Supporting Presbycusic Drivers in Detection and Localization of Emergency Vehicles: Alarm Sound Signal Processing Algorithms

Marco Paoloni and Andrea Zanela
Robotics Lab
ENEA
Rome, Italy
marco.paoloni@enea.it, andrea.zanela@enea.it

***Abstract*—The fatality rate for car drivers is more than five times higher for the 75 years and older than for the average for all ages. Presbycusis is a gradual hearing loss in both ears that commonly occurs as people age, it represent the third impairment for the elderly after arthritis and hypertension. One of the symptoms is the altered perception of the distance and the direction of a sound source especially for indoor and noisy environments. In this paper we present two algorithms, one for the early detection of alarm signal and the other for the localization of the emitting device, to be used in a car-embedded autonomous system aimed to support presbycusic and other car drivers to get a faster response in presence of an emergency vehicle in their surroundings.**

Keywords—presbycusis, car-embedded autonomous system; alarm signal detection; alarm source localization; steered response power, stochastic region contraction

I. Introduction

Age Sensitive ICT Systems for Intelligible City For All (I'CityForAll) is an ongoing project [1] partially funded by the European Ambient Assisted Living Joint Programme. The project aims at enhancing the sense of security and self-confidence of presbycusic people whose hearing degradation increases with age.

Presbycusis represent the third impairment for the elderly after arthritis and hypertension; it is a gradual hearing loss in both ears that commonly occurs as people age, it is progressive, bilateral and symmetrical and the loss is most marked at higher frequencies. Sounds appear less clear and sharp so that a severe difficulty in conversations understanding occurs and also the perception of distance and the direction of arrival of a sound source is altered.

The goal of I'CityForAll project is to design audio-aged-sensitive ICT systems enhancing self-confidence, mobility, safety, to improve social and mental well being.

The project targeted population corresponds to people older than 50 years in mobility situations and affected by presbycusis that induces a loss of sense of safety and self-confidence.

The ICT solutions consist of intelligent loudspeakers for better intelligibility of vocal messages in public confined spaces and systems embedded in vehicles for better localization of urban sound alarms like ambulances, police cars, fire trucks, etc..

The latter scenario is of particular relevance because older drivers have a relatively high fatality rate. Taking the distances travelled into account, the fatality rate for car drivers is more than five times higher for the 75 years and older than for the average for all ages [2]. Elderly are involved in 40% of fatal injuries (105000 deaths/year), and 1500 accidents/day require medical assistance according to the European Network for Safety. In a driving situation, a correct perception of the audio ambiance surrounding the car is necessary for a safe driving: it allows the driver to react in due time to a given event and even more, to anticipate on different possible scenarios, especially when are involved emergency vehicles because of their speed and behavior [3][4].

In this paper we present a set of algorithms to identify the presence of an alarm signal and to determine its direction of arrival. A first prototype solution foresees to place a microphone array on top of a car and to process these sensors data via the proposed algorithms, to extract meaningful information and displaying to the driver. Such a system will support presbycusic and other car drivers in detecting the presence of alarms in their surroundings, such as the sirens of an emergency vehicle, also helping the drivers to determine the position of the alarm. This on-board system will be "transparent" and embedded in mass products for people with presbycusic hearing without impacting normal hearing people, in a "Design for All" approach.

II. Alarm Detection

A. Alarm signal regulation

The alarm signals dealt within the project are constituted by alternations of two tone emissions of different frequencies and durations. For example, the Italian regulation for the alarm signals emitted by the Police vehicles [5] requires that a whole acoustic cycle must be formed by four tone emissions of the same duration: the odd at 466 Hz, the even at 622 Hz; for the alarm signals emitted by the National Fire Corps vehicles and by the ambulances (in the rest of the paper referred together as Rescue vehicles) the regulation [6] requires that a whole acoustic cycle must be formed by eight tone emissions: the first at 392 Hz lasting 1/3 of the cycle duration, the second at 660 Hz lasting 1/18 of the cycle duration, the third at 392 Hz also lasting 1/18 of the cycle duration, the fourth at 660 Hz lasting 1/18 of the cycle duration and the last four alternating as the first ones. The frequency alternation patterns for the two alarm signals are resumed in Fig. 1.

