BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES
Vol. 54, No. 4, 2006

Pitch shifter based on complex dynamic representation rescaling and direct digital synthesis

E. HERMANOWICZ* and M. ROJEWSKI

Multimedia Systems Department, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, 11/12 Narutowicza St., 80-952 Gdańsk, Poland

Abstract. In this paper a new pitch shifter using a complex instantaneous frequency rescaler and direct digital synthesizer is presented aimed at an application in a handset calling signal composer. The pitch shifter introduced here exhibits an excellent performance as a generator of different melodies, where the sound of each note in a melody, e.g., imitating a popular hit, is derived from a short recording of a voice of a chosen creature via complex dynamic representation processing.

Key words: pitch shifter, complex instantaneous frequency, direct digital synthesis.

1. Introduction

With the advent of commercial worldwide information systems an interest arises, especially among young mobile handset holders, in having at disposal unusual, original, and amusing individual calling signals. Our paper contains a proposal of creating such signals.

Up to now the most popular calling signals in use have been segments of digitized melodies generated by electronic musical instruments. Our proposal generalizes this for sounds generated by a chosen creature.

The core of the concept lies in composing a melody digitally from a set of sounds (the musical scale) derived on the basis of a short record of a living creature voice, e.g., a canary chirp [1]. The set is created using an original pitch-shifting method applied to the basic record. The method consists in complex dynamic representation (CDR) modification (scaling) and direct digital synthesis (DDS) [2], [3,4]. Ordering of the diatonic scale is performed using the MIDI (Musical Instrument Digital Interface standard) note number to frequency conversion chart. In this way a scale of sounds is created having the same shape of the CDR real-valued components but different magnitude introduced using different pitch-shifting factors. The semitones of our scale are assigned to the notes of the composed melody.

The utility of the proposed method is illustrated by an example of a calling signal record. The pitch shifter introduced here exhibits an excellent performance as a generator of different melodies, where the sound of each note in a melody is derived from a short recording of a voice of a chosen creature via the CDR processing.

2. The proposed pitch-shifting method

Figure 1 shows a block scheme of the processing aimed at pitch shifting of a real-valued bandlimited discrete-time sound record

$$\{x\} \triangleq \{x[n]\}_{n=0}^{N-1} = \{x[0], x[1], \ldots, x[N-1]\} \tag{1}$$

where

$$x[n] = a[n]\cos\varphi[n] \tag{2a}$$

represents a primary continuous-time signal

$$x(t) = a(t)\cos\varphi(t) \tag{2b}$$

uniformly discretised with a sampling period T in seconds. In (2a) $a[n]$ is the instantaneous amplitude and $\varphi[n]$ is the instantaneous phase of $x[n]$. The right-hand side of (2b) is the so-called AM · FM representation of an arbitrarily modulated signal $x(t)$, where $a(t)$ is the AM factor and $\cos\varphi(t)$ is the FM factor.

In order to convert the signal $x[n]$ into polar form and next to define its instantaneous angular frequency one has to create a complex Hilbertian signal (HS) [5–7], otherwise called the Hilbertian equivalent of $x[n]$

$$x_H[n] \triangleq x[n] + j\tilde{x}[n] \tag{3}$$

whose real part is $x[n]$. The imaginary part of $x_H[n]$

$$\tilde{x}[n] \triangleq H_H\{x[n]\} \tag{4}$$

is the Hilbert transform of $x[n]$. The linear operator H_H in (4) is known under the name of Hilbert transformer (HT). Its ideal frequency response is defined as

$$H_H(f) \triangleq -j\,\mathrm{sgn}(f), \qquad 0 < |f| < 0.5 \tag{5}$$

where $f = FT$ stands for the 'digital' (normalised) frequency, F stands for the physical frequency in Hz, T is the sampling period and $j^2 \triangleq -1$.

It is further assumed that the AM · FM factors of $x[n]$ fulfill the Bedrosian theorem which states that the Hilbert transform of the product of two signals with nonoverlapping spectra

*e-mail: hewa@eti.pg.gda.pl

equals the product of the low-pass term by the Hilbert transform of the high-frequency term. In other words, only the high-frequency term is transformed [6,7].

