Session ThAA Noise Mitigation, Speech Enhancement II

Chairperson Bayya Yegnanarayana IIT MADRAS, India

Home

NOISY SPEECH ENHANCEMENT BY FUSION OF AUDITORY AND VISUAL INFORMATION: A STUDY OF VOWEL TRANSITIONS

Authors: L. Girin, G. Feng & J.L. Schwartz

Institut de la Communication Parlée, UPRESA 5009 INPG/ENSERG/Université Stendhal B.P. 25, 38040 GRENOBLE CEDEX 09, FRANCE E-mail : girin@icp.grenet.fr

Volume 5 pages 2555 - 2558

ABSTRACT

This paper deals with a noisy speech enhancement technique based on the fusion of auditory and visual information. We first present the global structure of the system, and then we focus on the tool we used to melt both sources of information. The whole noise reduction system is implemented in the context of vowel transitions corrupted with white noise. A complete evaluation of the system in this context is presented, including distance measures, gaussian classification scores, and a perceptive test. The results are very promising.

SPECTRAL SUBTRACTION USING A NON-CRITICALLY DECIMATED DISCRETE WAVELET TRANSFORM

Authors: Andreas Engelsberg and Thomas Gulzow

Institute for Network and System Theory, Technical Department, Kiel University, Kaiserstrasse 2, D-24143 Kiel / Germany, E-mail: ae@techfak.uni-kiel.de and tg@techfak.uni-kiel.de

Volume 5 pages 2559 - 2562

ABSTRACT

The method of spectral subtraction has become very popular in speech enhancement. It is performed by modifying the spectral amplitudes of the disturbed signal. The spectral analysis of the signal is usually done by a Discrete Fourier Transformation (DFT). We propose a spectral transformation with nonuniform bandwidth to take into account the characteristics of the human ear. The spectral analysis and synthesis is performed by a non-critically decimated discrete wavelet transform. Critical subsampling is not performed to avoid errors due to aliasing. A significant drawback of spectral-subtraction methods are tonal residual noises in speech pauses with unnatural sound. The application of the proposed wavelet transform results in reduced residual noise with subjectively more comfortable sound.

BAYESIAN AFFINE TRANSFORMATION OF HMM PARAMETERS FOR INSTANTANEOUS AND SUPERVISED ADAPTATION IN TELEPHONE SPEECH RECOGNITION

Authors: Jen-Tzung Chien (a), Hsiao-Chuan Wang (a) and Chin-Hui Lee (b)

(a) Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan (b) Multimedia Communications Research Lab, Bell Laboratories, Murray Hill, USA chien@speech.ee.nthu.edu.tw hcwang@ee.nthu.edu.tw chl@research.bell-labs.com

Volume 5 pages 2563 - 2566

ABSTRACT

This paper proposes a Bayesian affine transformation of hidden Markov model (HMM) parameters for reducing the acoustic mismatch problem in telephone speech recognition. Our purpose is to transform the existing HMM parameters into its new version of specific telephone environment using affine function so as to improve the recognition rate. The maximum a posteriori (MAP) estimation which merges the prior statistics into transformation is applied for estimating the transformation parameters. Experiments demonstrate that the proposed Bayesian affine transformation is effective for instantaneous adaptation and supervised adaptation in telephone speech recognition. Model transformation using MAP estimation performs better than that using maximum-likelihood (ML) estimation.

INTEGRATED BIAS REMOVAL TECHNIQUES FOR ROBUST SPEECH RECOGNITION \Lambda

Authors: Craig Lawrence and Mazin Rahim (1)

University of Maryland, College Park, MD 20742 (1)AT&T; Labs-Research, Murray Hill, NJ 07974

Volume 5 pages 2567 - 2570

ABSTRACT

In this paper, we present a family of maximum likelihood (ML) techniques that aim at reducing an acoustic mismatch between the training and testing conditions of hid- den Markov model (HMM)-based automatic speech recognition (ASR) systems. We propose a codebook-based stochastic matching (CBSM) approach for bias removal both at the feature level and at the model level. CBSM associates each bias with an ensemble of HMM mixture components that share similar acoustic characteristics. It is integrated with hierarchical signal bias removal (HSBR) and further extended to accommodate for N-best candidates. Experimental results on connected digits, recorded over a cellular network, shows that the proposed system reduces both the word and string error rates by about 36% and 31%, respectively, over a baseline system not incorporating bias removal.

ACOUSTIC FRONT ENDS FOR SPEAKER-INDEPENDENT DIGIT RECOGNITION IN CAR ENVIRONMENTS

Authors: D. Langmann, A. Fischer, F. Wuppermann, R. Haeb-Umbach, T. Eisele

Philips GmbH Forschungslaboratorien Aachen P.O. Box 50 01 45 D-52085 Aachen Germany Email: flangmann,afischer,wupper,haeb,eiseleg@pfa.research.philips.com

Volume 5 pages 2571 - 2574

ABSTRACT

This paper describes speaker-independent speech recognition experiments concerning acoustic front end processing on a speech database that was recorded in 3 different cars. We investigate different feature analysis approaches (mel-filter bank, mel-cepstrum, perceptually linear predictive coding) and present results with noise compensation techniques based on spectral subtraction. Although the methods employed lead to considerable error rate reduction the error analysis shows that low signal-to-noise ratios are still a problem.

