elektrorevue
ISSN 1213 - 1539

Rok / Year:	Svazek / Volume:	Číslo / Number:
2011	13	6

Přehled a popis vybraných algoritmů pro posun výšky základního tónu (pitch shifting)

Review and description of selected pitch shifting algorithms

Aleš Křupka

xkrupk01@stud.feec.vutbr.cz

Fakulta elektrotechniky a komunikačních technologií VUT v Brně.

Abstrakt: Tento článek popisuje základní přístupy ke zpracování signálů, jež slouží k posunu základního tónu zpracovávaného zvukového signálu při zachování měřítka jeho časového průběhu. Je zde popsán princip funkce tří metod a poté jsou jednotlivé metody porovnány na základě provedeného testu a diskutovány problémy a omezení, které použití uvedených metod přináší.

Abstract: This paper describes some basic approaches to a signal processing, which are used for sound pitch shifting, when the time scale of the signal is preserved. The principle of three pitch shifting methods is described as well as their comparison based on the taken test. Also problems and constraints of using these methods are discussed.

elektrorevue ISSN 1213 - 1539

Přehled a popis vybraných algoritmů pro posun výšky základního tónu (pitch shifting)

Aleš Křupka

Fakulta elektrotechniky a komunikačních technologií VUT v Brně
xkrupk01@stud.feec.vutbr.cz

Abstrakt - Tento článek popisuje základní přístupy ke zpracování signálů, jež slouží k posunu základního tónu zpracovávaného zvukového signálu při zachování měřítka jeho časového průběhu. Je zde popsán princip funkce tří metod a poté jsou jednotlivé metody porovnány na základě provedeného testu a diskutovány problémy a omezení, které použití uvedených metod přináší.

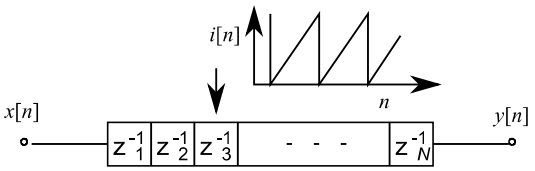
1 Úvod

Změna výšky tónu, tzv. pitch shifting, je prostředkem často využívaným pro úpravu kmitočtových vlastností zvukového signálu. Taková modifikace může být žádoucí například při zpracování nahrávek v hudebním studiu, při vytváření zvukových efektů pro hudební nástroje či při realizaci digitálního hudebního nástroje. Na metody, které posun výšky základního tónu provádějí, je kladen požadavek, aby zpracovaný signál nebyl zkrácen či prodloužen vzhledem k originálnímu signálu. Tím se především odlišují od prostého přehrávání originálního signálu s různou rychlostí, čehož lze u číslicového signálu docílit jeho převzorkováním při zachování vzorkovacího kmitočtu při přehrávání.

Obecně se metody pro posun výšky tónu dělí na metody se zpracováním v časové oblasti a na metody se zpracováním krátkodobých časových segmentů v kmitočtové oblasti.

2 Metoda využívající modulace zpožďovací linky

Jedná se o metodu se zpracováním v časové oblasti. Hlavním prvkem využívaným při této metodě je zpožďovací linka délky N, jejíž výstupem není vždy nejstarší vzorek zpožděný o N vzorkovacích period, ale vzorek určený ukazatelem i, platí tedy rovnice $y[n] = x[n - i]$, ukazatel i může nabývat hodnot v rozmezí <0; N>. Zpožďovací linku znázorňuje Obrázek 1.



Obrázek 1: Modulovaná zpožďovací linka

Je-li třeba dosáhnout efektu posunu výšky tónu, hodnota ukazatele i je modulována pilovou funkcí, jak zobrazuje Obrázek 1, [3]. Pozice vyčítání se tedy v čase mění, což si lze představit jako zrychlené resp. zpomalené přehrávání zvuku. Dosáhne-li ukazatel začátku či konce zpožďovací linky, svoji pozici skokově změní na opačný konec linky. Tato skutečnost způsobuje v průběhu zpracovaného signálu nespojitosti, které lze potlačit vhodným váhováním, kdy je zpracovaný signál v době nespojitosti potlačen při přenosu na výstup.

