THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.22, No.3E 2003. 9. pp. 124~132

A Simple Speech/Non-speech Classifier Using Adaptive Boosting

Oh-Wook Kwon*, Te-Won Lee**
*School of Electrical and Computer Engineering, Chungbuk National University
**Institute for Neural Computation, University of California, San Diego, USA
(Received April 1 2003; accepted July 21 2003)

Abstract

We propose a new method for speech/non-speech classifiers based on concepts of the adaptive boosting (AdaBoost) algorithm in order to detect speech for robust speech recognition. The method uses a combination of simple base classifiers through the AdaBoost algorithm and a set of optimized speech features combined with spectral subtraction. The key benefits of this method are the simple implementation, low computational complexity and the avoidance of the over-fitting problem. We checked the validity of the method by comparing its performance with the speech/non-speech classifier used in a standard voice activity detector. For speech recognition purpose, additional performance improvements were achieved by the adoption of new features including speech band energies and MFCC-based spectral distortion. For the same false alarm rate, the method reduced 20-50% of miss errors.

Keywords: *Speech/non-speech classification, Speech detection, Adaptive boosting, AdaBoost algorithm*

I. Introduction

Recent application of speech recognition technologies to portable devices (e.g., personal digital assistants, cellular phones with a hands-free car kit) in realistic noisy environments made robust speech detection one of the most critical components. Speech detection or endpoint detection has turned out to significantly influence word accuracy in case of cellular phones operating in a noisy automobile environment[1,2]. Conventional endpoint detection algorithms based on energy and zero crossing rate (ZCR) do not handle noisy speech signals in a proper manner especially in mobile communications. The ZCR feature is not suitable any more for noisy environments. Various other features including high-pass/low-pass energies[3], linear prediction coding (LPC) residual and auto-correlation of LPC residual information[4] have been used to improve robustness and accuracy of speech detectors. For noisy speech detection, speech enhancement stages are often adopted to reduce noise signals before speech/non-speech classification[5,6].

Usually, the noise spectrum is assumed known from the early part of the utterances. However, some methods estimate the noise spectrum continuously by using minimum statistics on a predefined time window[7], taking median values[8], or tracking power envelope dynamics[3]. A necessary and strong constraint for most speech/non-speech classification algorithms is that they should have as low computational complexity as possible to reduce computational burden on the entire speech recognition system. For that purpose, a decision tree was

Corresponding author: Oh-Wook Kwon (owkwon@chungbuk.ac.kr)
School of Electrical and Computer Engineering, Chungbuk National University, 48 Gaesin-dong, Heungdeok-gu, Cheongju, 361-763, Korea

used to combine multiple features for improved accuracy[4].

The purpose of this study was to design improved speech detectors robust to noise in various environments. The speech detectors should be designed automatically from data to relieve human efforts. The computation load should be as low as the G.729 voice activity detector (VAD) which uses multiple decision hyperplanes for speech/non-speech classification. To satisfy the constraints, we apply an adaptive boosting (AdaBoost) algorithm[9,10] to speech/non-speech classifier design.

Even though our purpose is to obtain speech detectors, we compared the performance with the G.729 VAD and illustrated the validity of the method just because it is a publicly available reference. The proposed algorithm is intended for robust speech recognition. Speech detection can be implemented in two stages: speech/non-speech classification and subsequent postprocessing. The performance of speech detection largely depends on speech/non-speech classification. Hence we focused on speech/non-speech classification in this paper.

To check the validity of speech/non-speech classifier design using the AdaBoost algorithm, we combine multiple features extracted from voice activity detectors in the G.729 Annex B speech coding recommendation[11] using the AdaBoost algorithm and show that the algorithm can be successfully applied to designing a new classifier and can achieve accuracy similar to the manually-optimized G.729 VAD with comparable computational complexity. After confirming the validity of the AdaBoost algorithm we proposed a new speech speech/non-speech classifier for speech recognition by using new features and weights for spectral-subtracted signals. Further analysis of the learned weights for the base classifiers revealed the contribution of each feature component.

II. Speech Detection Using the AdaBoost Algorithm

2.1. Speech/Non-speech Classifier Design

Input speech signals of voice activity detection with the sampling frequency of 8 kHz are preemphasized and blocked into frames of 10 ms. A speech signal frame is transformed into the frequency domain by the fast Fourier transform (FFT).