On the grounds of the Bedrosian theorem, the imaginary part $\tilde{x}[n]$ of $x_H[n]$ (3) can be written as

$$\tilde{x}[n] = a[n]\sin\varphi[n] \tag{6}$$

and the complex HS, $x_H[n]$, for $x[n]$ is given by

$$\begin{aligned} x_H[n] &= a[n]\{\cos\varphi[n] + j\sin\varphi[n]\} \\ &= a[n]\exp(j\varphi[n]) \end{aligned} \tag{7}$$

The instantaneous angular frequency (IF) [8] of the HS (7), as well as of $x[n]$, is defined as

$$\omega[n] \triangleq \left.\frac{\mathrm{d}\varphi(t)}{\mathrm{d}t}\right|_{t=nT} \tag{8}$$

In our experiments, in Sect. 3, we create the Hilbertian signal $x_H[n]$ by using a digital linear-phase FIR (finite impulse response) Hilbert filter (HF in Fig. 1), approximating the ideal frequency response

$$H(f) \triangleq \begin{cases} 2, & 0 < f < 0.5 \\ 0, & -0.5 < f < 0 \end{cases} \tag{9}$$

The impulse response of the HF is complex-valued. The IF $\omega[n]$ (8) of $x_H[n]$ is estimated digitally by the IFE, the instantaneous angular frequency estimator in Fig. 1, without any need to resort to the primary continuous signal $x(t)$. We have used the following estimator

$$\omega[n] = \mathrm{Arg}(x_H[n]x_H^*[n-1]) \ \in [-\pi,\ \pi) \ \frac{\text{rad}}{\text{sample}} \tag{10}$$

for the IFE block, where Arg() stands for the principal value of the argument of the product in parantheses and asterisk stands for the complex conjugate.

The instantaneous amplitude (otherwise called envelope) $a[n]$ of $x_H[n]$ is defined as

$$a[n] = |x_H[n]| \tag{11}$$

It is estimated in Fig. 1 by the IAE – an instantaneous amplitude estimator.

Further on we call the pair

$$\lambda[n] \triangleq \ln a[n] \quad \text{and} \quad \omega[n] \tag{12}$$

complex dynamic representation (CDR) of $x[n]$. The CDR has two real-valued components. The first of them, $\lambda[n]$, is the instantaneous level otherwise called log-envelope of the Hilbertian signal $x_H[n]$ and the second, $\omega[n]$, is the instantaneous angular frequency of $x_H[n]$. The CDR components represent the HS corresponding to the primary real-valued signal $x(t)$.

By the following remapping of the CDR

$$\lambda[n], \omega[n] \Rightarrow \lambda_\kappa[n] = \kappa\lambda[n] + \lambda_0, \omega_\kappa[n] = \kappa\omega[n] \tag{13}$$

we obtain a new CDR having the components: $\lambda_\kappa[n]$ and $\omega_\kappa[n]$. This CDR modification, which constitutes the core processing block in Fig. 1 indicated by broken line, results in pitch shifting of a given sound signal $x[n]$ with a pitch modification (scaling) factor $\kappa > 0$. As a result of the CDR rescaling in accordance with (13) not only the pitch but also the instantaneous level $\lambda[n]$ responsible for audibility of the sound is changed. In order to counteract the latter, the maximal value of the instantaneous level change has to be compensated for by adding to $\kappa\lambda[n]$ a correction term $\lambda_0(\kappa)$, as seen in Fig. 1, computed by using

$$\lambda_0(\kappa) \triangleq \lambda_{\max}(1-\kappa)$$

where $\lambda_{\max}$ stands for the maximal level of the primary signal.

Hence, in Fig. 1, firstly, the input signal $x[n]$ is filtered by the HF. Next the filtered signal is mapped into its CDR $\{\lambda, \omega\}$. The CDR components are extracted using the IAE, $\ln(\cdot)$ and IFE blocks. Further on both CDR components are multiplied by the same coefficient κ having a positive value as above and the instantaneous level $\kappa\lambda[n]$ undergoes the above-mentioned correction by $\lambda_0(\kappa)$. Finally, after this remapping, performed in accordance with (13), the new CDR, $\{\lambda_\kappa, \omega_\kappa\}$, is demapped into the target pitch-shifted sound signal record $\{x_\kappa\}$ using the DDS shown in Fig. 2a.

The main processing block of the DDS is phase accumulator (PA). Functioning of the PA is presented in Fig. 2b. Driven by $\omega_\kappa[n]$ the PA wraps the instantaneous phase to the interval $\varphi_p[n] \in [-\pi, \pi)\forall n$ (principal phase wind).

