A Low-Complexity Voice Activity Detector for Smart Hearing Protection of Hyperacusic Persons

Narimene Lezzoum, Ghyslain Gagnon and Jérémie Voix

École de technologie supérieure, 1100 Notre Dame West, Montréal (Qc) H3C 1K3 Canada

Abstract

In this paper, a Voice Activity Detector (VAD) is proposed for smart hearing protection applications where speech is to get through the hearing protector while ambient noise is to be blocked out. The VAD calculates a short-term statistical assessment of the temporal envelopes within different frequency bands. This assessment uses the Inter-Quartile Range (IQR) and reflects the dispersion of the envelopes' magnitudes. The VAD's decision is made using two threshold comparison rules and a hangover scheme triggered after a given number of observations. These four parameters have been optimized off-line using a genetic algorithm approach. The performance of the proposed VAD is compared to Sohn's VAD using a database of 90 speech signals corrupted by five real-world noise environments at Signal-to-Noise ratios (SNR) varying from 0 to +10 dB. Results show that the proposed VAD performs better than Sohn's VAD with an 85.9% (compared to 77.5%) F1 score averaged across all SNRs and also minimizes by a factor of three the mid-speech clipping rate. In addition, the evaluation of the proposed VAD's computational cost shows that its implementation on-board a low-power low-consumption DSP is very feasible and would enable smart hearing protection for hypersensitive persons.

Index Terms: Voice activity detection, inter-quartile range, genetic algorithms, temporal envelope

1. Introduction

Hyperacusis is defined as hypersensitivity and intolerance to ordinary environmental sounds [1]. It has been mentioned in [2] that one in 10 people report such sensitivity to sound. Over time, persons with hyperacusis begin to avoid social interaction, withdraw completely from environments that were once pleasant and become socially isolated [3]. The most common treatment for this hearing disorder is desensitization by careful presentation of sounds -limited in level and progressive in time-, as well as wearing passive hearing protection devices (HPDs) during daily activities to prevent the situation from worsening until the desensitization therapy has succeeded [1].

However, wearing passive HPDs is somewhat inconvenient for these patients because HPDs not only block unwanted noise signals, but also wanted speech signals. To palliate this problem, a $smart$ HPD i.e,. an active HPD that guarantees protection while discriminating between speech and noise to allow speech signals to get through to the protected ear is being worked on. For this purpose, the integration of a Digital Signal Processor (DSP) in the traditional passive HPD is required. The smartness of this HPD lies in its capability of transmitting speech signals while protecting the ear from environmental noise.

The discrimination between speech and noise signals is known in the literature as Voice Activity Detection (VAD). Numerous VAD algorithms have been developed; some require the extraction of features such as: the periodicity [4], zero crossing rate, full and low band energy and line spectrum frequencies [5] or pitch [6]. However, the performance of these VADs degrades when the SNR decreases [7]. To palliate this problem, other VADs have been developed and require the characterisation of noise depending on an estimate during noise periods such as the calculation of the a posteriori and a priori SNR [8]. Nevertheless, these VADs are sensitive to changes in the SNR [9]. Therefore, some researchers resort to learning techniques or modelling algorithms in their VAD [10], [11] and [12]. This however, leads to other problems when the intended application must operate in an embedded system with limited hardware resources.

In this paper, we propose the calculation of a short-term statistical assessment of the temporal envelope within different frequency bands. Extracting features from the temporal envelope has been widely used for hearing aids to detect the presence of speech and decide when gain should be reduced [13], [14], [15].

The VAD's decision is made after multiple observations using two thresholds in addition to a hangover scheme to take into consideration "long time" information, knowing that speech signals are highly time-correlated [16]. Thresholds, number of observations and hangover parameters are optimized off-line using a Genetic Algorithm (GA) [17]. The VAD's decision is set after multiple observations and using a hangover scheme to minimize false positives and mid-speech clipping knowing that for hyperacusis patients wearing smart hearing protection, perception of "short time" noise signals is unpleasant.

