A REAL-TIME SIREN DETECTOR TO IMPROVE SAFETY OF GUIDE IN TRAFFIC ENVIRONMENT

*F. Meucci, L. Pierucci, E. Del Re, L. Lastrucci**

Department of Electronics and Telecommunications, University of Florence
V.S.Marta,3 - 50139 Firenze, Italy
phone: + 390554796271, fax: + 39055494569, email: laura.pierucci@unifi.it
web: lenst.det.unifi.it
*Powersoft S.r.l.
Via Enrico Conti , 5 - Scandicci 50018 - Firenze - Italia
Ph: +39-055-7350230 Fax: +39-055-7356235

ABSTRACT

A system based on a modified pitch detection method is proposed that can be used for the detection of acoustical signals the frequency components of which vary according to specific periodic patterns. Usually, signals of this category are produced by the siren of an emergency vehicle. The detection of this type of signals can improve the safety of guide for hearing impaired people by alerting the driver with a visual indication for example by putting on a light point on the map of a navigator device. This paper discusses the development of a small, inexpensive Atmel microcontroller based siren detection system that is capable of discriminate the siren signal respect to other sound sources in the traffic. The siren detector is tested on real signals acquired and recorded in city streets with many traffic, different other sound sources (automobile horns, anti-theft alarms,motor noisy etc.) and the presence of a siren approaching. The performance of the detector in terms of false alarm rate and missing of the siren signal is analyzed varying different variable parameters .

1. INTRODUCTION

Emergency vehicles (including police, fire and paramedic) often use high amplitude screeching sirens to warn road users and pedestrians that the emergency vehicle is approaching. A visual alarm provided by the detection of the emergency veichle sound can improve the safety of guide for people with hearing impairements . The European normative only permits for emergency vehicles the use of high/low series of tones. The siren signal for ambulance and fire vehicles alternates two tones respectively at the frequencies of 392 Hz (Sol natural) e 660 Hz (Mi natural). A cycle includes a tone at 392 Hz for a period of 1/3 of the total cycle duration, a tone at 660 Hz for a period of 1/18 of the total duration, a tone at 392 Hz for a period of 1/18 of the total, a tone at 660 Hz for a period of 1/18 of the total duration, a tone at 392 Hz for a period of 1/3 of total duration, a tone at 660 Hz for a period of 1 /18 of the total cycle duration, a tone at 392 Hz for a period of 1/18 of the total duration, a tone at 660 Hz for a period of 1/18 of the total length cycle. The sounds must succession without interruptions and without appreciable overlap. The duration of thecomplete cycle is 3 sec. + / - 0.5 sec. Between a cycle and the next there may be a pause whose duration should not exceed 0.2 sec.

For the police vehicles the tones are equally interspaced at the frequencies of 466 Hz (La Diesis) and 622 Hz (Re Diesis). The range of frequencies of the siren for police services is contained in the range of frequencies for emergency vehicles. A cycle includes a tone at 466 Hz without interruptions and without appreciable overlap from a tone at 622 Hz, followed by yet another tone at 466 Hz and one at 622 Hz. The cycle must take place in a time equal to 3 sec. + / - 0.5 sec. This includes any interval between a complete cycle noise and the next. The pause should not exceed the period of 0.2 sec.

Therefore, two qualities are distinctive of a typical pure siren signal: the frequency content and the periodic repetition.

Although the siren sound consists of a number of harmonic spectral components, the one corresponding to the lowest frequency is the dominant. The periodic alteration of the frequency of this dominant component can be visualized as a curve that relates the current value of frequency with time. This curve, which is called the frequency characteristic curve of the siren (FCC), may be considered as pattern and thus the problem of siren identification reduces to a pattern recognition task.

Unfortunately, the electronically generated sirens are supplied from a siren generator whose output is a square wave from a saturating push-pull type output stage. The fundamental of the square wave is the required siren frequency, but many harmonics also exist. The spectral purity of the real siren signal is very poor and the frequencies generated depend also upon vehicle speed, siren age and moreover automobile horns present a frequency spectrum with dominant component.

