LOCAL BINARY PATTERNS FOR 1-D SIGNAL PROCESSING

Navin Chatlani and John J. Soraghan

Centre for Excellence in Signal and Image Processing (CeSIP), University of Strathclyde
Royal College Building, 204 George Street, Glasgow
phone: +(0044) 141 548 2205, email: navin.chatlani@eee.strath.ac.uk, j.soraghan@eee.strath.ac.uk
http://www.eee.strath.ac.uk

ABSTRACT

Local Binary Patterns (LBP) have been used in 2-D image processing for applications such as texture segmentation and feature detection. In this paper a new 1-dimensional local binary pattern (LBP) signal processing method is presented. Speech systems such as hearing aids require fast and computationally inexpensive signal processing. The practical use of LBP based speech processing is demonstrated on two signal processing problems: - (i) signal segmentation and (ii) voice activity detection (VAD). Both applications use the underlying features extracted from the 1-D LBP. The proposed VAD algorithm demonstrates the simplicity of 1-D LBP processing with low computational complexity. It is also shown that distinct LBP features are obtained to identify the voiced and the unvoiced components of speech signals.

1. INTRODUCTION

Local Binary Patterns (LBP) have been extensively used in 2-D image processing [1] [2]. LBP has been shown in [3] to be a computationally simple, discriminative descriptor of texture. The motivation for the above applications is that an image can be described by a combination of texture patterns. We aim to develop a 1-D LBP signal processing framework and demonstrate its applicability on a real problem. Real time systems such as hearing aids require fast processing of the input signal while maintaining low computational complexity. One common process in speech systems is Voice Activity Detection (VAD) which attempts to estimate periods of speech and non-speech. Different flavours of VAD base their decisions on statistical techniques [4] [8], energy level detection [5] or periodicity measures. VAD performance is affected by the SNR of the noisy speech and performance depends on computational complexity and parameter tuning.

In this paper, a novel 1-D LBP operator is developed as a signal processing tool. An LBP code for a neighbourhood of sampled data is produced by thresholding the neighbouring samples against centre samples of a processing window. This procedure is iteratively done across the entire signal and a segment of the 1-D signal is alternatively described by a sparser occurrence histogram of LBP codes. The paper is organized as follows. The novel 1-D LBP operator is presented in section 2. In section 3, a LBP-based segmentation of a 1-D signal is used to illustrate the processing capability of the 1-D LBP. A computationally simple LBP-based VAD is designed in section 4. This uses the occurrence histogram of the underlying signal to identify the voiced, unvoiced and non-speech components. The performance of the new VAD is demonstrated on a speech sample taken from the TIMIT database [6] contaminated with non-stationary noise from the NOISEX-92 database [7]. Finally, concluding remarks are presented in section 5.

2. 1-D LOCAL BINARY PATTERNS

The 1-D LBP operator is adapted from the 2-D LBP [3]. It examines a neighbourhood of data samples from a signal $x[i]$ and assigns an LBP code to each centre sample after thresholding them against the neighbouring samples. The 1-D LBP operating on a sample value $x[i]$ is defined as:

$$LBP_P\left(x[i]\right) = \sum_{r=0}^{P/2-1}\left\{S\left[x\left[i+r-P/2\right]-x[i]\right]2^r + S\left[x[i+r+1]-x[i]\right]2^{r+P/2}\right\} \quad (1)$$

where the Sign function $S[.]$ is given by:

$$S[x] = \begin{cases} 1 \; for \; x \geq 0 \\ 0 \; for \; x < 0 \end{cases} \quad (2)$$

and where the P neighbouring samples are thresholded around the centre sample from the neighbourhood of $P+1$ data samples from the signal $x[i]$ of length N for $i=[P/2 : N-P/2]$. The Sign function $S[.]$ transforms the differences to a P-bit binary code. The binomial weight applied to each thresholding operation converts the binary code into a unique LBP code.