The VAD used in the G.729 Annex B recommendation uses the following features for speech/non-speech classification[11]:

(1) Instantaneous full-band log energy (0 - 4 kHz)
(2) Low-band log energy difference (0 - 1 kHz)
(3) Full-band log energy difference (0 - 4 kHz)
(4) Spectral distortion measured by line spectral frequencies
(5) Zero-crossing rate difference (ZCR)

Each difference feature is obtained by the difference between the instantaneous parameter and the running average of the background noise.

In our design we can keep the simplicity of the speech/non-speech classifier of the ITU-T Recommendation G.729 Annex B while we obtain the hyperplanes in a principled and automatic manner by using the AdaBoost algorithm. The AdaBoost algorithm was chosen among many classifiers due to its simple use of base classifiers and its non-overfitting property.

All differential features were normalized to have zero mean and unit variance along each axis. We used a perceptron as the base classifier with the sigmoidal activation function $f(\mathbf{x})=\tanh(\gamma\mathbf{x})$ where $\gamma=4$ was used to control the range of boundary regions in our experiments. The base classifier $h_t(\mathbf{x})$ is real-valued rather than binary, which can be interpreted as confidence-rated prediction[9]. The sign of $h_t(\mathbf{x})$ is the predicted label and the magnitude denotes a measure of confidence. We use base classifiers with linear decision boundaries due to fast and simple learning procedures. For $h_t(\mathbf{x})$, only a single feature or whole features can be used. We chose to use a decision stump where only a single feature is used for each base classifier for its simplicity and trainability. The learning algorithm to combine weights is given in Appendix 5.1.

For spectral subtraction, the noise spectrum is estimated from the input signals and subtracted from the magnitude spectrum of input signals. In this work, we estimated the noise spectrum by tracking the minimum statistics of the magnitude spectrum[7], where the minimum of each frequency bin within the time window of 1 second was regarded as a noise component. This method does not require any other assumption on input speech utterances and can be used continuously without reinitialization. Note that there is no significant delay in using the minimum statistic because the last 1 second of windowed data is used. The spectral subtraction method adopted in this work and the selection of the relevant parameters are described in detail in Appendix 5.2.

Figure 1 illustrates the block diagram for the speech/non-speech classifier based on adaptive boosting[9]. The final classifier is given as (2) and the relevant weights are trained by using the AdaBoost algorithm described in Appendix 5.1. When only one feature is used for each classifier, it resembles the partitioning of the feature space into vertical or horizontal decision boundaries. By combining the results of the base classifiers using the signum function, the decision boundary of the final classifier can be non-linear.

It may be regarded analogous to the multi-layer perceptron where perceptrons are combined by the upper layer. But in this case, the base classifiers are usually very simple and sequentially learned. While a decision tree-based classifier propagates a decision error from the upper node to child nodes and thus errors in earlier stages cannot be recovered by the later stage, the AdaBoost-based classifier avoids this problem by using the result combined from all base classifiers.

2.2. New Features for Robust Speech Detection

The VAD in the G.729 Annex B is targeted for speech signals with rather low level of noise signals and its performance degrades as the signal-to-noise ratio (SNR) goes down to about 5 dB. Therefore a speech enhancement block is applied before feature extraction. In addition, it is advantageous to use mel-frequency cepstral coefficient (MFCC)-based features so that we can combine feature extraction and spectral subtraction to share the required computation with a speech recognizer.

Features for speech/non-speech classification are extracted as shown in Figure 2:

(1) Full-band speech log energy difference
(2) Low-band speech log energy difference (0 - 1.0 kHz)
(3) Pass-band speech log energy difference (1.0 - 2 kHz)
(4) Spectral distortion measured by MFCC
(5) Zero crossing rate difference
(6) Instantaneous total log energy

We note that in this case the band energies are for spectral-subtracted signals. The low-pass log energy and band-pass log energy are useful to reject high-frequency noise (e.g., drill noise) and low-frequency noise (car noise) [3]. The MFCC-based feature can reflect the human perception characteristics better than the line spectrum