Hodnota ukazatele i může pro obecný poměr změny výšky tónu nabývat reálných hodnot, ne tedy pouze celočíselných. Uvnitř zpožďovací linky jsou však k dispozici pouze vzorky $x[n]$ až $x[n - N]$ příslušející k okamžikům daným celočíselnými hodnotami n. Proto se hodnota ukazatele i rozloží na její celou část i_i a její zlomkovou část i_{frac}, takže

$$i = i_i + i_{frac} \quad . \tag{1}$$

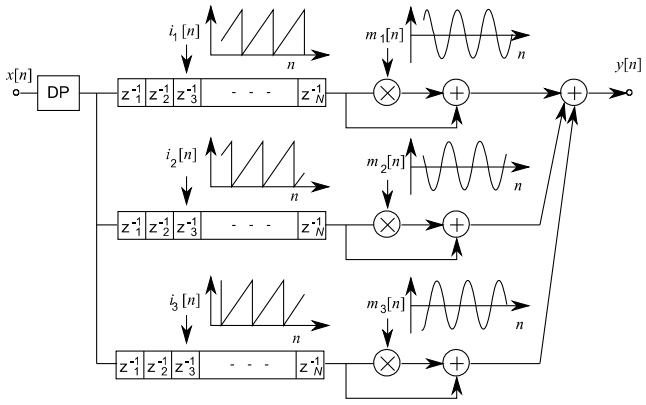
Výstupní vzorek je pak z nejbližších vzorků interpolován, jako nejjednodušší možnost lze například použít lineární interpolaci [3].

Váhování nespojitostí vznikajících skokovou změnou ukazatele i lze provést násobením výstupního signálu ze zpožďovací linky harmonickým signálem, jehož fáze je nastavena tak, aby byl přenos signálu ze zpožďovací linky nejvíce potlačen v okamžiku skokové změny. Toto váhování se ve výsledném signálu znatelně projeví jako amplitudová modulace. V případě malého poměru změny výšky je rychlost pohybu ukazatele i ve zpožďovací lince nižší, ke skokům ukazatele dochází s nižším kmitočtem, proto je i kmitočet váhovacího signálu nižší a váhování se tak projevuje jako tremolo efekt. Se zvětšujícím se poměrěm změny výšky roste rychlost pohybu ukazatele i a tedy i kmitočet váhovacího signálu. Se zvyšujícím se kmitočtem váhovacího signálu již váhování přestává být vnímáno jako tremolo efekt, ale jako samostatné disharmonické kmitočty vznikající amplitudovou modulací.

V [4] je navržena metoda využívající modulace zpožďovací linky, kdy je vstupní signál zpracován ve třech oddělených větvích, což má eliminovat artefakt způsobený amplitudovou modulací, viz Obrázek 2. Vstupní signál $x[n]$ je prvně filtrován antialiasingovou dolní propustí, poté je rozdělen do tří větví, z nichž každá obsahuje jednu zpožďovací linku provádějící převzorkování. Výsledný signál je pak tvořen součtem signálů z jednotlivých větví. Výstup ze zpožďovacích linek je modulován harmonickými signály pro potlačení nespojitostí. Zde si lze všimnout, že zpožďovací linky jsou modulovány pilovitými signály fázově vzájemně posunutými o $2\pi/3$. Výstupní signál je tedy tvořen zpracovanými vstupními signály překrývajícími se se vzájemným posunutím o třetinu délky zpožďovací linky. Podle nastavení modulačních signálů zpožďovacích linek jsou nastaveny i váhovací signály výstupů, které jsou také vzájemně fázově posunuty o $2\pi/3$. Důležité je, že sčítání stejných harmonických signálů vzájemně posunutých o $2\pi/3$ dává jako výsledek konstantní

elektrorevue ISSN 1213 - 1539

hodnotu. Proto ubývající příspěvek do výstupního signálu z jedné větve způsobený modulačním signálem je kompenzován narůstajícím příspěvkem z další větvě. Při zkoumání této konfigurace je vidět, že na výstupu dochází ke sčítání tří váhovaných signálů, které vlastně představují různě časově posunutý vstupní signál, což je dáno různou pozicí ukazatele vyčítajícího signál ze zpožďovací linky. Výsledný efekt posunu výšky tónu tedy bude závislý na spektrálním složení zpracovávaného vstupního signálu, jelikož si lze uvědomit, že sčítáním různě posunutého signálu mohou být některé kmitočtové složky eliminovány (lze připodobnit například k hřebenovému filtru).