The paper is organised as follows. Section 2 introduces the proposed VAD algorithm. Section 3 describes the off-line parameters optimization. Section 4 presents the validation and discussions and section 5 the conclusions.

2. Proposed VAD Algorithm

Figure 1 illustrates the detailed architecture of the proposed VAD where N is the number of observations, i the frame number and m the frequency band number.

2.1. Windowing

The entire signal is first cut into frames with a Hamming window. The length of each frame is 25 ms with an 80% overlap.

2.2. Feature Extraction

2.2.1. Filter Bank

Each frame is passed into a filterbank of 16 frequency bands using -for ease on device implementation- a 4^{th} order Butterworth filter. Cut-off frequencies are described in the Bark scale [18] and lie between 20 and 3150 Hz.

Lezzoum, Narimene, Ghyslain Gagnon and Jérémie Voix. 2013. "A low-complexity voice activity detector for smart hearing protection of hyperacusic persons". In 14th Annual Conference of the International Speech Communication Association (Lyon, France, Aug. 25-29, 2013), p. 4-8.

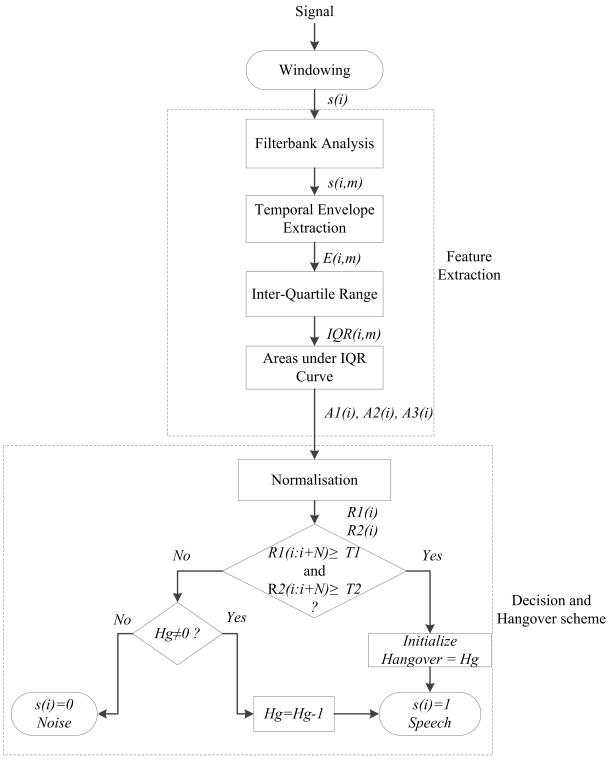

Figure 1: *Diagram block of the proposed VAD algorithm.*

2.2.2. *Temporal Envelope Extraction*

For each frame, the temporal envelope of each frequency band is extracted using the Hilbert Transform. Envelope extraction using the Hilbert transform involves the calculation of the analytic signal [19], as illustrated in Eq.1, where $E(t)$ is the Hilbert envelope of $x(t)$.

$$\mathrm{E(t)} = \sqrt{x(t)^2 + \tilde{x}(t)^2} \tag{1}$$

with $\tilde{x}(t)$ the Hilbert Transform of $x(t)$:

$$\tilde{x}(t) = x(t) * \frac{1}{\pi t} \tag{2}$$

2.2.3. *Statistical Assessment of Temporal Envelopes*

The statistical assessment is the Inter-Quartile Range (IQR) and is calculated within the temporal envelopes of the various frequency bands by using the 75^{th} percentile, or third quartile ($Q3$): the value below which 75% of the values in the distribution lie, and the 25^{th} percentile, or first quartile ($Q1$): the value above which 25% of the values lie. The IQR is calculated as shown in equation 3.

$$\mathrm{IQR} = Q3 - Q1 \tag{3}$$

Figure 2 illustrates an example of the IQR in all frequency bands for one signal's frame showing speech produced by a male speaker corrupted by noise with 5, 0 and -5 dB SNRs, speech and then noise. Figure 2 also shows that in the eighth frequency band (770-920 Hz) which represents the first formant of the speech segment (a voiced phoneme), the IQR of speech in a quiet setting is higher than that of the noise signal.