Therefore, techniques based on spectral analysis (such as FFT or Short Time Fast Discrete Fourier Transform (ST-FDFT)) besides being computational expensive often they do not detect the correct dominant frequency components of the siren sound because they can not be present or they can be overlapped by other components. Method based on the use of filter bank centered on the Hi and Low frequencies do not also detect all the siren power. The method of detection must take advantage of the long duration of the siren signal, lasting at least some seconds respect to other sound sources present in the traffic environment which are usually of short duration. The paper discusses the use of pitch detection algorithm capable of extracting the periodic (siren) signal from aperiodic (speech, automobile horn) ones. In the following the pitch detection algorithm modified for a detection in real time of the siren signal,is explained. In Section 3 the results obtained during a test validation trial in different condition of traffic are provided. The performance of the detector in terms of false alarm rate and missing of the siren signal is analyzed varying different variable parameters and considering different sound sources(speech, automobilke horn, music) which can stress the proper functioning of the pitch detector.

2. SIREN DETECTION PROCESSING

The problem of the siren detection has been attacked with a process working on two levels. First MDF (Module Difference Function), a time domain technique, aims to classify each portion of the audio signal as pitched or unpitched.This first step can be divided in: MDF calculation and Peak Searching. The peak searching gives us the estimation of the pitch frequency. At the end of the first phase we obtain a signal representing the pitch evolving over time, we call this $Pitch(t)$.

Secondly, $Pitch(t)$ is analysed in order to recognise a time pattern of the desired siren type and declare presence or absence of the siren.

2.0.1 Pitch Detection: MDF calculation and Peak Searching

The implementation of a pitch detection algorithm can be realised in various ways. Here we estimate the pitch finding the delay for which the verisimilitude between the signal and a delayed copy of the same signal is maximised. The delay can be found with the correlation function. Classical correlation function needs to be implemented with additions and multiplications:

$$R_{xx}(j) = \sum_{n=0}^{L-1} x_n x_{n-j}^* \tag{1}$$

Since a lightweight implementation for portable device was the focus of the research, we avoid multiplications implementing MDF. MDF obtain completely comparable results respect to the correlation function, at least for the aim of searching the lowest frequency periodicity of the signal. MDF has less computational load especially on power-limited devices without hardware multiplicators. MDF is defined as following:

$$MDF(l,m) = \sum_{n-m-L+1}^{m} |s(n) - s(n-l)| \tag{2}$$

where $s(n)$ is the audio sampled signal, $s(n-l)$ is the delayed audio signal, l is the delay (or *lag*), m is the time index which will be neglected afterwards, L is the length of the of the window on which MDF is calculated, [1-4].

MDF simply operates with absolute values and additions between specifically delayed copies of the signal, [1].

$MDF(l)$ is calculated varying the lag value. We obtain a function over the lag domain. The minimum value of l represent the inverse of the fundamental pitch of the audio signal.

The shape of the $MDF(l)$ function deeply depends on the ratio between the power spectrum of the pitched signal and the power of unpitched components summed with the noise.

$MDF(l)$ can be calculated only for a specific delay interval corresponding to the inverse of the frequency range searched:

$$\{l_{min}, l_{MAX}\} \approx \left\{ \frac{1}{f_{MAX}}, \frac{1}{f_{min}} \right\} \tag{3}$$

The approximate relation holds because $f_{min}, f_{MAX} \in \mathbb{R}$ where instead lag values are discrete with accuracy step of the inverse of the sampling frequency.

In Fig. 1, it is reported the analysis for a signal with a pitched component at $660Hz$.

