World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:7, No:7, 2013

A New Time-Frequency Speech Analysis Approach Based On Adaptive Fourier Decomposition

Liming Zhang

***Abstract*—**In this paper, a new adaptive Fourier decomposition (AFD) based time-frequency speech analysis approach is proposed. Given the fact that the fundamental frequency of speech signals often undergo fluctuation, the classical short-time Fourier transform (STFT) based spectrogram analysis suffers from the difficulty of window size selection. AFD is a newly developed signal decomposition theory. It is designed to deal with time-varying non-stationary signals. Its outstanding characteristic is to provide instantaneous frequency for each decomposed component, so the time-frequency analysis becomes easier. Experiments are conducted based on the sample sentence in TIMIT Acoustic-Phonetic Continuous Speech Corpus. The results show that the AFD based time-frequency distribution outperforms the STFT based one.

***Keywords*—**Adaptive fourier decomposition, instantaneous frequency, speech analysis, time-frequency distribution.

I. Introduction

THE representation of a signal in both the time (space) and frequency domain has been a challenge direction in signal processing areas, especially when the signals are time-varying non-stationary ones [1], [2]. Speech signal is an example of such signals. The frequency and amplitude of a speech signal varies with time. In many speech-analysis applications, it is important to identify the spectral contents of the speech signal. Challenging problems include speech enhancement and denoising, blind separation of two and more speakers and mixed audio sources, etc. [3].

Several time-frequency techniques have been used for the analysis of speech in the literature [4]-[7], including the Short-Time Fourier Transform (STFT) and the Wigner distribution. But these techniques have some limitations when being applied to speech signals.

The STFT and the associated spectrogram techniques are among the most commonly used methods in analyzing time-varying signals [4]. The STFT is a linear time-frequency representation that decomposes the signal into basic trigonometric functions by windowing the analyzed signal and applying the Fourier transform [7]. In STFT, the main difficulty is to find suitable short-time windows for signals with rapid spectral contents change, such as speech signals [1]. It needs a tradeoff between frequency and time resolution: good time resolution requires short windows, and good frequency resolution requires long ones [7].

The Wigner distribution is another popular time-frequency representation with a quadratic kernel. It distributes the energy of the signal over time and frequency, offering good time-frequency localization and preserving time-frequency shifts [5], [7]. There are two disadvantages being noted. In Wigner distribution, there is usually a smoothing process in frequency and/or time to alleviate the undesired effect of the quadratic cross-terms. It results in some important details of speech being hidden. In addition, the Wigner distribution has no easy inverse transform and is not suitable for filtering purposes [7].

A new signal decomposition theory – Adaptive Fourier Decomposition (AFD) has been proposed recently by Qian et al. with proven mathematical foundation [8], [9]. The theory is based on Gabor's analytic signal method [10]. By using Gabor's method, one can uniquely obtain an analytic quadrature signal and accordingly define the analytic amplitude and phase of the signal. It is a physical requirement that a frequency has to be non-negative. It is well accepted that only the analytic quadrature signals that possess non-negative phase derivatives can have instantaneous frequency (IF) with physical meanings. However, Gabor's method does not necessarily produce analytic signals of non-negative IF. Defining mono-components as signals of non-negative analytic phase derivatives, Qian et al. established a new mono-component function theory and correspondingly the concept adaptive mono-component decomposition [11]-[16]. They further proposed the mono-components bank consisting weighted Blaschke products to practically perform adaptive mono-component decomposition, called AFD [8], [9], [17].

AFD has the following advantages:

AFD has well-founded mathematical roots in harmonic analysis and analytic function theories, being similar with Fourier transform [8]. Indeed, Fourier transform is a special case of AFD when the corresponding parameters are all chosen to be 0. AFD has inverse transform that can reconstruct through the partial sums back to the signal in time domain. AFD keeps most good characteristics of Fourier transform.

