3 STOICA, P., and NEHORAI, A.: 'Performance study of conditional and unconditional direction-of-arrival estimation', *IEEE Trans.*, 1990, **ASSP-38**, pp. 1783–1795

4 OTTERSTEN, B., VIBERG, M., STOICA, P., and NEHORAI, A.: 'Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing' *in* HAYKIN, S., LITVA, J., and SHEPHERD, T. (Eds.): 'Radar: Array Processing' (Springer Verlag, Berlin, 1993) ch. 4, pp. 99–151

5 MARCOS, S. (Ed.): 'Les méthodes à haute résolution: traitement d'antenne et analyse spectrale' (Hermès, Paris, 1998)

6 CADZOW, J.: 'Direction-of-arrival estimation using signal subspace modeling', *IEEE Trans.*, 1992, **AES-28**, pp. 64–79

7 LI, J., HADLER, B., STOICA, P., and VIBERG, M.: 'Computationally efficient angle estiniation for signals with known waveforms', *IEEE Trans.*, 1995, **SP-43**, pp. 2154–2163

8 XU, G., and KAILATH, T.: 'DOA estimation via exploitation of cyclostationarity - A combination of spatial and temporal processing', *IEEE Trans.*, 1992, **SP-40**, pp. 1775–1786

9 VAN DER VEEN, A., and PAULRAJ, A.: 'An analytical constant modulus algorithm', *IEEE Trans.*, 1996, **SP-44**, pp. 1136–1155

10 SHYNK, J., and GOOCH, R.: 'The constant modulus array for cochannel signal copy and direction finding', *IEEE Trans.*, 1996, **44**, pp. 652–660

11 LESHEM, A., and VAN DER VEEN, A.: 'Direction-of-arrival estimation for constant modulus signals', *IEEE Trans.*, 1999, **SP-47**, pp. 3125–3129

12 STOICA, P., and NEHORAI, A.: 'MUSIC, maximum likelihood and Cramér-Rao bounds', *IEEE Trans.*, 1989, **ASSP-37**, pp. 720–741

Signal bias removal with orthogonal transform for adverse Mandarin speech recognition

Wern-Jun Wang and Sin-Horng Chen

A new method for applying orthogonal transforms in signal bias removal (SBR) for adverse Mandarin speech recognition (MSR) is proposed. The orthogonal transform process is performed in a moving window manner to extract features from the input speech. Codewords are then obtained by matching high-order, bias-free features with pre-trained codebooks for bias estimation. The effectiveness of the method has been confirmed by an experiment involving multi-speaker adverse continuous MSR. Significant improvements in the recognition accuracy and computation time were achieved as compared with the conventional SBR method.

Introduction: Signal bias removal (SBR) has been shown to be effective at eliminating multiplicative spectral bias or equivalently additive cepstral bias [1]. A popular approach is to use a two-step iterative procedure to remove the signal bias. The first step involves estimating the signal bias by calculating the average encoding residual of the testing utterance using pre-trained codebooks and the second step involves subtracting the bias estimate from every frame of the testing utterance. There are two problems with this approach. One is that it requires a sufficient number of iterations to attain better results and this is always in conflict with the real-time requirements for system implementation. The other difficulty is that some input frames may be erroneously encoded to improper codewords so as to seriously deteriorate the signal bias estimation. To overcome these two drawbacks, we propose a novel SBR approach in which orthogonal transforms are used to improve the accuracy of bias estimation for adverse Mandarin speech recognition. Instead of applying the conventional frame-based process, this method uses a segment-by-segment process. The basic idea is to represent the feature trajectories of each speech segment by using orthogonal transform coefficients. Owing to the characteristics of orthogonal transforms, only the zeroth order coefficients are bias-corrupted and all high order coefficients are bias-free. We can therefore use these high order, bias-free coefficients to find the optimum codeword and then estimate the biases from the zeroth order coefficients. This will improve the accuracy and the speed of the bias estimation.

Proposed orthogonal transform-based SBR method: The orthogonal transform technique has been widely used in waveform coding for data compression [2]. It is employed to decompose an input data sequence into mutually orthogonal components in the transform domain. The input data sequence can therefore be represented by a smooth curve formed by orthogonal expansion using some low order transform coefficients. Basis functions used in an orthogonal transform need to comply with the orthogonality property. In this study, the following four basis functions are used [3]:

$$\Phi_0\left(\frac{i}{N}\right) = 1 \tag{1}$$

$$\Phi_1\left(\frac{i}{N}\right) = \left[\frac{12\times N}{(N+2)}\right]^{\frac{1}{2}} \times \left[\left(\frac{i}{N}\right) - \frac{1}{2}\right] \tag{2}$$

$$\Phi_2\left(\frac{i}{N}\right) = \left[\frac{180\times N^3}{(N-1)(N+2)(N+3)}\right]^{\frac{1}{2}} \times \left[\left(\frac{i}{N}\right)^2 - \left(\frac{i}{N}\right) + \frac{N-1}{6\times N}\right] \tag{3}$$