It's evident that choosing the sampling frequency f_c has a direct impact on the f_{MAX} which is detectable by the process. Virtually, in presence of a $SNR = \infty$ we could state that the minimum $f_{MAX} = f_c$. However, noise and unpitched components degrades the MDF output because each output value is calculated considering all L samples inside the analysis window. In section 3.2, designing parameters are reported for the MDF audio buffer length based also on various audio sampling frequencies.

As seen in Fig. 1, $MDF(l)$ reports several local minimums. The number of minimums inside the output is related to the length of the analysis window.

The pitch detection phase thus needs a peak searching algorithm. The peak searching algorithm concentrates in the frequency band of the desired signal. Different siren specification are available depending on Country Regulatory Bodies. However, in the case where different pitches are to be searched, parallel peak searching algorithms can work with different band settings. The smaller the searching frequency range, more immunity to noise and unpitched signal but higher probability to miss the classification of a pitched signal when the PUNR is low, where we define PUNR as the ratio between the desired signal and interference and noise as: PUNR, Pitched to Unpitched and Noise Ratio.

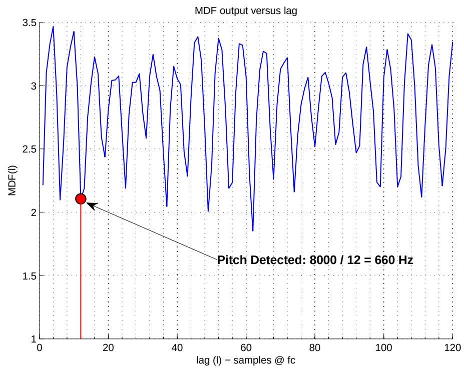

Figure 1: Pitched audio signal with 660 Hz tone

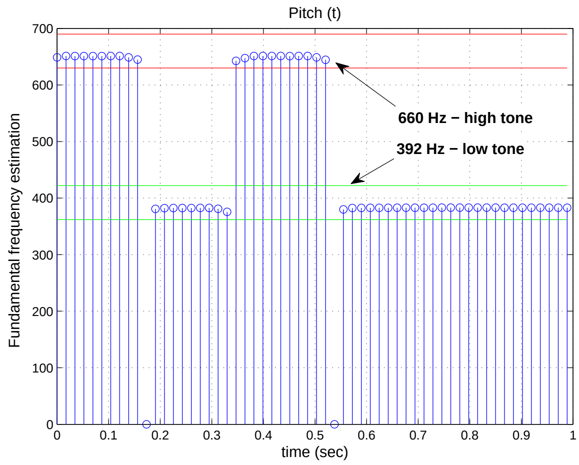

Figure 2: Pitch of a two-tones siren with high and low tones over 1 second. It's possible to see the signal marked unpitched in the transitions between high and low tones: the signal has no dominant pitch in this regions.

After MDF calculation, the peak searching algorithm is searching for a local minimum limiting the search in the band of interest: l_{min}, l_{MAX}. Some more robust peak search approach can be found in [1]. The signal is declared pitched if a local minimums is found in the fundamental band.

2.0.2 Siren Detection

The previous section was explaining the construction of the signal $Pitch(t)$. Based on this signal a second simple estimator is calculating the probability of the presence of a siren. This value is compared with a fixed threshold in order to asset a presence flag. The siren here analysed is a two-tones type: low tones is $392Hz$ and high tone is $660Hz$.

The rate of the $Pitch(t)$ signal is much lower than f_c which is the operating frequency of MDF. Here the sample rate can calculated as:

$$Rate_{Pitch(t)} = \frac{f_c}{L \cdot (1 - OL)} \tag{4}$$

where OL is the overlap of the audio portions entering the MDF function.

For example, with a $f_c = 8000Hz$, MDF is operating at a sample rate of $125\mu s$ while the pattern recognition sample rate is around $16ms$, using MDF windows of 512 samples with 75% overlapping.