AFD can represent time-varying non-stationary signals into the physically realizable basic signals with meaningful time-varying IFs. Meaningful IF means the phase derivative is non-negative function. The difference between Fourier transform and AFD is that basic trigonometric functions decomposed by Fourier transform are with constant IFs, but mono-components decomposed by AFD are with time-varying IFs. So Fourier transform can only provide constant time-frequency distribution, but AFD can provide time-varying time-frequency distribution. Meanwhile it doesn't cause parameter selection problems, such as window size selection in

L.Zhang is with University of Macau, Av.Padre Tomas Pereira, Taipa, Macao (phone: +853 83978467; fax: +853 28838314; e-mail: lmzhang@umac.mo).

STFT.

In contrast with the Fourier transform, which decomposes any signals into the same basic trigonometric functions, AFD decomposes the given signal into different mono-components. The mono-components decomposed are selected based on the given signal by using Maximal Selection Principle (MSP) [14], [15]. MSP means that for a given signal, the AFD algorithm starts with selecting a mono-component that is in the energy sense closest to the given signal. Then at each consecutive selection, it applies the same energy principle to find a mono-component that is the closest in the energy scale to the reduced remainder. It is the reason that the decomposition is said to be adaptive. Therefore the decomposition leads to fast convergence than what Fourier decomposition does.

To any given signal, AFD can represent it by using a bunch of plural numbers. For example, a signal is decomposed by AFD into weighted Blaschke products. The energy of the reconstructed partial sum from those n mono-components can be over 99% close to the energy of the original signal. In this situation, the given signal can be fully represented by 2n plural numbers. It has potential applications in speech signal compression and decompression.

The detailed AFD theory has been provided in [8], [9]. The AFD algorithm is given in [14]. The present paper is the very first one that introduces AFD theory and the corresponding time-frequency distribution into the speech signal analysis area. It focuses on the spectrogram analysis, which takes advantage of the clear time-varying frequency distribution provided by AFD. Experiments are conducted in comparison with STFT. AFD based time-frequency distribution performs more promising than STFT.

The paper is organized as follows. The AFD theory is briefly introduced in Section II. The principle of AFD based time-frequency distribution is presented in Section III. The experiment results are shown in Section IV. The conclusions are drawn in Section V.

II. Brief Introduction to AFD

For a real-valued signal f defined on the unit circle, AFD is to be applied to its Hardy space projection

$$f^+ = \frac{1}{2}(f + iHf) + \frac{c_0}{2} \tag{1}$$

where Hf is the circular Hilbert transform of f of the expression

$$f^+(e^{it}) = \sum_{k=1}^{\infty} < f,\ B_k > B_k(e^{it}), \tag{2}$$

where $a_k = argmax\{< f, B_{\{a_1,\cdots,a_{k-1},a\}} > | a \epsilon the unit disc\}$, for $a_1, \cdots, a_{k-1}$ being previously determined,

$$B_k(e^{it}) = B_{\{a_1,\cdots,a_k\}}(e^{it}) = \frac{\sqrt{1-|a_k|^2}}{1-\overline{a}_k e^{it}} \prod_{l=1}^{k-1} \frac{e^{it}-a_l}{1-\overline{a}_l e^{it}} \tag{3}$$

The algorithm stops at the step such that

$$\| f^+ \| - \sum_{k=1}^{N} |< f, B_k >|^2 < \varepsilon \tag{4}$$

for the first time. The approximation to the original real-valued function is

$$\begin{aligned} f &= 2Re\{\sum_{k=1}^{N} < f, B_k >\} - c_0 \\ &= \sum_{k=1}^{N} \rho_k(t) \cos\theta_k(t) - c_0, \end{aligned} \tag{5}$$

where $\rho_k(t)e^{i\theta_k(t)} = 2 < f, B_k > B_k(e^{it})$, $c_0 = \frac{1}{2\pi}\int_0^{2\pi} f(e^{it})dt.$

Practically we take the initial value. We note both the decompositions (2) and (4) are orthogonal.

