

# 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.

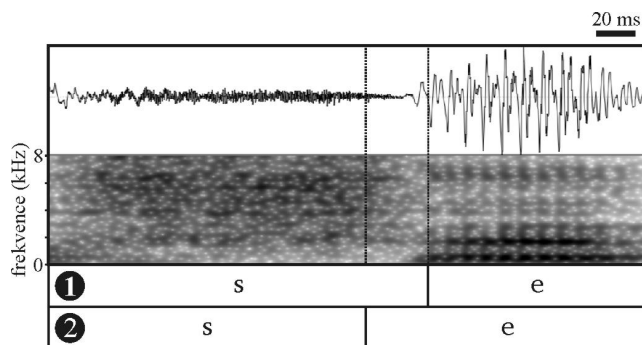
Podívejme se nejprve, co to vlastně znělost je a jakými způsoby ji můžeme měřit. Znělost je v mnoha jazycích světa využívána pro signalizaci fonologického kontrastu mezi hláskami jako [t]–[d] či [s]–[z]. Různé jazyky se znělostí zacházejí různými způsoby; čeština patří k jazykům, u nichž pozorujeme převážně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

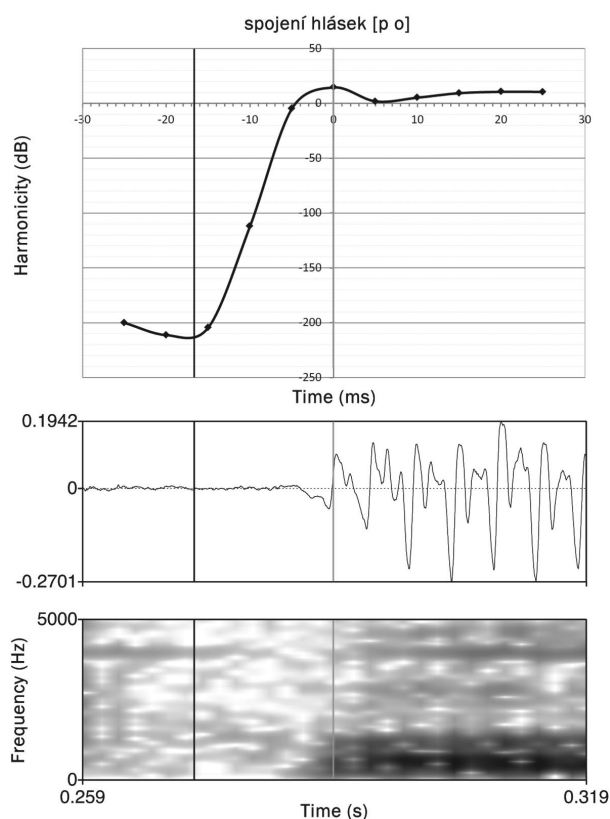
Pro tuto studii byly použity nahrávky 12 rodilých mluvčích češtiny (8 žen a 4 mužů) pořízené v nahrávací kabině Fonetického ústavu FF UK. Jednalo se o dospělé nekuřáky (věkové rozmezí se pohybovalo mezi 30 a 62 lety), kteří po krátké přípravě přečetli přibližně minutový text. Nahrávky byly zpracovány 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] 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-

