohybová rovnice lineárního mechanického buzeného oscilátoru. Nalezená soustava rovnic má tu výhodu, že ji lze poměrně snadno numericky řešit v kmitočtové oblasti a rovněž skýtá možnost nalezení přibližných analytických řešení pomocí vhodně zvolených perturbančních metod.

2. Modelové rovnice

Pro nalezení soustavy modelových rovnic uvažujeme rezonanční soustavu vázaných rezonátorů, která je zachycena na obrázku 1. V nelineární teorii druhého řádu je možné



Obrázek 1: Dva rezonátory svázané mechanickou lineární vazbou

popsat nelineární akustické pole v rezonátorech jako superpozici dvou proti sobě postupujících nelineárních vln. Tyto dvě vlny jsou spolu svázané pouze okrajovými podmínkami na podstavách válcového rezonátoru. V případě, že uvažujeme pouze vlivy, které jsou maximálně druhého řádu, lze rovněž zanedbat skutečnost, že se stěny tvořené budičnými písty pohybují. Na základě těchto předpokladů můžeme pro nelineární vlny pohybující se proti sobě v uvažovaných rezonátorech ($i = 1, 2$) odvodit následující modelové rovnice (viz např. [3], [1])

$$c_0 \frac{\partial v_i^{(+)}}{\partial x_i} - \frac{\partial v_i^{(+)}}{\partial t} - \frac{\beta}{c_0} v_i^{(+)} \frac{\partial v_i^{(+)}}{\partial \tau_i^{(+)}} = \frac{b}{2\rho_0 c_0^2} \frac{\partial^2 v_i^{(+)}}{\partial \tau_i^{(+)^2}} \quad (1)$$

a

$$-c_0 \frac{\partial v_i^{(-)}}{\partial x_i} - \frac{\partial v_1^{(-)}}{\partial t} - \frac{\beta}{c_0} v_i^{(-)} \frac{\partial v_i^{(-)}}{\partial \tau_i^{(-)}} = \frac{b}{2\rho_0 c_0^2} \frac{\partial^2 v_i^{(-)}}{\partial \tau_i^{(-)2}}. \quad (2)$$

Zde x_i je prostorová souřadnice ve směru osy uvažovaného rezonátoru, t je čas, c_0 je rychlost šíření zvuku v lineárním přiblížení, ρ_0 je klidová hustota tekutiny, b reprezentuje parametr související s tepelnou vodivostí a vnitřním třením v tekutině, $\tau_i^{\pm} = t \mp x_i/c_0$ je retardovaný čas, β je koeficient nelinearity a v_i^{\pm} je akustická rychlost pohybujících se vln, znaménko $+/-$ souvisí s vlnou pohybující se v kladném/záporném směru osy x_i .

Řešením rovnic (1) a (2) můžeme pro akustickou rychlost stojaté vlny uvnitř příslušného rezonátoru psát

$$v_i = v_i^{(+)} - v_i^{(-)}. \quad (3)$$

Vlastní kruhové kmitočty (rezonanční kruhové kmitočty) rezonátorů jsou dány následujícími výrazy

$$\omega_i^{(n)} = \frac{n\pi c_0}{L_i} = n\omega_i, \quad n = 1, 2, 3, \dots \quad (4)$$

V případě, že uvažujeme harmonické buzení stojatých vln píсты v místech $x_i = L_i$, můžeme vyjádřit okrajové podmínky následujícím způsobem

$$v_i = \left(v_i^{(+)} - v_i^{(-)} \right)_{x_i=0} = w(t), \quad v_i = \left(v_i^{(+)} - v_i^{(-)} \right)_{x_i=L_i} \\ = v_{pi}(t) = \sum_{n=1}^N (-1)^n v_{mi}^{(n)} \sin(\Omega_i^{(n)} \tau_i^{\pm} + \varphi_i^{(n)}), \quad (5)$$

kde $v_{mi}^{(n)}$, $\Omega_i^{(n)}$, $\varphi_i^{(n)}$ je amplituda, kruhový kmitočet a fáze n -té oscilace i -tého pístového zdroje ($i = 1, 2$) a $w(t) = du(t)/dt$ je rychlost kmitání elasticky připevněné stěny, která představuje oscilátor s jedním stupněm volnosti. Pohybovou rovnicí tohoto oscilátoru můžeme napsat jako

$$\frac{d^2 u(t)}{dt^2} + 2\delta \frac{du(t)}{dt} + \omega_0^2 u(t) = \\ \frac{S}{m} [p_1(x_1 = 0, t) - p_2(x_2 = 0, t)], \quad (6)$$

kde $u(t)$ je výchylka mechanického oscilátoru, δ je parametr útlumu, S je vnitřní průřez rezonátoru, m je hmotnost pružně se pohybující stěny, p_1 a p_2 představují akustický tlak v prvním a druhém rezonátoru, $\omega_0^2 = s/m$ je charakteristický kmitočet oscilátoru, kde s je jeho tuhost.

Rovnice (1) a (2) spolu s podmínkami (5) mohou být řešeny metodou postupných aproximací, viz např. [3]. Pomocí této metody získáme následující modelové rovnice (nehomogenní modifikovaná Burgersova rovnice), které

jsou použitelné v blízkosti rezonančních kmitočtů rezonátorů

$$\frac{\partial \bar{v}_i}{\partial t} + q_i \frac{\partial \bar{v}_i}{\partial \tau_i} - \frac{\beta}{c_0} \bar{v}_i \frac{\partial \bar{v}_i}{\partial \tau_i} - \frac{b}{2\rho_0 c_0^2} \frac{\partial^2 \bar{v}_i}{\partial \tau_i^2} = \\ (-1)^{i-1} w(\tau_i) + \frac{c_0}{2L} \sum_{n=1}^N (-1)^{n+1} v_{mi}^{(n)} \sin(\Omega_i^{(n)} \tau_i + \varphi_i^{(n)}), \\ i = 1, 2, \quad (7)$$

kde $q_i = \Delta L_i/L_i$ představuje rozladění. Znaménka \pm byla vynechána, protože rovnice (7) mají stejný tvar pro obě proti sobě pohybující se vlny.

Rovnice (6) a (7) představují soustavu rovnic popisující nelineární stojaté vlny v uvažovaných rezonátorech. Akustické tlaky p_i jsou v rezonátorech svázány s pohybem oscilátoru pomocí následující podmínky

$$\left. \frac{\partial p_i}{\partial x_i} \right|_{x_i=0} = -\rho_0 \frac{dw}{dt}. \quad (8)$$

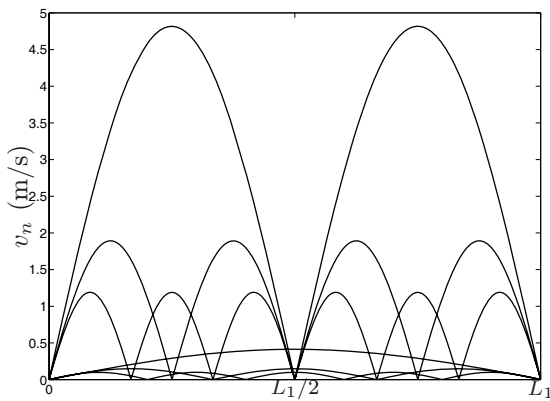
Rovnice (7) představují nehomogenní Burgersovy rovnice, kde

$$\bar{v}(t, \tau_i^{\pm}) = v_{\pm}(t, \tau_i^{\pm}) \pm \frac{x_i - L}{2L} w(\tau_i^{\pm}) \\ \mp \sum_{n=1}^N \frac{v_{mi}^{(n)}}{2L_i} \sin(\Omega_i^{(n)} \tau_i^{\pm} + \varphi_i^{(n)}). \quad (9)$$

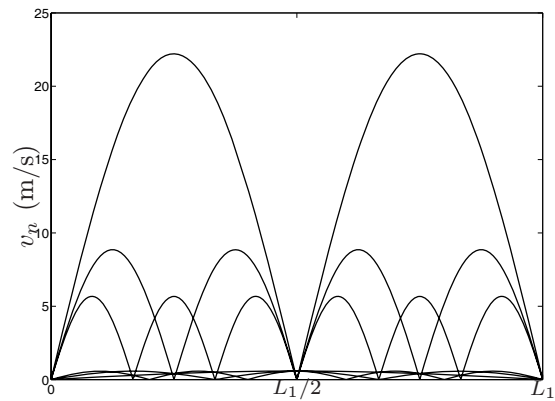
3. Výsledky

Soustava rovnic (6) a (7) obsahuje řadu parametrů, které mohou být měněny, což umožňuje vytvořit konfigurace, které nabízejí zajímavé výsledky, někdy na první pohled překvapivé. Jako první příklad můžeme představit konfiguraci s následujícími zvolenými parametry: $v_{p1}(t) = 0,01 \sin(\omega_1 t)$, $v_{p2}(t) = 0$, $L_2 = L_1/2$, $\omega_0 = 2\omega_1$. Soustava rovnic byla s těmito parametry řešena numericky v kmitočtové oblasti pomocí Rungeovy-Kuttovy rovnice 4. řádu. Výsledky jsou zachyceny na obrázcích 2 a 3. Z těchto obrázků můžeme vidět, že v prvním rezonátoru sudé harmonické složky převažují (co do amplitudy) nad složkami lichými. Jejich převaha roste s klesajícím parametrem útlumu δ , jak je možné vidět na obrázcích 4 a 5. S ohledem na skutečnost, že druhý rezonátor má poloviční délku oproti rezonátoru prvnímu, jsou v druhém rezonátoru vybudeny jak liché, tak sudé harmonické složky nelineární stojaté vlny. Akustické pole v druhém rezonátoru odpovídá akustickému poli, které je vybudeno v rezonátoru s dokonale tuhými stěnami pístem kmitajícím na jeho základním rezonančním kmitočtu. V případě, že parametr útlumu δ je nízký, potom svázané akustické rezonátory pracují v této konfiguraci jako kmitočtový konvertor.