A presence probability variable Θ is constructed counting the number of pitches that are inside the specification bands. In Fig. 2

two bands are reported on the vertical axis. These bands are cantered on the specification frequencies (392-660 Hz) plus a range which depends on the maximum Doppler frequency related to the maximum vehicle velocity and to manufacturing tolerances. We determined a range of $50Hz$ accounting for a maximum vehicle velocity of $100Km/h$ and construction tolerances.

Θ value is calculated on a portion of the $Pitch(t)$ signal. In the following we reports results for an observation window of $1sec$ with an overlapping factor of 50%. Furhter consideration about these parameters are given in section 3.5.

Complex pattern recognition algorithms are not more effective then the one here proposed, especially in presence of unpitched or pitched noise. The contemporary presence of a siren and of unpitched not desired signal (music, voice, radio) poses great challenges: in this case, FAR and MISS rates are deeply related to PUNR ratio.

3. RESULTS

In this section we report results obtained both with digitally synthesised and real-time audio signals.

The digitally synthesised signals are obtained with a siren audio sample registered from an emergency vehicle with zero velocity and adding AWGN noise or unpitched and pitched registered samples.

Then the algorithm has been tested when operating on an low power Atmel microprocessor. The audio signal has been samples with a low-price condenser microphone in urban scenario.The sample have been classified as an average of 20 independent listeners. In this section, parameters have been varied in order to optimise the algorithm performance.

3.1 MDF window length: L

A first fundamental parameter for the performance is the length of the buffer containing audio samples to be processed by the MDF function. Design requires a trade-off between:

- long buffers in order to mitigate the effects of noise and not desired signals
- short buffers to limit the computational load of the MDF calculation

However, the signal portion analysed by the MDF needs to contain a sufficient number of replicas of the tone at the fundamental frequency in order to asset a stable estimation. The number of replica in the audio portion is also connected to number of multiplicity checks we can afford in the peak searching algorithm, as explained in section 2.0.1.

In this case, the lowest frequency is specified in the Regulatory body as $392Hz$. The design however considers a lowest frequency of $342Hz$ accounting for 50 Hz due to Doppler negative frequency when the vehicle is moving away from the device and also due to manufacturing tolerances of the siren itself. Minimum buffers size then is expressed as:

$$L_{min} = C \cdot \frac{1}{f_{low}} \cdot f_c \qquad (5)$$

where C is the number of replica included in the buffer.

In Fig. 3 is reported the value of the FAR rate with a $f_c = 8000Hz$. The sample contains a low frequency pitch signal (a car starting the engine and leaving). The presence variable Θ is constructed with the pattern recognition process as specified in section 2.0.2 over the whole sample duration. The performance is gaining a 60% when passing from 128 to 1024 samples.

In Fig. 4 we report the variable Θ in the case where we have an emergency vehicle in urban scenario moving along a straight line with a velocity of around 40 Km/h and 150 meters away from the sensing device in the closest point. Also in this case there is a deep increase in the probability of detection.

These two example shows that, in presence of noise and other disturbing signal, the buffer has to be increased much over the minimum theoretical need when analysing clean signals ($L > 121$ samples).

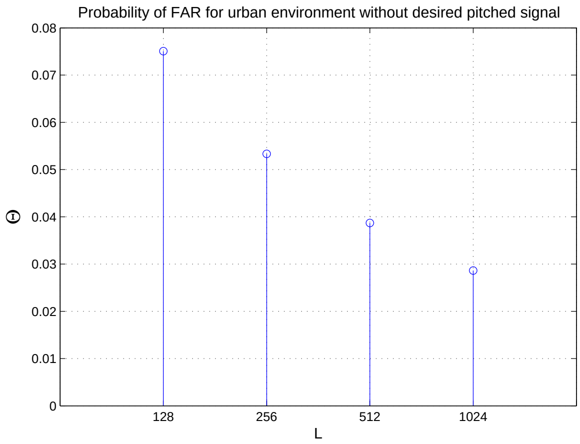

Figure 3: False Alarm Rate over variable MDF buffer.