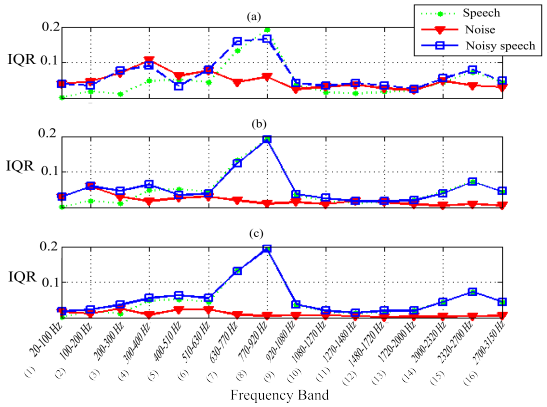

Figure 2: *IQR calculated in the frequency bands of one signal's frame with (a): -5 dB, (b) 0 dB and (c) 5 dB SNR.*

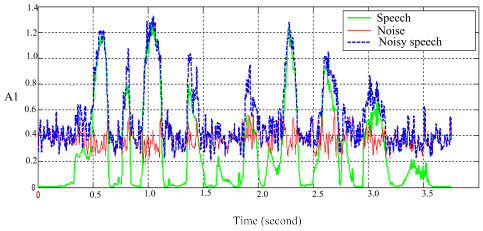

Figure 3: *The area $A1$ calculated for speech in a quiet setting, noisy setting with 0 dB SNR, and separate noise signal.*

2.2.4. *Area under IQR Curve*

The ascertainment noted from Figure 2 is confirmed by the studies conducted in [20] on the influence of noise on vowels and consonants, which concluded that when the speech signal is corrupted by noise, the first formant can be reliably detected compared to the second formant, which is heavily masked by noise in low SNRs. Based on this conclusion, we choose not to consider only one frequency band to characterize a speech signal, but rather the area under the IQR curve from the third to the ninth frequency band. This area is named $A1$ and its choice is based on the frequency region containing the largest amount of speech information. The area under the IQR curve is calculated to take into consideration both spectral (first formant) and temporal (IQR) characteristics. Figure 3 illustrates $A1$ for a speech signal corrupted by a 'Car' noise in 0 dB SNR and $A1$ for noise and speech separately. Figure 3 illustrates that $A1$ is typically high when a speech signal is present, whereas when only noise is present this area is low. This ascertainment has been validated on different utterances.

2.3. VAD's Decision

2.3.1. *Normalization*

Figure 3 shows that the VADs' decision could be performed using a decision threshold upon $A1$. However, this procedure is not applicable directly on $A1$ since $A1$ depends on the IQR, which itself depends on the scale of the temporal envelope. This yields us to normalize the data by using two other areas under the IQR curve that reflect the noise signal.

Speech and noise signals differ in their frequency compo-

nents: noise signals have generally more energy in the lower frequencies than speech signals, which have a lower energy in these frequencies [21]. This ascertainment yields us to calculate the area $A2$ under the IQR curve from the first frequency band (20-100 Hz) to the second (100-200 Hz). The choice of this area is based on the frequency region containing the most noise information and the least speech information, it has been found empirically to be the most reliable for noise assessment. In addition to the $A2$ area, we added another area under the IQR curve ($A3$) that characterizes high frequency noises. This additional area is calculated in the high frequency bands and represents an alternative choice in the decision.