For both of the two signals the tones must alternate without any appreciable interruption or overlap and the whole cycle duration must be 3 s; this time must also include a possible silence interval between a complete cycle of emissions and the successive one (but the duration of this silence interval must be less than 0.2 s).

When a cycle is started it must be performed entirely. The maximum tolerance for the frequencies values is required to be 5% and the same tolerance is required for the cited cycle time fractions; for the whole acoustic cycle is required a tolerance of ±0.5 s.

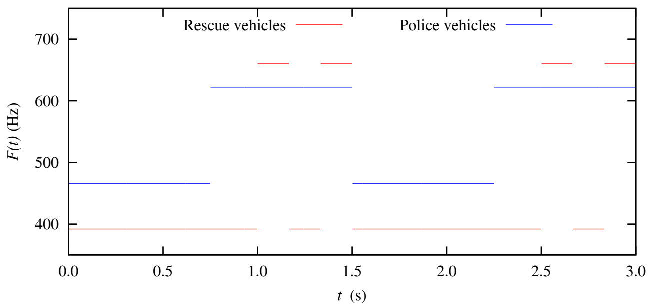

Fig. 1. Frequency alternation patterns for the Italian alarm signals.

B. Detection algorithm

The proposed algorithm for the alarm signals detection is composed by two blocks, the first dedicated to the tone detection that provides a logical "true" value when a tone with a frequency close to those regulated is present in the microphones signals and a second that measures the timing of the logical true signals to confirm their conformity to the alarm signals regulation. At the current stage of the project development, the greater effort of the ENEA activities is devoted to the first block, while the second will be developed later.

In the proposed detection algorithm the signals from microphones array after the A/D conversion are processed to estimate the power carried in a narrow band centered on the tones regulated frequencies or in other terms the amplitude of the first harmonic of the tone; the employed technique in the process is the AM synchronous quadrature demodulation [7].

AM synchronous demodulation is a technique commonly used in many AM radio receivers (these receivers are also known as direct-conversion, homodyne, synchrodyne, or zero-IF receiver); it uses beats between the signal and a wave generated by a local oscillator at a frequency very close to the carrier frequency; quadrature demodulation uses two waves (phased by 90°) to implement the envelope detection of AM signals and is usually employed in conjunction with DSP to process more complex modulation methods.

The adopted process is sketched in Fig. 2.

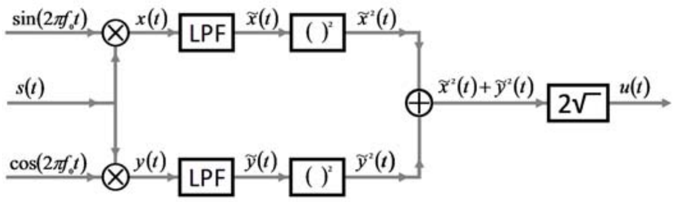

Fig. 2. AM synchronous quadrature demodulation process representation.

As it can be seen, the incoming signal $s(t)$ is mixed with two sinusoidal waves with frequency f_0 (the tuning frequency) and then are submitted to two identical low-pass filter with cut frequency f_c; in presence of a tone with frequency f in the incoming signal the outputs of the filters are virtually null if $|f - f_0| > f_c$ or two (quadrature) sinusoids with frequency $|f - f_0|$ if $|f - f_0| \le f_c$; from these two sinusoids the amplitude of the first harmonic of the incoming tone is then reconstructed by simple mathematical operations, giving as output the signal $u(t)$.

The algorithm uses four demodulation blocks: two for the two tones detection of the Police signals and two for the Rescue ones. After the demodulation the signals obtained are compared with a threshold of suitable value to obtain the logical value asserting the presence of one of the four tones.