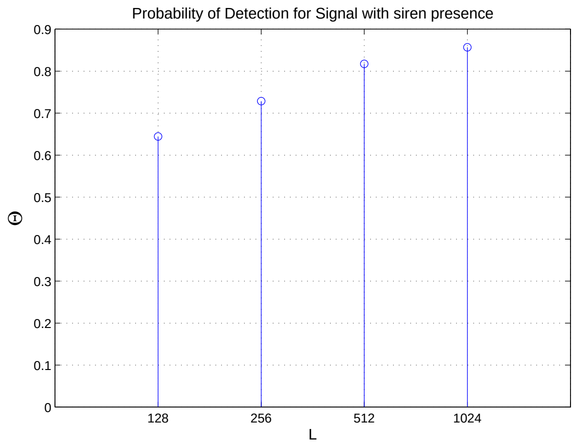

Figure 4: Probability of detection over variable MDF buffer.

3.2 Combined influence of MDF windows length and sampling frequency

Buffer length and sampling frequency are strictly related when analysing the performance of the MDF function.

In Fig. 5 results are shown for the combined effect of f_c and L. Sampling frequencies have been chosen as $4.000 - 8.000 - 11.025Hz$. For each of these, buffers lengths of $256 - 512 - 1024$ are shown. The probability variable Θ is shown for different SNR ratio. The signal is composed with an audio sample of a siren with no Doppler frequency added with AWGN noise.

From the figure, we can see that a sampling frequency of $4KHz$ is leading to very poor performance. This is mainly due to aliasing effect; the siren signal has very high energy concentrated on harmonics, due to the amplifier stage usually used in this equipment. Since the power to be delivered is high and quality is not of concern, the equipment use amplifiers with relative high Total Harmonic Distortion, like class D and E amplifiers, sometimes also drived by a square wave.

The design phase lead to the choice of $f_c = 8KHz$ and $L = 512$ samples. Decision variable Θ is reported in Fig. 5 for a siren signal with additive AWGN noise with various values of MDF buffer length and sampling frequencies. SNR ratios are reported in Decibels and are negative (more noise than siren signal).

In Table 1 are reported the values of Θ for SNR ranging from +10dB to -15 dB. The human hear is almost not able to hear the siren signal at SNR=-10dB.

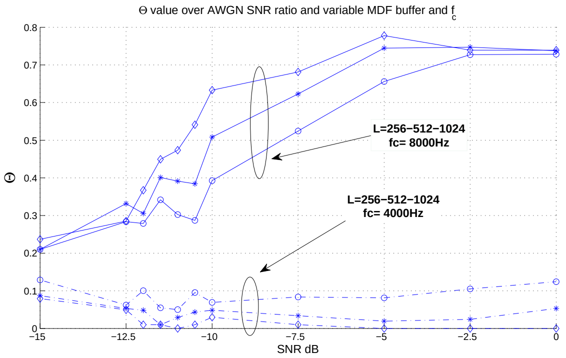

Figure 5: Θ variable for siren signal and AWGN noise for various value of MDF buffers and audio sampling frequencies.

SRN (dB)	Θ
+10	0.983
0	0.979
-10	0.716
-15	0.463

Table 1: Performance in presence of AWGN

3.3 Improving the performance with an high pass audio filter

MDF is very sensitive to low frequency noise. High frequency noise also in the interest band is easily masked by lower frequencies components, even if the power ratio between the two spectrum components is much favourable to the noise component. This is not the case for low frequency noise. This noise tends to mask the real pitch of the signal also when the power ratio is favourable to the pitched signal. This can be explained looking at Fig. 6. Low frequency noise is biasing the output of MDF function thus causing much less reliable minimum detection, since the succession of local minimums is almost monotonically crescent. Thus the peak searching tends to underestimate the lag position of the minimum, especially when high tones are to be detected in the range of interest: $[l_{min}, l_{MAX}]$.