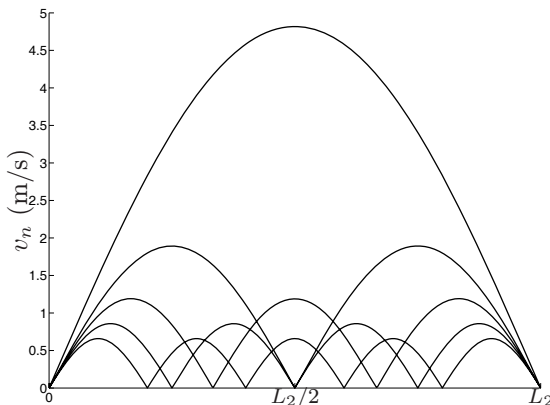
Další konfigurace s parametry $v_{p1}(t) = 0,01 \sin(\omega_1 t)$, $v_{p2}(t) = 0$, $L_2 = L_1$, $\delta = 4 \text{ s}^{-1}$, $2\omega_0 = \omega_1$ umožňuje budít akustické pole, kterému odpovídají harmonické složky zachycené na obrázcích 6 a 7. Z těchto obrázků je zřejmé, že



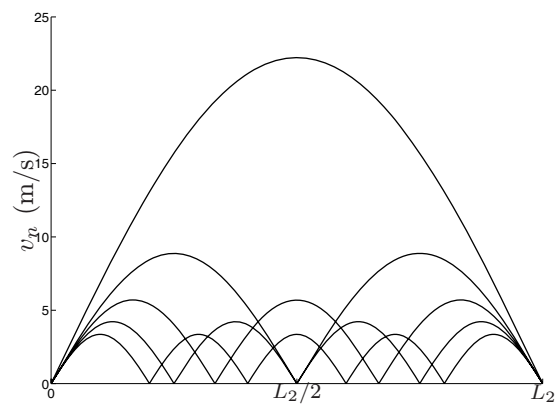
Obrázek 2: Distribuce amplitud jednotlivých harmonických složek podél osy prvního rezonátoru – potlačení lichých harmonických složek, $\delta = 4 \text{ s}^{-1}$



Obrázek 4: Distribuce amplitud jednotlivých harmonických složek podél osy prvního rezonátoru – potlačení lichých harmonických složek, $\delta = 2 \text{ s}^{-1}$



Obrázek 3: Distribuce amplitud jednotlivých harmonických složek podél osy druhého rezonátoru v případě potlačování lichých harmonických složek v prvním rezonátoru, $\delta = 4 \text{ s}^{-1}$



Obrázek 5: Distribuce amplitud jednotlivých harmonických složek podél osy druhého rezonátoru v případě potlačování lichých harmonických složek v prvním rezonátoru, $\delta = 2 \text{ s}^{-1}$

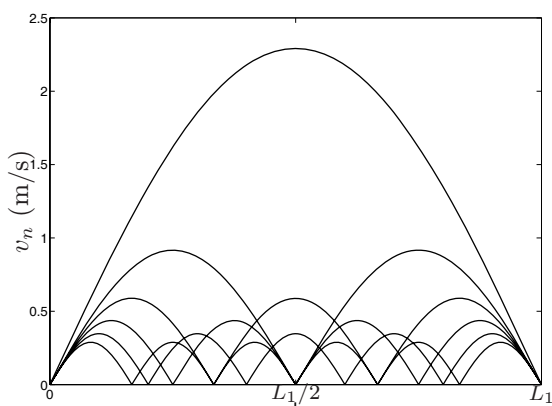
pružně upevněná stěna se chová téměř jako stěna dokonale tuhá. Druhá harmonická složka budí v druhém rezonátoru výrazně slabší pole, které obsahuje jen sudé harmonické složky.

Poslední příklad představuje následující konfiguraci: $v_{p1}(t) = 0,01 \sin(\omega_1 t)$, $v_{p2}(t) = 0,25 \sin(\omega_2 t)$, $L_2 = L_1/2$, $\delta = 2 \text{ s}^{-1}$, $\omega_0 = 2\omega_1$. Tato konfigurace umožňuje výrazně potlačit vyšší harmonické složky v prvním rezonátoru, viz obrázky 8 a 9.

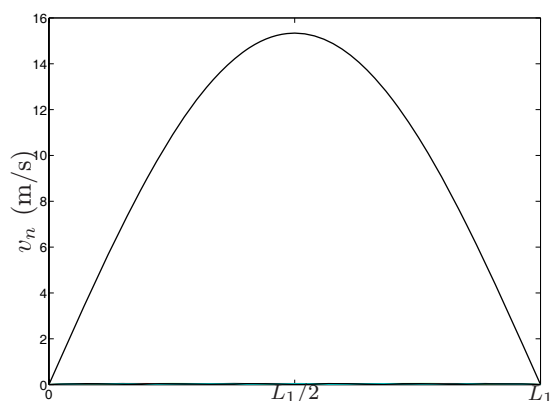
4. Závěr

Prezentovaná práce má teoretický charakter a je věnována popisu nelineárních stojatých vln v cylindrických akustických rezonátorech. Předpokládali jsme, že se akustická pole v obou rezonátorech vzájemně ovlivňují pomocí nelineární mechanické vazby. Tato vazba je reprezentována

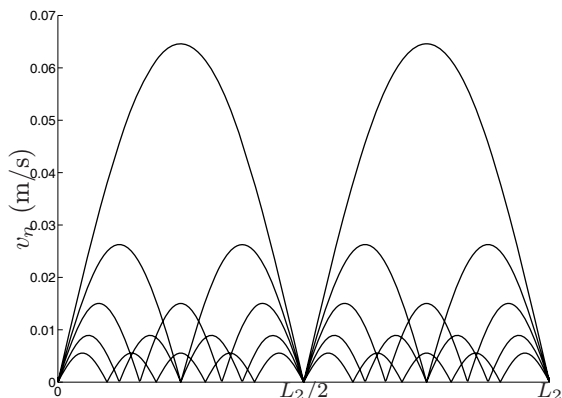
dokonale tuhous stěnou, která je umístěna mezi uvažované rezonátory. Za účelem modelování akustických polí v rezonátorech jsme odvodili nehomogenní Burgersovy rovnice obsahující vazební člen. Dvě nehomogenní Burgersovy rovnice spolu s pohybovou rovnicí vazebního prvku představují soustavu rovnic, které umožňují modelovat chování vyšetřované rezonanční soustavy. Tato rezonanční soustava má celou řadu parametrů, které je možné ve fyzikálně přípustných mezích měnit. Tímto je možné nastavit mnoho nejrůznějších konfigurací. V této práci byly popsány celkem tři konfigurace. První z konfigurací demonstrovala možnost využití rezonanční soustavy jako konvertoru kmitočtu. Druhá z uvažovaných konfigurací ukazovala na případ, kdy žádný z rezonančních kmitočtů není v koincidenci s charakteristickým kmitočtem uvažované mechanické vazby. V tomto případě se pružně upevněná tuhá stěna chová, jako by byla upevněna nepružně. Třetí



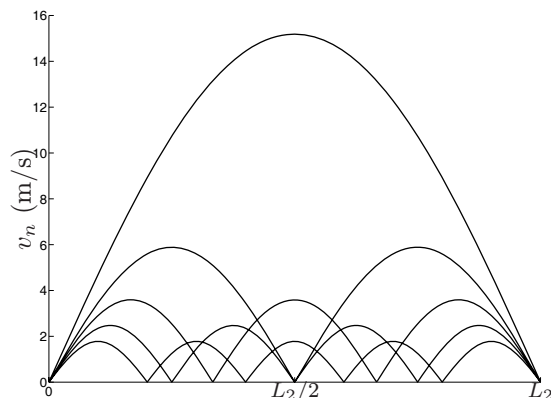
Obrázek 6: Distribuce amplitud jednotlivých harmonických složek podél osy prvního rezonátoru – druhá konfigurace



Obrázek 8: Distribuce amplitud jednotlivých harmonických složek podél osy prvního rezonátoru – potlačení vyšších harmonických složek



Obrázek 7: Distribuce amplitud jednotlivých harmonických složek podél osy druhého rezonátoru – druhá konfigurace



Obrázek 9: Distribuce amplitud jednotlivých harmonických složek podél osy druhého rezonátoru v případě potlačení vyšších harmonických složek v prvním rezonátoru

prezentovaná konfigurace ukazuje na možnost využití rezonanční soustavy pro potlačení vzniku vyšších harmonických složek v prvním rezonátoru. Dosažené výsledky jsou lepší než v případě multifrekvenčního buzení, které bylo popsáno např. v [2], [4].

