様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成　24　年　5　月　18　日現在

機関番号：82636
研究種目：若手研究（B）
研究期間：2010～2011
課題番号：22700193
研究課題名（和文）Ｈｉｌｂｅｒｔ再生核空間の正規法による頑健音声処理
研究課題名（英文）Regularization in a reproducing kernel Hilbert space for robust speech processing

研究代表者
ルー　シュガン（Lu Xugang）
独立行政法人情報通信研究機構
ユニバーサルコミュニケーション研究所 音声コミュニケーション研究室 専攻研究員
研究者番号：20362022

研究成果の概要（和文）：従来の音声特徴抽出では、線形で次数の低い音声統計構造を扱っている。このように抽出されたものの特徴は、ノイズと音声が混在する場合、ロバスト性が不足することである。本プロジェクトの目的は、再生核ヒルベルト空間（Kernel Hilbert Space: RKHS）における非線形マッピング関数を用いて、ノイズによる干渉を受けにくい非線形で高次の音声統計情報を抽出する、信号処理手法を提案することである。音声増強、音声認識、音声有無の検出に本提案手法を適用し有効であることで確認した.

研究成果の概要（英文）：
In most current speech feature extraction methods, only low-order statistical information is extracted. In order to improve the robustness, we proposed a speech signal processing method in a reproducing kernel Hilbert space (RKHS). We first proved a theoretical analysis of the proposed framework, and showed a connection of the trade-off problem in machine learning and noise reduction. (From theoretical aspect, we proved that the trade-off problem in machine learning and noise reduction is essentially the same). Based on the theoretical analysis, we applied the framework for speech enhancement, and voice activity detection problems. Based on our application experiments, we showed that the proposed framework can be well used on speech processing and obtained improvement compared with several traditional noise reduction methods.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	1,900,000	570,000	2,470,000
2011 年度	1,200,000	360,000	1,560,000
年度			
年度			
年度			
総　計	3,100,000	930,000	4,030,000

研究分野：知覚情報処理・知能ロボティクス
科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス
キーワード：知能ロボット

１．研究開始当初の背景

Feature extraction can be regarded as to find a transform function based on optimizing (minimizing) a cost function. Usually, the cost function is measured by the reconstruction error which is a function of the input signal and the transform function. Generally speaking, the feature extraction is an ill-posed inverse problem since two different acoustic signals may have the same representation (i.e., one to many inverse problem). In order to make the problem to be well-posed, constraints on the transform function should be given. For example, the transform function is chosen as a linear operator to estimate the second-order correlation information of the signals. Most currently used feature representations, for example, the linear prediction coefficient (LPC), Mel frequency cepstral coefficient (MFCC) are such kinds of representations. These representations, however, are not robust to noise interferences. We show an example in Fig.1. Left panel is the representation for clean speech while the right one is for the noise condition (train noise with SNR 5dB). From this figure, we can see that the speech and noise can not be separated in this representational space, i.e., the speech is easily distorted by the noise interference.

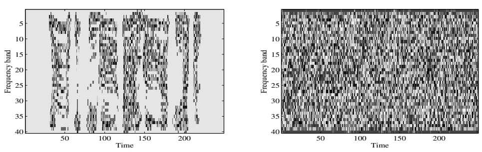

Fig.1 MFCC spectrum of a clean speech signal (left panel) and of the noisy one (right panel)

２．研究の目的

Speech signals are a special type of acoustic signals. They are produced by the speech production organs carrying linguistic information. They have special statistical regularities which are different from other kinds of acoustic signals. The traditional linear transform can not reflect their nonlinear and high-order statistical information, hence is easy to be distorted by noise. In this project, we try to design a new type of signal processing framework which explores the nonlinear and high-order statistical information of speech, and is not easy to be distorted by noise interferences.

３．研究の方法

We propose a new type of feature extraction method that uses a nonlinear transform function to explore the nonlinear and high-order statistical information. But the estimation of the nonlinear transform function from noisy and limited observations is an ill-posed problem. We need to put constraints on the functional space for solving the ill-posed problem. The regularization theory is widely used to solve the ill-posed inverse problem. In this study, we propose to estimate the nonlinear transform function in a regularized reproducing kernel Hilbert space (RKHS). The detailed procedures are:

- Recording the clean speech signal corpus as well as the noisy one based on which the nonlinear transform function is estimated in the framework.
- Designing the data-dependent kernel functions which are suitable for noisy robust speech feature extraction.
- Designing the constraints on the nonlinear function which can bring robustness in speech feature extraction.
- Optimal selecting the regularization parameter by consideration of the trade-off

between the noise reduction and speech distortion.