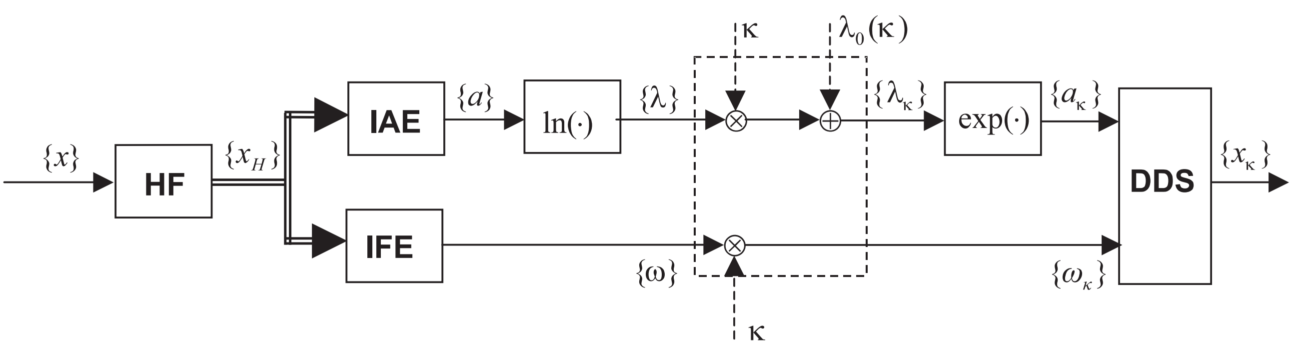

Fig. 1. Block scheme of the processing of a discrete-time sound record aimed at pitch-shifting with pitch scaling factor κ

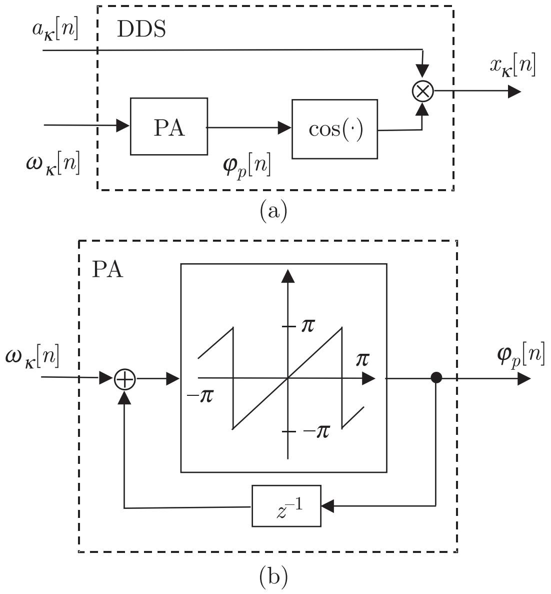

Fig. 2. DDS basic architecture (a) with the phase accumulator operation (b) revealed

It is worth emphasizing that the fundamentals of the HS application to sound processing were formulated in [9] and [10]. Also in [11] examples of shifting the spectrum of an analytic audio signal are given, where, however, similarly as in [9] and [10], the instantaneous amplitude (IA) remained unaffected. Opposite to that the original method and algorithm conceived here do not violate the amplitude-phase relationships typical of natural signals (speech sounds, animal voices, etc.). This is achieved by simultaneous modification via remapping of both: the IA and IF.

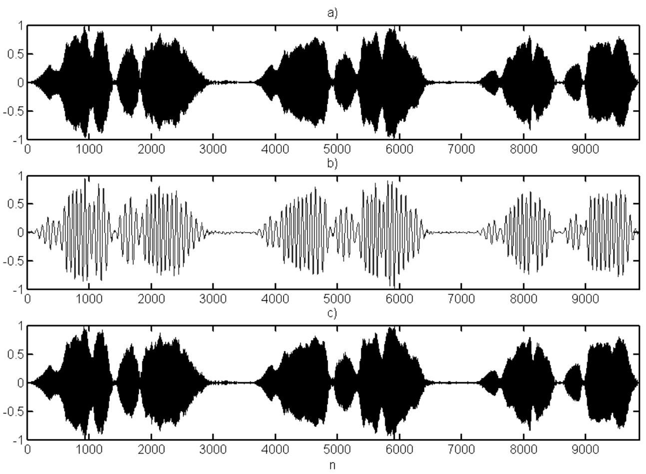

Fig. 3. Exemplary waveforms of a canary song processed via CDR remapping aimed at pitch down shifting with $\kappa = 1/10$: (a) input waveform $\{x\}$, (b) remapped waveform $\{x_\kappa\}$ and (c) demapped waveform $\{\hat{x}\}$

Our approach applied in the aim of pitch rescaling, guarantees invertibility of the proposed pitch-shifter even for very small values of the pitch modification factor κ. Such a favourable result has been reached yet neither in a phase vocoder nor in the processing of the IA or the IF solely.