$$\Phi_3\left(\frac{i}{N}\right) = \left[\frac{2800}{(N-1)(N-2)(N+2)}\right]^{\frac{1}{2}} \times \left[\frac{N^5}{(N+3)(N+4)}\right]^{\frac{1}{2}} \times\left[\left(\frac{i}{N}\right)^3 - \frac{3}{2}\left(\frac{i}{N}\right)^2 + \frac{6N^2-3N+2}{10\times N^2}\left(\frac{i}{N}\right) - \frac{(N-1)(N-2)}{20\times N^2}\right] \tag{4}$$

for $0 \le i \le N$ where $N + 1$ is the length of the contour and $N \ge 3$. These basis functions are, in fact, discrete Legendre polynomials. The contour of the kth feature element, $f_k(i)$, of a segment with length of $N + 1$ frames can thus be approximated by

$$f_k(i) \simeq \sum_{j=0}^{3} c_j(k) \times \Phi_j\left(\frac{i}{N}\right) \tag{5}$$

for $0 \le i \le N$, where

$$c_j(k) = \frac{1}{N+1}\sum_{i=0}^{N} \Phi_j\left(\frac{i}{N}\right) \times f_k(i) \tag{6}$$

is the jth order orthogonal transform coefficient. It is noted that the zeroth order coefficient represents the mean of the contour, and the other three represent its shape. According to the additive bias assumption in the cepstral domain for SBR, the bias-corrupted feature $f_k{}^b(i)$ can be modelled by

$$f_k^b(i) = f_k(i) + b_k \tag{7}$$

where b_k is the bias. The orthogonal transform coefficients $c_j{}^b(k)$ of $f_k{}^b(i)$ can then be expressed by

$$c_j^b(k) = \frac{1}{N+1}\sum_{i=0}^{N} \Phi_j\left(\frac{i}{N}\right) \times f_k^b(i) \tag{8}$$

From the characteristics of these four basis functions, it is straightforward to determine that

$$c_j^b(k) = \begin{cases} c_0(k) + b_k & \text{for } j = 0 \\ c_j(k) & \text{for } j \neq 0 \end{cases} \tag{9}$$

From the above analysis, the orthogonal transform coefficients of order greater than 0 of the bias-corrupted speech are the same as those of the clean speech. Such high order coefficients are bias-free and therefore can be used to determine the optimum codeword without interference by the corrupted bias. After determining the best-matched codeword, the bias can then be obtained by subtracting the zeroth order component of the codeword from $c_0{}^b(k)$. The orthogonal transform operation is realised in a moving window process with consecutive windows being overlapped by several frames. In the training phase, all orthogonal transform coefficients of each feature element in the clean-speech training set are collected and used to train a codebook by the LBG algorithm [4]. In the testing phase, the orthogonal transform coefficients of the bias-corrupted testing utterance are calculated and compared with these pre-trained codebooks in the above-mentioned way to find the bias estimates. By subtracting the corresponding bias estimates from the features of every frame, we obtain the biasremoved speech features for recognition. It is worth noting that the bias estimation process of the proposed method is non-iterative, so it is computationally efficient.

Experimental results: The effectiveness of the proposed orthogonal transform-based SBR (OTSBR) method was examined by simulations using a multi-speaker continuous Mandarin speech recognition task. The database was generated by ten speakers including eight males and two females. It contained, in total, 3050 utterances including 2572 training utterances and 478 testing utterances. Each utterance comprised several syllables and was uttered in such a way that every syllable was clearly pronounced. All speech signals were digitally recorded into a PC with a SoundBlaster card through a microphone and sampled at 16kHz. An adverse testing speech database was constructed artificially by passing each utterance of the clean-speech testing set through a filter which simulated a telephone channel. A set of 32 simulated filters generated from a large telephone-speech database was used in this study. All speech signals were divided into 20ms frames with 10ms frame shifts for feature extraction. A set of 25 features, including 12 MFCCs, 12 delta MFCCs, and a delta log-energy was extracted for each frame. A sub-syllable-based hidden Markov model (HMM) recogniser was constructed from the clean-speech training set by the maximum likelihood training algorithm. It consists of 100 three-state right-final-dependent initial models, 39 five-state context-independent final models, and a single-state non-speech model. The baseline SBR method used three separate codebooks for the three feature sets containing 12 MFCCs, 12 delta MFCCs, and a delta log-energy, respectively [1]. For the proposed OTSBR method, the orthogonal transform coefficients of these 25 features were calculated for all utterances in the training set and used to create 25 codebooks.