The three areas show in our testing the same trends: when the signal's level increases, $A1$, $A2$, $A3$ increase and similarly, when the signal's level decreases, $A1$, $A2$ and $A3$ decrease. This trend leads us to calculate the ratios $R1$ and $R2$ (see Eq. 4 and 5), upon $R1$ and $R2$ the first and second decision thresholds $T1$ and $T2$ are determined using the genetic algorithm approach.

$$\text{R1} = \frac{A1}{A2} \tag{4}$$

$$\text{R2} = \frac{A1}{A3} \tag{5}$$

The use of $T1$ and $T2$ as a decision rule eliminates the need for an adaptive decision threshold or an SNR estimator.

$T1$ and $T2$ must be optimized in addition to two other parameters: first, the number of observations that represents the number of consecutive frames having $R1$ and $R2$ higher than $T1$ and $T2$ respectively and after which the decision might be set to 1 (speech) and second, the hangover parameter, which represents the time after which the VAD is reset to 0.

3. Off-Line Parameters Optimization

3.1. Start of Speech Confirmation and Hangover Scheme in Smart Hearing Protection

The start of speech confirmation is defined as the number N of consecutive frames having $R1$ and $R2$ higher than $T1$ and $T2$ and after which the decision is set to 1. They have been used in Ramirez et al's VAD [22], where it was demonstrated that taking several frames into account in the VAD improves the reliability of the decisions.

The value N cannot exceed a certain number of consecutive frames, otherwise lip-sync errors may occur. Lip sync errors are defined by the ITU [23] as the errors between lip movement and the perceived speech signal, and a lip-sync error of 40 ms was considered acceptable. Thus, the maximum number of consecutive frames after which the decision might be set to one in the proposed VAD is eight consecutive frames, which represents a delay of 40 milliseconds.

The hangover scheme or end of speech confirmation has been widely used in VADs to minimize the false rejection rate caused by the non-detection of low energy speech frames containing consonants such as fricatives and unvoiced stops.

3.2. Objective Function

To optimize the thresholds ($T1$ and $T2$), hangover, and number of observations, an objective function should be minimized. This function's role is to evaluate the performance of the VAD algorithm. For this purpose, we used the F1 score measure [24]. This score combines the FPR (False Positive Rate), TPR (True Positive Rate) and FNR (False Negative Rate). Knowing that FPR, TPR, FNR are based on maximum of 100%.

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

with

$$\text{precision} = \frac{\text{TPR}}{\text{TPR} + \text{FPR}} \tag{7}$$

$$\text{recall} = \frac{\text{TPR}}{\text{TPR} + \text{FNR}} \tag{8}$$

For the smart hearing protection application, we calculated the TPR and FNR for the noisy speech signals and the FPR for the noise signals. This evaluation method focuses on the fact that once the speech signal has been detected it must be transmitted in its entirety to the Smart HPD's wearer -possibly with a few seconds extra duration- while continuing to protect the wearer from noise when no speech signals are present.

The objective function to be minimized is shown in Equation 9:

$$\text{Penalty} = 1 - F1 \tag{9}$$

3.3. Genetic Algorithm for Off-Line Parameter Optimization

Genetic Algorithms (GAs) [17] are randomized search and optimization techniques based on the mechanisms of natural selection and natural genetics. They are used to optimize the four parameters using an optimization database. For the purposes of this application, we have used a limited samples of five speech signals corrupted by 'Subway' noise, with a 0 dB SNR knowing that many hyperacusis patients are exposed daily to this type of noise. The speech signals are from the TIMIT database [25] and the noise signals from the AURORA database [26], both sampled to 16 kHz. The hangover's duration tends to vary between 50 and 250 frames which represents 0.25 to 1.25 seconds and the number of observations varies from 4 to 8 consecutive frames. The upper boundary of the hangover seems long when compared to the hangover durations used in telecommunications or speech recognition [27] and [16]. However, it was considered that this was not to be included in the objective function since it theoretically does not reduce the performance of the algorithm as it could reduce the performance of VADs for telecommunications or speech recognition. It will not affect the process but permit entire speech signal transmission without interruption.

After 10 generations, the GA reached an optimal solution with a best penalty value of 9%, which is equal to an F1 score of 91%. The optimization process gave a hangover of 250 frames and 6 consecutive frames.