S ohledem na teoretický charakter této práce je nutné provést řadu následných experimentů, které by ověřily platnost teoretických závěrů. Dále je možné rozšířit některé teoretické výsledky nalezením přibližných analytických řešení prezentované soustavy modelových rovnic.

Poděkování

Tato práce vznikla z podpory SGS 10/265/OHK3/3T/13.

Reference

- [1] M. Bednařík, M. Červenka and P. Koníček: Parametric excitation of nonlinear standing waves in acoustic resonator, *Ultrasonic Symposium (IUS)*, IEEE International, 2009.
- [2] V. E. Gusev: Buildup of forced oscillations in acoustic resonators, *Sov. Phys. Acoust.*, 30, p. 121–125, 1984.
- [3] M. Bednařík and P. Koníček: Description of quasi-plane nonlinear standing waves in cylindrical resonators, *Czech. J. Phys.*, 54, p. 349–355, 2004.
- [4] M. Bednařík and P. Koníček: Asymptotic solutions of the inhomogeneous Burgers equation, *J. Acoust. Soc. Am.*, 115, p. 91–98, 2004.

Nízkofrekvenční zvukové svazky

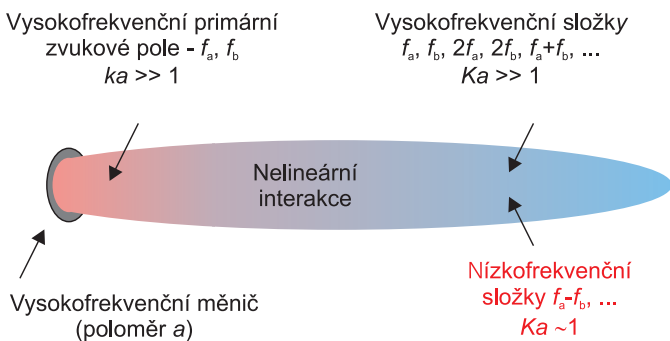
Milan Červenka, Michal Bednařík, Jaroslav Plocek a Martin Šoltés

České vysoké učení technické v Praze, Fakulta elektrotechnická,
Technická 2, 166 27 Praha 6
e-mail: milan.cervenka@fel.cvut.cz

This paper is concerned with study of low-frequency sound beams generated as difference-frequency secondary field in parametric array. As model equations for theoretical investigation, KZK equation and Higher-order parabolic equation (HOPE) [Kamakura, T., Masahiko, A., Kenicij, A, *Acoust. Sci. & Tech.* 25, 2, (2004)] were used. Efficient numerical algorithm capable of massive parallelization was proposed for numerical integration of the model equations. Numerical results obtained using KZK equation and HOPE show that the KZK equation overestimates amplitude of the difference-frequency secondary wave in the near-field of the primary wave at the axis of symmetry. Both the equations provide the same results in the far-field and at the off-axis for both low- and high-frequency secondary fields.

1. Úvod

V současné době je z teorie nelineárních zvukových interakcí již velmi dobře znám jev tzv. parametrického pole (orig. parametric array), viz např. [1], [2], pomocí něhož lze generovat nízkofrekvenční, a přesto vysoce směrová zvuková pole.



Obrázek 1: Ilustrační obrázek

Jestliže jsou pomocí vhodného měniče v prostředí vygenerovány dvě směrová vysokofrekvenční zvukové vlny konečných amplitud a blízkých frekvencí (tzv. primární pole), nelineární interakce zapříčiní vznik dalších vysokofrekvenčních složek (vyšší harmonické, součtové složky atp.), ale i nízkofrekvenčních kmitočtových složek, zejména pak složky s rozdílovým kmitočtem. Všechny produkty nelineárních interakcí si do jisté míry zachovávají směrovost primárního pole a budeme je dále v textu nazývat sekundárním polem.

Nízkofrekvenční směrová zvuková pole nacházejí využití v akustických aplikacích, jako je například přenos akustického signálu vzduchem pouze k vybraným posluchačům, vytváření virtuálních (i pohybujících se) zdrojů zvuku a podobně, viz např. [3], [4].

K teoretickému studiu těchto zvukových polí byla dosud nejčastěji používána paraxiální KZK rovnice, viz např. [2], [5], zahrnující nelineární jevy, difrakci a tepelně-viskózní

útlum. Tato rovnice velmi dobře popisuje kvazirovinné zvukové vlny za splnění podmínky $ka \gg 1$, kde k je vlnové číslo a a je poloměr zdroje zvukových vln. I když je tato podmínka splněna pro primární zvukové pole a vysokofrekvenční produkty jeho nelineárních interakcí, nemusí být (a velmi často není) splněna pro nízkofrekvenční produkty jeho nelineárních interakcí, zejména pak pro rozdílovou kmitočtovou složku.

Tato práce je zaměřena na počítačové modelování nízkofrekvenčních sekundárních polí, pro něž není podmínka $ka \gg 1$ striktně splněna. Jako modelová rovnice je použita jednak KZK rovnice, jednak přesnější paraxiální rovnice vyššího řádu (HOPE – higher-order parabolic equation), viz [6], přičemž numerické výsledky jsou navzájem porovnány.

Jelikož je při numerických výpočtech v tomto případě nutné pokrýt široký rozsah kmitočtů s velkým rozlišením (jak vyšší harmonické primárního pole, tak rozdílové složky), bylo zapotřebí navrhnout efektivní výpočetní algoritmus umožňující provést danou analýzu v rozumném čase.

2. Modelové rovnice

Jako výchozí modelová rovnice byla použita HOPE rovnice, viz [6], ve tvaru

$$\frac{\partial^2 p'}{\partial z \partial \tau} - \frac{c_0}{2} \nabla_{\perp}^2 p' - \frac{c_0^3}{8} \mathcal{I}_{\tau}^2 (\nabla_{\perp}^4 p') = \frac{\delta}{2c_0^3} \frac{\partial^3 p'}{\partial \tau^3} + \frac{\beta}{2\rho_0 c_0^3} \frac{\partial^2 p'^2}{\partial \tau^2}, \quad (1)$$

kde p' je akustický tlak, c_0 je rychlost zvuku, ρ_0 je rovnovážná hustota prostředí, δ je koeficient difúze spojený s tepelně-viskózním útlumem, β je parametr nelinearity, z je prostorová souřadnice ve směru šíření zvukových vln, $\tau = t - z/c_0$ je retardovaný čas a ∇_{\perp}^2 je příčný laplacián (vyjádřený v rovině kolmé ke směru šíření vln). Člen

$\mathcal{I}^2_\tau(\nabla_\perp^4 p')$ je reprezentován integrodiferenciálním operátorem

$$\mathcal{I}^2_\tau(\nabla_\perp^4 p') = \int_{-\infty}^\tau \left\{ \int_{-\infty}^{\tau_1} [\nabla_\perp^4 p'(x, y, z, \tau_2)] d\tau_2 \right\} d\tau_1,$$

kde $\nabla_\perp^4 = \nabla_\perp^2 \nabla_\perp^2$ je příčný bilaplacián.

Jestliže je člen $\mathcal{I}^2_\tau(\nabla_\perp^4 p')$ vynechán, rovnice (1) se redukuje na KZK rovnici, viz např. [2], [5].

V práci [6] bylo ukázáno, jak teoreticky, tak i experimentálně, že rovnice (1) popisuje přesněji blízké pole i širší zvukové svazky než KZK rovnice. V závislosti na stupni aproximace může být rovnice (1) doplněna dalšími členy s vícenásobnými příčnými laplaciány, viz [6].

3. Numerický algoritmus

3.1. Normování a diskretizace

Za účelem numerické integrace rovnice (1) je s ohledem na minimalizaci zaokrouhlovacích chyb výhodné zavést bezrozměrné veličiny, definované například jako

$$P = \frac{p'}{A}, \quad R = \frac{r}{a}, \quad Z = \frac{z}{a}, \quad T = \omega\tau,$$

kde A je amplituda tlaku primární zvukové vlny v blízkosti zdroje, r je radiální souřadnice (pro redukcí výpočetní náročnosti je uvažována osová symetrie), a je poloměr zdroje zvuku (měniče) a ω je kruhový kmitočet nejnižší uvažované kmitočtové složky. Protože tato práce se zaměřuje na periodická zvuková pole, je výhodné akustický tlak reprezentovat komplexní Fourierovou řadou

$$P(Z, R, T) \approx \sum_{k=-N}^N P_k(R, Z) e^{ikT}, \quad (2)$$

kde N je uvažovaný počet harmonických složek.