Table 1: Performance of baseline SBR method

Codeword number	Bias deviation	Syllable accuracy	Relative bias estimation time
		%	
128	132.4	63.0(57.2)	0.5
256	114.6	64.6(58.7)	1.0
512	140.2	62.4(56.6)	2.0
1024	122.8	64.1(58.3)	4.0

Table 2: Performance of proposed OTSBR method

Window length/ window shift	Bias deviation	Syllable accuracy	Relative bias estimation time
		%	
4/1	46.4	70.1	0.21
4/3	46.3	70.3	0.08
6/1	46.0	70.0	0.21
6/3	45.8	69.9	0.08
8/1	46.7	70.1	0.21
8/3	46.0	70.4	0.08

Table 1 shows the performance of the baseline SBR method. The average bias deviation of the bias estimation is defined by

$$D_{bias} = \frac{1}{N_{utt}} \sum_{u=1}^{N_{utt}} \left(\sum_{k=1}^{24} \left(bias_{est}(u,k) - bias_{des}(u,k) \right)^2 \right) \qquad (10)$$

where N_{utt} is the total number of utterances in the testing set, $bias_{est}(u, k)$ and $bias_{des}(u, k)$ are, respectively, the estimated and desired biases of utterance u and feature element k. Here $bias_{des}(u, k)$ was obtained by taking the average of the differences between the kth features of bias-corrupted speech and of clean speech of all frames in utterance u. It is noted that the calculation of D_{bias} only involves 12 MFCCs and 12 delta MFCCs. In Table 1, the numbers within the parentheses and outside the parentheses for syllable accuracy are the results of the first and tenth iterations, respectively. As to the bias deviation and the relative bias estimation time, only the results of the tenth iteration are shown in Table 1. The performance of the OTSBR method is shown in Table 2. Here, the bias estimation times are normalised to that of the baseline SBR method with 256 codewords. It can be seen from Table 2 that all cases using a different window length and window shift have comparable recognition performances. They are all much better than those achieved by using the baseline SBR method. They also all have smaller bias estimation times.

Conclusions: We have proposed a new SBR method using orthogonal transforms to improve the accuracy of bias estimation for adverse Mandarin speech recognition. Experimental results have confirmed that the proposed method outperformed the conventional SBR method significantly both in terms of the recognition performance and the computation speed.

© IEE 2000 *6 March 2000*
Electronics Letters Online No: 20000622
DOI: 10.1049/el:20000622

Wern-Jun Wang and Sin-Horng Chen (*Department of Communication Engineering, National Chiao Tung University, Taiwan, Republic of China*)

E-mail: wjwang@ms.chttl.com.tw

Wern-Jun Wang: Also with the Applied Research Laboratory, Chunghwa Telecommunication Laboratories, Taiwan, Republic of China

References

1 RAHIM, M., and JUANG, B.-H.: 'Signal bias removal by maximum likelihood estimation for robust telephone speech recognition', *IEEE Trans.*, 1996, **SAP-4**, (1), pp. 19–30

2 JAYANT, N.S., and NOLL, P.: 'Digital coding of waveforms' (Prentice-Hall, Englewood Cliffs, NJ, 1984)

3 CHEN, S.H., HWANG, S.H., and WANG, Y.R.: 'A RNN-based prosodic information synthesizer for Mandarin text-to-speech', *IEEE Trans.*, 1998, **SAP-6**, (3), pp. 226–239

4 LINDE, Y., BUZO, A., and GRAY, R.M.: 'An algorithm for vector quantizer design', *IEEE Trans.*, 1980, **COM-28**, (1), pp. 84–95

Threshold-type call admission control in wireless/mobile multimedia networks using prioritised adaptive framework

Taekyoung Kwon, Sooyeon Kim, Yanghee Choi and M. Naghshineh

Limitations to the bandwidth of wireless links has motivated the development of adaptive multimedia services where the bandwidth of a call can be dynamically adjusted. A threshold-type call admission control algorithm is proposed for quality of service provisioning; a nonlinear programming model is formulated for determining the optimal threshold values.

Introduction: Limitations to the bandwidth of wireless links has motivated the development of adaptive multimedia services which can operate over a wide range of available bandwidths [1]. That is, it is possible to overcome the link overload condition by reducing the bandwidth of individual calls. For example, handoff blocking due to bandwidth limitations can be avoided. A bandwidth adaptation algorithm (BAA) that reduces/expands the bandwidth of individual calls is invoked in the event of a new call arrival, a call completion, or an incoming/outgoing handoff call.

Under this adaptive framework, the quality of service (QoS) parameters are expressed in terms of the call blocking probability and the call degradation probability. The call degradation probability is the probability that a call will be allocated less than its maximum bandwidth at a given time. Call admission control (CAC) is required to satisfy the above QoS parameters. Recently, prioritisation (or 'differentiation') in the Internet has become of extreme importance. We believe that this concept will be reflected in wireless/mobile networks in the near future. Thus, we take prioritisation into consideration in our adaptive multimedia framework.