This article makes use the algorithm code in the homepage of Qian: http://www.fst.umac.mo/en/staff/fsttq.html. There are two AFD algorithms. The algorithm used in this paper is the one with the main function, corresponding to best rational approximation.

III. Principles of AFD Based Time-Frequency Analysis

AFD has two salient characteristics. One is the well defined IF in each composing mono-component; and the other is time-frequency distribution that is densities represented in the time and frequency simultaneously.

A. IF Analysis of the Signal

Let f be a given real-valued signal and series (2) is its AFD. The instantaneous amplitude $\rho_k(t)$ is explicitly given in (5). Calculation gives

$$\begin{aligned} \theta_k'(t) &= \frac{|a_k|\cos(t-\theta_{a_k}) - |a_k|^2}{1 - 2|a_k|\cos(t-\theta_{a_k}) + |a_k|^2} \\ &+ \sum_{l=1}^{k-1} \frac{1-|a_l|^2}{1 - 2|a_l|\cos(t-\theta_{a_l}) + |a_l|^2}, \end{aligned} \tag{6}$$

where $a_k = |a_k| e^{i\theta_{a_k}}$. We can easily show $\theta_{k'}(t) < \theta_{k+1}'(t), \quad for\ each\ k.$

B. Frequency Spectrum Analysis of the Signal

The frequency spectrum is defined to be

$$E(\omega) = \sum_{all\ t,k\ such\ that\ \theta_k'(t)=\omega} \rho_k^2(t). \tag{7}$$

It accumulates all the amplitude values corresponding to a given frequency. The analysis in frequency spectrum is similar with that of Fourier frequency spectrum. In the latter ω has to

be integers, and, corresponds to each admissible ω there is only one spectrum that is a constant.

C. Spectrogram or Time-Frequency Distribution Analysis

Now we have a time-varying and frequency-varying spectrum. Combine them together; we get the time-frequency distribution.

IV. Experiment and Results

A. Decomposition Comparison between AFD and FD

A sample speech sentence "She had your dark suit in greasy wash water all year" is used in the experiment. It is chosen from TIMIT Acoustic-Phonetic Continuous Speech Corpus [18]. The TIMIT corpus of read speech is designed to provide speech data for acoustic-phonetic studies and for the development and evaluation of automatic speech recognition systems. It includes a 16-bit, 16kHz speech waveform file for each utterance. The word "water" is used to illustrate the experiment results. The "water" signal has total4096 sampling points. The original signal is shown in Fig. 1.

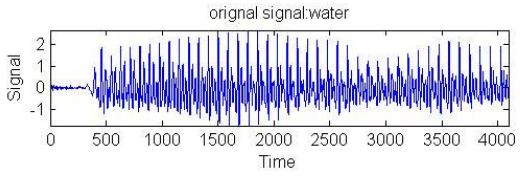
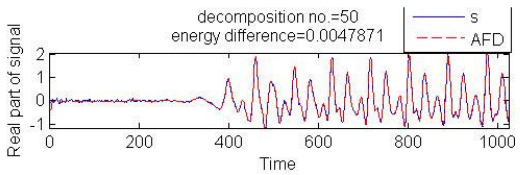


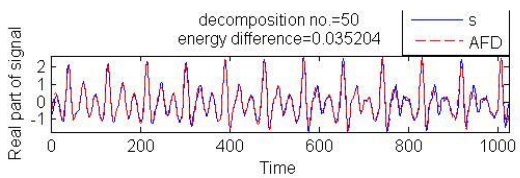
Fig. 1 Original speech signal

To get clearer spectrogram of the signal, the original signal is divided into 4 parts with 1024 sampling points in each part. The AFD decomposition and reconstruction signals of the 4 parts are illustrated in Fig. 2. The Fourier decomposition (FD) and reconstruction signals of the 4 parts are illustrated in Fig. 3.