Jak lze zřejmě očekávat, nízkofrekvenční sekundární pole má oproti poli primárnímu poněkud nižší směrovost. Z tohoto důvodu, aby bylo zamezeno nefyzikálním numerickým artefaktům souvisejícím s odrazem zvukových vln od hranice výpočetní oblasti, je výhodné zavést neortogonální transformaci souřadnic, viz např. [5], ve tvaru $R = R'(1 + \sigma Z)$, kde $\sigma > 0$, a formulovat akustický tlak pomocí souřadnic Z, R' . Výpočetní oblast může být následně diskretizována jako

$$\begin{aligned} Z &\rightarrow Z_m = m\Delta Z, & m &= 0, 1, 2, \dots, M_Z, \\ R' &\rightarrow R'_n = n\Delta R', & n &= 0, 1, 2, \dots, M_R, \end{aligned}$$

takže akustický tlak je počítán v jednotlivých bodech výpočetní mříže $P_k(Z, R') \rightarrow P_k(m\Delta Z, n\Delta R') = P_k^{m,n}$.

3.2. Dělení operátoru

Ukazuje se, že výpočetní náročnost algoritmu může být zásadním způsobem redukována při použití techniky tzv. dělení operátoru (orig. operator-splitting), viz např. [7].

Modelová rovnice (1) může být po provedení normalizace a reprezentace Fourierovou řadou zapsána ve tvaru

$$\frac{\partial P_k}{\partial Z} = \mathcal{L}_{\text{dif}} P_k + \mathcal{N}_{\text{na}}(1, \dots, P_N), \quad k = 1, \dots, N, \quad (3)$$

kde lineární operátor \mathcal{L}_{dif} reprezentuje pouze difrakci a nelineární operátor \mathcal{N}_{na} reprezentuje pouze nelineární jevy a disipaci. V rámci metody dělení operátoru je každá část rovnice (3) integrována podél osy Z v každém kroku algoritmu nezávisle jedna po druhé, viz [7].

3.3. Difrakční operátor

Příslušnou část rovnice (3) lze zapsat ve tvaru

$$\begin{aligned} \frac{\partial P_k}{\partial Z} &= \frac{\sigma R'}{1 + \sigma Z} \frac{\partial P_k}{\partial R'} - \\ &- \frac{iD_0}{k(1 + \sigma Z)^2} \left(\frac{\partial^2 P_k}{\partial R'^2} + \frac{1}{R'} \frac{\partial P_k}{\partial R'} \right) + \\ &+ \frac{iD_0^3}{k^3(1 + \sigma Z)^4} \left(\frac{\partial^4 P_k}{\partial R'^4} + \frac{2}{R'} \frac{\partial^3 P_k}{\partial R'^3} - \right. \\ &\left. - \frac{1}{R'^2} \frac{\partial^2 P_k}{\partial R'^2} + \frac{1}{R'^3} \frac{\partial P_k}{\partial R'} \right), \quad (4) \end{aligned}$$

kde $D_0 = c_0/2a\omega$. Rovnice (4) je diskretizována pomocí centrálních konečných diferencí, viz [8], Crankovým-Nicolsonovým schématem, viz např. [7], [8]. Výsledné rovnice pro výpočet kroku $m+1$ za použití kroku m lze zapsat ve tvaru

$$\begin{aligned} a_{m+1,n} P_k^{m+1,n-2} + b_{m+1,n} P_k^{m+1,n-1} + c_{m+1,n} P_k^{m+1,n} + \\ + d_{m+1,n} P_k^{m+1,n+1} + e_{m+1,n} P_k^{m+1,n+2} = \\ = -a_{m,n} P_k^{m,n-2} - b_{m,n} P_k^{m,n-1} + f_{m,n} P_k^{m,n} - \\ - d_{m,n} P_k^{m,n+1} - e_{m,n} P_k^{m,n+2}, \quad (5) \end{aligned}$$

kde

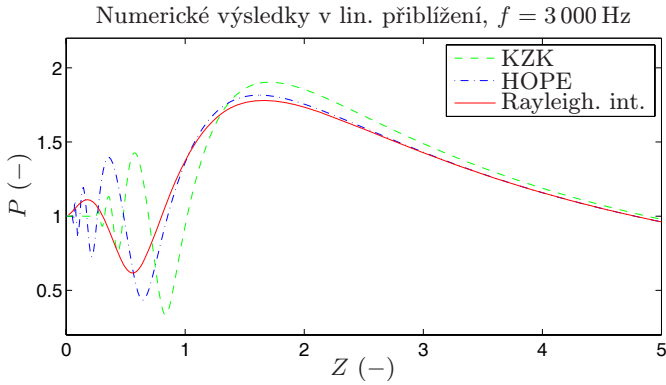
$$\begin{aligned} a_{m,n} &= -n^2(2n-1)\epsilon_m''', \\ b_{m,n} &= n\epsilon_m' + n^2(2n-1)\epsilon_m'' + (8n^3 - 2n^2 + 2n + 1)\epsilon_m''', \\ c_{m,n} &= -4n[n^2\epsilon_m'' + (3n^2 + 1)\epsilon_m'''] + 1, \\ d_{m,n} &= -n\epsilon_m' + n^2(2n+1)\epsilon_m'' + (8n^3 + 2n^2 + 2n - 1)\epsilon_m''', \\ e_{m,n} &= -n^2(2n+1)\epsilon_m'', \\ f_{m,n} &= 4n[n^2\epsilon_m'' + (3n^2 + 1)\epsilon_m'''] + 1 \end{aligned}$$

a

$$\begin{aligned} \epsilon_m' &= \frac{\sigma\Delta Z}{4(1 + \sigma Z_m)}, & \epsilon_m'' &= \frac{iD_0\Delta Z}{4kn^3(1 + \sigma Z_m)^2(\Delta R')^2}, \\ \epsilon_m''' &= \frac{iD_0^3\Delta Z}{4k^3n^3(1 + \sigma Z_m)^4(\Delta R')^4}. \end{aligned}$$

Soustava (5) může být rozdělena na N navzájem nezávislých soustav o M_R lineárních rovnicích¹ takže tyto

¹Jedná se pouze o M_R rovnic, neboť na ose symetrie ($R = 0$) může být využita okrajová podmínka $\partial P_k/\partial R = 0$ pro výpočet P_k . Dále se předpokládá, že akustický tlak vně integrační oblasti je nulový.



Obrázek 2: Amplituda normalizovaného akustického tlaku v ose symetrie, $f = 3000$ Hz, $ka \approx 11$, porovnání numerických výsledků získaných řešením Rayleighova integrálu, KZK a HOPE ($A_0 = N_0 = 0$)

rovnice mohou být řešeny nezávisle a paralelně. Matice jednotlivých soustav jsou pásmově diagonální, v případě HOPE pětidiagonální a v případě KZK třídiagonální ($\epsilon_m''' = 0$ a proto $a_{m,n} = e_{m,n} = 0$). Jelikož koeficienty ϵ_m' , ϵ_m'' , ϵ_m''' závisí na souřadnici Z , je třeba v každém kroku numerické integrace počítat inverzní matice soustavy.

3.4. Nelineární a absorpční operátor

Příslušná část rovnice (3) má tvar

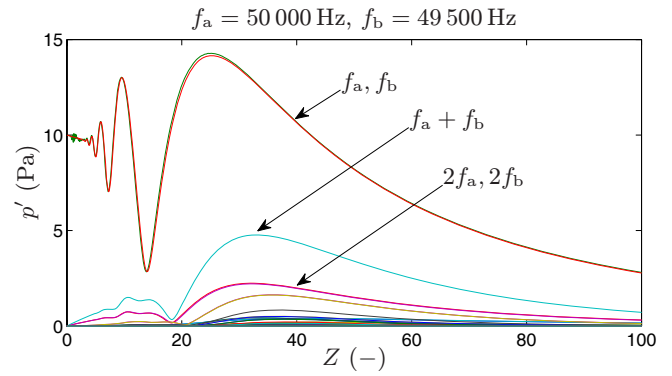
$$\frac{dP_k^n}{dZ} = -k^2 A_0 P_k^n + ikN_0 \sum_{l=-N+k}^N P_l^n P_{k-l}^n, \quad (6)$$

kde $n = 1, 2, \dots, M_R$, $N_0 = \beta\omega a A / 2\rho_0 c_0^3$ a $A_0 = \delta a \omega^2 / 2c_0^3$. Soustava (6) může být rozdělena na M_R navzájem nezávislých soustav N nelineárních obyčejných diferenciálních rovnic, takže tyto soustavy mohou být integrovány paralelně. Pro přechod od kroku m ke kroku $m+1$ je použita Rungova-Kuttova metoda třetího řádu.

3.5. Implementace numerického algoritmu

Výše popsáný algoritmus byl implementován v jazyce C++. K řešení soustav lineárních rovnic (5) byla využita knihovna NR 3.0 (Numerical Recipes), konkrétně algoritmy pro řešení soustav s pásmově-diagonálními maticemi. Jelikož soustavy rovnic pro jednotlivé harmonické (difrakční část rovnice) jsou navzájem nezávislé, bylo možné optimalizovat příslušnou část algoritmu pro paralelní běh na více procesorových jádrech s využitím knihovny OpenMP.

Rungův-Kuttův algoritmus pro integraci nelineární části rovnice (6) bylo možné přímočaře implementovat pro masivně paralelní GPGPU platformu s využitím Nvidia CUDA C, jelikož rovnice pro jednotlivá n jsou navzájem nezávislé a týž jednoduchý algoritmus je prováděn stejně, pouze na různých datech.