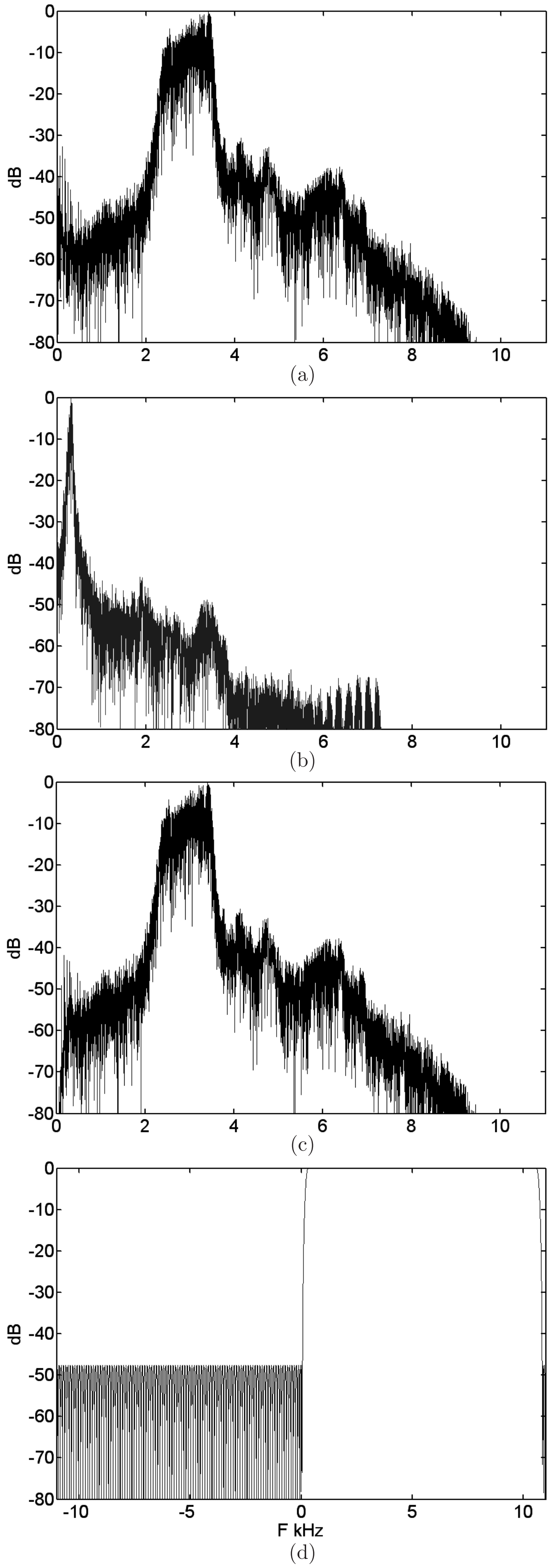

Fig. 4. Normalized magnitude spectra of signals from Figs. 3a, 3b and 3c, respectively, and magnitude response of an FIR Hilbert filter of length 229 (d) used here

An important issue is that only the complete CDR processing assures the recovery of the original signal from its pitch-shifted counterpart. However, the key for an implementation to be a success lies in using, as a primary signal for generation of the scale, a signal which, at least approximately, fulfills the Bedrosian theorem.

3. Experiments

Firstly, we experimented with an exemplary waveform fulfilling the above-mentioned Bedrosian theorem, in order to show the abilities of the algorithm from Fig. 1. The algorithm operates on-line. A recording of three consecutive chirps of a canary song shown in Fig. 3a was processed. The results are presented in Figs. 3b and 3c in the discrete-time domain, and in Fig. 4 in the frequency domain. The processing was aimed at pitch down shifting using the pitch modification factor $\kappa = 1/10$. It results in waveform expansion as well as simultaneous spectrum compression and down shifting, both by a factor of κ.