The High Pass filter on audio input has been very effective since various noise sources are concentrated in the lowest part of the frequency spectrum , such as engine frequencies, some wind components, wheels rolling on the road and some part of human voice and music. In Fig. 7 we can see the spectrum of noise, composed by voice from radio station recorded inside a vehicle while in movement. The peak is around 110 Hz. Fig. 7 illustrates also the spectrum components of the siren signal, located around $1300Hz$, the PUNR is $+1.43dB$.

3.4 Performance in urban environment in presence of pitched and unpitched noise

The algorithm has been tested also in real operating environment. Results herein reported are referred to our reference database. This database represents various operating conditions and includes samples of pitched noise (radio, voice, music, car horns) and unpitched noise.

False Alarm and Miss rates have been calculated off-line from the detection history recorded in the device. Following results are obtained with multiplicity minimum search as reported in [1]. The device was operating at the same time with various setting regarding the multiplicity check threshold th. The variable Θ has been stored. Results are thus reporting FAR and MISS rates when Θ is compared to a variable threshold variable between 0 and 1 as reported on the x axis.

Fig. 8 reports the FAR. FAR is low when the threshold th is kept low ($th = 0.05$). However, in this case, MISS rate is higher,

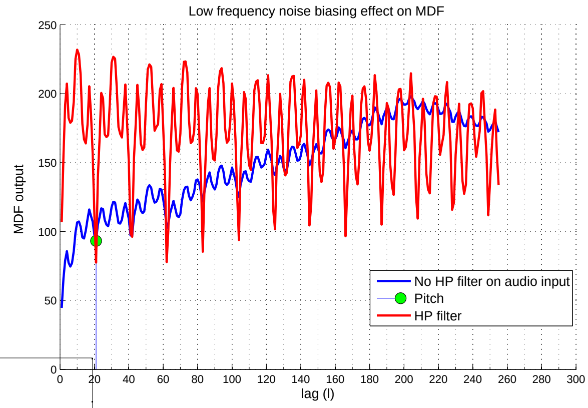

Figure 6: Comparison of MDF performance with and without High Pass Filter on audio input.

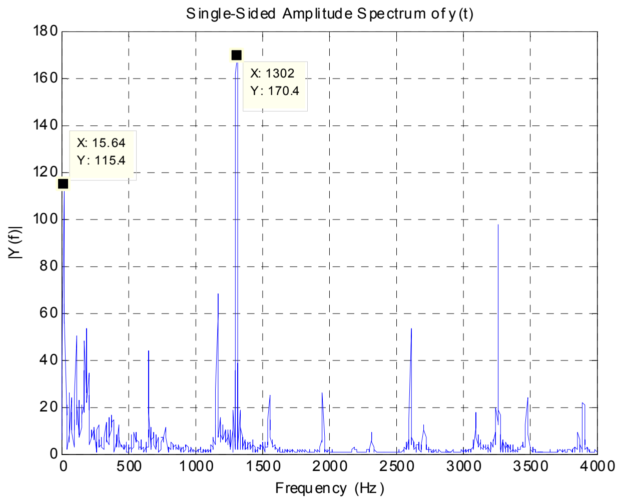

Figure 7: Spectrum of voice from radio inside the car during movement and siren signal with a PUNR of +1.43 dB.

meaning that we could need more then one single pattern observation window to asset siren presence. On the other hand, the MISS rate is decreased if the threshold is higher, as evident in Fig. 9.

3.5 Influence of the $Pitch(t)$ signal window length

The pattern observation window influences the response of the presence indicator. Throughout this paper, the pattern was analysed with a window covering one second and overlapping of 50%. Longer windows will lead to less FAR but higher MISS. Furthermore, if the dynamic of the response is of main concern, the observation window should be kept as short as possible, in order to be able to asset siren presence as quickly as possible. Of course this can generate much more False Alarms. Again, here the design has to chose the trade-off between velocity and accuracy of the presence estimator. To improve the performance without sacrificing the output rate of the presence estimator, high overlap values are needed.