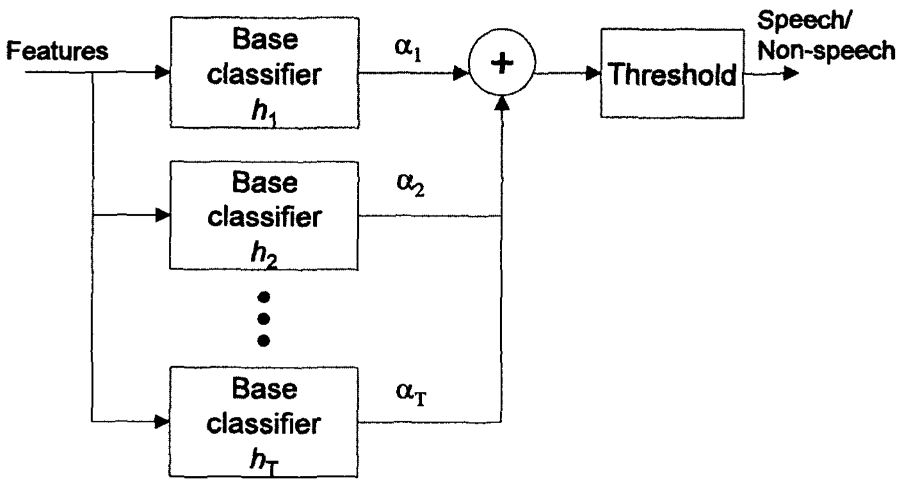

Figure 1. Speech/Non-speech classifier using the AdaBoost algorithm.

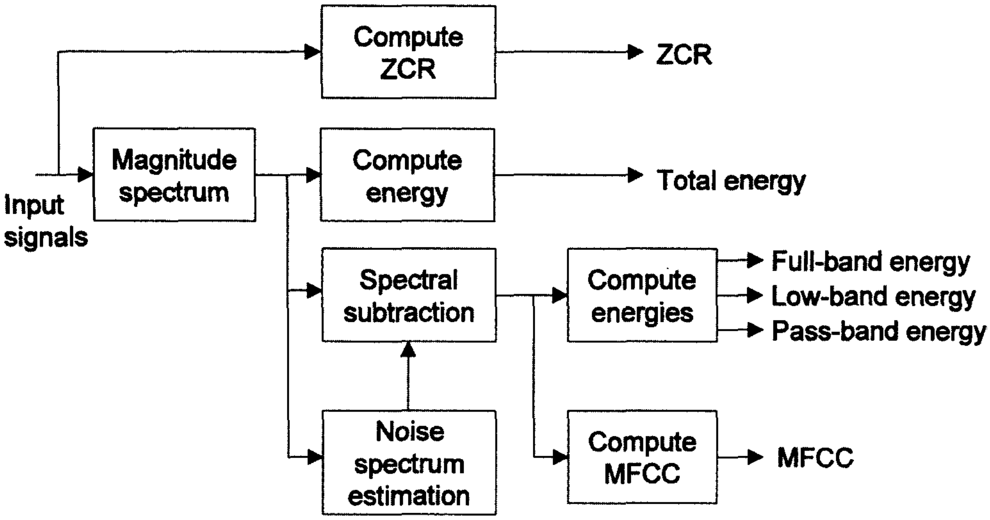

Figure 2. Feature extraction in combination with spectral subtraction-based speech enhancement.

frequency feature used in the G.729 VAD.

III. Experimental Results and Discussions

3.1. Toy Experiment

To validate the simplicity of the classifier as well as the accuracy in data classification problems, we designed a small toy experiment to illustrate the performance of the AdaBoost algorithm. The training data were generated by a mixture of Gaussians with means located at 4 different quadrants. For this kind of data, two hyperplanes parallel to x- and y-axis combined with an OR logic can perform good classification. Figure 3 shows the experimental results with 25 decision stumps. The figure shows the classification results and the first 5 hyperplanes labeled in the appearing order. The error rate curves showed that the final combined classifier can classify the two classes successfully with as small as 3 hyperplanes.

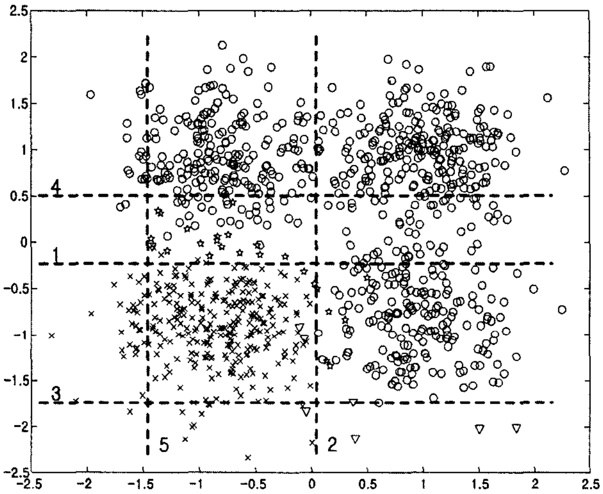

Figure 3. Classification results with decision stumps on toy data. The circle and cross denote the correctly classified samples with +1 and labels, respectively. The triangle denotes a missed sample and the star denotes a false alarm.