Obrázek 2: Schéma struktury pro posun výšky tónu [4]

3 Metoda PICOLA

Tato metoda se taktéž řadí mezi metody se zpracováním v časové oblasti a je prvotně určena pro změnu délky signálu při zachování jeho kmitočtových vlastností, tzv. time stretching. Postup pro posun výšky tónu je pak takový, že originální signál je s příslušným poměrem prodloužen, viz [6], [9]. Původní délky je pak dosaženo pomocí převzorkování s daným poměrem, což požadovaným způsobem ovlivní kmitočtové vlastnosti zvukového signálu.

Metoda PICOLA patří mezi metody, u kterých je pro docílení požadovaného efektu třeba odhadnout základní periodu p zpracovávaného signálu. Tato informace je pak využita při procesu prodlužování či zkracování signálu.

Následuje popis postupu metody, kdy dochází k posunu kmitočtu základního tónu směrem k vyšším hodnotám. Zároveň ji ilustruje Obrázek 3.

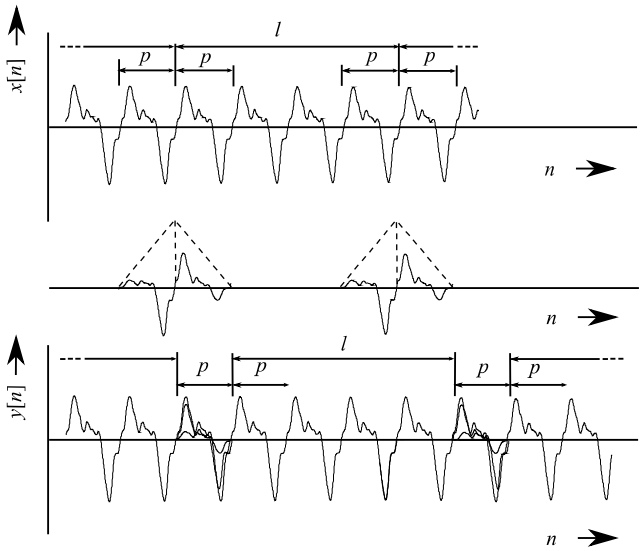
- Pro vstupní signál $x[n]$ je určen odhad počtu vzorků p základní periody.
- Na základě hodnoty p a požadovaného poměru změny výšky v určíme délku bloku l podle vztahu

$$l = \frac{p}{v-1} \quad , \tag{2}$$

při úpravě vztahu na

$$v = \frac{p+l}{l} \tag{3}$$

lze vidět, že pro dosažení požadovaného poměru roztažení signálu je k původnímu signálu délky l nutné přidat p vzorků.
- Ze vstupního signálu $x[n]$ jsou vybrány dva po sobě jdoucí bloky o délce p, z nichž první končí s předchozím blokem o délce l a druhý začíná zároveň s následujícím blokem o délce l. Tyto dva bloky jsou následně vynásobeny váhovacími okny a vzájemně překryty a sečteny. Vznikne tak nový blok o délce p.
- Tento blok je do původního signálu vložen před následující blok o délce l, čímž dojde k prodloužení signálu o p vzorků. Tímto způsobem je tedy mezi každých l vzorků vloženo p vzorků, vznikne tak prodloužený signál $y[n]$.
- Prodloužený signál $y[n]$ je převzorkován s poměrem $1/v$, čímž je dosaženo jeho původní délky.



Obrázek 3: Prodloužení signálu metodou PICOLA

Podobným způsobem lze dosáhnout snížení základního tónu, následující postup znázorňuje Obrázek 4.

- Pro vstupní signál $x[n]$ je určen odhad počtu vzorků p základní periody.
- Na základě hodnoty p a požadovaného poměru změny výšky v je určena délka bloku l podle vztahu