4. Validation and Discussions

In the first part of this section, we present the VAD's performance assessment and in the second part, we quantify the computational cost of the proposed VAD.

4.1. Performance Assessment

The validation database is composed of 90 speech signals corrupted by five everyday noise environments with 3 SNRs (10, 5 and 0 dB). The speech signals are from the TIMIT database and the noise signals from the AURORA database. The average length of each speech signal corrupted by noise is 3.06 seconds and 83.6% of the signal comprises speech.

Noise environment		Sohn VAD		Proposed VAD		
Noise	SNR	F1	MSC	F1	MSC	F1*
Exhibition	0 dB	75.0	15.4	80.0	10.1	85.9
	5 dB	78.9	10.8	91.3	2.3	94.5
	10 dB	79.2	6.7	94.9	0.7	98.1
Babble	0 dB	73.2	9.9	78.6	1.6	76.3
	5 dB	74.9	7.3	82.8	0.9	82.7
	10 dB	76.1	5.6	82.1	0.4	81.0
Subway	0 dB	74.8	13.3	79.1	4.1	84.5
	5 dB	76.2	8.5	91.3	1.8	88.9
	10 dB	78.1	6.4	94.9	0.6	90.9
Airport	0 dB	76.2	9.4	77.5	4.5	77.7
	5 dB	77.7	6.9	86.2	1.5	87.1
	10 dB	79.2	5.2	87.5	0.2	85.5
Car	0 dB	79.7	15.6	77.0	13.2	79.8
	5 dB	81.5	10.8	91.5	2.6	96.0
	10 dB	83.0	7.4	95.1	0.5	98.7
Average	Average	77.5	9.2	85.9	3.0	87.3

Table 1: *Performance evaluation of the proposed VAD compared to Sohn's VAD using the F1 score and the MSC rates.*

As mentioned previously, the F1 score is used to evaluate the VAD's performance. Sohn's VAD [8] has been implemented from the VoiceBox [28]. The proposed VAD is compared to Sohn's VAD, which has proven its effectiveness with standard G729.B [5] AMR1, AMR2 [6] as demonstrated in [22] and [8].

In addition, we calculated the Mid-Speech Clipping rate (MSC) which represents the rate of speech frames classified as noise in the middle of the utterance. This measure is very important for speech intelligibility. The lower it is the more the speech segment is intelligible.

Table 1 illustrates the comparison of the two VADs.

As shown in Table 1, the F1 score of the proposed VAD is higher than the F1 score of Sohn's VAD in all noise environments and SNRs except for the 'Car' noise in 0 dB SNR which gives a F1 score of 77% instead of 79.5% for Sohn's VAD. The performance of the proposed VAD is more noticeable in the range of 5 and 10 dB SNR where the F1 score average in these SNRs has an increase of 11.2% for the proposed VAD.

Furthermore, we note from this table that the proposed VAD minimizes about three times the mid-speech clipping rate in comparison to Sohn's VAD. This leads us to say that the hangover scheme described in this paper is not only simpler but also more efficient than Sohn's hangover.

Moreover, we evaluated the proposed VAD using one speech signal of 150 seconds duration with 77.4% of speech (46 signals concatenated into one signal without additive noise periods between the 46 speech signals) corrupted by five noise environments at three SNR levels. This evaluation was conducted to validate the proposed algorithm with a signal of long duration to ensure that the performance of the proposed VAD is not only due to the hangover's duration. F1 scores are illustrated in the last part of Table 1(F1*). F1* shows almost the same F1 scores found earlier which enables us to validate the proposed VAD for its further implementation.

4.2. Computational Cost

The required hardware resources for the smart hearing protector are quite similar to those presently used in hearing aids and cochlear implants. The first two steps used in the feature extraction stage of the proposed VAD are already optimized to work in DSPs with limited hardware resources. For instance, DSPs for hearing aids are provided with an integrated filterbank coprocessor: the WOLA (Weighted Overlap Add) filterbank coprocessor [29], which allows the splitting of the signal in different frequency bands using an optimized architecture. For this purpose, we evaluated the additional computational cost arising from the IQR and areas calculation, to calculate by how much these two steps increase the number of instructions per second in the entire process.