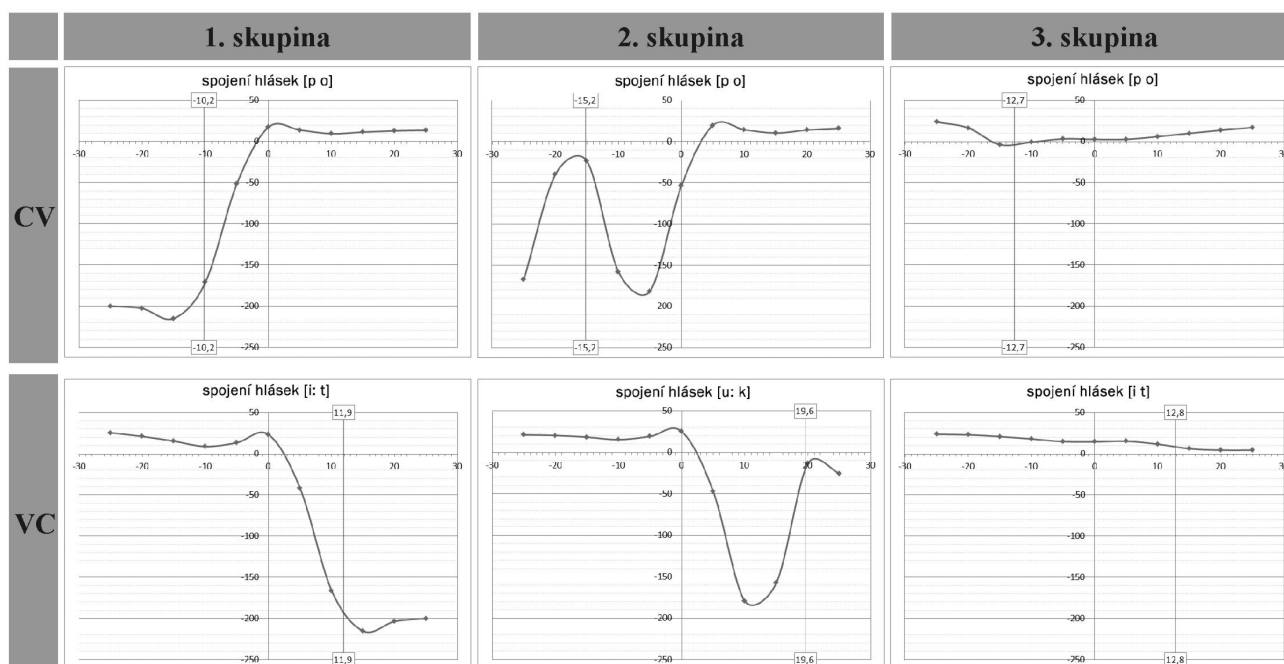
rizována sledem zřetelných formantových sloupků. Byla-li mezi segmenty přítomna přechodová oblast, umístili jsme hranici do temporálního středu této oblasti. Všechny hranice byly umísťovány do průchodu nulou (bod v oscilogramu, kde zvuková vlna protíná nulovou hodnotu amplitudy). Ukázkou srovnání automatické 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  $\pm 25$  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  $\pm 25$  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. Protiopolem 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

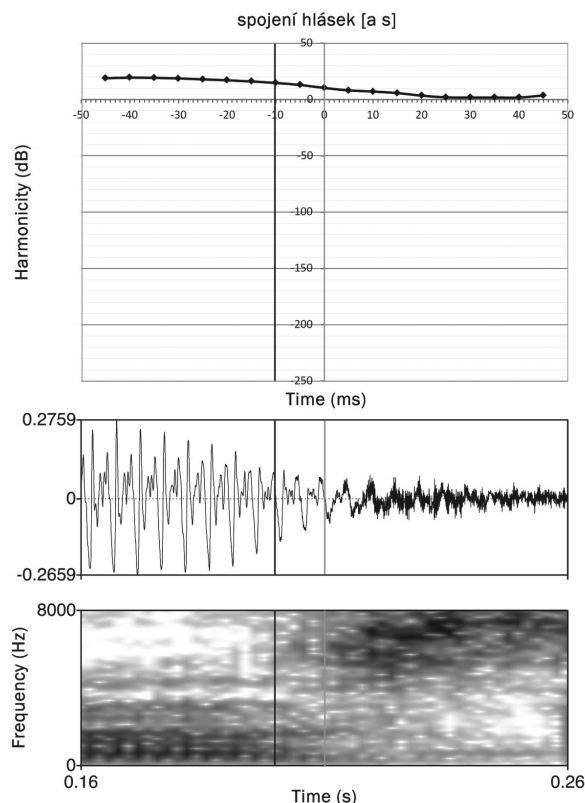
Tabulka 1: Procentuální zastoupení jednotlivých skupin průběhů HNR ve spojeních CV a VC

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 5000 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 zřetelně 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 dříve, 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 ( $\pm 5$  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  $\pm 10$  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>

# AKUSTICKÉ LISTY

České akustické společnosti  
www.czakustika.cz

ročník 17, číslo 4

prosinec 2011

## Obsah

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