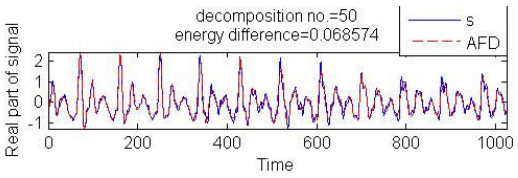
The decomposition number of both the AFD and FD are 50 times. Figs. 2 and 3 show that with the same decomposition number, the energy difference between the original signal and the reconstructed sum of AFD is over 10 less than FD is. The results for other selected words in the sentence are shown in Table I. This means that AFD has faster convergence than FD does.



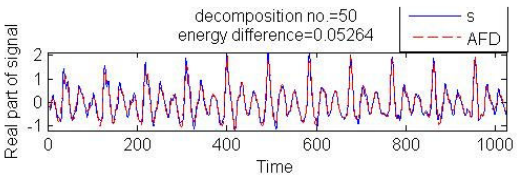
(a)



(b)

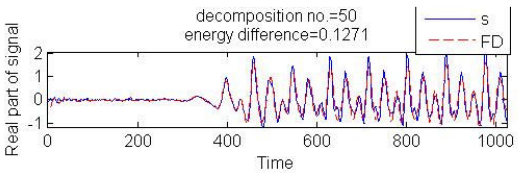


(c)

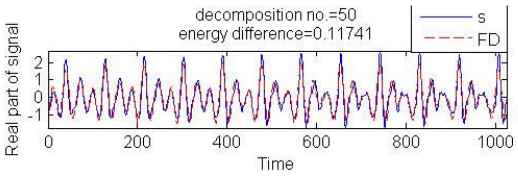


(d)

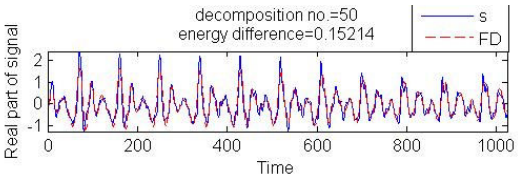
Fig. 2 AFD decomposed and reconstructed signal of (a) part 1, (b) part 2, (c) part 3, (d) part 4



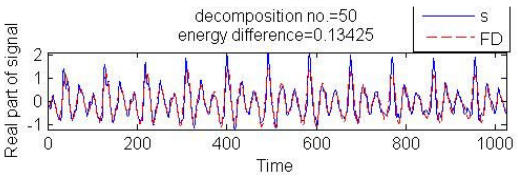
(a)



(b)



(c)



(d)

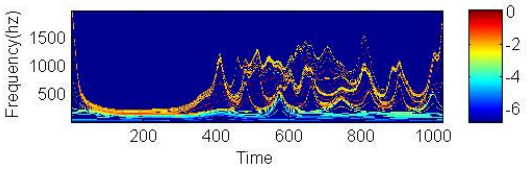
Fig. 3 FD decomposed and reconstructed signal of (a) part 1, (b) part 2, (c) part 3, (d) part 4

TABLE I
ENERGY DIFFERENCE

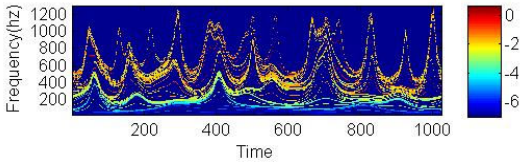
Tested words	AFD	FD	Comparison (FD/AFD)
she	0.0056	0.2435	43.5
your	0.0047	0.0768	16.3
greasy	0.0002	0.1086	543
wash	0.0004	0.0433	108.3

B. Time-Frequency Distribution (TFD) Comparison between AFD and STFT

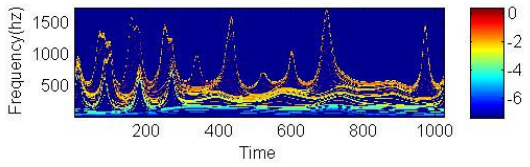
AFD based TFD and STFT based TFD of the 4 parts of "water" are illustrated in Figs. 4 and 5, respectively. Fig. 4 shows clear time-varying frequencies; however, time-varying characteristics are not so obvious in STFT based TFD. The time-varying properties provide the potential to analysis the speech difference between two persons. It also provides the potential filtering possibility for different frequency bands.