C. Detection test

The proposed algorithm has been tested with simulated and real signals. Simulated signals were tones generated by the simulation program using the alternation in time of two tones according to the Italian regulation for Police and Rescue emergency vehicles. As waveforms for the tone generation were employed sinusoidal, square and saw-tooth waveforms, the latter two appropriately clipped in frequency to simulate the presence of the anti-aliasing low pass filter typical of every digital recording device. Simulated signals with Doppler effect were obtained simulating rectilinear uniform motions for the emergency vehicles and the car equipped with the microphones.

Typical results obtained for the simulated stationary emissions are represented in Fig. 3 where are reported the output signals $u(t)$ for the two Rescue demodulation channels (expressed in arbitrary units) and in Fig. 4 where are reported the same signals for the two Police demodulation channels.

The result obtained for the signals with Doppler effect are represented in Fig. 5 where are reported the signals $u(t)$ for the two Rescue demodulation channels for a Police vehicle surpassing at 60 km/h the car with microphones while it is moving at 30 km/h; the Police vehicle starts 100 m behind the car and passes at a lateral distance of 5 m from it. In the same figure are also reported the frequency of the emitted signals and the frequency of the received ones (differing for the Doppler effect).

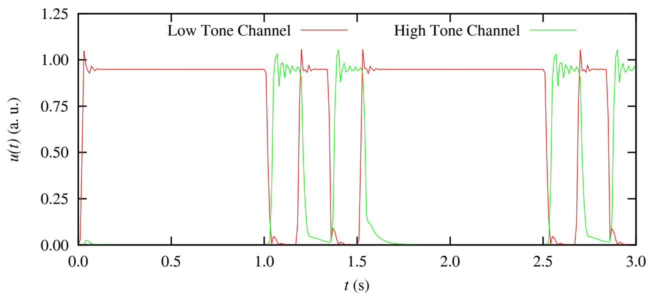

Fig. 3. Output signals of the two Rescue demodulation channels for a simulated stationary emission.

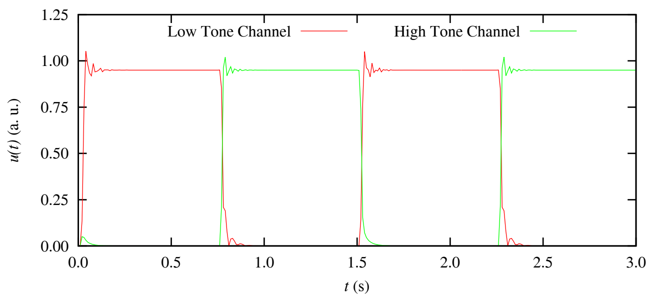

Fig. 4. Output signals of the two Police demodulation channels for a simulated stationary emission.

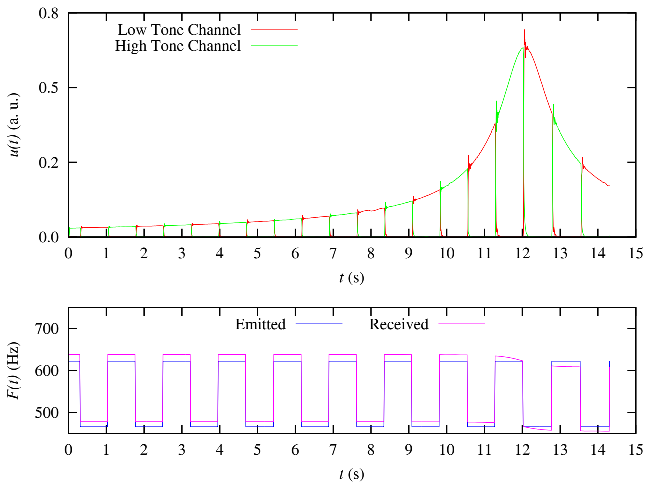

Fig. 5. Output signals of the two Police demodulation channels for a simulated emission by a moving vehicle with the representation of the frequency of the emitted and received tones.