Obrázek 3: Amplituda akustického tlaku jednotlivých harmonických podél osy symetrie, $f_a = 50000$ Hz, $f_b = 49500$ Hz, $A_a = A_b = 10$ Pa

Masivní paralelizace algoritmu pro nelineární část modelové rovnice urychlila numerické výpočty cca $30\times$ ($N = 640$, $M_R = 4000$, procesor Intel Core i7 (čtyři jádra), 3 GHz, grafická karta Nvidia GeForce GTX 580 – 512 stream procesorů).

4. Numerické výsledky

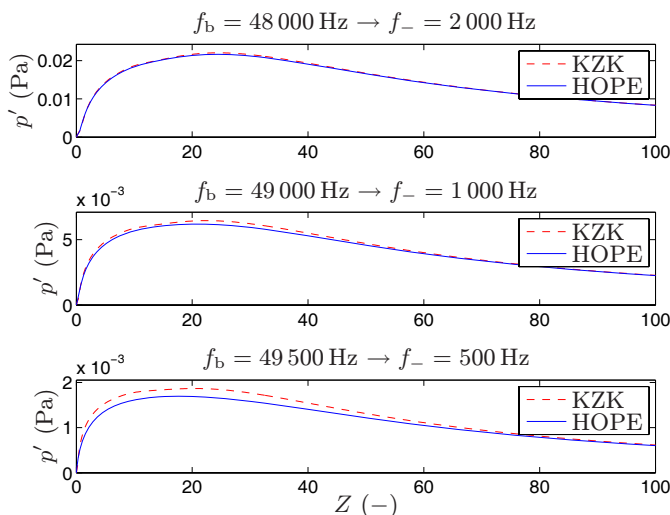
V tomto odstavci jsou prezentovány některé výsledky získané numerickým řešením rovnice (1) pomocí výše popsaného algoritmu, ve kterém byly nastaveny tyto parametry: $N = 640$, $M_R = 4000$, $\Delta Z = \Delta R' = 10^{-2}$, $\sigma = 0,09$, což znamená, že pro $Z = 0$ je $R_{\max} = 40$ a např. pro $Z = 100$ je $R_{\max} = 400$. Takto rychlé rozšiřování integrační oblasti je nutné kvůli zamezení nežádoucích odrazů nízkofrekvenčních složek od její hranice. Konvergence algoritmu pro dané parametry byla otestována porovnáním numerických výsledků (pro primární pole, druhou harmonickou a rozdílovou složku) s výsledky získanými v kvazilineárním přiblížení založeném na metodě rozvoje zvukového pole do Gaussových svazků, viz např. [9].

V modelu je zvukové pole buzeno ve vzduchu za normálních podmínek. Akustický tlak v rovině zdroje $z = 0$ je popsán pomocí vztahu

$$p'(r, t) = e^{-(r/a)^{16}} (A_a \cos 2\pi f_a t + A_b \cos 2\pi f_b t),$$

kde A_a a A_b jsou amplitudy dvou primárních vln o kmitočtech f_a a f_b , $a = 0,2$ m.

Pro posouzení validity algoritmu a vyšší přesnosti HOPE oproti KZK rovnici v blízkém poli jsou na obrázku 2 porovnány numerické výsledky v lineárním a bezdisipativním přiblížení ($A_0 = N_0 = 0$) s výsledky získanými výpočtem Rayleighova integrálu pro případ $A_b = 0$, $f_a = 3000$ Hz ($ka \approx 11$). Z obrázku je patrné, že výsledky získané pomocí HOPE korespondují s výsledky Rayleighova integrálu mnohem lépe než v případě KZK rovnice. Ve vzdáleném poli (v tomto případě pro $Z > 5$) jsou všechny výsledky shodné.



Obrázek 4: Amplituda tlaku rozdílové kmitočtové složky podél osy symetrie pro různé kmitočty f_b , $f_a = 50\,000$ Hz, $A_a = A_b = 10$ Pa

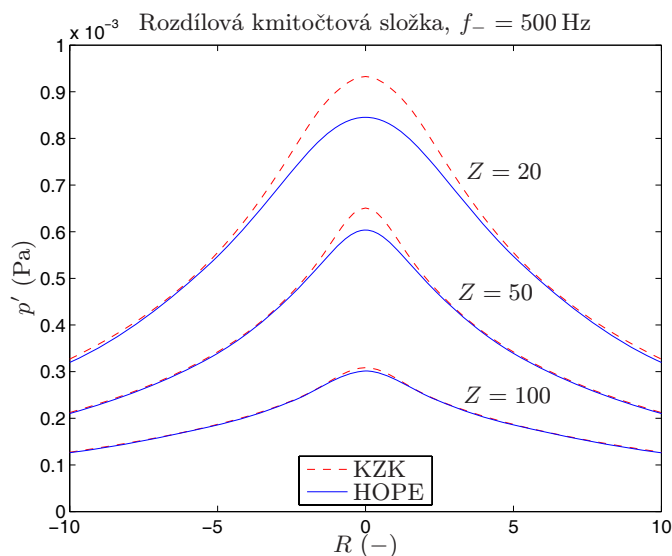
Pro posouzení chování nízkofrekvenčního sekundárního pole byl uvažován dvoufrekvencní zdroj primárního zvukového pole s kmitočty $f_a = 50\,000$ Hz a $f_b = 48\,000$ Hz, $49\,000$ Hz a $49\,500$ Hz. Ve všech případech $A_a = A_b = 10$ Pa. Na obrázku 3 je zachyceno amplitudové spektrum akustického tlaku podél osy symetrie. Je patrné, že díky nelineárním interakcím jsou vybudeny vyšší harmonické složky. Jelikož pro primární pole a vyšší kmitočtové složky pole sekundárního v tomto případě platí $ka > 180$, KZK rovnice a HOPE v tomto případě poskytují stejné výsledky.

Odlišná situace však nastává pro nízkofrekvenční složky sekundárního pole (které nejsou díky nižší amplitudě na obrázku 3 patrné). Obrázek 4 zachycuje rozložení amplitudy akustického tlaku rozdílové složky o kmitočtu $f_- = f_a - f_b$ pro případ různých kmitočtů f_b , vypočtené pomocí KZK rovnice a HOPE. Z obrázku je patrné, že KZK rovnice předpovídá oproti (přesnější) HOPE poněkud vyšší hodnoty, tento rozdíl, jak lze předpokládat, klesá se vzrůstajícím kmitočtem nízkofrekvenčního pole. Rozdíl rovněž mizí ve vzdáleném poli (v tomto případě pro $Z > 90$).

Obrázek 5 zachycuje rozložení amplitudy rozdílové kmitočtové složky ve směru kolmém k ose symetrie v různých vzdálenostech od zdrojové roviny pro kmitočtet $f_- = 500$ Hz. Z obrázku je patrné, že rozdíl mezi výsledky předpovězenými KZK rovnicí a HOPE klesá se vzdáleností od osy symetrie.

5. Závěr

V rámci této studie byl navržen a implementován masivně paralelní algoritmus pro numerickou integraci KZK rovnice a HOPE v kmitočtové oblasti. S jeho využitím bylo zkoumáno chování rozdílové kmitočtové složky sekundár-



Obrázek 5: Amplituda tlaku rozdílové kmitočtové složky v závislosti na vzdálenosti od osy symetrie, $f_a = 50\,000$ Hz, $f_b = 49\,500$ Hz, $A_a = A_b = 10$ Pa

ního zvukového pole, pro niž není splněna podmínka $ka \gg 1$. Numerické výsledky ukazují, že KZK rovnice předpovídá vyšší hodnoty tlaku nízkofrekvenčních kmitočtových složek oproti HOPE. Tyto rozdíly se projevují pouze v blízkém poli primární vlny, klesají se vzdáleností od osy symetrie a zanikají ve vzdáleném poli primární vlny.

Poděkování

Tato práce byla podpořena grantem SGS 10/265/OHK3/3T/13.

Reference

- [1] P. J. Westervelt: Parametric Acoustic Array, *J. Acoust. Soc. Am.*, 35(4), p. 535–537, 1963.
- [2] M. F. Hamilton and D. T. Blackstock: *Nonlinear Acoustics*, Academic Press, 1997.
- [3] M. Yoneyama, J. Fujimoto, Y. Kawamo and S. Sasabe: The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design, *J. Acoust. Soc. Am.*, 73, p. 1532–1536, 1983.
- [4] F. J. Pompei: The use of airborne ultrasonics for generating audible sound beams, *J. Audio Eng. Soc.*, 47, p. 726–731, 1999.
- [5] N. S. Bakhvalov, Y. M. Zhileikin and E. A. Zabolot'skaya: *Nonlinear Theory of Sound Beams*, American Institute of Physics, 1987.

- [6] T. Kamakura, M. Akiyama and K. Aoki: A higher-order parabolic equation for describing nonlinear propagation of ultrasound beams, *Acoust. Sci. & Tech.*, 25(2), p. 163–165, 2004.
- [7] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery: *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge University Press, 2007.
- [8] W. F. Ames: *Numerical Methods for Partial Differential Equations*, 3rd Edition, Academic Press, 1992.
- [9] D. Ding: A simplified algorithm for the second-order sound fields, *J. Acoust. Soc. Am.*, 108, p. 2759–2764, 2000.