Data must be sorted to calculate the IQR by using a sorting algorithm. Among the existing sorting algorithms, the Merge sort requires $Nlog_2N$ operations per frame [30]. Furthermore, to calculate $A1$, $A2$ and $A3$, 30 additions and 10 multiplications per frame are required. Table 2 shows the overall resource requirements for these two steps.

Processing step	Op. per frame	Op. per second
IQR	55,337	11,067,400
Areas	40	8,000
Global	55,377	11,075,400

Table 2: *Resource requirements for the 3rd and 4th steps in the feature extraction stage of the proposed VAD (abbreviation Op. defines the number of operations).*

The targeted DSP for smart hearing protection offers typically 60 MIPS (Million Instructions Per Second). Thus, the number of instructions per second required for the IQR and areas is 18.4% of the entire available number of instructions per second. This is reasonable since 81.6% of the entire computational cost could be dedicated to the filterbank, the Hilbert envelope extraction, and other operations such as noise reduction and dynamic range adaptation.

5. Conclusions

In this paper we proposed a new VAD particularly suited for smart hearing protection for hyperacusis patients. The proposed VAD uses a short term statistical assessment of the temporal envelope within different frequency bands. The VAD's decision is made after multiple observations using two decision thresholds and a hangover scheme, all optimized off-line using a genetic algorithm. Experiments conducted using speech signals corrupted by five real-world noise environments show that coupling the multiple observations and the hangover scheme in the decision process permits the maximization of the VAD's performance. Results show that the proposed VAD is more efficient than Sohn's VAD which by itself is more efficient than the Standards G.729b and AMR1, AMR2. This leads us to assume that the proposed VAD outperforms these standards as well. In addition to these satisfactory results, the proposed VAD requires neither assumption nor noise estimation depending on the first signal's frames, and is sufficiently simple to be implemented in a DSP of limited hardware resources. In future work, we intend to validate the proposed VAD with subjective tests, work on noise reduction to render the speech signals intelligible and adapt the dynamic range of the incoming speech signals to send them to the protected ear without damaging it.

6. Acknowledgements

The authors would like to thank Sonomax Technologies Inc. and its "Industrial Research Chair in In-ear Technologies" for its financial support.

7. References

[1] J. Vernon, "Pathophysiology of tinnitus: a special case hyperacusis and proposed treatment," *The American Journal of Otology*, vol. 8, pp. 201–202, 1987.

[2] G. Andersson, N. Lindvall, T. Hursti, and P. Carlbring, "Hypersensitivity to sound (hyperacusis): a prevalence study conducted via the Internet and post," *International Journal of Audiology*, vol. 41, pp. 545–554, 2002.

[3] M. Valente, J. Goebel, D. Duddy, B. Sinks, and J. Peterein, "Evaluation and Treatment of severe Hyperacusis." *Washington University School of Medecine in St. Louis. Paper 15*, vol. 11, no. 6, pp. 295–9, Jun. 2000.

[4] R. Tucker, "Voice activity detection using a periodicity measure," *IEE Proceedings I Communications, Speech and Vision*, vol. 139, no. 4, pp. 377–380, 1992.

[5] ITU T, "Annex B: A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70," *International Telecommunication Union*, 1996.

[6] ETSI, "Digital cellular telecommunications system (Phase 2+); Voice Activity Detector (VAD) for Adaptive Multi-Rate (AMR) speech traffic channels; General description (GSM 06.94 version 7.1.0 Release 1998)," Tech. Rep., 1999.

[7] F. Beritelli, S. Casale, G. Ruggeri, and S. Serrano, "Performance evaluation and comparison of G.729/AMR/fuzzy voice activity detectors," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 85–88, Mar. 2002.

