Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 55 (2013) 40–50

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Speech recovery based on the linear canonical transform ☆

Wei Qiu, Bing-Zhao Li *, Xue-Wen Li

Mathematics Department of Beijing Institute of Technology, Beijing 100081, China

Received 16 May 2011; received in revised form 1 June 2012; accepted 4 June 2012
Available online 9 July 2012

Abstract

As is well known, speech signal processing is one of the hottest signal processing directions. There are exist lots of speech signal models, such as speech sinusoidal model, straight speech model, AM–FM model, gaussian mixture model and so on. This paper investigates AM–FM speech model by the linear canonical transform (LCT). The LCT can be considered as a generalization of traditional Fourier transform and fractional Fourier transform, and proved to be one of the powerful tools for non-stationary signal processing. This has opened up the possibility of a new range of potentially promising and useful applications based on the LCT. Firstly, two novel recovery methods of speech based on the AM–FM model are presented in this paper: one depends on the LCT domain filtering; the other one is based on the chirp signal parameter estimation to restore the speech signal in LCT domain. Then, experiments results are presented to verify the performance of the proposed methods. Finally, the summarization and the conclusion of the paper is given.

Keywords: Linear canonical transform; Chirp signal; The AM–FM model; Speech; Signal reconstruction

1. Introduction

As is known, a speech signal is a kind of typical non-stationary signal whose frequency components vary with time, which is consisting of vowels, consonants and transient parts. It is very important to detect its spectral components in many speech processing applications. The traditional tool for this estimation is spectral analysis by means of Fourier transform, however, this is a tool for stationary signals processing only and it simply indicates the global frequency components, but does not tell us when they occurred. This defection has seriously hindered the development of the speech signal processing. Focus on the speech signal processing, many time–frequency analysis tools have been proposed, for example, the short-time Fourier transform (STFT), the wavelet transform (WT), the fractional Fourier transform (FRFT) (Yin et al., 2008).

The speech signal, as a special kind of signal by some harmonic structures consisted, based on the mechanism of the production of speech, people has proposed some speech models, such as speech sinusoidal model, straight speech model, gaussian mixture model and so on (Yin et al., 2008; Teager and Teager, 1989). Besides, the speech also can be seen as a special kind of multicomponent chirp signals (Bovik et al., 1993). A chirp is an important signal in which the frequency increases (‘up-chirp’)or decreases (‘down-chirp’) with time. Chirp signal is a typical non-stationary signal, which widely appears in systems such as communication, radar, sonar, biomedicine and earthquake exploring system and also be used as a signal model for a lot of natural phenomena. Therefore, it is very important for chirp signal processing. The detection, the parameters estimation and the time–frequency representation for the multicomponent chirps are still an important research topic. In 1980s Teager and Teager (1989) discoveres by experiments that vortices could be the secondary source to excite the channel and produce the speech signal. Therefore, speech should be composed of the plane-wave-based linear part and a vortices-based non-linear part. According

☆ This work was supported by the National Natural Science Foundation of China (Nos. 60901058 and 61171195), and also supported partially by Beijing Natural Science Foundation (No. 1102029).
* Corresponding author. Tel.: +86 10 68912131 16.
E-mail address: li_bingzhaobit@bit.edu.cn (B.-Z. Li).

0167-6393/$ - see front matter © 2012 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.specom.2012.06.002

to such theory, Dimitriadis and Maragos (2005), Bovik et al. (1993), Santhanam and Maragos (2000), and Huang (1998) proposed an AM–FM modulation model for speech analysis, synthesis and coding. During the last thirty years, the four milestones among the best practices methods associated with the speech detection and recovery are MESA (Bovik et al., 1993), PASED (Santhanam and Maragos, 2000), Hilbert Huang Transform (Huang, 1998) and Gianfelici Transform (Gianfelici et al., 2007).

The paper Bovik et al. (1993) develops a multiband or wavelet approach for capturing the AM–FM components of modulated signals immersed in noise. It is demonstrated that the performance of the energy operator/ESA approach is vastly improved if the signal is first filtered through a bank of bandpass filters, and at each instant analyzed using the dominant local channel response. In the paper Santhanam and Maragos (2000), authors present a nonlinear algorithm for the separation and demodulation of discrete-time multicomponent AM–FM signals, it avoids the shortcomings of previous approaches and works well for extremely small spectral separations of the components and for a wide range of relative amplitude/power ratios. The paper Gianfelici et al. (2007) developes another new method for analyzing nonlinear and non-stationary signal processes. This decomposition method is adaptive, and, therefore, highly efficient. The proposed approach presented in the paper Gianfelici et al. (2007) is on the basis of a rigorous mathematical formulation, and the validity of this approach has been proven by some applications to both synthetic signals and natural speech.