4. CONCLUSIONS

We have presented a pitch detection algorithm suited for Emergency Vehicle detection in presence of uncorrelated, pitched and not pitched noise. Optimal parameters have been designed based on measurements obtained in real operating scenario.

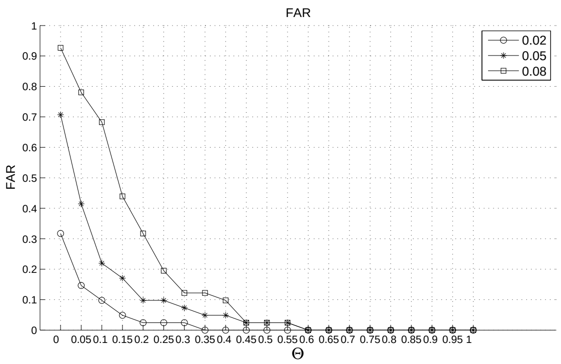

Figure 8: False Alarm Rate: various thresholds *th* for the peak searching with multiplicity check.

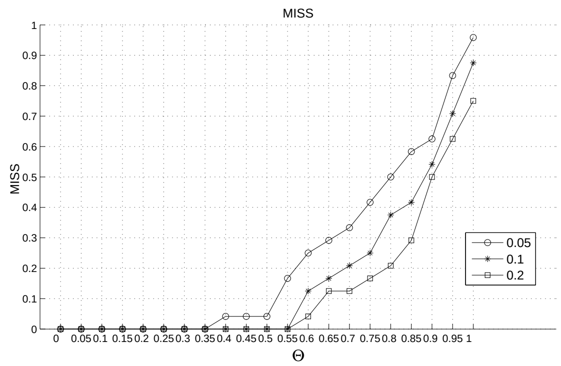

Figure 9: Miss Rate: various thresholds *th* for peak searching with multiplicity check.

The process has been implemented on an low power Atmel microprocessor. The process indicates the presence of a siren comparing a decision variable with a fixed threshold. As future development, aiming at a Constant FAR (CFAR), the presence detection can operate with dynamically adjusted threshold depending on parameters of the audio signal at the microphone. In the case a very noisy audio signal is detected we compare Θ with an higher value respect to the case when a clean audio is available. In this second case, with CFAR and adaptive threshold, we can greatly reduce the probability of an undetected siren signal portion thus leading to a quicker response of the algorithm when the emergency vehicle is far away from the driver.

REFERENCES

[1] W. C. Chu, *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*. New York, NY, USA: John Wiley & Sons, Inc., 2003.

[2] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *Acoustical Society of America Journal*, vol. 111, pp. 1917–1930, Apr. 2002.

[3] S. Clontz and R. Adhami, "Long-duration signal detection in a noisy environment," *System Theory, 1989. Proceedings., Twenty-First Southeastern Symposium on*, pp. 527–532, 26-28 Mar 1989.

[4] H. Y. Kim, J. S. Lee, M.-W. Sung, K. H. Kim, and K. S. Park, "Pitch detection with average magnitude difference function using adaptive threshold algorithm for estimating shimmer and jitter," *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, vol. 6, pp. 3162–3164 vol.6, 29 Oct-1 Nov 1998.

[5] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley, "Average magnitude difference function pitch extractor," *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on*, vol. 22, no. 5, pp. 353–362, Oct 1974.

[6] R. Sherratt, D. Townsend, and C. Guy, "Cancellation of siren noise from two way voice communications inside emergency vehicles," *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*, vol. 4, pp. 2395–2398 vol.4, 15-19 Mar 1999.

[7] J. A. Zakis, H. J. McDermott, and A. E. Vandali, "A fundamental frequency estimator for the real-time processing of musical sounds for cochlear implants," *Speech Commun.*, vol. 49, no. 2, pp. 113–122, 2007.