The Wayback Machine - https://web.archive.org/web/20000311093614/http://www.wcl2.ee.upatras.gr:80/eurt…

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-

Sony v. Jawbone
U.S. Patent No. 8,019,091
Sony Ex. 1017

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

email: pjm@crl.dec.com, bse@crl.dec.com Digital Equipment Corporation Cambridge Research Laboratory

Volume 5 pages 2599 - 2602

ABSTRACT

In this paper we present a new model-based compensation technique called Delta Vector Taylor Series (DVTS). This new technique is an extension and improvement over the Vector Taylor Series (VTS) approach [7] that addresses several of its limitations. In particular, we present a new statistical representation for the distribution of clean speech feature vectors based on a weighted vector codebook. This change to the underlying probability density function (PDF) allows us to produce more accurate and stable solutions for our algorithm. The algorithm is also presented in a EM-MAP framework where some the environmental parameters are treated as random variables with known PDF's. Finally, we explore a new compensation approach based on the use of convex hulls. We evaluate our algorithm in a phonetic classification task on the TIMIT [5] database and also in a small vocabulary size speech recognition database. In both databases artificial and natural noise is injected at several signal to noise ratios (SNR). The algorithm achieves matched performance at all SNR's above 10 dB.

ROBUST ENHANCEMENT OF REVERBERANT SPEECH USING ITERATIVE NOISE REMOVAL

Authors: David Cole (d.cole@qut.edu.au) Miles Moody (m.moody@qut.edu.au) Sridha Sridharan (s.sridharan@qut.edu.au)

Speech Research Lab, Signal Processing Research Centre School of Electrical and Electronic Systems Engineering Queensland University ofTechnology GPO Box 2434 Brisbane, Australia

Volume 5 pages 2603 - 2606

ABSTRACT

We suggest a new technique for the enhancement ofsingle channel reverberant speech. Previous methods have used either waveform deconvolution or modulation envelope deconvolution. Waveform deconvolution requires calculation of an inverse room response, and is impractical due to variation with source or receiver movement. Modulation envelope deconvolution has been claimed to be position independent, but our research indicates that envelope restoration in fact degrades intelligibility of the speech. Our method uses the observation that the smoothed segmental spectral magnitude of the room response is less variable with position. This is used to estimate the reverberant component of the signal, which is removed iteratively using conventional noise reduction algorithms. The enhanced output is not perceptibly affected by positional changes.

A NETWORK SPEECH ECHO CANCELLER WITH COMFORT NOISE

Authors: D.J.Jones*, S.D.Watson* ,K.G.Evans*, B.M.G.Cheetham* and R.A.Reeves#.

***Department of Electrical Engineering, The University of Liverpool, Liverpool, L69 3BX, UK. #BT Laboratories, Martlesham Heath, Ipswich, IP5 3RE. Tel: +44 (0)151 708-7724 E-mail: davej@liv.ac.uk**

Volume 5 pages 2607 - 2610

ABSTRACT

This paper describes a proposed comfort noise system for a network echo canceller. In this system, any residual echo is suppressed using a single threshold centre-clipper, but instead of transmitting silence to the far-end of the network, a synthetic version of the background sounds is sent. This masks any 'noise modulation' or 'noise pumping' that may otherwise occur. The background sounds are characterised using linear prediction. Periods when only background sounds are present are identified by a modified GSM Voice Activity Detector (VAD). Informal listening tests have shown that this 'synthetic background' is preferable to the transmission of silence or pseudo-random noise that is not spectrally shaped to match the original background.