[8] J. Sohn, N. S. Kim, and W. Sung, "A Statistical Model-Based Voice Activity Detection," *IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 1998–2000, 1999.

[9] M. H. Moattar and M. M. Homayounpour, "A Simple But Efficient Real-Time Voice Activity Detection Algorithm," in *17th European Signal Processing Conference*, 2009, pp. 2549–2553.

[10] E. Chuangsuwanich and J. Glass, "Robust Voice Activity Detector for Real World Applications Using Harmonicity and Modulation frequency," *INTERSPEECH*, pp. 2645–2648, 2011.

[11] X. Liu, Y. Liang, Y. Lou, H. Li, and B. Shan, "Noise-Robust Voice Activity Detector Based on Hidden Semi-Markov Models," *IEEE, 20th International Conference on Pattern Recognition*, pp. 81–84, Aug. 2010.

[12] J. Wu and X. Zhang, "Efficient Multiple Kernel Support Vector Machine Based Voice Activity Detection," *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 466–469, 2011.

[13] L. C. R. Bentler, "Digital noise reduction : An Overview," *Trends in Amplification*, vol. 10, no. 3, pp. 67–82, 2006.

[14] G. Mueller and T. Ricketts, "Digital noise reduction : Much ado about something?" *The Hearing Journal*, vol. 58, no. 1, pp. 10–17, 2005.

[15] K. Chung, J. Tufts, and L. Nelson, "Modulation-Based Digital Noise Reduction for Application to Hearing Protectors to Reduce Noise and Maintain Intelligibility," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 51, no. 1, pp. 78–89, May 2009.

[16] A. Davis, S. Nordholm, and R. Togneri, "Statistical Voice Activity Detection Using Low-Variance Spectrum Estimation and an Adaptive Threshold," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 412–424, 2006.

[17] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, 1989.

[18] E. Zwicker, "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)," *The Journal of the Acoustical Society of America*, vol. 33, no. 2, p. 248, 1961.

[19] S. L. Marple, "Computing the Discrete-Time "Analytic" Signal via FFT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2600–2603, 1999.

[20] G. Parikh and P. Loizou, "The influence of noise on vowel and consonant cues," *The Journal of the Acoustical Society of America*, vol. 118, no. 6, pp. 3874–3888, 2005.

[21] H. Levitt, "Noise reduction in hearing aids: a review." *Journal of Rehabilitation Research and Development*, vol. 38, no. 1, pp. 111–121, 2001.

[22] J. Ramírez, J. C. Segura, C. Benítez, L. García, and A. Rubio, "Statistical Voice Activity Detection Using Multiple Observation Likelihood Ratio Test," *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 689–692, 2005.

[23] ITU, *International Telecommunication Union Document 11A/47-E*, 1993.

[24] van Rijsbergen, *Information Retrievial*, 2nd ed., Butterworths, Ed., 1979.

[25] S. V.Zue and J.Glass, "Speech Database Development: TIMIT and Beyond," *Speech Communication*, vol. 9, no. 4, pp. 351–356, 1990.

[26] H.-g. Hirsch and D. Pearce, "The Aurora Experimental Framework for the Performance Evaluation of Speech Recognition Systems Under Noisy Conditions," *ISCA ITRW ASR2000 "Automatic Speech Recognition : Challenges for the Next Millennium"*, 2000.

[27] D. Vlaj, M. Kos, M. Grašič, and Z. Kačič, "Influence of Hangover and Hangbefore Criteria on Automatic Speech Recognition," in *16th International Conference on Systems, Signals and Image Processing, 2009. IWSSIP*, 2009.

[28] M. Brookes, "VoiceBox," 2004. [Online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[29] ON Semiconductors, "Introduction to Audio Processing Using the WOLA Filterbank Coprocessor," pp. 1–10, 2009.

[30] D. E. Knuth, *The Art of Computer Programming, Sorting and Searching*, 1998.