As it can be seen, the amplitudes of the output signals of the two demodulators increase while the Police vehicle is approaching the car and decrease as the Police vehicle moves away after having passed the car; in the same way the frequencies of the received tones are greater than the ones of the emitted tones in the approaching phase and smaller in the other phase.

The proposed algorithm was then applied to real signals recorded by the Centro Ricerche FIAT, one of the project partners; these signals were recorded in their silent room with the stationary emitting device posed at 1 m from microphones and at their «Centro Sicurezza» test track using the same device mounted on a vehicle moving at 30 km/h.

Typical results obtained for these signals are reported in Fig. 6 and in Fig. 7 for the Rescue case and in Fig. 8 and in Fig. 9 for the Police case.

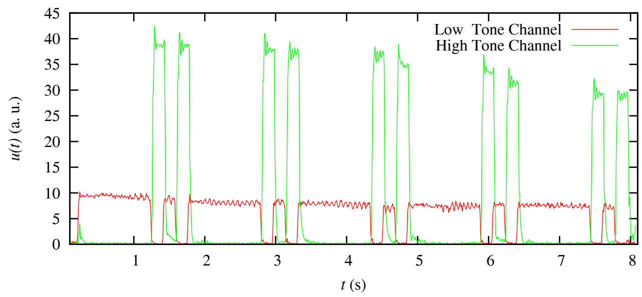

Fig. 6. Output signals of the two Rescue demodulation channels for the stationary emission of a real alarm signal.

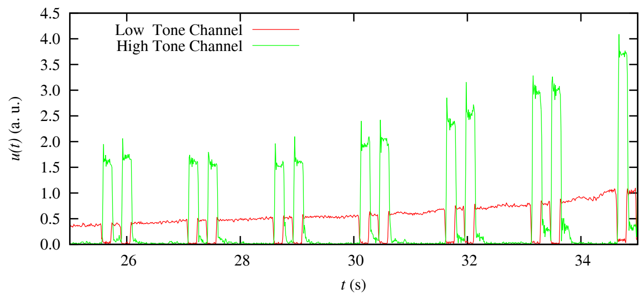

Fig. 7. Output signals of the two Rescue demodulation channels for the emission of a real alarm signal by a vehicle moving at 30 km/h.

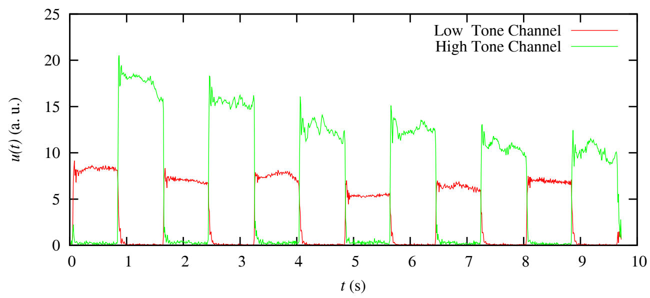

Fig. 8. Output signals of the two Police demodulation channels for the stationary emission of a real alarm signal.

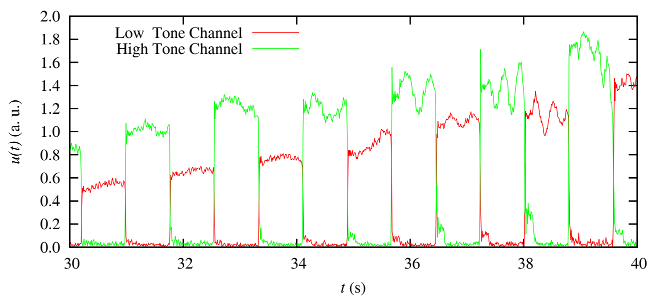

Fig. 9. Output signals of the two Police demodulation channels for the emission of a real alarm signal by a vehicle moving at 30 km/h.