An illustration of the 1-D LBP operator is given in Figure 1 where P is set to 8 and the centre sample C is circled. As in Eq. (1), the 8 neighbouring samples are thresholded against C to produce a binary code of 1111_0000. This code is then multiplied by the binomial weights given to the corresponding samples and the obtained values are summed to give the resulting LBP code of 15. The LBP codes can locally describe the data using the difference between a sample and its neighbours. For a constant or slowly varying signal, these differences cluster near zero. At peaks and troughs, the difference will be relatively large, whereas at edges, the differences in some directions will be larger than those from other directions. The local patterns formed from $x[i]$ can be described by the distribution of the LBP codes:

$$H_k = \sum_{\frac{P}{2} \leq i \leq N-\frac{P}{2}} \delta\left(LBP_P\left(x[i]\right), k\right) \quad (3)$$

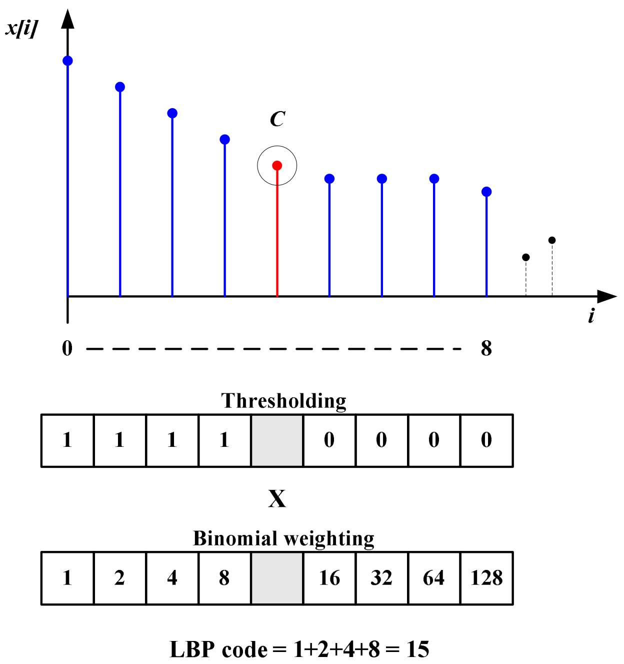

Figure 1 - Computation of 1-D local binary pattern (1-D LBP)

where $k=1..n$ and n is the number of histogram bins and each bin corresponds to an LBP code. $\delta(i,j)$ is the Kronecker delta function.

The standard LBP_P operator produces 2^P different LBP codes. Extensions of the LBP_P are presented in [3] for rotation invariant patterns LBP_P^r, uniform patterns LBP_P^u, and rotation invariant uniform patterns $LBP_P^{r,u}$. LBP_P^r is produced by shifting the LBP code for the P neighbouring samples until its minimum value is found. In this way, LBP_P^r of the processed window produces the same code for all shifted versions of that code and it is therefore invariant to rotation. A uniform pattern is defined by an LBP code which has at most two one-to-zero or zero-to-one transitions. The remaining non-uniform patterns are assigned to a single histogram bin and each uniform pattern is assigned to a separate bin. LBP_P^u gives a histogram with $P(P-1)+3$ bins. $LBP_P^{r,u}$ shifts the uniform codes until they attain their minimum values and results in a histogram with $P+1$ bins for uniform patterns plus one bin for non-uniform patterns. The LBP code evaluated earlier in this section is an example of a code that is uniform and is already rotation-invariant. The choice of which LBP to use depends on the need for either a more resolved representation or for a sparser histogram. In the presented work, normalized histograms will be used for $LBP_P^{r,u}$ resulting in histograms with $P+2$ bins.

3. UNSUPERVISED SIGNAL SEGMENTATION USING 1-D LBP

The 1-D LBP operator is used to produce a histogram of LBP codes which can be used as an alternative representation of the signal. In signal segmentation, the histogram can be used as a non-parametric estimator of the empirical LBP feature histogram. Resistor Average Difference (RAD) [2] can be used for measuring the similarity of adjacent LBP histograms. RAD is derived from the non-symmetric Kullback-Leibler Distance (KLD) [2] which is used for measuring the difference between two histograms p and q. KLD is given by:

$$D_{KL}\left(p\|q\right)=\sum_{k=1}^{n}p(k)\left\{\lg\left(p(k)\right)-\lg\left(q(k)\right)\right\} \qquad (4)$$

where n is the number of histogram bins and $p(k)$ and $q(k)$ are the number of occurrences in histograms p and q respectively at bin k. The RAD is defined as:

$$D_{RAD}\left(p,q\right)=\left[\left(D_{KL}\left(p\|q\right)\right)^{-1}+\left(D_{KL}\left(q\|p\right)\right)^{-1}\right]^{-1} \qquad (5)$$

$D_{RAD}(p,q)$ between the two histograms p and q increases with dissimilarity and in contrast to KLD, RAD is symmetric [9].