3.2. Speech Database

To evaluate the performance of the proposed speech detector, we used the AURORA speech database[12]. We trained the speech classifiers with speech data in all environments and SNR levels available in the AURORA database: clean, 20, 15, 10, and 5 dB. For the test set, we used the speech data with the same noise environments in the same range of SNRs. To train speech classifiers, we need speech/non-speech labels for the speech data. Because the AURORA database did not provide the speech/non-speech labels, we obtained the label information by running a Viterbi aligner obtained through multi-style training with various environments and noise levels. We manually corrected by viewing the spectrogram. We randomly sampled 12000 frames for the training data set and another 3000 samples for the validation data set. We used the test data set with babble and car noise to evaluate the performance of speech/non-speech classification.

3.3. Voice Activity Detection

We used the decision stump and the perceptron as the base classifier to keep the classifier as simple as possible. With the decision stump, the base classifiers have axis-parallel decision boundaries and the final classifier used in speech detection can be implemented by comparing with threshold values determined by the parameters of the base classifiers. Experimental results showed that performance difference between the decision stump and the perceptron as the base classifier was not significant. In both cases, the classifiers converged sufficiently with 100 base classifiers and the difference between the training data set and the validation data set was small. To be precise, the decision stump showed slightly lower validation error rate for the training set but both of the base classifiers yielded a similar level of error rates for the test set. Since the implementation and computation of decision stumps are much simpler, we pursued only the decision stump case.

We evaluated the performance of the classifier under babble noise environments in different SNR conditions. The number of base classifiers used in the test was decided to give the minimum error rate for the validation test set. In this case the number of the base classifiers used in the test was 95. To obtain Figure 4, we normalized α_t to have unity sum of absolute value of α_t and varied δdiscretely from -0.3 to 0.5 with a step size of 0.05. Each symbol in the ROC curve denotes the condition with a certain control parameter. The filled symbols denote the case of the G.729 VAD without any additional processing, where the performance is shown as a point for each SNR condition because the G.729 VAD has a fixed parameter. In both cases we did not use the hang-over scheme[11] to obtain a fair comparison. The performance points of the G.729 VAD were located near or on the ROC curves of our proposed method. The AdaBoost-based classification has the advantage that it can provide a flexible trade-off between the hit rate and the false alarm rate by controlling only one parameter δ depending on applications.

We analyzed the learned weight values of the AdaBoost classifier as the number of base classifiers increases and the relative weight for each feature component as shown in Figure 5. The relative weight for each feature index was obtained by adding the weights for the base classifiers and normalizing the result. The results imply that the instan-

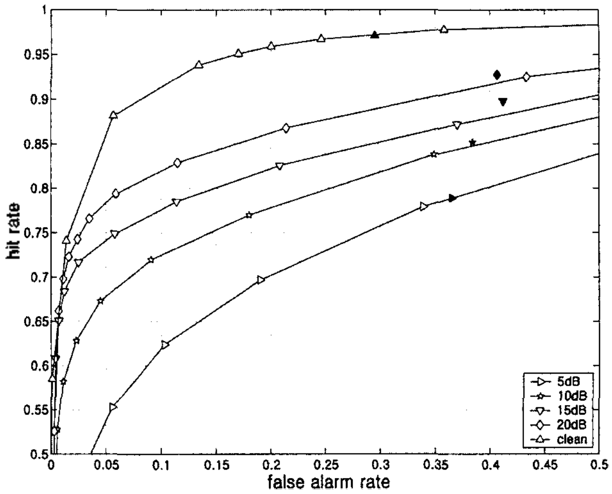

Figure 4. ROC curves of voice activity detection of original speech signals using the AdaBoost classifier in babble noise environments. The filled symbols denote the performance of the G.729 VAD without any additional processing, where the performance is shown as a point for each SNR condition because there is no control parameter in the G.729 VAD. Only the features from the current frame were used.