As can be seen in these figures the amplitude of the demodulators output signals for the low tones is smaller than the high ones in both the two cases; this is caused by the fact that real low and high tones have very different spectral characteristics and so the respective fundamental harmonics have very different amplitude.

III. ALARM LOCALIZATION

A. Localization algorithm

A microphone array has the capability of focusing on signals generated from a specific location or direction. Such capability is referred to as a beam-former. A steered beam-former based algorithm hypothesizes a spatial, point-source location in a predefined region and computes a beam-forming functional at that location, the one yielding the maximum values of the functional is the estimated source locations [8][9].

The output of it is known as the steered response. When the point of focus matches the true source location, the steered response power (SRP) will peak. This method has been shown to be more robust under high noise and reverberation than the approaches that are based on intersecting or least-squares fitting the time-differences-of-arrival (TDOA) of the signal over the array. Also avoids the process of making early, and often incorrect, decisions of selecting "good" TDOA's as done in other methods.

In a *M*-microphones array system, a beam-former can be created by delaying the microphone signals $s_m(t)$ (with $m = 1,\ldots,M$) by appropriate steering delays to make them aligned in time, and then summing all these time-aligned signals together. Naming $\tau_m(\mathbf{r})$ the time delay from the focusing point (located at $\mathbf{r}$) to the microphone m and $\tau_0(\mathbf{r})$ the minimum of all these time delays, the steering delay for the microphone m will be $\delta_m(\mathbf{r}) = \tau_m(\mathbf{r}) - \tau_0(\mathbf{r})$.

When the phase transform (PHAT) pre-filtering is applied before computing the cross-correlations, to pre-whiten the signals, we obtain the SRP-PHAT algorithm. The SRP-PHAT for each point $\mathbf{r}$ in the space is defined as follows:

$$P(\mathbf{r}) \equiv \sum_{k=1}^{M}\sum_{m=1}^{M}\int_{-\infty}^{\infty}\frac{S_k(\omega)S_m^*(\omega)}{\left|S_k(\omega)S_m^*(\omega)\right|}\exp\left(j\omega[\delta_m(\mathbf{r}) - \delta_k(\mathbf{r})]\right)d\omega\,,$$

where $S_m(\omega)$ is the Fourier Transform of the signal $s_m(t)$.

As previously described, to find the source locations, we steer the beam-former over all possible points in a focal volume containing the source. The points that give the maximum weighted output power of the beam-former will be the source locations. For a single source, the location estimate $\mathbf{r}_s$ is the argument that maximize $P(\mathbf{r})$.

The robust performance of the beam-forming approach and its bias-free property [10][11] comes at the price of high computational cost, because the SRP surface to be searched has many local maxima. In order to cut the computational cost and thus making the SRP-PHAT single-source locator more practical in real-time, a stochastic region contraction (SRC) approach [12] has been implemented in our work.

The basic idea of the SRC algorithm is, given an initial rectangular search volume containing the desired global optimum and perhaps many local maxima or minima, gradually, in an iterative process, contract the original volume until a sufficiently small sub volume is reached in which the global optimum is trapped. The contraction operation is based on a stochastic exploration of the SRP-PHAT functional. It can be shown that the missing probability can be made negligible throwing sufficient random points.

B. Localization test

To test the proposed algorithm, it has been simulated a setup where a microphone array was positioned in the center of a reference frame and a single alarm source with spherical emission could be placed in stationary known positions. The microphone array was composed by eight omni-directional microphones with a 20 cm radius circular geometry. Alarm source and microphones laid in the same plane.

The emergency signals employed in the simulation adopted as alternating tones clipped square waves with an amplitude normalized so that at the minimum distance microphone-source of a simulation set, the signal could cover the full range of a 16 bit analog to digital conversion. For all the other distances the amplitude was normalized to the former according to a propagation law proportional to the inverse distance between source and microphones. In this way none of the signals in the same simulation set was in a saturation condition, but the signals coming from larger distances were more attenuated.

In the first scenario was simulated a lateral pass-by, as depicted in Fig. 10.