A NEW METRIC FOR SELECTING SUB-BAND PROCESSING IN ADAPTIVE SPEECH ENHANCEMENT SYSTEMS

Authors: Amir Hussain, Douglas R. Campbell and Thomas J. Moir

Department of Electronic Engineering and Physics, University of Paisley, High St., Paisley PA1 2BE, Scotland U.K. Corresponding author's email: huss_ee0@paisley.ac.uk

Volume 5 pages 2611 - 2614

ABSTRACT

A multi-microphone adaptive speech enhancement system employing diverse sub-band processing is presented. A new robust metric is developed, which is capable of real-time implementation, in order to automatically select the best form of processing within each sub-band. It is based on an adaptively estimated inter-channel Magnitude Squared Coherence (MSC) relationship, which is used to detect the level of correlation between in-band signals from multiple sensors during noise-alone periods in intermittent speech. This paper reports recent results of comparative experiments with simulated anechoic data extended to include simulated reverberant data. The results demonstrate that the method is capable of significantly outperforming conventional noise cancellation schemes.

ESTIMATION OF LPC CEPSTRUM VECTOR OF SPEECH CONTAMINATED BY ADDITIVE NOISE AND ITS APPLICATION TO SPEECH ENHANCEMENT

Authors: Hidefumi KOBATAKE and Hideta SUZUKI

Graduate School of Bio-Applications and Systems Engineering Tokyo University of Agriculture and Technology Koganei, Tokyo 184, JAPAN Tel. +81 423 88 7147, FAX: +81 423 85 5395, E-mail: kobatake@cc.tuat.ac.jp

Volume 5 pages 2615 - 2618

ABSTRACT

This paper presents a new method for speech enhancement. It is well known that Wiener filtering is effective in reducing additive noises and the proposed method is based on it. This paper focuses on the design of Wiener filter, where we place emphasis on the recovery of original formant characteristics and the smooth transition of speech spectrum. Transformation method of LPC cepstrum vector extracted from noisy speech to reduce noise effects is given, which gives an estimated LPC cepstrum vector of original speech. Sharpening of formant peaks and eliminating false spectral peaks are necessary for high quality speech restoration and they are realized by the proposed method. Experiments of noise reduction have been performed, whose results show the effectiveness of the proposed method.

MULTI-BAND AND ADAPTATION APPROACHES TO ROBUST SPEECH RECOGNITION

Authors: Sangita Tibrewala (1) and Hynek Hermansky (1),(2)

(1) Oregon Graduate Institute of Science and Technology, Portland, Oregon, USA. (2) International Computer Science Institute, Berkeley, California, USA. Email: sangita,hynek@ee.ogi.edu

Volume 5 pages 2619 - 2622

ABSTRACT

In this paper we present two approaches to deal with degradation of automatic speech recognizers due to acoustic mismatch in training and testing environments. The first approach is based on the multi-band approach to automatic speech recognition (ASR). This approach is shown to be inherently robust to frequency selective degradation. In the second approach, we present a conceptually simple unsupervised feature adaptation technique, based on recursive estimation of means and variances of the cepstral parameters to compensate for the noise effects. Both techniques yield significant reduction in error rates.

NON-QUADRATIC CRITERION ALGORITHMS FOR SPEECH ENHANCF NT

Authors: Enrique Masgrau, Eduardo Lleida, Luis Vicente

Communication Technologies Group (GTC). Depart.ment of Electronic Engineering & Communications Centro Politecnico Superior. C/Maria de Luna 3, 50015-Zaragoza. Spain Universidad de Zaragoza Tel: +34-976-761930, FAX: +34-976-762111, E-mail: masgrau@posta.unizar.es

Volume 5 pages 2623 - 2626

ABSTRACT

A new algorithm for speech enhancement based on the iterative Wiener filtering method due to Lim-Oppenheim [1] is presented. We propose the use of a generalized non-quadratic cost function in addition to the classical MSE term (quadratic term). The proposed cost function includes two signal-error cross- correlation terms and a L2 norm term of the filter weights. The signal-error cross- correlation terms reduce both the residual noise and the signal distortion in the enhanced speech. The L2 norm term of the filter weights reduces the overall gain of the filter, decreasing the weight noise variance and removing the side lobe of the filter response. Two solutions to the new cost function are presented: the classical non-causal type (ideal Wiener), working in the frequency

domain; and a causal finite length in the time domain. In both cases, as Lim's algorithm, the filter output of each iteration is used as "noiseless" speech signal for the following one. Simulation results demonstrate the effectiveness of these algorithms.