Porovnání dvou spektrálních metod pro analýzu akustických signálů

Václav Turoň^a, Ján Janík^a, Radim Špetík^a, Pavel Sovka^a a Miroslav Vlček^b

^aČeské vysoké učení technické v Praze, Fakulta elektrotechnická, Technická 2, 166 27 Praha 6,

^bČeské vysoké učení technické v Praze, Fakulta dopravní, Konviktská 20, 110 00 Praha 1

Spectral analysis of acoustic signals is a complex discipline of the digital signal processing. Several methods are used for these purposes and the results are influenced by different factors. In this paper we focus on the comparison of two methods. The first one was designed for audio signal analysis. The method utilizes the Discrete Fourier Transform with varying window length which is based on minimal energy of spectral leakage. The second method is a new Approximated Discrete Zolotarev Transform (ADZT) which seems to give good results in non-stationary signal analysis.

1. Úvod

Krátkodobá Fourierova transformace (STFT) [1] patří mezi nejpoužívanější metody spektrální analýzy signálů. Hlavní předností této transformace je snadná reprezentace výsledků ve formě spektrogramu. Časové a frekvenční rozlišení je ovlivněno několika parametry, z nichž nejvýznamnější je délka okna. Tato vlastnost je spojena s Heisenbergovým-Gaborovým principem nejistoty, který říká, že nelze získat maximální frekvenční rozlišení bez ztráty časového rozlišení a obráceně. Proto musí být zvolen kompromis mezi časovým a frekvenčním rozlišením pro daný typ analyzovaného signálu. Z tohoto důvodu byly odvozeny metody, které adaptivně nastavují délku okna podle zpracovávaného signálu a tím optimalizují časové a frekvenční rozlišení STFT, např. [2], [3]. Mezi další metody, které přistupují k transformaci signálu z časové do frekvenční oblasti komplexněji, patří například Hilbertova-Huangova transformace, jež rozkládá signál do tzv. vnitřních funkcí [4]. Díky tomuto rozkladu lze komplexní obálku signálu zobrazit pomocí Hilbertova spektra přesněji než při použití Hilbertovy transformace přímo na signál [1], [5]. Její nevýhodou je zcela odlišná interpretace výsledků než ta, kterou používáme u metod založených na Fourierově transformaci. Proto metody s adaptivním nastavováním délky okna umožňující interpretaci výsledků ve Fourierově smyslu jsou pro zpracování akustických signálů vhodnější. V tomto příspěvku je popsána jedna z metod STFT s proměnnou délkou okna [2] a porovnána s metodou spektrální analýzy využívající Zolotarevovy polynomy (ADZT) [6]. Obě metody umožňují fourierovskou interpretaci výsledků, a proto je lze snadno porovnat.

Tento článek porovnává dvě spektrální metody zaměřené na analýzu signálů. Jsou rovněž uvedeny dva příklady analýzy audio signálů (hudba, rec). Metoda MESP byla speciálně k tomuto účelu navržena, metoda ADZT se rovněž jeví jako dostatečně silný nástroj k tomuto účelu. Srovnání s dalšími metodami je možné najít v citované literatuře.

2. Adaptivní metody spektrální analýzy

2.1. Metoda minimální energie spektrálního prosakování

Během posledních let bylo odvozeno několik adaptivních metod, které využívají různá kritéria pro volbu optimální délky segmentů (oken) STFT. Jednou z publikovaných metod vhodných pro analýzu audio signálů je metoda minimální energie spektrálního prosakování (MESP) [2]. Tato metoda využívající kritéria MDL [7] (metoda pro stanovení minimální délky dat s maximálním zachováním informace) je založena na opakovaném výpočtu DFT s rostoucí délkou okna. Pomocí vhodného kritéria, v tomto případě je použito kritérium minimální energie spektrálního prosakování, je vybrána optimální délka okna, která poskytne výsledné spektrum použité ve spektrogramu. Prosakování vzniká porušením ortogonalitě mezi analyzovaným signálem a bázovou funkcí transformace (komplexní exponenciála) a jeho velikost závisí na vzájemném poměru délky okna a velikostí period složek, ze kterých se signál skládá. Vlastní algoritmus výpočtu optimální délky okna navržený v článku [2] lze rozdělit do několika kroků, které byly pro naši implementaci realizovány takto:

1. Výpočet spektrogramů S_r s rozdílnou délkou okna $N_r = 2^c$, které je centrované kolem aktuálního vzorku a s rostoucí délkou se symetricky rozšiřuje do stran. Parametr c je přirozené číslo z intervalu $\langle c_{\min}, c_{\max} \rangle$. Tím získáme sadu spekter s různým časově-frekvenčním rozlišením, z nichž to, které minimalizuje energii prosakování, je vybráno do výsledného spektrogramu.
2. Výpočet energie spektrálního prosakování je realizován podle vztahu [2]

$$L_r(k) = \frac{\sum_{i=1}^{N_r} N_r i |a_{i,r}|^2}{\sqrt{\sum_{i=1}^{N_r} N_r |a_{i,r}|^2 + \epsilon}}, \quad (1)$$

kde k je index segmentu r -tého spektrogramu s příslušnou délkou okna N_r a $a_{i,r}$ označuje i -tý spektrální

koeficient k -tého segmentu. Koeficienty $a_{i,r}$ musí být seřazeny sestupně podle velikosti. Výraz (1) představuje první moment distribuce proměnných $|a_{i,r}|^2$ (čitatel) normovaný celkovou energií signálu (jmenovatel). Volba $\epsilon > 0$ ve jmenovateli ošetřuje případné dělení nulou.

- Optimální délka okna STFT je rovna délce okna spektrogramu s nejmenší energií spektrálního prosakování L_r .
- Spektrogram s optimálním časově-frekvenčním rozlišením je složen z jednotlivých spekter realizovaných spektrogramů $S_r(k)$.

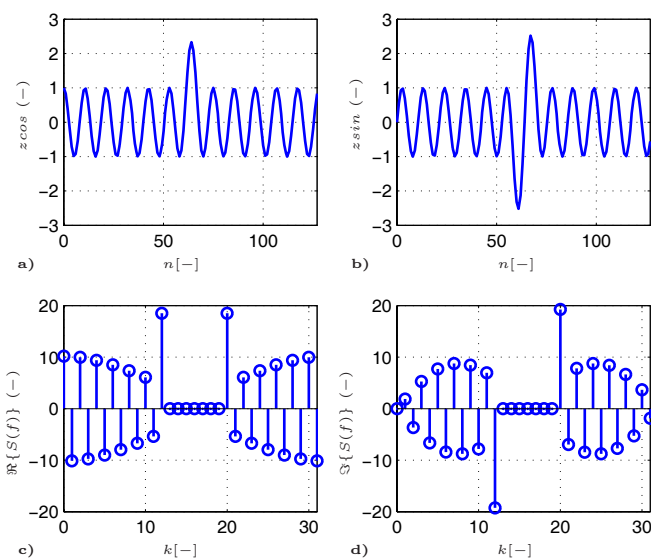
Pozn.: Původní článek [2] nediskutoval vliv použitého tvaru okna na velikost spektrálního prosakování. Z provedených experimentů jsme se rozhodli použít při implementaci algoritmu Hammingovo okno, které dosahovalo nejlepších výsledků. Jednotlivá spektra S_r (použitá pro výpočet optimální délky okna v 1. a 2. bodě výše popísaného algoritmu) jsou tedy realizována Hammingovým oknem a bez doplnění nulovými vzorky. Pro sestavení finálního spektrogramu byla použita interpolovaná spektra o maximálním počtu čar 2_{\max}^c . Interpolace spolu s posunem okna o 1 vzorek umožňuje sestavený spektrogram porovnat s výsledkem ADZT a sledovat chování obou metod s vyloučením vlivu decimace (překryv oken menší než $N - 1$ vzorků), která je běžně při výpočtu STFT používána.

2.2. Aproximovaná diskretní Zolotarevova transformace

Diskretní Zolotarevova transformace (DZT) byla navržena R. Špetíkem v roce 2009 [8]. DZT je časově-spektrální transformace Fourierova typu. Její báze je tvořena selektivním sinem (z_{sin}) a kosinem (z_{cos}), které jsou řešením první a druhé Zolotarevovy aproximační úlohy [9]. Selektivní siny a kosiny se vyznačují schopností měnit velikost centrálních laloků. Příklad průběhu těchto funkcí spolu s jejich spektrem je ilustrován na obrázku 1. Na obrázku jsou patrné laloky (převýšení) u obou funkcí, jejichž velikost odpovídá nestacionaritě signálu. Dále je ze spodních obrázků zřejmé, že každé spektrum selektivního sinu nebo kosinu lze rozdělit na stacionární a nestacionární část. Stacionární část je tvořena čarou s nejvyšší frekvencí, nestacionární část tvoří všechny čáry s nižšími frekvencemi. Pokud je signál stacionární, lalok zdegeneruje a Zolotarevovy polynomy přejdou na Čebyševovy polynomy (kosinus a sinus). Ve spektru se v tomto případě objeví pouze stacionární část a zmizí celá nestacionární část.