In this paper, we introduce two novel methods to restore the speech signal with background noise based the AM–FM model associated with the linear canonical transform (LCT). The LCT is also known ABCD transform, generalized Fresnel transform (Aizenberg and Astola, 2006), Collins formula (Fan and Lu, 2006), generalized Huygens integral (Kostenbauder, 1990). It was proposed early in the seventies by Moshinsky and Quesne (2006) and Collins (1970) the special case with complex parameters was proposed by Bargmann (1961). LCT can be considered as a generalization of traditional Fourier transform and fractional Fourier transform. As same as the fractional Fourier transform, the LCT is used for analyzing Optical Systems and solving Differential Equations in the first. With 1990s' development of fractional Fourier transform, LCT has began to be taken seriously in the field of signal processing community. In such circumstances, we can analysis the signal in LCT domain, because from Pei and Ding (2002) and Sharma and Joshi (2008) we know that the one-dimensional LCT is linear integral transform of a three-parameter class, LCT can be more general and flexible than the Fourier transform as well as the fractional Fourier transform in some properties (Sharma and Joshi, 2008), and it can solve problems that cannot be dealt with well by the latter. It is shown in Tao et al. (2009), Li et al. (2012), Sharma and Joshi (2008), and Maragos etal. (1993) that the LCT is one of the powerful non-staionary signal processing tools, especially suitable for time-varying chirp signal processing. So it is most hopeful by use of LCT for this speech signal processing based AM–FM model. So far, LCT has many applications in signal processing community. For example, it has been applied to filter design, communication signals against multi-path and so on (Li et al., 2012; Sharma and Joshi, 2006; James and Agarwal, 1996).

The rest of the paper is organized as follows. In Section 2, the definition and some basic properties of LCT are briefly introduced. In Section 3, single chirp signal model is produced, then the AM–FM model of speech is described. In Section 4, based the AM–FM model we introduce two novel methods to restore the speech signal according to LCT. Section 5 presents the experimental results and discussion. Some conclusions and future research are given in Section 6.

2. Preliminary

When parameters (a, b, c, d) are real numbers, the LCT of a signal $f(t)$ defines as follows (Tao et al., 2009):

$$F_{(a,b,c,d)} = \begin{cases} \sqrt{\frac{1}{j2\pi b}} \int_{-\infty}^{\infty} K_{a,b,c,d}(u,t) f(t)\, dt & \text{if } b \neq 0 \\ \sqrt{d} \exp\left(\frac{jcd}{2} u^2\right) f(du) & \text{if } b = 0 \end{cases} \tag{1}$$

and $ad - bc = 1$, kernel function is defined as

$$K_{a,b,c,d} = \exp\left(\frac{jd}{2b} u^2\right) \exp\left(-\frac{j}{b} ut\right) \exp\left(\frac{ja}{2b} t^2\right), \tag{2}$$

Eq. 1 can be written as $F_{(a,b,c,d)} = L^{a,b,c,d}(f(t)) = L^A[f](u)$, the parameters

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

By the above definition it is proved in paper Moshinsky and Quesne (2006) that the LCT satisfies the additivity property of the parameters, that is

$$L^{a_2,b_2,c_2,d_2}[L^{a_1,b_1,c_1,d_1}(f(t))] = L^{e,f,g,h}(f(t)), \tag{3}$$

where

$$\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}.$$

Besides, the reversibility property is derived (Tao et al., 2009)

$$L^{d,-b,-c,a}[L^{a,b,c,d}(f(t))] = f(t). \tag{4}$$

When these parameters reduces to

$$(a, b, c, d) = (\cos\alpha, \sin\alpha, -\sin\alpha, \cos\alpha),$$

then LCT becomes into the FRFT by a fixed phase factor multiplied (Moshinsky and Quesne, 2006):

$$L^{\cos\alpha,\sin\alpha,-\sin\alpha,\cos\alpha}(f(t)) = \sqrt{\exp(-j\alpha)} F^{\alpha}(f(t)). \tag{5}$$