3.1 Noise Onset Identification

In this example, the onset of noise is detected for a noise source switched on at some time τ. The signal x is first split into segments $x_a[j]$ of length W by applying a window $w[j]$ of length W as:

$$x_a\left[j\right]=x\left[aR+j\right]w\left[j\right] \text{ for } 0\le j\le W-1 \qquad (6)$$

where a is the segment number, $R<W$ for overlapping segments and $R=W$ for contiguous segments. W is chosen to be small enough to capture transitions in the LBP feature histograms. $D_{RAD}(p,q)$ is measured for the segments of the adjacent histograms and similar segments are merged. When two adjacent segments are merged, their histograms are summed and normalized to produce the histogram of the new segment. This procedure continues until the segment does not expand and the previously merged segments are considered as a component of the signal with similar underlying LBP features.

This procedure was performed for an artificially generated sinusoidal signal of length 768 samples which was contaminated by Additive White Gaussian Noise (AWGN) in the middle portion of the signal as shown in Figure 2(a). The signal was split up as in Eq. (6) with $W=128$ and $R=128$ and a rectangular window $w[j]$. The 1-D $LBP_8^{r,u}$ extension was used with $P=8$ to give a LBP histogram with $P+2=10$ bins as shown in Figure 2(b) for each segment. The D_{RAD} values for adjacent segments are shown for illustrative purposes. The results of the segmentation are shown in Figure 2(c) and Figure 2(d). It can be seen that the algorithm exactly separates the sinusoidal components from the noise affected portion based on the similarity of the underlying signal features. No overlap was used in this example, however overlapping the segments will improve fidelity.

4. VOICE ACTIVITY DETECTION USING 1-D LBP

Traditional VAD detects speech activity in the presence of noise. VAD does not usually distinguish between voiced and unvoiced components [4][5][8]. Unvoiced speech contains high occurrences of non-uniform patterns and use of the uniform LBP extension, $LBP_P^{r,u}$, can distinguish between these two speech components. The speech utterance

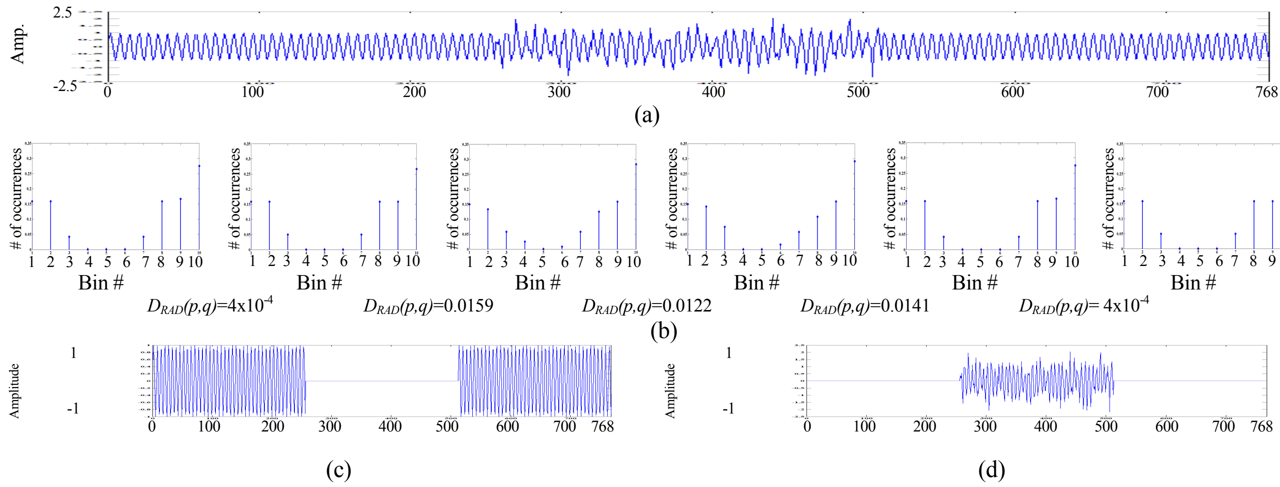

Figure 2 - Segmentation of a sinusoidal signal contaminated by AWGN (a) Original noisy signal (b) $LBP_8^{r,u}$ histograms of the 6 segments formed and $D_{RAD}(p,q)$ measure for adjacent histograms (c) Segmented sinusoidal components (d) Noise affected segment