$$l = \frac{p}{1-v} \quad , \tag{4}$$

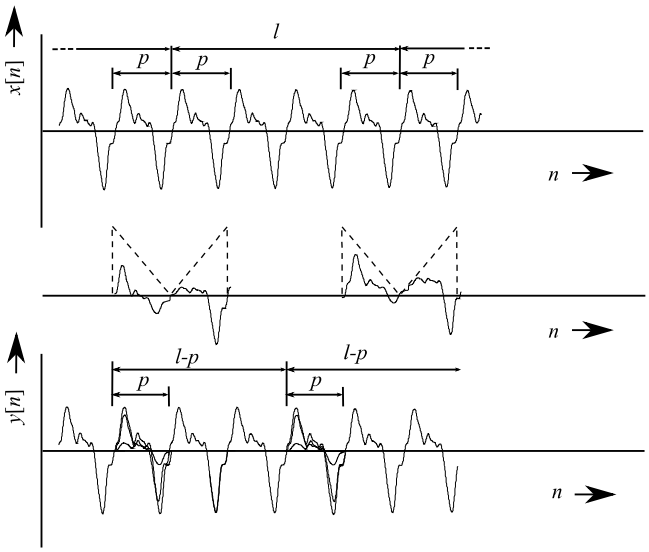
po úpravě na

$$v = \frac{l-p}{l} \tag{5}$$

je zřejmé, že pro zkrácení původního signálu $x[n]$ se zkrátí blok délky l o p vzorků.
- Ze vstupního signálu $x[n]$ jsou vybrány dva po sobě jdoucí bloky o délce p, z nichž první končí s předchozím blokem o délce l a druhý začíná zároveň s následujícím blokem o délce l. Tyto dva bloky jsou následně vynásobeny váhovacími okny a vzájemně překryty a sečteny. Vznikne tak nový blok o délce p.
- Tímto novým blokem je nahrazen první z váhovaných bloků, druhý váhovaný blok již není ve výsledném signálu zastoupen, tudíž je za první blok doplněno $l - p$ zbývajících vzorků. Tímto způsobem je každý blok o délce l zkrácen o p vzorků,

vznikne tak signál $y[n]$.

- Zkrácený signál je převzorkován s poměrem $1/v$, čímž je dosaženo jeho původní délky.



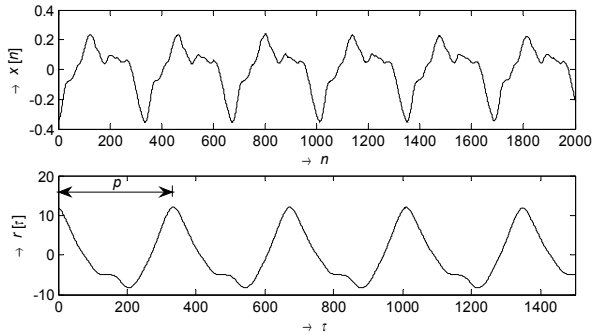
Obrázek 4: Zkrácení signálu metodou PICOLA

Výše uvedený postup tedy každých l vzorků vstupního signálu $x[n]$ prodlouží, resp. zkrátí vždy o p vzorků. Problémem zůstává určení odhadu základní periody p.

Nejznámější metodou pro detekci základní periody je autokorelační metoda [5]. Autokorelační funkce je dána vztahem

$$r_t[\tau] = \sum_{j=t+1}^{t+W} x[j]x[j+\tau] \quad . \tag{6}$$

Udává míru podobnosti signálu $x[n]$ se sebou samým pro různé hodnoty posunu τ (tzv. lag). Parametr W určuje, pro jak dlouhý úsek signálu $x[n]$ bude autokorelační funkce počítána. Je-li signál periodický, pak lze v průběhu autokorelační funkce hledat maxima, pro která byla míra shody vzájemně posunutého signálu $x[n]$ největší, jejich hodnoty posunu τ pak odpovídají násobkům periody p. Obrázek 5 ukazuje průběh signálu a jeho autokorelační funkci určenou podle vztahu (6). Základní periodu p lze z grafu určit jako pozici τ prvního maxima autokorelační funkce, což je v konkrétním případě 334 vzorků.



Obrázek 5: Autokorelační funkce

Kromě autokorelační funkce existují i další metody pro odhad základní periody, např. spektrální metoda, kepstrální [1], yin [5] a další.

Z uvedeného požadavku na detekci základní periody p vyplývá omezení, že tato metoda je vhodná pouze na jednohlasý periodický signál.

Kromě metody PICOLA existují i jiné metody pro prodloužení a zkrácení signálu se zpracováním v časové oblasti, které jde na popsaném principu využít i pro změnu výšky základního tónu, např. metody SOLA a PSOLA, viz [10]. Obecně se však dá říct, že všechny tyto metody jsou založeny na principu správného odhadu základní periody, liší se jen ve způsobu zpracování. Fáze odhadu základní periody a fáze zpracování lze tak od sebe oddělit. Jednotlivé metody odhadu základní periody a metody zpracování je pak možné vzájemně podle konkrétních požadavků kombinovat.