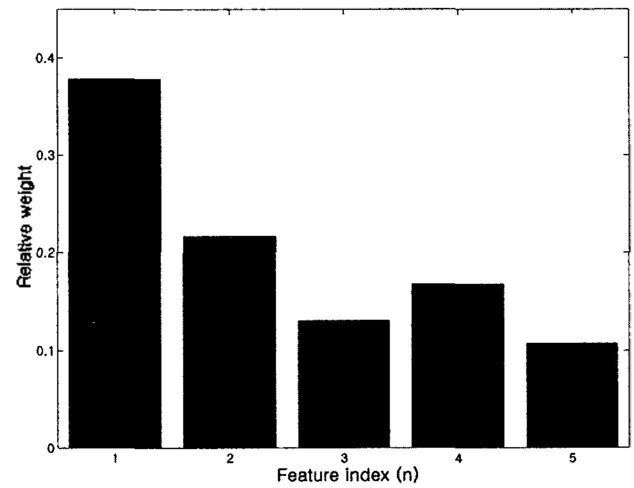

Figure 5. Relative weight of each feature component averaged over 200 base classifiers. The relative weight for each feature index in the bottom figure was obtained by adding the weights for 95 base classifiers and normalizing the result by the sum along the index.

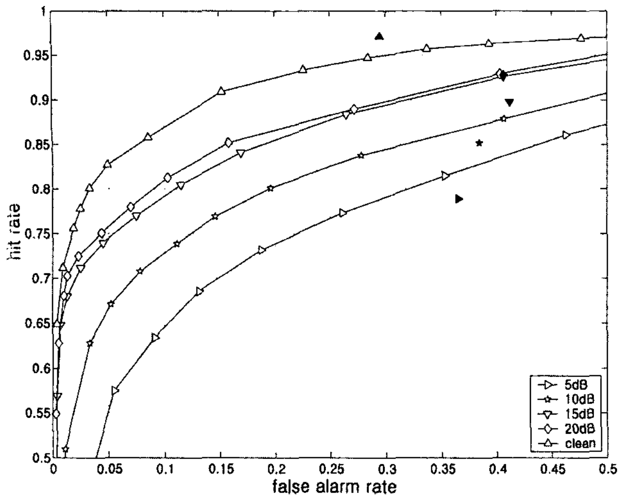

Figure 6. ROC curves of voice activity detection of spectral-subtracted signals by the AdaBoost classifier in the babble noise environments. The filled symbols have the same meaning as Figure 4.

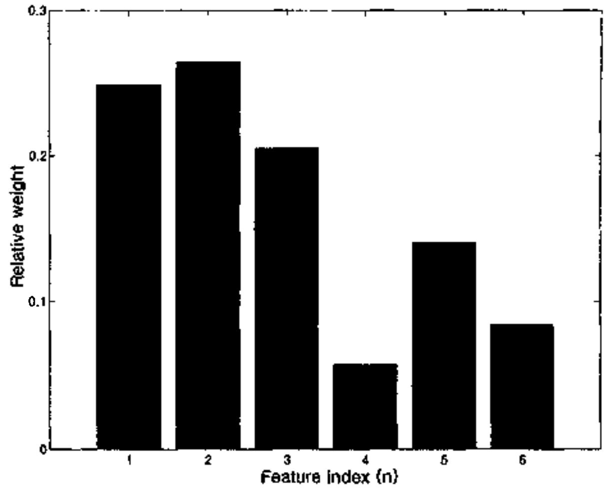

Figure 7. Relative weight of each feature component averaged over 107 base classifiers.

taneous full-band log energy (1) and the low-band energy (2) are the most important two features in speech/non-speech classification.

We trained the AdaBoost classifier by using the speech signals enhanced by spectral subtraction. Using 107 base classifiers, we compared the performance with the previous case. Figure 6 shows the ROC curves for the spectral-subtracted signals. Although its performance degraded slightly in the clean speech case, the hit rate at the same false alarm rate was improved in the noisy cases. One important advantage in using the AdaBoost-based classifier is that we can automatically obtain a simple classifier with performance comparable to the manually optimized classifier.

We did not plot the performance of the G.729 with spectral subtraction because the operating points were mostly out of the current plot range: the false alarm rate was over 0.75 and the hit rate was over 0.95 for noisy speech. This is due to the changes in the speech signal characteristic induced by the spectral subtraction algorithm and the G.729 VAD cannot adapt to the distorted signals.

3.4. Using New Features for Robust Speech Detection

For speech detection purposes we attempted to use better features derived from feature extraction for speech recognition as described in Section 2.2. As shown in Figure 7, we analyzed the learned weight values and found that the low- and pass-band energies contribute to the performance of the classifier compared with the G.729-based features. In particular, the low-band speech log energy (2) has a relatively large weight and the spectral distortion feature (4) has a lower weight than the G.729 VAD because spectral subtraction introduced spectral distortion. However, the MFCC-based spectral distortion feature did not have a large weight because the spectral subtraction caused nonlinear distortion on the spectrum of the input signals.