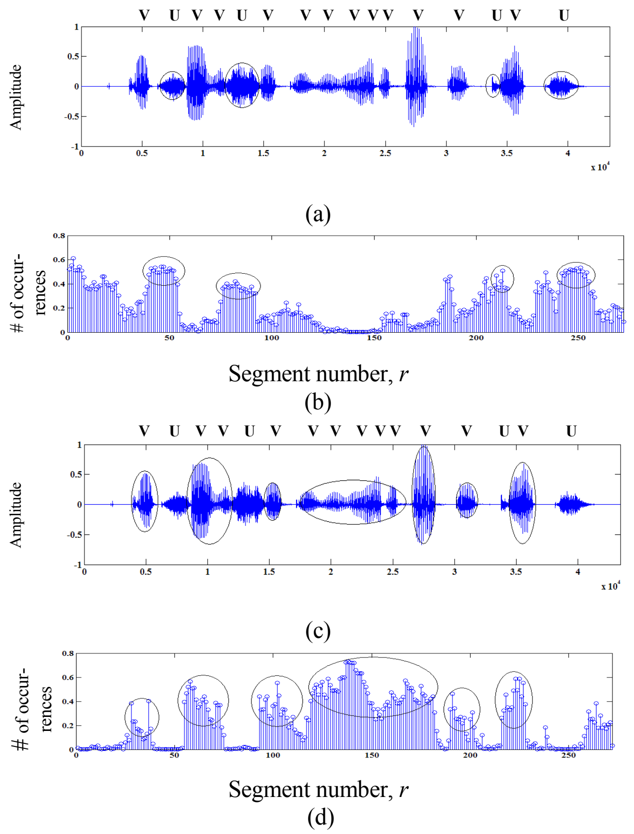

Figure 3 – $LBP_8^{r,u}$ results for clean speech utterance (a) Clean speech with unvoiced segments circled (b) Occurrence results in non-uniform bin 10 from LBP feature histograms (c) Clean speech with voiced segments circled (d) Occurrence results in central uniform bin 5 from LBP feature histograms

"*Good service should be rewarded by big tips*" was taken from the TIMIT database [6] and is plotted in Figure 3(a). A sampling frequency of 16 kHz was used and the signal was segmented according to Eq. (6) with a rectangular window of length $W=160$ samples and no overlap. The $LBP_8^{r,u}$ for each segment was measured to give LBP histograms with 10 bins. Any non-uniform patterns are separated into a single bin. Figure 3(b) shows the plot for the non-uniform bin (bin 10) for each speech segment. This illustrates that the higher frequency unvoiced speech circled in Figure 3(a) and labelled "U" produce higher occurrences of non-uniform patterns. Non-uniform patterns occur in other portions of the signal. This is due to low-power recording noise from the speech sample used. This distinctive non-uniform marker can be used to identify unvoiced speech segments of the analyzed signal that have an increased number of occurrences in the non-uniform histogram bin.

The lower frequency voiced components are highlighted in circles and labelled "V" in Figure 3(c). These produce uniform patterns with the resulting plot shown in Figure 3(d). This shows the number of occurrences in the central uniform bin 5 for the segmented signal. The distribution of the patterns for speech signal shows peak activity in the uniform bin 5 at segments corresponding to voiced speech. This LBP feature relates to a particular rotation-invariant feature of the voiced components. It can be seen that during voiced speech activity there is significant activity in this central bin. Therefore, the occurrence histograms of these speech components can distinguish these two regions based on their extracted LBP features. Noise may contain non-uniform patterns and for noisy speech signals, the bin 5 features can also distinguish unvoiced speech components from weaker voiced speech components that have been more affected by the added noise. A higher resolved histogram such as LBP_P^u can be used if this criterion to distinguish unvoiced speech from noise or weak speech components affected by noise is required. LBP_P^u distributes the occurrences in the histogram over a larger number of bins and thus keeps activity low in any particular uniform bin for unvoiced speech.