4 Metoda fázového vokodéru

Další možnou metodou pro posun výšky tónu je metoda fázového vokodéru. Tato metoda se vyznačuje zpracováním časových segmentů signálu v kmitočtové oblasti. Hlavní myšlenka spočívá v transformaci určitého úseku časového průběhu vstupního signálu do průběhu vyjadřujícího jeho kmitočtové složení. Pro dosažení změny výšky tónu jsou v kmitočtovém vyjádření jednotlivé složky patřičným způsobem modifikovány a posléze je toto modifikované spektrum transformováno zpět do časového vyjádření, vznikne tak požadovaný výstupní signál.

Pro vyjádření spektrálního složení určitého segmentu signálu se využívá tzv. krátkodobé spektrální analýzy [10], jejímž představitelem je diskrétní krátkodobá Fourierova transformace diskrétního signálu daná vztahem

$$\begin{aligned} X(n,k) &= \sum_{m=-\infty}^{\infty} x[m]h[n-m]e^{-jk\frac{2\pi}{N}m} \\ &= |X(n,k)|e^{j\varphi(n,k)}, \end{aligned} \tag{7}$$

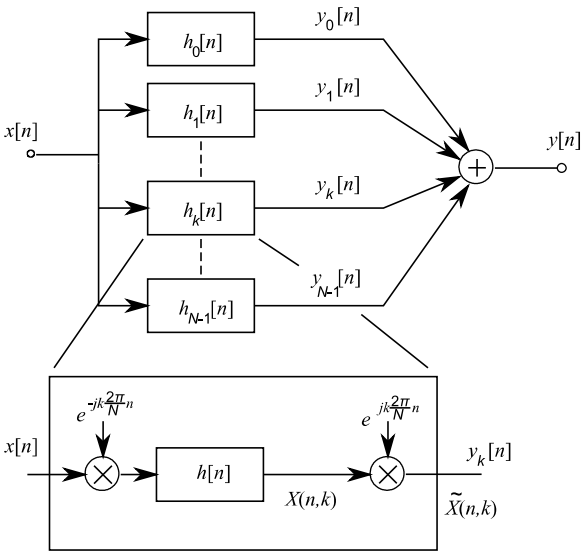
kde $X(n,k)$ je časově proměnným spektrem signálu $x[m]$, které pro daný časový okamžik n udává amplitudy $|X(n,k)|$ a fáze $\varphi(n,k)$ jednotlivých složek k v rozmezí celých čísel <0, N-1>, je tedy vidět, že se jedná o transformaci definovanou v oboru komplexních čísel. Výraz $h[n-m]$ slouží jako posuvné okno délky N, které z teoreticky nekonečně dlouhého signálu $x[m]$ vybere segment délky N, který je podroben zpracování.

Vztah (7) lze realizovat jako banku filtrů znázorněnou na Obrázku 6, [10]. Vstupní signál je rozdělěn do N větví. V každé větvi je signál filtrován pomocí pásmové propusti, která vznikne modulací impulzní charakteristiky $h[n]$ funkcí $e^{jk\frac{2\pi}{N}n}$, což je zobrazeno ve spodní části Obrázku 6. Výstupy jednolivých větví jsou dány vztahem

$$\begin{aligned} y_k[n] &= \sum_{m=-\infty}^{\infty} x[m]h_k[n-m] \\ &= \sum_{m=-\infty}^{\infty} x[m]h[n-m]e^{jk\frac{2\pi}{N}(n-m)} \\ &= e^{jk\frac{2\pi}{N}n} \sum_{m=-\infty}^{\infty} x[m]h[n-m]e^{-jk\frac{2\pi}{N}m} \\ &= e^{jk\frac{2\pi}{N}n} X(n,k) \\ &= \tilde{X}(n,k) = |X(n,k)|e^{j\tilde{\varphi}(n,k)}, \end{aligned} \tag{8}$$

elektrorevue ISSN 1213 - 1539

kde $\tilde{X}(n,k)$ představuje spektrum pásmově omezeného signálu se středním úhlovým kmitočtem Ω_k. V případě, že posloupnost $h[n]$ představuje pravoúhlé okno, jsou výstupem banky přímo koeficienty diskrétní Fourierovy tranformace. Nebude-li po konvoluci s impulzní charakteristikou $h[n]$ signál násoben funkcí $e^{jk\frac{2\pi}{N}n}$, je získán jeho nízkofrekvenční ekvivalent $X(n,k)$.