Figure 8 shows the ROC curves in the babble noise environments. We used spectral subtraction and the new features. For every SNR condition the AdaBoost-based speech/non-speech classifier yielded improved performance. For the same false alarm rate condition as in the G.729 VAD, the miss rate (=1-hit rate) decreased by 20-50 percent. This improvement mainly results from spectral subtraction and the proper design of the classifier as the features are changed.

We compare the performance of the G.729 VAD in the babble noise environments with 10 dB SNR as shown in Figure 9: the AdaBoost classifier with G.729 features, the AdaBoost classifier with G.729 features and spectral subtraction, and the AdaBoost classifier with the new

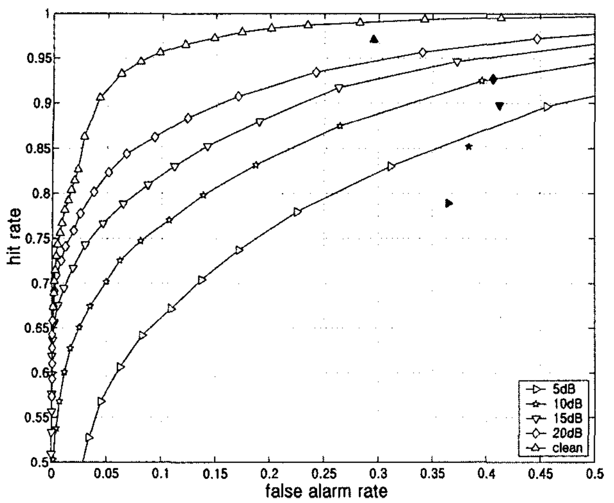

Figure 8. ROC curves in case of new features in the babble noise environments. The AdaBoost classifier was used for voice activity detection and spectral subtraction was applied before feature extraction.

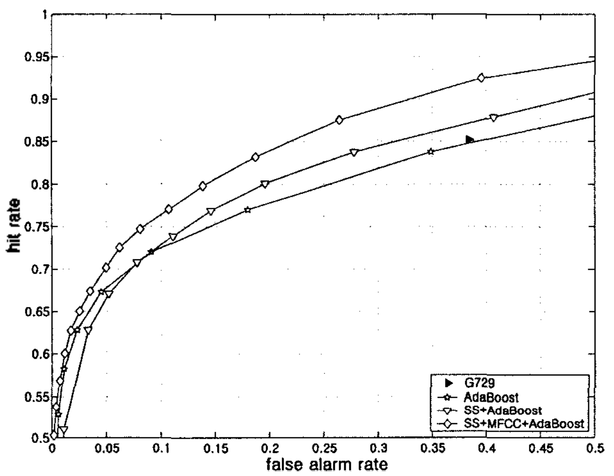

Figure 9. Performance comparison of the G.729 VAD in the babble noise environments with 10 dB SNR: the original G.729 VAD ('G729'), the AdaBoost classifier with G.729 features ('AdaBoost') the AdaBoost classifier with G.729 features and spectral subtraction ('SS + AdaBoost'), and the AdaBoost classifier with the new features and spectral subtraction ('SS + MFCC + AdaBoost').

features and spectral subtraction. Figure 9 shows that the AdaBoost-based classifier yields the same level of performance similar to the manually-optimized classifier. The AdaBoost classifier was systematically designed by multiple decision stumps, which requires a small number of computations. For better performance we changed the feature and improved the performance with minimal increase of computational load.

The performance in the noisy car environment was also evaluated with the same configuration as the babble noise environment. Experimental results showed that the AdaBoost classifier made similar performance improvements. Due to the nature of the car noise, the false alarm rates were relatively small compared to the babble noise case.

Examining errors in the experiment, we found that in noisy cases most of the misclassification was made at the utterance boundary beginning or ending with a unvoiced sound (e.g., /t/ in eight, /f/ in five and so on). With this regard, a more elaborate speech enhancement algorithm can be designed to further improve the VAD accuracy.

3.5. Considerations on Computational Complexity

For reference, we investigate the computational complexity of the G.729 VAD. The speech/non-speech classifier for the VAD uses 14 hyperplanes with 2 features involved for each hyperplane and the final decision is made through the sequential test of the hyperplanes. Each hyperplane needs one multiplication, one addition and one comparison.

On the contrary, the decision stump-based speech/non-speech classifier requires 1 addition of the bias term, 1 table lookup for $\tanh(\cdot)$ and 1 multiplication by the weight for each base classifier. The required computation is linearly proportional to the number of base classifiers, which can be tuned with additional experiments.