Environmental sounds may contain low-frequency noise and periodic components whose spectra overlap with the voiced components of the speech signal. Therefore, discrimination of features that produce similar histograms from

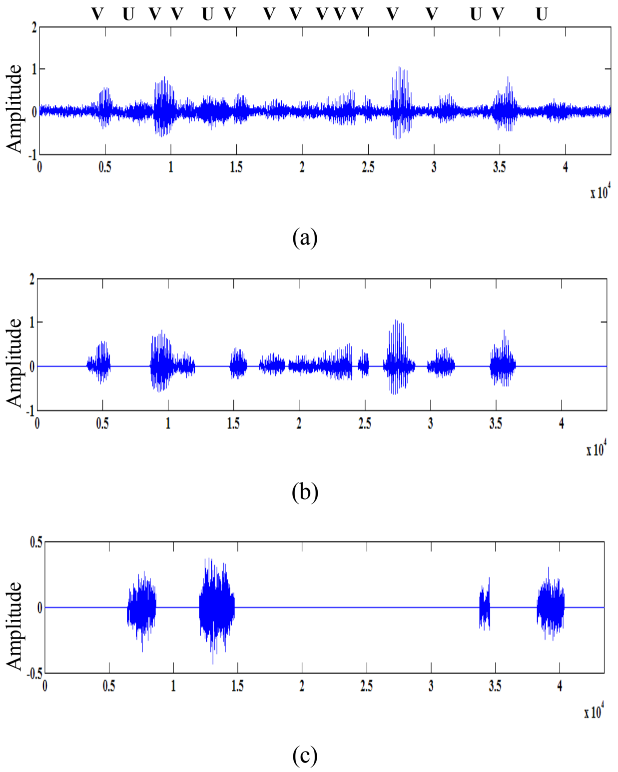

Figure 4 - VAD results for speech contaminated with F16 cockpit noise at 5 dB SNR (a) Noisy speech (b) Voiced speech components identified (c) Unvoiced speech components identified

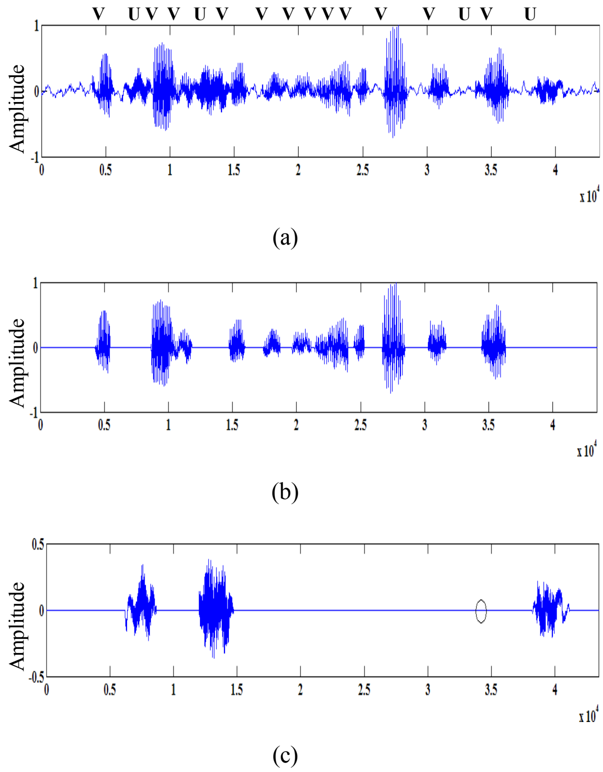

Figure 5 - VAD results for speech contaminated with car interior noise at 5 dB SNR (a) Noisy speech (b) Voiced speech components identified (c) Unvoiced speech components identified

different sound sources is performed by incorporating a local power measure of the analyzed signal segment $x_a[j]$ to give the joint operator $LBP_P^{r,u}/VAR_{seg}$ where $VAR_{seg}(x_a[j])$ is given by:

$$VAR_{seg}\left(x_a[j]\right)=\frac{1}{W}\sum_{j=0}^{W-1}(x_a[j]-\bar{X}_a)^2 \text{ where } \bar{X}_a=\frac{1}{W}\sum_{j=0}^{W-1}(x_a[j]) \qquad (7)$$

4.1 1-D LBP-based VAD algorithm

The algorithm presented below uses the 1-D LBP to separate noisy speech into voiced, unvoiced and non-speech components by the following steps:

1. Segment the input noisy speech signal x to give segments $x_a[j]$
2. Perform LBP_8^{riu2} for each segment $x_a[j]$ to obtain the normalized occurrence histogram for that segment
3. Separate all segments which have the normalized histogram bin $p(10)>0.3$ and label as unvoiced speech segments. $p(k)$ is the occurrence probability in histogram bin k
4. Measure $VAR_{seg}(x_a[j])$ for each segment and separate the LBP features with $VAR_{seg}(x_a[j])<thresh$. Label as non-voiced speech segments
5. Label remaining segments as voiced speech segments
6. Perform final grouping by assigning contiguous speech segments $T_{NV}<$ 50 ms to non-voiced speech label

The value of *thresh* must be chosen to distinguish voiced speech from non-speech with similar LBP features. A value of 0.003 for *thresh* was selected empirically from experimental studies for speech contaminated with different noise types ranging down to 0dB SNR. The value of T_{NV} was chosen as in [10] to remove the influence of noise intrusion.