Obrázek 6: Znázornění banky filtrů pro krátkodobou spetrální analýzu [10]

Reprezentace nízkofrekvenčích ekvivalentnů $X(n,k)$ pásmových signálů $\tilde{X}(n,k)$ má tu výhodu, že bude-li mít vstupní signál $x[n]$ přesně úhlový kmitočet Ω_k, fáze $\varphi(n,k)$ bude v čase konstantní. Naopak její změna v čase značí odchylku od středního úhlového kmitočtu Ω_k. k-tého pásma. Okamžitý úhlový kmitočet Ω_i se pak určí jako derivace fáze podle vztahu

$$\begin{aligned}\Omega_i(n,k) &= \Omega_k(n,k) + \frac{d}{dn}\varphi(n,k) \\ &= \Omega_k(n,k) + \varphi(n,k) - \varphi(n-1,k).\end{aligned} \tag{9}$$

Z tohoto vztahu pak lze odvodit výpočet okamžitého absolutního kmitočtu f_i jako

$$f_i(n,k) = \left(\frac{k}{N} + \frac{\varphi(n,k) - \varphi(n-1,k)}{2\pi}\right) f_{vz}, \tag{10}$$

kde f_{vz} je vzorkovací kmitočet, který v předchozím vztahu nevystupoval, jelikož úhlové kmitočty Ω_k *a* Ω_i jsou relativními úhlovými kmitočty. Jak je vidět, okamžitý kmitočet $f_i(n,k)$ k-tého pásma je určen na základě rozdílu dvou fází $\varphi(n,k)$ a $\varphi(n-1,k)$ k-tých složek obsažených v po sobě následujících krátkodobých spektrech.

Je-li zjištěn okamžitý kmitočet $f_i(n,k)$ pro každé z N pásem, je jeho hodnota vynásobena požadovaným poměrem změny výšky v. Výsledný zvuk je poté realizován pomocí banky oscilátorů, kdy každý oscilátor generuje jemu odpovídající okamžitý kmitočet $f_i(n,k)$ vynásobený poměrem změny výšky v. Sečtením signálů z jednotlivých oscilátorů vznikne výsledný zpracovaný zvuk.

U metody s využitím fázového vokodéru se také lze setkat se specifickými artefakty danými principem metody, které jsou způsobeny principem neurčitosti platným pro disktrétní Fourierovu transformaci. Analýzou delšího segmentu signálu lze získat více kmitočtových složek a tedy přesnější spektrum, avšak toto spektrum je platné pro dlouhý analyzovaný segment, chybí tak informace, jaké jsou časové poměry mezi jednotlivými spektrálními složkami v rámci analyzovaného segmentu. Navíc je toto spektrum zkresleno váhovacím oknem $h[n]$, které ze vstupního signálu vybírá tu část, která je pro daný časový okamžik převáděna do kmitočtového vyjádření. Při syntéze signálu z upraveného spektra se pak uvedená omezení projevují jako artefakty připomínající určité rozmazávání zvuku [2]. Výhoda této metody provádějící modifikaci signálu v kmitočtové oblasti je ta, že je aplikovatelná na signál s vícehlasým obsahem, jelikož poloha každé kmitočtové složky vstupního signálu je s požadovaným poměrem změněna.

5 Srovnání metod

V následující části je provedeno srovnání uvedených technik na základě vykonaného subjektivního testu. Pro testování byly použity různé typy zvuků ze sady EBU SQAM, které slouží k účelům subjektivních poslechových testů [8]. Při testu byl sledován vliv použité metody na výskyt rušivých disharmonických tónů, přítomnost nespojitostí způsobujících rušivé zvuky (připomínající lupání či praskání) a dále celkový vliv metody na změnu charakteru zpracovaného zvuku vzhledem k originální nahrávce. Jako předloha pro ohodnocení jednotlivých přístupů posloužila technika z doporučení [7], která zavádí pět stupňů míry účinku algoritmu na výsledný vjem zvuku. Tyto stupně popisuje Tabulka 1.

Tabulka 1: Stupně účinku metody

Ohodnocení	Vliv metody na výskyt nežádoucích prvků
5	Nepostřehnutelný
4	Postřehnutelný, ale ne rušivý
3	Lehce rušivý
2	Rušivý
1	Zcela degradující

Tabulka 2 udává jednohlasé zvuky ze sady EBU SQAM, které byly použity pro porovnání. Tabulka 3 pak uvádí vícehlasé zvuky, na nichž také proběhl test metod.