IV. Conclusions

We proposed a new speech/non-speech classification algorithm based on the AdaBoost algorithm and a set of new speech features. Our experimental results indicate that a nearly optimal classifier can be designed automatically with computational complexity comparable to the manually-optimized classifier. We investigated the contribution of each feature for speech/non-speech classification and the effects of spectral subtraction. After evaluating the performance of the AdaBoost method to design a speech/non-speech classifier, we suggested a new efficient speech detection method using different kinds of features including estimated speech band energies and MFCC-based spectral distortion intended for speech recognition. When spectral subtraction is used, the low-band speech log energy has a relatively large weight and the spectral distortion feature has a small weight. For speech signals in babble noise environments, the method reduced 20-50% of miss errors for the same false alarm rate. Our proposed method gives good classification accuracy by combining simple linear base classifiers.

V. Appendix

5.1. The AdaBoost Algorithm

Consider that m samples $\mathbf{x}_i \in \mathbf{X}$, $i=1,\cdots,m$ and the corresponding labels $y_i \in Y$ are given where $\mathbf{X}$ denotes the N-dimensional feature space and Y is given as {-1+1}. First initialize the weight of each sample uniformly as $D_1(i) = 1/m$. For each round t =1, ...,T, add a new base classifier to the final classifier by performing the following steps[9]:

* Train a new base classifier using distribution D_t.
* Get the base classifier $h_t : \mathbf{X} \rightarrow \Re$.
* Choose $\alpha_t \in \Re$.
* Compute a new weight for each sample $D_{t+1}(i)$ by

$$D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(\mathbf{x}_i))}{Z_t} \tag{1}$$

where Z_t is a normalization factor chosen so that D_{t+1} is a probability distribution.

The final classifier is given by

$$H(\mathbf{x}) = \operatorname{sign}(\sum_{t=1}^{T} \alpha_t (h_t(\mathbf{x}) + \delta)) \tag{2}$$

where δ is a parameter used to control the hit and false alarm rates. According to[9], the α_t is chosen as

$$\alpha_t = \frac{1}{2}\log(\frac{1+r_t}{1-r_t}) \tag{3}$$

where

$$r_t = \sum_i D_t(i) y_i h_t(\mathbf{x}_i) \tag{4}$$

5.2. Spectral Subtraction Based on Minimum Tracking

Speech signals are windowed by a Hamming window of length 30 ms shifting every 10 ms and transformed into the frequency domain to produce N coefficients $X(t,k)$ where is the frame index and k is the frequency bin index running from 0 to N/2+1. For each frequency bin, the signal spectrum $\tilde{S}(t,k)$ is obtained by smoothing temporally as

$$\tilde{S}(t,k) = \alpha_1 \tilde{S}(t-1,k) + (1-\alpha_1) \| X(t,k) \|^2 \tag{5}$$

The spectrum is then smoothed in the spectral domain as

$$\bar{S}(t,k) = \sum_{i=-I/2}^{I/2} h(i) \tilde{S}(t,k-i) \tag{6}$$

where $h(i)$ is a smoothing filter to prevent abrupt changes along the frequency axis and I is the filter length. The noise spectrum for spectral subtraction is estimated by tracking the minimum of the smoothed signal spectrum within a specified frame period T_f[7].

$$\tilde{N}(t,k) = \beta \min_{0 \le i \le T_f} \bar{S}(t-i,k) \tag{7}$$

where β is a constant to estimate the average from the minimum statistics. Then the speech spectrum is estimated by subtracting the noise spectrum from the signal spectrum

$$\hat{X}(t,k) = X(t,k) \max\{1 - (\bar{g}(t,k)\tilde{N}(t,k)/\tilde{S}(t,k))^{\xi}, \quad \eta(\tilde{N}(t,k)/\| X(t,k) \|^2)^{\xi}\} \tag{8}$$

where η is introduced to prevent negative spectra. The ξ is set to 0.5 so that we can use spectral magnitude subtraction, which is known to yield better performance [13]. $\bar{g}(t,k)$ The is a smoothed over-subtraction factor obtained by filtering SNR-dependent instantaneous over-subtraction factors

$$\tilde{g}(t,k) = \alpha_2 \tilde{g}(k)(t-1,k) + (1-\alpha_2)(1+f(k)\frac{\tilde{N}(t,k)}{\tilde{N}(t,k)+\tilde{S}(t,k)}) \quad (9)$$

where $f(k)$ is the predefined over-subtraction curve. In this work we use the following function

$$f(k) = (1+\lambda \frac{kf_s}{Nf_c})^{-1} \quad (10)$$

where λ is an over-subtraction scale factor, f_s is the sampling frequency and f_c is the cut-off frequency. For our experiments we used the following set of parameters $\alpha_1 = 0.8187, \alpha_2 = 0.9048$, $\beta = 1$, $\xi = 0.5$, $\eta = 0.05$, $\lambda = 4$, $f_s = 8000$, $f_c = 400$, $N = 256$, $T_f = 125$, $I = 5$, $\mathbf{h} = [1/9\ 2/9\ 3/9\ 2/9\ 1/9]$.