As depicted in Fig. 3b the remapped waveform became expanded with a factor of κ, relative to the input waveform from Fig. 3a. Accordingly, the input signal spectrum magnitude, presented in Fig. 4a, changed into a compressed by κ, as shown in Fig. 4b. This waveform was then demapped back to $\kappa = 1$ and resynthesised. The demapped waveform is presented in Fig. 3c and the magnitude spectrum of this waveform is shown in Fig. 4c. It may be interesting to note that the invertible operations of mapping and demapping work effectively even for relatively great values of κ, depending on the oversampling ratio of the signal under processing, as well as for very small κ values, for example, for $\kappa = 1/1000$.

In our further experiments we processed a single chirp, the first one from Fig. 3a. Fig. 5a shows this chirp of 3000 samples having 22.050 kHz sample rate. Figs. 5b and 5c present the chirp AM $\cdot$ FM representation factors and Fig. 5d depicts the chirp instantaneous angular frequency waveform. Next, we computed amplitude spectra (Fig. 6) as well as spectrograms (Fig. 7) in order to have better insight into the properties of the chirp. The value of spectrograms lies in that they bring to light the dependence of the signal spectrum properties on time.

Fig. 6 is very important. It reveals that the magnitude spectra of the AM and FM factors of the canary chirp are practically non-overlapping in their information bearing (useful) parts with the approximate accuracy of – 30 dB, as seen in Fig. 6a.

The spectrograms in Figs. 7b and 7c disclose that it holds for every instant of time. Moreover, Figs. 6 and 7 confirm that the chirp after the processing by the algorithm from Fig. 1 can be inverted back to the original one. The accuracy of this inversion, see Fig. 6b, depends on the accuracy of approximation of the ideal Hilbert filter frequency response (9) by its realisable counterpart. Further on, using the proposed pitch-shifting method, incorporated into the algorithm from Fig. 1, and data gathered in Table 1, we created the canary scale. Its performance is shown in Figs. 8. Finally, on this basis we composed a melody whose waveform, spectrogram scale and CDR components are presented in Figs. 9, 10 and 11.

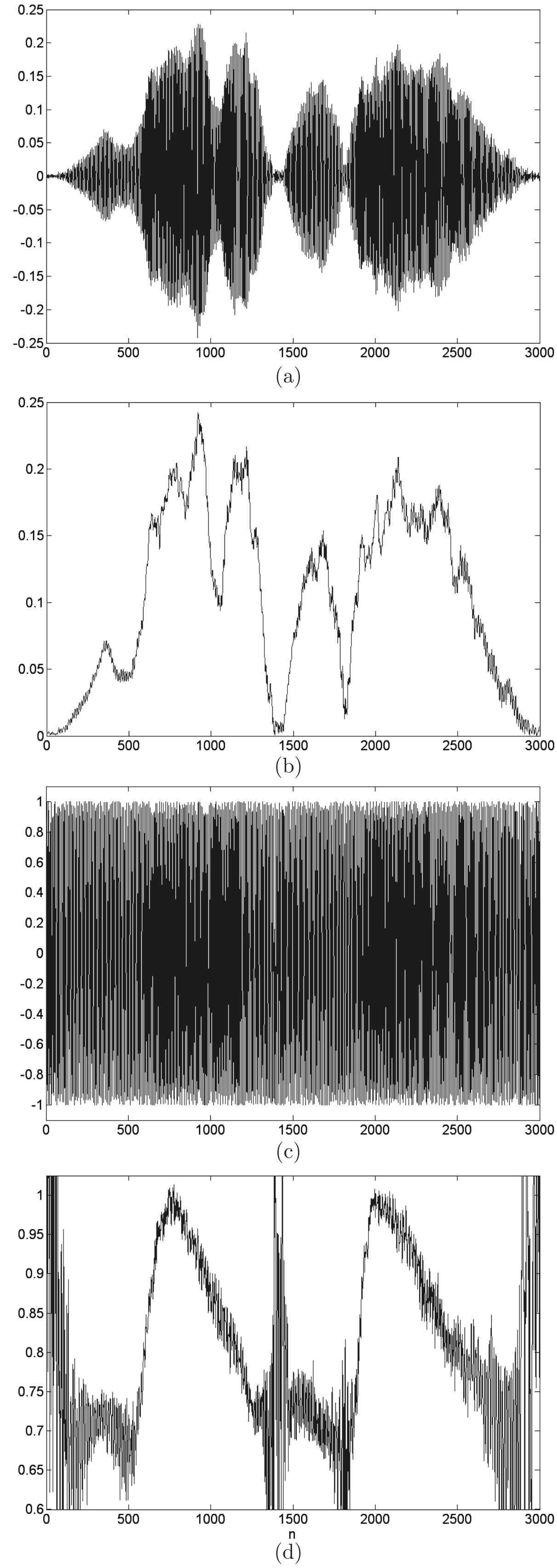

Fig. 5. The original canary sound record of 3000 samples (a), its AM (b) and FM (c) factors, and instantaneous angular frequency (d) of the chirp obtained by processing according to Fig. 1 with $\kappa = 1$