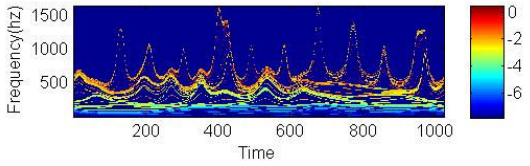
(a)



(b)

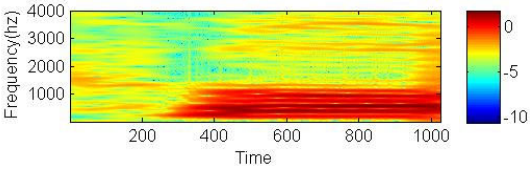


(c)

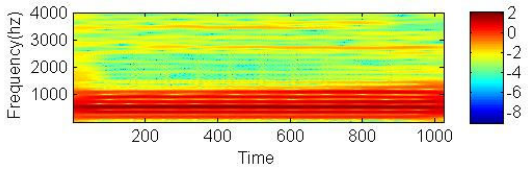


(d)

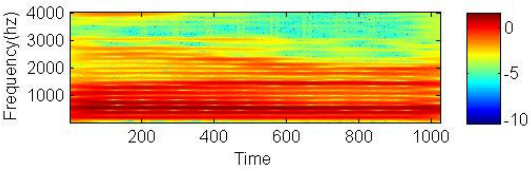
Fig. 4 AFD based TFD of (a) part 1 (b) part 2 (c) part 3 (d) part 4



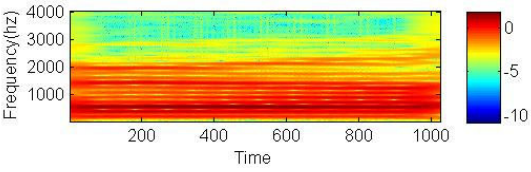
(a)



(b)



(c)



(d)

Fig. 5 STFT based TFD of (a) part 1 (b) part 2 (c) part 3 (d) part 4

The experiments are also conducted to other words in the sample sentences. The AFD based and STFT based TFD for the whole word "greasy" are illustrated in Figs. 6 and 7, respectively. The window sizes of STFT are changed between $\frac{N}{2} \sim \frac{N}{8}$, STFT cannot provide high resolution TFD no matter what window widths are chosen.

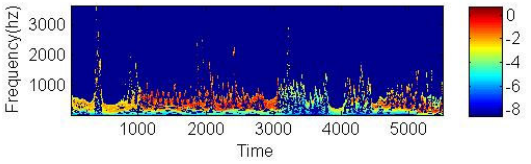


Fig. 6 AFD based TFD of "greasy"

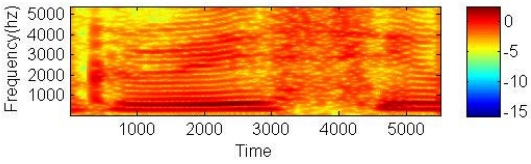


Fig. 7 STFT based TFD of "greasy"

V. CONCLUSIONS

AFD is a new signal decomposition theory. It has two salient characteristics which are distinguished from other signal decomposition approaches. One is the well-defined IF in each composing mono-component; and the other is the time-varying frequency distribution. This paper seeks to introduce AFD as a new time-frequency analysis approach into the speech signal analysis area. The experiment results demonstrate that the AFD based TFD provides more detailed time-varying frequency

distribution than STFT based TFD does.