References

1. H. K. Kim, and R. C. Cox, "Evaluation of robust speech recognition algorithms for distributed speech recognition in a noisy automobile environment," *Proc. ICSLP 2002*, 233-236, Sept. 2002.
2. C.-P. Chen, K. Filali, amd J. F. Bilmes, "Frontend post-processing and backend model enhancement on the Aurora 2.0/3.0 databases," *Proc. ICSLP 2002*, 241-244, Sept. 2002.
3. M. Marzinzik, and B. Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," *IEEE Trans. Speech and Audio Processing*, **10** (2), 109-110, Feb. 2002.
4. W.-H. Shin, B.-S. Lee, Y.-K. Lee, and J.-S. Lee, "Speech/Non-speech classification using multiple features for robust endpoint detection," *Proc. ICASSP 2000*, Ⅲ. 1399-Ⅲ. 1402, 2000.
5. B. Kotnik, Z. Kacic, and B. Horvat, "A computational efficient real time noise robust speech recognition based on improved spectral subtraction method," *Proc. EUROSPEECH 2001*, 1123-1126, 2001
6. L. Karray, and A. Martin, "Towards improving speech detection robustness for speech recognition in adverse conditions," *Speech Communication*, 40 (3), 261-276, May 2003.
7. R. Martin, "Spectral subtraction based on minimum statistics," *Signal Processing VII, Theories and Applications*. Proc. EUSIPCO-94, 1182-1185, 1994.
8. N. W. D. Evans, and J. S. Mason, "Noise estimation without explicit speech, non-speech detection: A comparison of mean, modal and median based approaches," *Proc. EUROSPEECH 2001*, 893-896, 2001.
9. R. E. Schapire, and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, 37 (3), 297-336, 1999.
10. Y. Freund, and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, **55** (1), 119-139, 1997.
11. A. Benyassine, E. Shlomot, and H.-Y. Su, "ITU-T Recommendation G.729 Annex B: A silent compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications," *IEEE Communications Magazine*, 64-73, Sept. 1997.
12. H. G. Hirsch, and D. Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions," *ISCA ITRW ASR2000 Automatic Speech Recognition: Challenges for the Next Millennium*, Paris, France, Sept. 18-20, 2000.
13. A. Acero, *Acoustical and Environmental Robustness in Automatic Speech Recognition*, Kluwer Academic Publishers, Boston, 1993.

[Profile]

● Oh-Wook Kwon

Oh-Wook Kwon received the B. S. degree in electronic engineering from Seoul National University, Korea in 1986, and the M. S. and Ph. D. degrees from Korea Advanced Institute of Science and Technology (KAIST) in 1988 and 1997, respectively. From 1988 he was with Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea. In 2000, he joined Brain Science Research Center at KAIST as a research professor. From 2001 to 2003, he worked in Institute for Neural Computation at University of California, San Diego as a postgraduate research associate. Since 2003, he has been with Chungbuk National University, Cheongju, Korea and he is now an assistant professor in School of Electrical and Computer Engineering. His research interests include speech and audio processing, language processing, and communication signal processing.

● Te-Won Lee

Te-Won Lee is an associate research professor at the Institute for Neural Computation at the University of California, San Diego and a collaborating professor in the Biosystems Department at the Korea Advanced Institute of Science and Technology (KAIST). Dr. Lee received his diploma degree in March 1995 and his Ph. D. degree in October 1997 with summa cum laude in electrical engineering from the University of Technology Berlin. He was a Max-Planck Institute fellow (1995-97) and a visiting professor KAIST in 1999. During his studies he received the Erwin-Stephan prize for excellent studies from the University of Technology Berlin and the Carl-Ramhauser prize for excellent dissertations from the Daimler-Chrysler Corporation. Dr. Lee's research interests include machine learning algorithms with applications in signal and image processing and inference. Recently, he has worked on variational Bayesian methods for independent component analysis, algorithms for speech enhancement and recognition, models for computational vision, and classification algorithms for medical informatics.