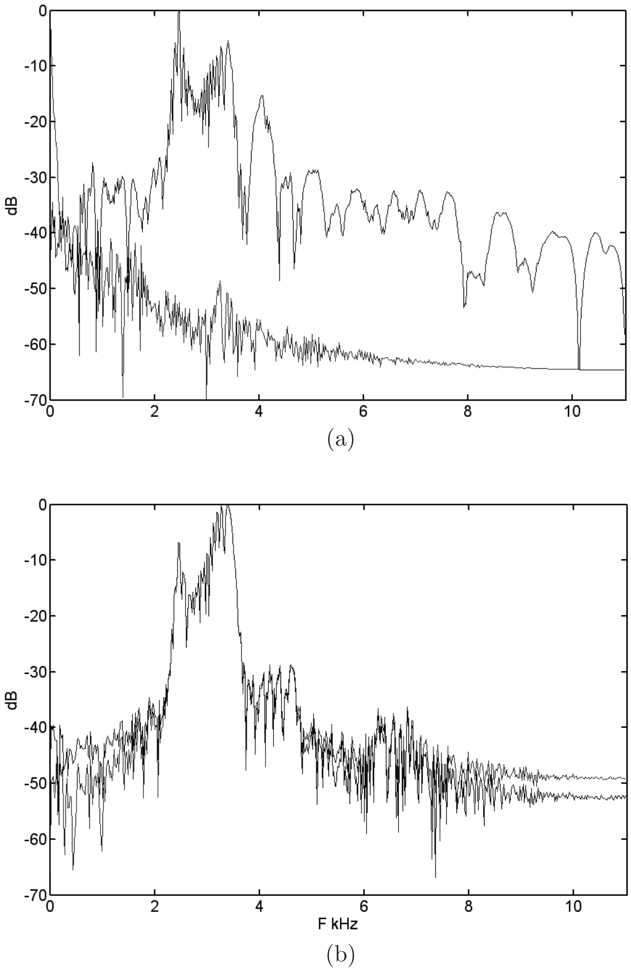

Fig. 6. Magnitude spectra of AM (upper) and FM (lower) factors of a canary chirp (a) and of original (upper) and recovered (lower) chirp (b) after direct, according to Fig. 1, and inverse processing with $\kappa = 1$

Table 1
ISO note names, frequencies in Hz and MIDI note numbers; the boldface number stands for the frequency relative to which all the values of pitch scaling factor κ were computed

Pitch	Frequency [Hz]	MIDI
A4	440.000	69
A#4	466.164	70
B4	493.883	71
C5	523.251	72
C#5	554.365	73
D5	587.330	74
D#5	**622.254**	75
E5	659.255	76
F5	698.457	77
F#5	739.989	78
G5	783.991	79
G#5	830.609	80
A5	880.000	81

All our experiments were performed in the MATLAB environment.

4. Conclusions

The CDR processing introduced in this paper appears to be a powerful means for pitch shifting of chirp-like sound signals fulfilling practically the assumption of the Bedrosian theorem. The quality of the CDR processing depends strongly on the quality of the Hilbert filter. The role of this filter in Fig.1 is twofold. Firstly, it creates a Hilbertian signal from a given real-valued signal. Secondly, it serves as an antialiasing filter.

The CDR remapper as a pitch shifter can serve, e.g., for entertainment. It exhibits an excellent performance as a generator of different melodies of calling signals, where the sound of each note in a melody is derived from a short recording of a voice of a chosen creature via the CDR processing.