Tabulka 2: Použité jednohlasé zvuky EBU SQAM

Číslo zvuku	Název zvuku
1	Sinusoida
4	Elektronický gong
8	Housle
21	Trubka
31	Činel
39	Piáno
44	Ženský zpěv - soprán
49	Ženský hlas

Tabulka 3: Použité vícehlasé zvuky EBU SQAM

Číslo zvuku	Název zvuku
25	Harfa
39	Piáno
48	Vokální kvartet
55	Haydn
69	ABBA

Test proběhl pro různé hodnoty posunu základního tónu, výsledný zvuk byl ohodnocen vzhledem k originálnímu zvuku patřičným stupněm. Bylo využito principu průměrování stanovených stupňů přes testovací zvuky pro jednotlivé hodnoty posunu výšky.

Tabulka 4, Tabulka 5, Tabulka 6 obsahuje výsledky pro metodu modulace zpožďovací linky, PICOLA a metodu fázového vokodéru při testování jednohlasých zvuků. Jednotlivé řádky odpovídají průměrným hodnotám nespojitosti, disharmonie a celkového výskytu artefaktů metody, sloupce určují posun výšky v půltónech.

Tabulka 4: Ohodnocení metody s jednohlasými zvuky

Počet půltónů	Modulovaná zpožďovací linka							
	-24	-12	-5	-1	+1	+2	+12	+24
Nespoj.	5,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0
Dish.	2,9	2,9	3,5	4,5	4,5	3,3	2,3	2,0
Celk.	2,5	2,6	3,1	4,4	4,4	3,0	2,1	1,8

Tabulka 5: Ohodnocení metody s jednohlasými zvuky

Počet půltónů	PICOLA							
	-24	-12	-5	-1	+1	+2	+12	+24
Nespoj.	3,3	3,8	4,0	4,5	4,6	4,3	4,3	4,1
Dish.	4,8	4,9	5,0	5,0	5,0	5,0	4,6	4,5
Celk.	3,6	4,3	4,5	4,8	4,9	4,8	4,0	3,4

Tabulka 6: Ohodnocení metody s jednohlasými zvuky

Počet půltónů	Fázový vokodér							
	-24	-12	-5	-1	+1	+2	+12	+24
Nespoj.	5,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0
Dish.	4,8	4,8	4,8	4,8	4,8	4,7	4,7	4,5
Celk.	4,1	4,4	4,6	4,5	4,5	4,4	4,0	3,6

Tabulka 7, Tabulka 8, Tabulka 9 podává výsledky při testování vícehlasých zvuků.

Tabulka 7: Ohodnocení metody s vícehlasými zvuky

Počet půltónů	Modulovaná zpožďovací linka							
	-24	-12	-5	-1	+1	+2	+12	+24
Nespoj.	5,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0
Dish.	2,2	2,4	3,0	4,8	4,8	3,0	1,8	1,6
Celk.	1,4	2,0	3,0	4,6	4,6	3,0	1,4	1,0

Tabulka 8: Ohodnocení metody s vícehlasými zvuky

Počet půltónů	PICOLA							
	-24	-12	-5	-1	+1	+2	+12	+24
Nespoj.	2,4	3,4	3,8	4,2	4,2	3,8	3,2	3,4
Dish.	2,4	2,8	3,6	5,0	5,0	3,4	2,2	1,4
Celk.	1,8	2,4	3,6	5,0	5,0	3,4	2,2	1,4

Tabulka 9: Ohodnocení metody s vícehlasými zvuky

Počet půltónů	Fázový vokodér							
	-24	-12	-5	-1	+1	+2	+12	+24
Nespoj.	5,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0
Dish.	4,4	4,6	5,0	5,0	5,0	4,8	4,2	2,4
Celk.	3,6	4,2	4,6	5,0	5,0	4,2	3,8	2,4

Výsledky testu potvrdily očekávané teoretické předpoklady. U metody využívající modulace zpožďovací linky si lze všimnout (Tabulka 4, Tabulka7), že dobrých výsledků dosahuje jen pro velmi malé hodnoty posunu a to jak pro jednohlasé, tak i vícehlasé zvuky. Ačkoliv jsou v signále nespojitosti vznikající skoky ukazatelů eliminovány váhováním, délka zpožďovací linky má stále konstantní velikost. Při sčítání signálů z jednotlivých větví pak dochází ke sčítání různě posunutého vstupního signálu, čímž dále dochází k úpravě kmitočtového spektra vstupního signálu. Kvalita posunu zvuku u této metody je tedy závislá na spektrálním složení zvuku vzhledem k délce zpožďovací linky. Při vyšších hodnotách posunu jsou ve zvuku více slyšitelné disharmonické složky. Výhodou této metody je však velmi nízká výpočetní náročnost.