Lze ukázat [8], [9], že Zolotarevovy funkce jsou tvořeny váženým součtem sinů (kosinů) nižších řádů. Z toho vyplývá, že DZT lze chápat jako reorganizaci, resp. filtraci spektra diskretní Fourierovy transformace (DFT) podle vztahu [6], [8], [10]

$$\mathbf{S}_Z = \mathbf{Z} \cdot \mathbf{S}, \quad (2)$$



Obrázek 1: a) Selektivní kosinus z_{cos} řádu 12. b) Selektivní sinus řádu 12. c) Reálná část DFT spektra z_{cos} . d) Imaginární část DFT spektra z_{sin}

kde \mathbf{S} je sloupcová matice DFT spektra, \mathbf{S}_Z je sloupcová matice DZT spektra a \mathbf{Z} je maticí váhovacích koeficientů DZT transformace. Tím se DZT výrazně liší od všech zmíněných metod adaptivní spektrální analýzy. Použití Čebyševových polynomů (sinů a kosinů) jako bazových funkcí DFT vede k nutnosti opakovaného výpočtu DFT spekter pro různé délky oken a následného výběru spektra s optimálními vlastnostmi. Naproti tomu DZT využívá obecnější bazové funkce (Zolotarevovy polynomy), jejichž parametry jsou optimálně určeny z analyzovaného segmentu signálu bez nutnosti opakovaného výpočtu spektra. Jedním z důležitých parametrů je výška laloku odpovídající selektivitě Zolotarevova polynomu. Jelikož nastavování selektivity podle aktuální nestacionarity signálu je výpočetně náročný proces, R. Špetík navrhl efektivní algoritmus pro výpočet DZT označovaný jako aproximovaná DZT (ADZT) [6]. Tento algoritmus vytváří matici \mathbf{Z} analýzou DFT spektra \mathbf{S} podle následujících kroků¹ (detaily viz [10]):

- Výpočet DFT spektrálních koeficientů \mathbf{S} analyzovaného signálu pomocí rychlé Fourierovy transformace (FFT).
- Separace stacionární $\mathcal{S}(k)$ a nestacionární části $\mathcal{N}(k)$ aktuálního spektrálního koeficientu $\mathbf{S}(k)$.
- Výpočet vážených faktorů a koeficientů tak, aby modul S_Z byl minimální.
- Sestavení regulární matice \mathbf{Z} .

¹Implementace metody je dostupná na internetových stránkách http://amber.feld.cvut.cz/selective_transforms.

Metoda ADZT nastavuje bazové funkce podle aktuálních vlastností analyzovaného signálu, čímž se spíše podobá metodě analýzy hlavních komponent (PCA) [1], jejíž bazové funkce jsou signálově závislé, než adaptivním metodám STFT s bazovými funkcemi tvořenými siny a kosiny. Adaptivní STFT nemění tvar bazových funkcí, zatímco ADZT a PCA ano. Pro dosažení nejlepších výsledků používáme pro DZT pravoúhlé okno. Použitím jiného okna vkládáme do metody cizorodý prvek, který degraduje výsledky analýzy [6].

3. Ukázky časově-spektrální analýzy

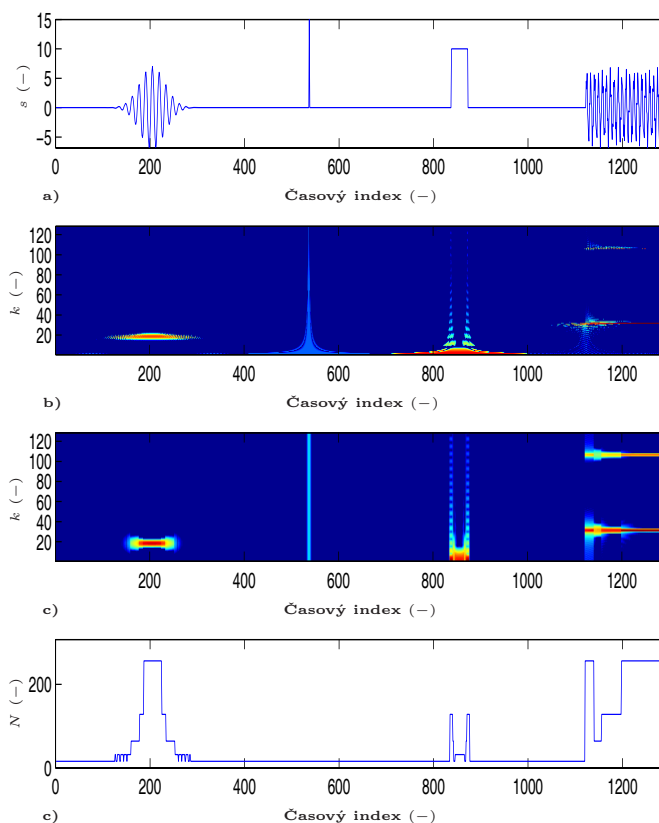
V této části článku se věnujeme porovnání výsledků obou výše zmíněných metod aplikovaných jak na umělém signálu, tak i na reálných akustických signálech.

3.1. Umělý signál

Prvním analyzovaným signálem je umělý signál s typickými příklady nestacionárního chování: gausovsky modulovaný sinusový impuls (S1), jednotkový impuls (S2), obdélníkový puls (S3) a součet dvou harmonických složek $\sin(2\pi nk_1/N)$ a $\sin(2\pi nk_2/N)$ (S4), s parametry $k_1 = 30,72$ a $k_2 = 104,96$. Pro analýzu signálu jsme použili pravoúhlé okno s posunem jedna. Optimální délka okna STFT byla nastavována algoritmem MESP v rozmezí 16 až 256 vzorků. Bylo použito Hammingovo okno s posunem o jeden vzorek. Pro ADZT spektrogram (zologram) bylo použito pravoúhlé okno délky 256 s posunem rovněž jedna.

Výsledky analýzy jsou zobrazeny na obrázku 2. Přesnost časové detekce na vyšších frekvencích je srovnatelná u obou metod. ADZT vykazuje horší časové rozlišení než MESP pro nízké frekvence (viz detekce signálů S2 a S3 a počátek signálu S4). Narazí-li MESP metoda na rychlý přechod (silnou nestacionaritu), vybere krátké okno, které ve spektrogramu implikuje konstantní časové rozlišení nezávislé na frekvenci. Naproti tomu ADZT používá okno konstantní délky a pro výpočet čar s nízkým indexem Zolotarevovy polynomy nízkých řádů, které neumožňují výraznou reorganizaci DFT spektra. Proto první dvě čáry ADZT jsou rovny prvním dvěma čarám DFT spektra [6]. Čáry s vyšším indexem jsou získány reorganizací DFT spektra, nicméně ta stále není dostatečná pro získání vysokého časového rozlišení.

Frekvenční rozlišení je znatelně lepší u ADZT, zejména u signálu S1 a S4. Vysoké frekvenční rozlišení ADZT je způsobené použitím dlouhým oknem, pro které jsou nalezeny optimální parametry Zolotarevových polynomů vedoucí k efektivní reorganizaci DFT spektra. MESP může poskytnout dobré frekvenční rozlišení pouze v úsecích, kde je signál stacionární, a proto může dojít k výběru delšího okna (viz střední část signálu S4). Optimální délka okna vybraná metodou MESP pro jednotlivé segmenty signálu je vynesena v obrázku 2d), který ilustruje princip této metody.



Obrázek 2: a) Umělý analyzovaný signál. S1 – gausovsky modulovaný sinusový puls, S2 – jednotkový impuls, S3 – obdélníkový puls a S4 – součet dvou harmonických složek. b) Absolutní hodnota STFT spektrogramu s dynamicky nastavovaným optimálním oknem. c) Absolutní hodnota ADZT zologramu. d) Optimální délka okna N vybraná metodou MESP

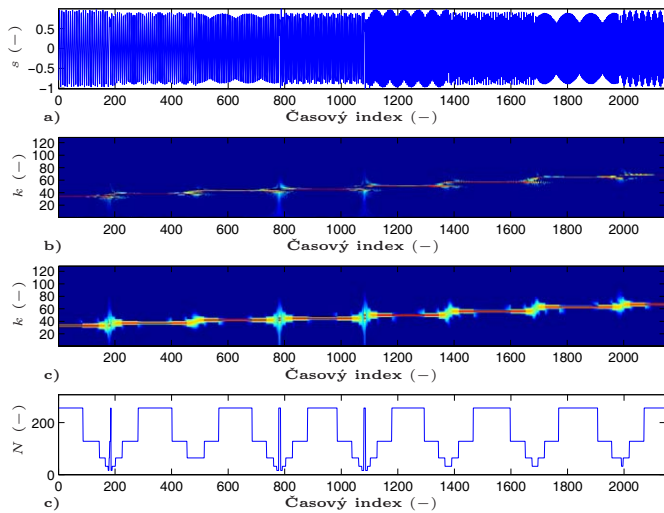
3.2. Hudební signál

Dalším analyzovaným signálem je reálná nahrávka hudební stupnice klavíru. Spektrální analýza byla provedena s pravoúhlým (ADZT) a Hammingovým (MESP) oknem s posunem o jeden vzorek pro obě metody. Délka okna pro ADZT byla nastavena pevně na 256 a MESP volil optimální délku okna v rozsahu 16 až 256.