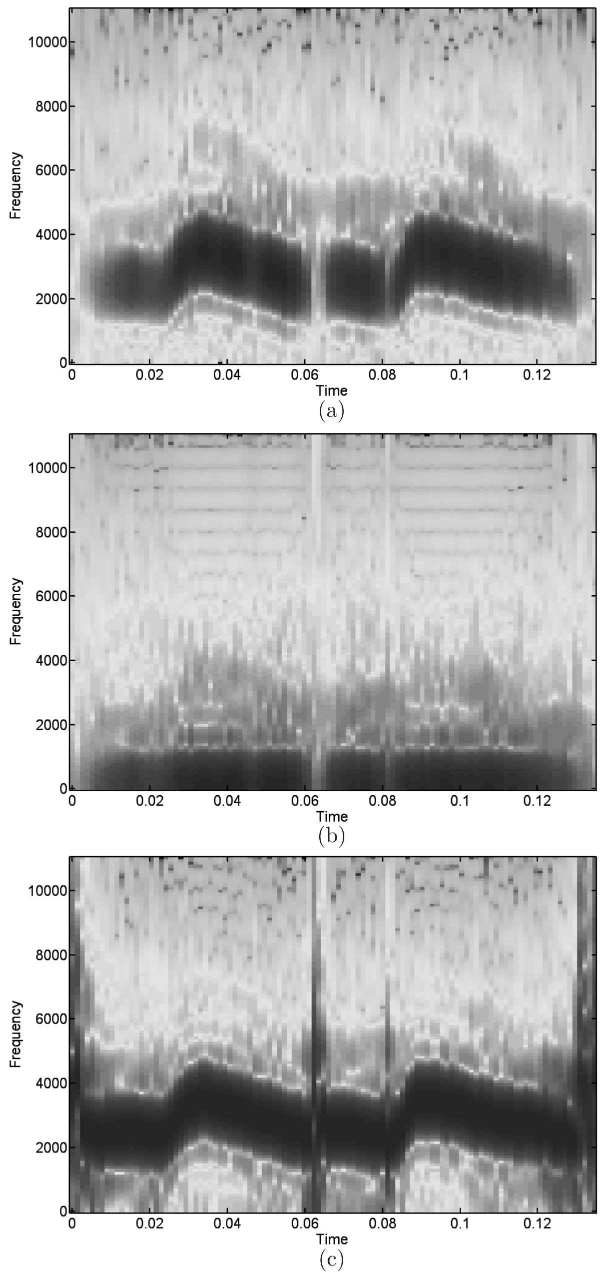

Fig. 7. Spectrograms of a canary chirp (a), its AM factor (b) and FM factor (c); frequency in Hz, time in seconds