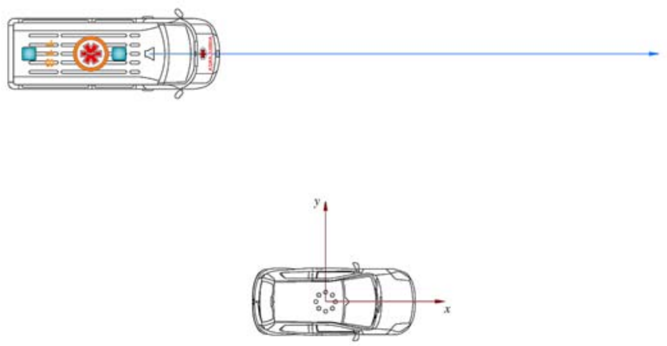

Fig. 10. Scenario 1 (Lateral Pass-By).

The simulation was performed placing the source on a set of positions starting from (-270; 2; 0) to (180; 2; 0) with steps of 10 m. For every position were processed in MATLAB framework four signals segments of length equal to 65536 samples (at a sampling frequency equal to 44110 Hz) from four microphones, the ones collocated on the coordinate axes.

The obtained results are resumed in Fig. 11 where are reported the errors on the estimate of the signals Direction of Arrival (DoA) for different x-axis source positions. As it can be seen the errors are about 1°.

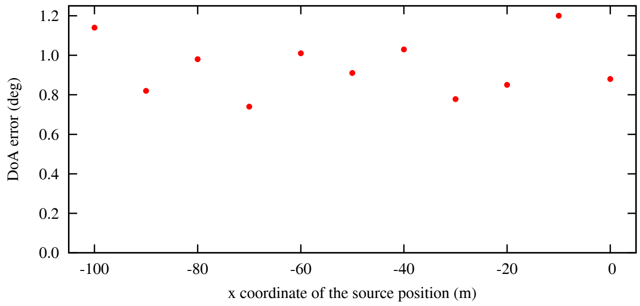

Fig. 11. Results for a simulated lateral pass-by.

In the second scenario was simulated a circular pass-by, as depicted in Fig. 12.

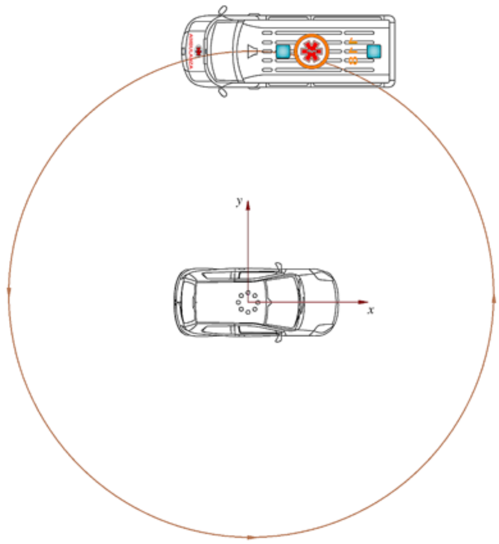

Fig. 12. Scenario 2 (Circular Pass-by)

For this scenario were performed two tests, the first with a radius of 20 m and the second with a radius of 30 m, placing the source every 20°. The obtained results are resumed in Fig. 13 where it can be seen that even in this case the errors on the estimate are about 1°.

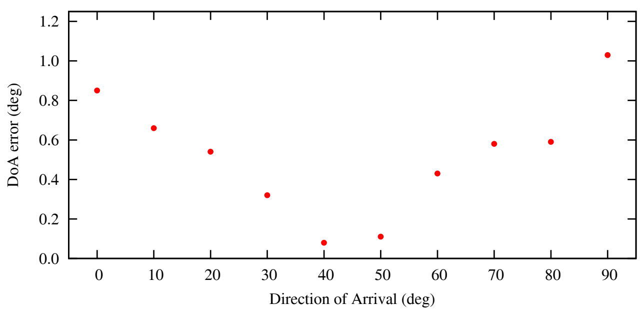

Fig. 13. Results for a simulated 20 m radius circular pass-by.