4.2 Performance Evaluation

The 1-D LBP-based VAD algorithm from section 4.1 was tested on the previous clean speech utterance from Figure 3(a) degraded with F16 cockpit noise and car interior noise. These noise sources were obtained from the Noisex92 [7] database. Figure 4 shows the results obtained for the speech utterance contaminated with F16 cockpit noise at SNR level of 5 dB with the voiced and unvoiced components labelled “V” and “U” respectively. Figure 4(b) and Figure 4(c) demonstrate that the LBP-based VAD algorithm is able to correctly identify all of the voiced and unvoiced components from the noisy speech. Figure 5 shows the results obtained for the speech utterance contaminated with car interior noise at SNR level of 5 dB. Figure 5(b) and Figure 5(c) demonstrate that the LBP-based VAD algorithm is able to correctly identify all of the voiced speech components. However, it does not identify one weak portion of the unvoiced speech since its LBP feature was affected by the stronger low-frequency noise component for low SNR values. The LBP feature for this unvoiced portion was not significantly modified in the previous case with the higher frequency components of the F16 cockpit noise and therefore resulted in correct identification in that situation.

5. DISCUSSION

The histogram of the 1-D LBP codes of a signal gives a sparser, alternative signal representation. The LBP operation is fast and computationally inexpensive. It was shown to be a distinctive marker of certain features of the underlying signal. This property has been applied in preliminary work for simple signal segmentation and fast and accurate VAD. The 1-D LBP is able to distinguish the unvoiced and the voiced components of speech signals using the distinguishing features of higher activity in certain characteristic histogram bins. The use of an overlapping factor will yield improved results and give better identification of the onset of distinct signal features. Future work will involve application of the 1-D LBP to signal enhancement and noise estimation techniques. Multi-resolution 1-D LBP will be developed to achieve improved results, especially for analysis of noisy signals. Further work will also involve the inclusion of a joint local variance measure on the samples that produce an LBP code to give improved fidelity.

REFERENCES

[1] T. Ojala, M. Pietikäinen, “Unsupervised texture segmentation using feature distributions”, in *Pattern Recognition 32*, pp. 477-486, 1999.

[2] S. He, J. J. Soraghan et al, “Quantitative Analysis of Facial Paralysis Using Local Binary Patterns in Biomedical Videos”, in *IEEE Transactions on Biomedical Engineering*, vol. 56(7), pp. 1864-1870, Jul 2009.

[3] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution Gray Scale and Rotation Invariant Texture Analysis with Local Binary Patterns”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24(7), pp. 971-987, 2002.

[4] J. Ramirez, J. C. Segura, J. M. Gorriz, L. Garcia, “Improved voice activity detection using contextual multiple hypothesis testing for robust speech recognition”, in *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15(8), pp. 2177–2189, 2007.

[5] C. Hsieh, T. Feng, P. Huang, “Energy-based VAD with grey magnitude spectral subtraction”, in *Speech Communication*, vol. 51, pp. 810-819, 2009.

[6] TIMIT speech database, Speech Enhancement and Assessment Resource, <http://cslu.cse.ogi.edu/nsel/data/SpEAR_noisyspeech.html> [accessed Sep 2009]

[7] NOISEX – 92 Database, <http://spib.rice.edu/spib/select_noise.html> [accessed Sep 2009]

[8] J. Ramirez, J. C. Segura, C. Benitez, A. de la Torre, A. Rubio, “Efficient Voice Activity Detection Algorithms using Long-Term Speech Information”, in *Speech Communication*, vol. 42, pp. 271-287, 2004

[9] D. H. Johnson, S. Sinanovic, “Symmetrizing the Kullback–Leibler Distance”, in *Technical Report, Rice University*, 2001

[10] G. Hu, D. Wang, “Monaural Speech Segregation based on Pitch Tracking and Amplitude Modulation”, in *IEEE Transactions on Neural Networks*, vol. 15(4), pp. 1135-1150, Sep 2004.