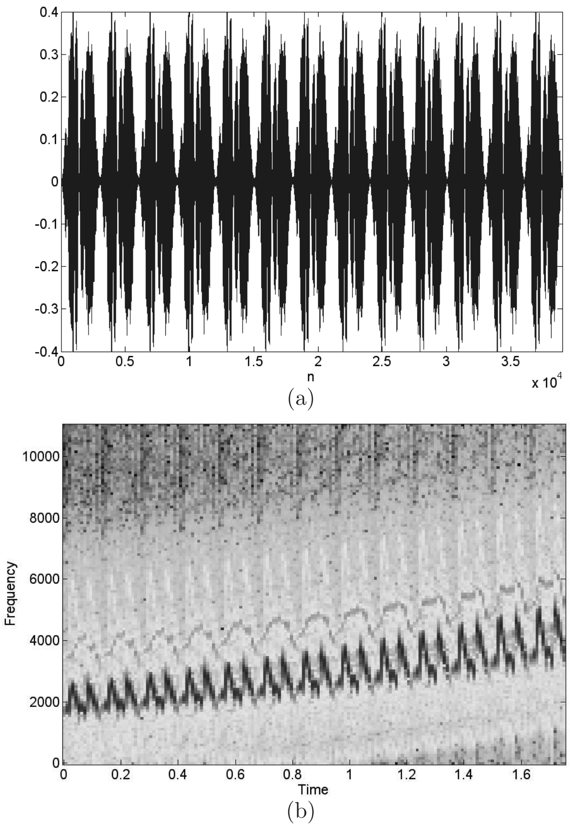

Fig. 8. Synthesised (a) canary scale and (b) its spectrogram

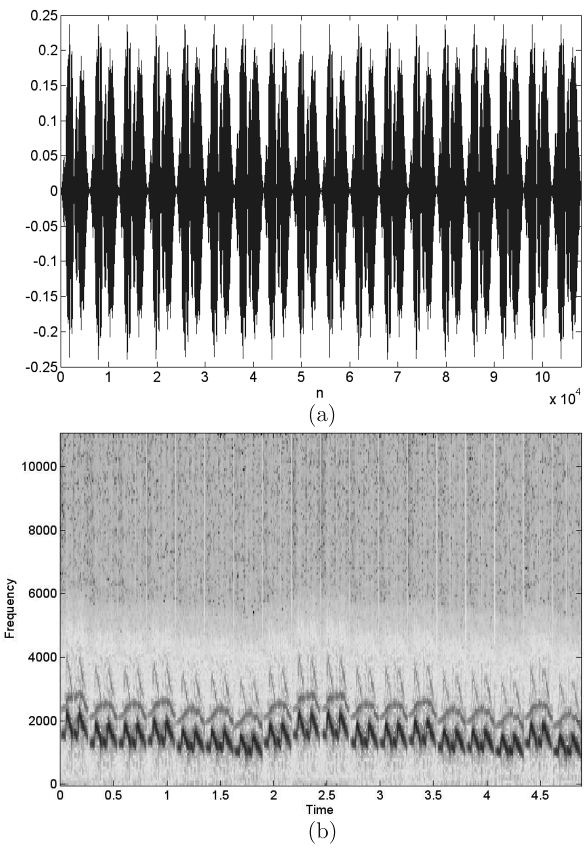

Fig. 9. Calling signal (a) synthesized by remapping original sound and (b) its spectrogram

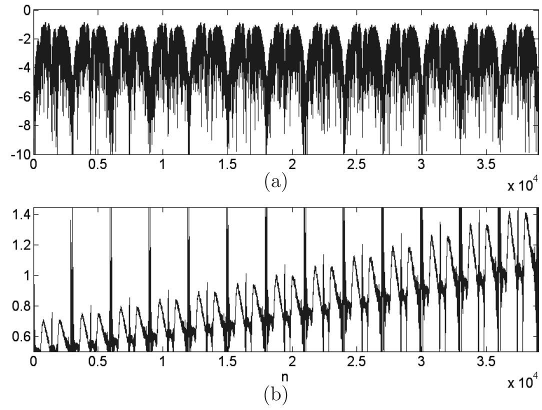

Fig. 10. Canary scale CDR components: (a) logenvelope and (b) instantaneous angular frequency waveform

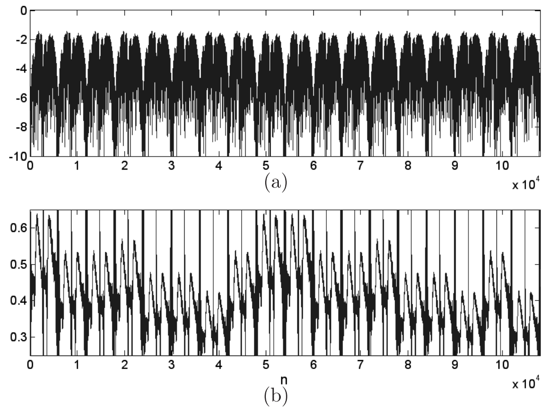

Fig. 11. Calling signal CDR components: (a) logenvelope and (b) instantaneous angular frequency waveform

REFERENCES

[1] E. Chassande-Mottin and P. Flandrin, "On the time-frequency detection of chirps", *Applied and Computational Harmonic Analysis* 2 (6), 252–281 (1998).

[2] V.F. Kroupa (ed.) *Direct Digital Frequency Synthesizers*, IEEE Press, 1999.

[3] J. Vankka, *Direct Digital Synthesizers: Theory, Design and Applications*, PhD Thesis, Helsinki Univ. of Technology, 2000.

[4] E. Hermanowicz, M. Rojewski, and E. Blok, "A memoryless direct digital synthesizer based on Taylor series approach", *Proc. of III National Conference on Electronics* I, 303–308 (2004).

[5] E. Hermanowicz, M. Rojewski, G.D. Cain, and A. Tarczynski, "On an instantaneous frequency estimator with FIR filters having maximally flat frequency response error magnitude", *Signal Processing* 81, 1491–1504 (2001).

[6] S.L. Hahn, *Hilbert Transforms in Signal Processing*, Artech House, 1996.

[7] D. Vakman, *Signals, Oscillations and Waves. A Modern Approach*, Artech House, 1998.

[8] L. Cohen, *Time-frequency Analysis*, Prentice Hall, 1995.

[9] J. H. Justice, "Analytical signal processing in music computation", *IEEE Trans.* 6 (27), 670–684 (1979).

[10] S. Wardle, "A Hilbert transformer frequency shifter for audio", *First Workshop on Digital Audio Effects DAFx*, Barcelona, Spain, 1998.

[11] S. Disch and U. Zolzer, "Modulation and delay line based digital audio effects", *2nd Workshop on Digital Audio Effects DAFx*, Trondheim, Norway, 1999.