ACKNOWLEDGMENT

This study is supported by the UM research grant MYRG144(Y3-L2)-FST11-ZLM.

REFERENCES

[1] L. Cohen, Time-Frequency Analysis: Theory and Applications, Prentice, Hall, 1995.
[2] S. Qian and D. Chen, "Joint Time-Frequency Analysis", IEEE Signal Processing Magazine, pp. 53-67, March, 1999.
[3] H. Choi and W.J. Williams, "Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernel," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.37, No.6, pp. 862-871, June 1989.
[4] L. Cohen, "Time-Frequency Distributions -- A Review", Proceedings of the IEEE, vol.77, No.7, pp.941-981, 1989.
[5] T.A.C.M. Claasen and W.F.G. Mecklenbrauker, "The Wigner distribution-a tool for time-frequency signal analysis, Part III--Relations with other time-frequency signal transformations", Philips Journal of Research, Vol. 35, No. 6. pp.372-389, 1980.
[6] B. Boashash, Time-Frequency Signal Analysis and Processing – A Comprehensive Reference, Elsevier Science, Oxford, 2003.
[7] M.Kepesi and L. Weruaga, "Adaptive chirp-based time–frequency analysis of speech signals", Speech Communication, 48, pp. 474-492, 2006.
[8] T. Qian, "Intrinsic Mono-components Decomposition of Functions: An Advance of Fourier Theory", Math.Meth.Appl.Sci. 33, pp.880-891, 2010.
[9] T. Qian and Y. Wang, "Adaptive Fourier Series – a Variation of Greedy Algorithm", Advances in Computational Mathematics, 34, no.3, pp.279-293, 2011.
[10] D. Gabor, "Theory of Communication", Journal of the IEE, vol.93, pp.~429-457, 1946.
[11] T. Qian, Q. H. Chen and L.Q. Li, "Analytic unit quadrature signals with non-linear phase", Physica D: Nonlinear Phenomena, 303, 80-87 2005.
[12] T. Qian, "Characterization of boundary values of functions in Hardy spaces with applications in signal analysis", Journal of Integral Equations and Applications, Volume 17, Number 2, pp 159-198, 2005.
[13] T. Qian, "Analytic Signals and Harmonic Measures", J. Math. Anal. Appl. 314, pp.526-536, 2006.
[14] T. Qian, L. Zhang and Z. Li, "Algorithm of Adaptive Fourier Transform", IEEE Transactions on Signal Processing, vol.59(12),5899-5906, Dec.,2011.
[15] T. Qian, "Mono-components for decomposition of signals", Mathematical Methods in the Applied Sciences, 29, pp. 1187-1198, 2006..
[16] T. Qian, "Boundary Derivatives of the Phases of Inner and Outer Functions and Applications", Mathematical Methods in the Applied Sciences, 32, pp. 253-263, 2009.
[17] T. Qian and E. Wegert, "Optimal Approximation by Blaschke Forms", Complex Variables and Elliptic Equations, preprint Available online: 20 Jun 2011.
[18] TIMIT Acoustic-Phonetic Continuous Speech Corpus, http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1 (Retrieved on 27/11/2012)

L. Zhang(M'10) received the B.S. degree from Nankai University, China, in 1987. She received the M.S. degree from the East China Institute of Technology, China, in 1990, and the PhD in computer science from the University of New England, Australia, 2002.

She worked in Tianjin Institute of Technology, China, from 1990 to 1998 as lecturer and associate professor. She participated in a number of national projects, including 863 project, major in Robert version. She joined the Faculty of Education, University of Macau in 2001 as lecturer and assistant professor till 2010 and worked in the direction of IT in education. From 2007 to 2010, she acted as the Director of the Information and Communication Education Technology Center. She organized several international conferences and workshops in IT in education area and published a series of papers, book chapters, and a co-edited book. She started to work in the Faculty of Science and Technology, University of Macau as an assistant professor from 2010. Her recent research interests including signal processing and image processing.