The same scenario was also considered for a real environment, thanks to the data collected by Centro Ricerche Fiat using a circular 8-microphone array realized by the laboratories of Ecole Polytechnique Fédérale de Lausanne.

Four circular passages around the car equipped with the microphone array were performed, the first and second on a circumference of about 30 m radius, the third and fourth on a circumference of about 20 m radius; during all the four passages the emergency vehicle speed was 20 km/h.

For tests in real environment, continuous data streams must be process and, unfortunately, a scenario ground truth was not given. The obtained results can be seen in Fig. 14, where are reported the estimated Direction of Arrival during the data stream.

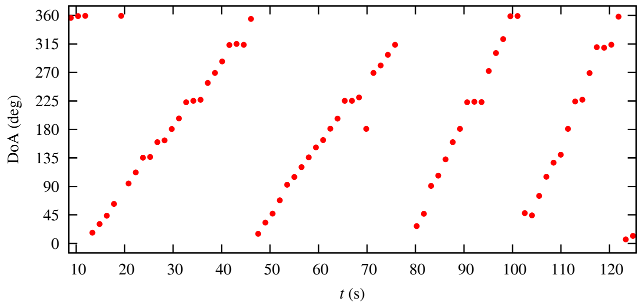

Fig. 14. Results for circular pass-by in a real scenario.

Real signals acquisitions also include a third scenario, the frontal pass-by, depicted in Fig. 15; for this scenario an emergency vehicle has been moved at 30 km/h, from the position (10; -65; 0) to the position (10; 65; 0).

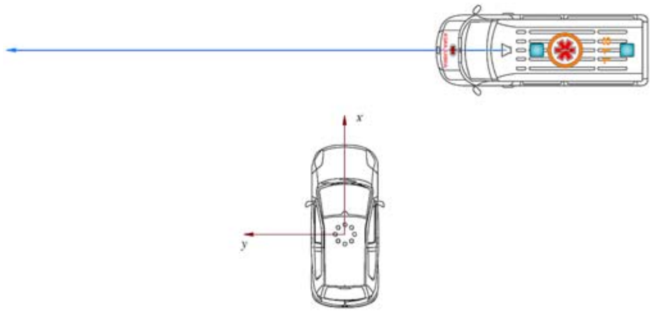

Fig. 15. Scenario 3 (Frontal Pass-By).

The obtained results processing the continuous data stream for this test can be seen in Fig. 16.

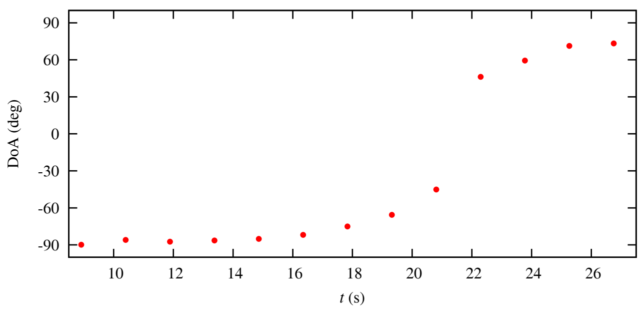

Fig. 16. Results for front pass-by in a real scenario.

The preliminary results on real signals showed in Fig. 14 and in Fig. 16 confirm that the adopted method is very effective in estimating the alarm direction of arrival.

As well known, however, this method is much less useful in estimating the source distance; thus we propose to use an interface where the information of the approaching emergency vehicle is presented to the driver as in Fig. 17.