SIGNAL BIAS REMOVAL USING THE MULTI-PATH STOCHASTIC EQUALIZATION TECHNIQUE

Authors: Lionel Delphin-Poulat and Chafic Mokbel

FT.CNET/DIH/RCP 2 av. Pierre Marzin, 22307 Lannion cedex, France. Tel. +33 2 96 05 13 47 FAX: +33 2 96 05 35 30 e-mail : delphinp@lannion.cnet.fr

Volume 5 pages 2575 - 2578

ABSTRACT

We propose using Hidden Markov Models (HMMs) associated with the cepstrum coefficients as a speech signal model in order to perform equalization or noise removal. The MUlti-path Stochastic Equalization (MUSE) framework allows one to process data at the frame level: it is an on-line adaptation of the model. More precisely, we apply this technique to perform bias removal in the cepstral domain in order to increase the robustness of automatic speech recognizers. Recognition experiments on two databases recorded on both PSN and GSM networks show the efficiency of the proposed method.

SUBBAND ECHO CANCELLATION IN AUTOMATIC SPEECH DIALOG SYSTEMS

Authors: Andrej Miksic and Bogomir Horvat

Laboratory for Digital Signal Processing Faculty of Electrical Engineering and Computer Science University of Maribor, Smetanova 17, 2000 Maribor, Slovenia Tel. +386 62 221112, E-mail: andrej.miksic@uni-mb.si

Volume 5 pages 2579 - 2582

ABSTRACT

Echo cancellation has been most widely studied for hands-free telephony and for cancelling line echos in telephone central offices. The problem of echo cancelling in speech dialog systems is similar, however it has some specific requirements. In this contribution, a subband echo cancellation structure is proposed which can be integrated in the feature extraction part of a recognizer. A NLMS gradient-based adaptation is performed in frequency subbands that can either be derived directly from FFT analysis of input speech signal, or by using a proposed reduced-subband approach where the number of subbands is reduced in order to lessen the aliasing effect of the FFT. A double-talk detector is proposed based on the estimated error function for decision on stopping the adaptation. Finally, a new approach of combining echo cancellation and noise reduction is proposed.

Speech Enhancement via Energy Separation

Authors: Hesham Tolba and Douglas O'Shaughnessy

Institut National de la Recherche Scientifique, INRS-Telecommunications, Quebec, Canada. E-mail: tolba@inrs-telecom.uquebec.ca and dougo@inrs-telecom.uquebec.ca.

Volume 5 pages 2583 - 2586

ABSTRACT

This work presents a novel technique to enhance speech signals in the presence of interfering noise. In this paper, the amplitude and frequency (AM- FM) modulation model [7] and a multi-band analysis scheme [5] are applied to extract the speech signal parameters. The enhancement process is performed using a time-warping function B(n) that is used to warp the speech signal. B(n) is extracted from the speech signal using the Smoothed Energy Operator Separation Algorithm (SEOSA) [4]. This warping is capable of increasing the SNR of the high frequency harmonics of a voiced signal by forcing the the quasiperiodic nature of the voiced component to be more periodic, and consequently is useful for extracting more robust parameters of the signal in the presence of noise.

A Method of Signal Extraction from Noisy Signal

Authors: Masashi UNOKI and Masato AKAGI

unoki@jaist.ac.jp akagi@jaist.ac.jp School of Information Science, Japan Advanced Institute of Science and Technology 1-1 Asahidai, Tatsunokuchi, Ishikawa 923-12, Japan

Volume 5 pages 2587 - 2590

ABSTRACT

This paper presents a method of extracting the desired signal from a noise-added signal as a model of acoustic source segregation. Using physical constraints related to the four regularities proposed by Bregman, the proposed method can solve the problem of segregating two acoustic sources. Two simulations were carried out using the following signals: (a) a noise-added AM complex tone and (b) a noisy synthetic vowel. It was shown that the proposed method can extract the desired AM complex tone from noise- added AM complex tone in which signal and noise exist in the same frequency region. The SD was reduced an average of about 20 dB. It was also shown that the proposed method can extract a speech signal from noisy speech.

MULTI-CHANNEL NOISE REDUCTION USING WAVELET FILTER BANK

Authors: SIKA Jiri - DAVIDEK Vratislav

Faculty of Electrical Engineering Czech Technical University Prague, Czech Republic. Tel. +420 2 24352291, FAX: +420 2 24310784 , E-mail: sika@feld.cvut.cz

Volume 5 pages 2591 - 2594

ABSTRACT

This paper deals with the problem of estimation of a speech signal corrupted by an additive noise when observations from two microphones are available. The basic method for noise reduction using the coherence function is modified by using wavelets. The both observations are splitted by filter bank in five narrow bands through the whole used bandwidth (0...4kHz). The coherence functions are then computed for each band and the output speech estimation is reconstructed.

SPEECH SIGNAL DETECTION IN NOISY ENVIRONEMENT USING A LOCAL ENTROPIC CRITERION

Authors: I. Abdallah, S. Montrésor and M. Baudry

Laboratoire d'Informatique de l'Université du Maine Email : imad@lium.univ-lemans.fr

Volume 5 pages 2595 - 2598

ABSTRACT

This paper describes an original method for speech/non-speech detection in adverse conditions. Firstly, we define a time-dependent function called Local Entropic Criterion [1] based on Shannon's entropy [2]. Then we present the detection algorithm and show that at Signal to Noise Ratio (SNR) above 5 dB, it offers a segmentation comparable to the one obtained in clean conditions. We finally, describe how at very low SNR (< 0 dB) , it permits to detect speech units masked by noise.

A New Algorithm for Robust Speech Recognition: The Delta Vector Taylor Series Approach

Authors: Pedro J. Moreno and Brian Eberman

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.