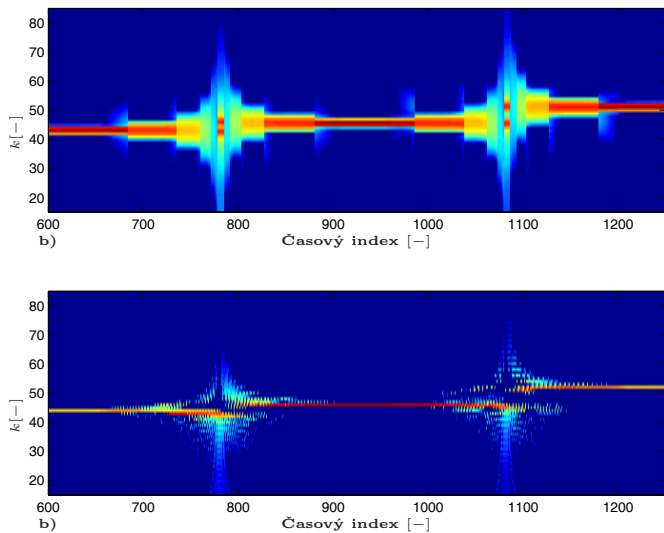
Z výsledků na obrázku 3 vidíme, že oba algoritmy detekují přechody mezi tóny s téměř stejnou přesností. Zásadní rozdíl je však ve frekvenčním rozlišení, kde ADZT dosahuje znatelně lepších výsledků. To je dobře vidět zejména na detailu dvou frekvenčních přechodů na obrázku 4.

3.3. Řečový signál

Posledním z analyzovaných signálů je řečový signál slova „osm“. Nastavení parametrů obou analýz je shodné s nastavením pro analýzu hudebního signálu z předchozího příkladu.

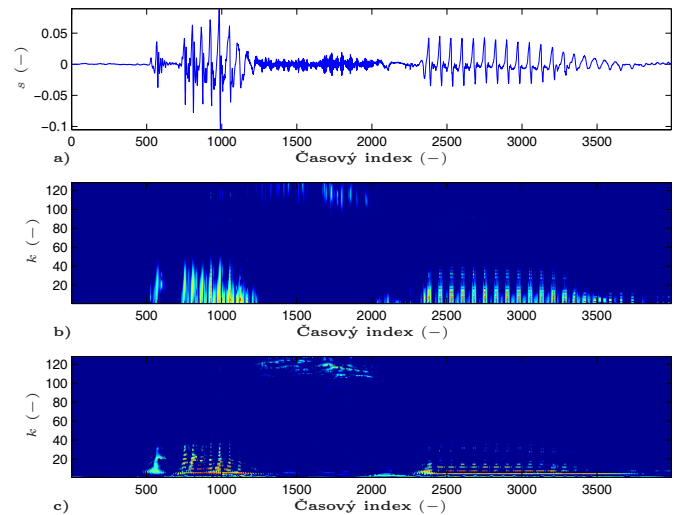


Obrázek 3: a) Hudební signál ($f_s = 2$ kHz). b) Absolutní hodnota ADZT zologramu. c) Absolutní hodnota STFT spektrogramu s dynamicky nastavovaným optimálním oknem. d) Optimální délka okna N vybraná metodou MESP

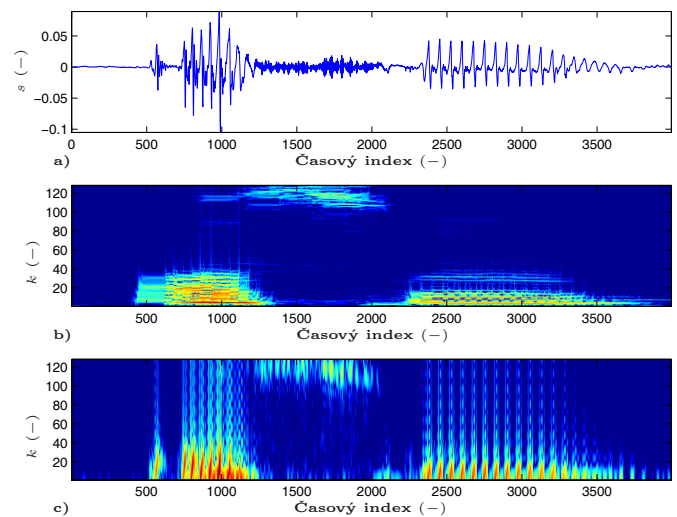


Obrázek 4: Detail obrázku 3 a) Absolutní hodnota STFT spektrogramu s dynamicky nastavovaným optimálním oknem. b) Absolutní hodnota ADZT zologramu

Výsledky jsou zobrazeny na obrázku 5. V obou případech lze pozorovat detekci glotálních rázů ve znělých hláskách (viz indexy 700–1200 a 2400–3500). ADZT opět dosahuje lepšího frekvenčního rozlišení, proto jsou patrné jednotlivé formantové kmitočty. Pro srovnání s klasickými metodami spektrální analýzy uvádíme na obrázku 6 širokopásmový a úzkopásmový spektrogram. Tyto spektrogramy byly realizovány použitím Hammingova okna o délce 16 a 256 vzorků, což představuje nejkratší a nejdelší možné okno spektrogramu s optimální délkou okna MESP (viz obrázek 5).



Obrázek 5: a) Řečový signál ($f_s = 8$ kHz). b) Absolutní hodnota STFT spektrogramu s dynamicky nastavovaným optimálním oknem. c) Absolutní hodnota ADZT zologramu



Obrázek 6: a) Řečový signál ($f_s = 8$ kHz). b) Absolutní hodnota úzkopásmového STFT spektrogramu. c) Absolutní hodnota širokopásmového STFT spektrogramu

4. Závěr

V článku byla představena spektrální analýza nestacionárních signálů pomocí dvou adaptivních metod s odlišným přístupem. První metoda využívá výpočet DFT spektra pro rozdílné délky oken a vybírá pomocí kritéria minimální energie spektrálního prosakování optimální délku okna. Druhá metoda je založená na bázi tvořené Zolotarevovými polynomy prvního a druhého druhu, které dynamicky mění svou selektivitu podle analyzovaného signálu při zachování délky okna.

Z uvedených výsledků lze dovést, že nová selektivní transformace ADZT dosahuje porovnatelných vlastností s adaptivní metodou spektrální analýzy MESP. Důsledkem použití nových typů polynomů umožňujících použití dlouhého okna konstantní délky nabízí ADZT podstatně lepší frekvenční rozlišení než MESP při zachování dobrého časového rozlišení, zejména pro vyšší frekvence. Časové rozlišení ADZT na nízkých frekvencích nemá takovou přesnost jako STFT s adaptivní volbou délky okna. Nicméně ADZT, na rozdíl od adaptivních metod spektrální analýzy, vychází ze zcela nového a obecnějšího konceptu analýzy nestacionárních signálů, který je v současné době podrobně analyzován a dále rozvíjen.

Poděkování

Tento výzkum byl podporován granty GAČR P102/11/1795 *Nové selektivní transformace pro číslicové zpracování nestacionárních signálů*, GAČR 102/08/H008 *Analýza a modelování biomedicínských a řečových signálů* a grantem SGS 10/181/OHK3/2T/13 *Spektrální vlastnosti Zolotarevovy transformace*. Dále bychom rádi poděkovali doc. Ing. Romanovi Čmejlovi, CSc., za poskytnutí reálných signálů k analýze.

Reference

- [1] Oppenheim, A. V., Schaffer, R. W., Buck, J. R.: *Discrete-Time Signal Processing* (2nd Edition), Prentice-Hall Signal Processing Series, Jan 10, 1999.
- [2] Lukin, A., Todd, J.: Adaptive Time-Frequency Resolution for Analysis and Processing of Audio, *Audio Engineering Society 120th Convention*, Paris, May 2006.
- [3] Katkovnik, V., Stankovic, L: Periodogram with Varying and Data-Driven Window Length, *Signal Processing*, vol. 67, no. 3, June 1998, p. 345–358.
- [4] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. London, Ser. A*, 454, p.903–995, 1998.
- [5] Jolliffe, I. T.: *Principal Component Analysis*, Springer, New York, 2002.
- [6] Janík, J., Turoň, V., Sovka, P., Špetík, R., Vlček, M.: A way to a new multi-spectral transform, *11th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision*, Florence, 2011.
- [7] Grunwald, P: A Tutorial Introduction to the Minimum Description Length Principle, *Advances in Minimum Description Length: Theory and Applications*, MIT Press, April 2005.
- [8] Špetík, R.: The Discrete Zolotarev Transform, *Doctoral thesis*, CTU in Prague, FEE, Prague, February 2009.
- [9] Vlček, M., Unbehauen, R.: Zolotarev Polynomials and Optimal FIR Filters, *IEEE Transaction on Signal Processing*, vol. 47, no. 3, March 1999, p. 717–730.
- [10] Turoň, V., Janík, J., Sovka, P., Špetík, R., Vlček, M.: Study of ADZT properties for spectral analysis, *11th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision*, Florence, 2011.