- Applications and evaluations of the proposed algorithm, i.e., extracting a new type of feature for robust speech applications. In this project, we will test the algorithm on noisy robust automatic speech recognition (ASR) and voice activity detection (VAD) since these two applications in noise environments are still challenge problems in speech engineering.

４．研究成果

The tradeoff between noise reduction and speech distortion is a key concern in designing noise reduction algorithms. This study proposes a novel framework for noise reduction by considering this tradeoff. We regard speech estimation as a function approximation problem in a regularized reproducing kernel Hilbert space (RKHS). In the estimation, the objective function is formulated to find an approximation function by controlling the tradeoff between approximation accuracy and function complexity. For noisy observations, this is equivalent to controlling the tradeoff between noise reduction and speech distortion. Since the target function is approximated in an RKHS, either a linear or nonlinear mapping function can be naturally incorporated in the estimation by a "kernel trick". Traditional signal subspace and Wiener filtering based noise reduction can be derived as special cases when a linear kernel function is applied in this framework. We first provided a theoretical analysis of the tradeoff property of the framework in noise reduction. Then we applied our proposed noise reduction method in speech enhancement and noisy robust speech recognition experiments. Compared with several classical noise reduction methods, our proposed method showed promising advantages.

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計３　件）

① X. Lu, M. Unoki, S. Nakamura, "Subband temporal modulation envelopes and their normalization for automatic speech recognition in reverberant environments," Elsevier, Computer Speech &Language, vol. 25, no. 3, pp. 571-584, 2011. (査読有)

DOI: http://dx.doi.org/10.1016/j.csl.2010.10.002

② X. Lu, S. Matsuda, M. Unoki, S. Nakamura, "Temporal modulation normalization for robust speech feature extraction and recognition," Springer, Journal of Multimedia Tools and Applications, vol. 52, no.1, pp. 187-199, 2011. (査読有)

DOI: http://dx.doi.org/10.1007/s11042-010-0465-7

③ X. Lu, S. Matsuda, M. Unoki, S. Nakamura, "Temporal modulation contrast normalization and edge-preserved smoothing for robust speech recognition," Elsevier, Speech Communication, vol. 52, no. 1, pp. 1-11, 2010. (査読有)

DOI: http://dx.doi.org/10.1016/j.specom.2009.08.006

〔学会発表〕（計５件）

① X. Lu, M. Unoki, R. Isotani, H. Kawai, and S. Nakamura, "Adaptive regularization framework for robust voice activity detection," INTERSPEECH, pp.2653-2656, Firenze Fiera venue, Florence, Italy, Aug. 17, 2011.

② X. Lu, M. Unoki, R. Isotani, H. Kawai, and S. Nakamura, "Functional approximation in a reproducing kernel Hilbert space for speech estimation in noisy environments," the Second Asia-Pacific Signal and Information Association Annual Summit and Conference (APSIPA), pp. 76-81, Singapore, Biopolis venue, Singapore, Dec. 16, 2010.

③ X. Lu, M. Unoki, R. Isotani, H. Kawai, and S. Nakamura, "Speech enhancement as a functional approximation and generalization," International Symposium on Chinese Spoken Language Processing. pp. 18-22, National Cheng Kung University, Tainan, Taiwan, Dec. 1, 2010.

④ X. Lu, M. unoki, R. Isotani, H. Kawai, and S. Nakamura, "Regularization in a reproducing kernel Hilbert space for noisy robust voice activity detection," the 10th International Conference on Signal Processing (ICSP), pp. 585-588, Northern Jiaotong University, Beijing, China, Oct. 26, 2010.

⑤ X. Lu, M. Unoki, R. Isotani, H. Kawai, and S. Nakamura, "Voice activity detection in a regularized reproducing kernel Hilbert space," INTERSPEECH, pp. 3086-3089, International Conference Hall, Makuhari Messe International Convention Complex, Chiba, Japan, Sept. 29, 2010.

６．研究組織

(1)研究代表者

ルー シュガン (Lu Xugang)

独立行政法人情報通信研究機構 ・ ユニバーサルコミュニケーション研究所 音声コミュニケーション研究室 ・専攻研究員

研究者番号：20362022

(2)研究分担者

( )

研究者番号：

(3)連携研究者

( )

研究者番号：