Srovnatelných výsledků stupně výskytu celkového artefaktu dosáhly metody PICOLA a metoda fázového vokodéru pro jednohlasé signály, viz. Tabulka 5 a Tabulka 6. U metody PICOLA byla pro odhad základní periody použita metoda yin, která je výpočetně poměrně náročná a zabírala převážnou část výpočetního času při zpracování. Na druhou stranu přesnost určení základní periody do značné míry ovlivňuje kvalitu výsledného zpracování, lze tedy říct, že optimalizace algoritmu odhadu základní periody je zde klíčovým problémem. Tabulka 8 však ukazuje, že metoda PICOLA není vhodná pro zpracování vícehlasého zvuku, protože její princip je založen na odhadu jedné základní periody, avšak vícehlasý zvuk může sestávat z více zvuků o různých základních periodách. Dále ve zpracovaném signále mohou být slyšet určité nespojitosti v místech změny základní periody ve vstupním signále.

Tabulka 9 ukazuje použitelnost metody fázového vokodéru i pro vícehlasé signály, což je její velkou výhodou. U této metody je u některých zvuků typický artefakt určitého rozmazání zvuku.

6 Závěr

Cílem článku bylo podat informaci o základních přístupech k posunu výšky tónu zvukových signálů. Byly zmíněny problémy, které se u jednotlivých metod z principu jejich funkce vyskytují. Nejprve byla popsána metoda využívající modulace zpožďovací linky, která je poměrně jednoduchá, ale u které je problematická závislost kmitočtového složení

vstupního signálu vzhledem k délce zpožďovací linky. Dále byla popsána metoda PICOLA, která díky odhadu základní periody umožnuje produkovat výškově posunuté zvuky, avšak pouze v případě jednohlasých periodických signálů. Se zpracováním v kmitočtové doméně byla uvedena metoda fázového vokodéru, kdy byl použit přístup využívající banky filtrů. Tato metoda je aplikovatelná i na vícehlasé zvukové signály, artefakty metody jsou dány omezeními vyplývajícími z vlastností diskrétní Fourierovy transformace, tedy jejím principem neurčitosti a zkreslováním spektra signálu váhovacím oknem.

Literatura

[1] ATASSI, H. Metody detekce základního tónu řeči. *Elektrorevue* [online]. 2008 .Dostupný z WWW: <www.elektrorevue.cz>.

[2] BERNSEE, S.: *DSPDimension* [online]. 1999. Time Stretching And Pitch Shifting of Audio Signals – An Overview. Dostupné z WWW: <www.dspdimension.com>.

[3] DATTORRO, J.: *AES* [online]. 1997. Effect Design: Part 2: Delay-Line Modulation and Chorus. Dostupný z WWW: <www.aes.org>.

[4] DISCH, S; ZÖLZER, U.: *Workshop on Digital Audio Effects*. 1999. Modulation and delay based digitalaudio effects.

[5] CHEVEIGNE, A; KAWAHARA, H. YIN, a fundamental frequency estimator for speech and music. In *J. Acoust. Soc. Am., Vol. 111, No. 4, April 2002* [online] Dostupné z WWW:<audition.ens.fr/adc/pdf/2002_JASA_YIN.pdf>.

[6] IKEDA, M. *PICOLA and TDHS* [online]. 2006 PICOLA and TDHS. Dostupné z WWW: <http://keizai.yokkaichi-u.ac.jp/~ikeda/research/picola.html>.

[7] ITU-T P.800. ITU, 1996

[8] Sound Quality Assessment Material recordings for subjective tests , 2008, Geneva. Dostupné z WWW: <http://tech.ebu.ch/publications/tech3253>.

[9] Yamaha corporation, Time-Scale Modification Method And Apparatus For Digital Audio Signals, United States Patent 6 519 567 B1, 2003.

[10] ZÖLZER, U.,et al. *DAFX : Digital Audio Effects* Willey, 2002.