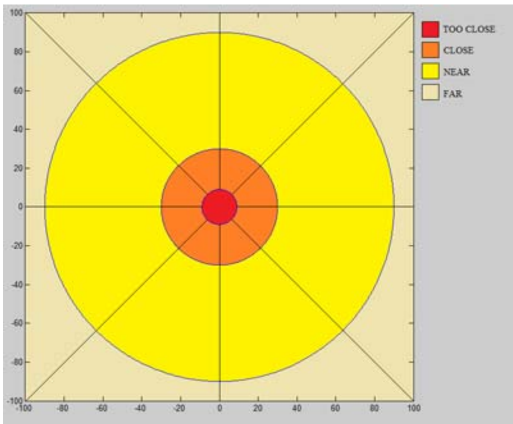

Fig. 17. A possible information rendering to the user. The car is positioned in the center.

With the car in the center of the graph we use four different zones (marked with a color code) where the approaching emergency vehicle can be localized; three zone (far/near/close) are further divided into eight sectors for the recovering of the alarm direction of arrival.

IV. CONCLUSIONS

In this paper two algorithms are presented, one for the early detection of alarm signal coming from an emergency vehicle, the other for the localization of the emitting device.

These algorithms will be implemented in a car-embedded autonomous system aimed to support presbycusic and other car drivers to get a faster response in presence of an emergency vehicle in their surroundings, enhancing the sense of safety while driving. This is of primary importance because Europe is ageing increasingly and faster, by 2020 around 25% of the Europe population will be over 65, and aged driver (75+) has a fatality rate more than five times higher than for the average for all ages.

Future activities will address the presence of the acoustic reverberations. This effect does last alarm signals more than their regular duration; when the emergency vehicle stops the first tone and switches to the second, the first tone decoder will continue to output signals for the presence of the retarded echoes. This should not be a problem if the processing of the second tone logical signal is triggered by the rising edge of the first tone rather than the falling edge.

For the source localization, as it can be seen in the results previously reported, the algorithm proposed is very effective in the recovery of sound direction of arrival, even if it is much less useful in finding the source distance, and can be a valuable aid for presbycusic drivers or other people suffering of age-related factors affecting their driving ability.

ACKNOWLEDGMENT

This work was partially funded by the Ambient Assisted Living Joint Programme.

REFERENCES

[1] http://www.icityforall.eu/

[2] http://ec.europa.eu/transport/road_safety/index_en

[3] C.B. Custalow, C.S. Gravitz. Emergency medical vehicle collisions and potential for preventative intervention. Prehospital Emergency Care 2004, 8, pp 175–184.

[4] M.G. Lenné, T.J. Triggs, C.M. Mulvihill, M.A. Regan and B.F. Corben. Detection of Emergency Vehicles: Driver Responses to Advanced Warning in a Driving Simulator. Human Factors: The Journal of the Human Factors and Ergonomics Society 2008, 50, pp 135-144.

[5] Decreto Ministeriale 20 Marzo 1979, Supplemento ordinario alla Gazzetta Ufficiale N. 206 del 28 Luglio 1979

[6] Decreto Ministeriale 11 Ottobre 1980, Gazzetta Ufficiale N. 310 del 12 Novembre 1980

[7] P. Horowitz and W. Hill, The art of electronic. Cambridge, NY: Cambridge University Press, 1989.

[8] J. Dmochowski, J. Benesty, and S. Affes. A generalized steered response power method for computationally viable source localization. IEEE Trans. Audio, Speech, Lang. Process., 15(8), pp 2510–2526, November 2007.

[9] C. Zhang, D. Florencio, D. Ba, and Z. Zhang. Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings. IEEE Trans. Multimedia, 10(3), pp 538–548, March 2008.

[10] J. H. DiBiase. A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays. PhD thesis, Brown University, Providence, RI, May 2000.

[11] S. T. Birchfield. A unifying framework for acoustic localization. In Proc. of European Signal Processing Conference (EUSIPCO 2004), pp 1127–1130, Vienna, Austria, September 2004.

[12] H. Do, H.F. Silverman, and Y. Yu. A Real-Time SRP-PHAT Source Location Implementation using Stochastic Region Contraction on a Large-Aperture Microphone Array, in Proc. ICASSP 2007 pp 121-124.