Sound Localization in Urban Areas using the Ambisonic Concept

C. Almeida[1], Joel Preto Paulo[1,2,4], Miguel Félix[1]

[1]Instituto Superior de Engenharia de Lisboa – ISEL, Lisboa, Portugal,

[2]CAPS - Instituto Superior Técnico, TULisbon, Lisboa, Portugal,

[4]CEDET – ISEL, Lisboa, Portugal,

A38415@alunos.isel.pt jpaulo@deetc.isel.ipl.pt miguel.felix@fi-sonic.com

Abstract — **The environmental sound monitoring in city is a practice increasingly used in order to have an objective sound image of what is happening over time. Usually the noise levels are measured only in certain places to build the so-called noise maps. These measurements of sound do not come into consideration with the location of sound sources, which makes it impossible to perceive what is the most important source in noisy environments from multiple sources. This project aims to estimate the direction of sound events in urban environment (cities) with the use of arrays of microphones based on the ambisonic concept.**

Keywords: sound localization, microphone arrays, ambisonic, smart cities.

I. Introduction

Many technics capable of detecting the direction of the sound use the concept of difference of time between sound arrival, TOA (Time Of Arrival). We look for a new approach on locating sound events.

This project was developed considering the concept of the ambisonic microphones which estimates direction of sound based on the difference of energies instead of time delay. This approach is only possible since ambisonic contains 4 microphones organized in a tetrahedral geometry. Therefore, one can assume that the sound arrives at all the four microphones without phase difference, as a point microphone. Moreover, with these devices the 3D sound field can be represented completely in a Cartesian system by applying the signals from each microphone to the ambisonic decoding matrix [1-2]. Additionally, the sum of the four signals from each microphone capsules can be used to a build a virtual microphone with omnidirectional directivity pattern.

This project measures, among other features, the noise levels in a certain place at a certain time or day, detect how much traffic exists in a certain street, detect a gun fire, people screaming or a vehicle accident.

In future phases of this project this system can notify the fire department or police by generating notifications of the more relevant sound events in the city.

II. Description

The goal of this work is to be able to capture sound through an array of ambisonic microphones which allows us to perceive the sound in a 2D plane or 3D space.

This way, it is possible to detect sound events in horizontal and/or vertical planes as well as its distance. In our case, we will only calculate the horizontal angle (azimuth). The angle calculation is obtainable using only one ambisonic microphone but the distance is only possible to obtain with more than one microphone.

These sound devices capture sound in 4 audio channels. This signals are processed in a way that allows us to perceive sound direction in a 3D space (although as mentioned before we will be looking only at data in a 2D plane – horizontal plane) as well as signal of omnidirectional amplitude. The resulting signals are w, x and y, where w means the sum of all signals (an omnidirectional signal) and x and y represent the signals in terms of the corresponding axis in a horizontal plane [3-4].

This way we will used difference of energy in signals based algorithms.

These devices have microcomputers able to run algorithms capability of locating events calculating in real time. Since these microprocessors are not very powerful we have to take in consideration optimized algorithms in order for the ambisonic being able to make the calculation in real time.

As mentioned before, this project goal is to detect events in real time. There can be events that aren´t relevant in real time but if we consider gun shots, car accidents or a person screaming we realize the importance of this algorithm running in real time.

This project is divided in 3 parts. The first step consists of remove background noise in order to improve the signal-to-noise ratio, SNR, get a “clean” signal or as clean as possible. This first step is very important since with low SNR the algorithm shows poor results in estimating the direction of sound, the calculations return many false positives.

We tested 2 different approaches to this first step. We tried both the Spectral Subtraction and the Wiener filter for algorithm of noise reduction. The first one is considerably faster than the latter but also less effective for some sound events. By definition, these filters are not adaptive in terms of noise profile, meaning the noise sample is not updated through time. We had to implement an algorithm that identifies the current noise profile for each iteration, each sliding time window (frame).

The second step is to detect what is a sound event, a kind of segmentation method. This method allows us to analyse only the signal frames that correspond to sound events. This step is essential to improve better the efficiency of the algorithm not only considering time processing but also in reducing the false positives estimated angles.

The first step of this second method is to generate the STFT (short time Fourier transform) of the signals. This way, we get a matrix of amplitudes through time and frequency. We used a Spectral Subtraction between adjacent frames to detect events. This step uses only the omnidirectional signal.

The third step is to calculate the angle at which the sound event came from. In this method, we, again, calculated the STFT for every signal. This way we were able to obtain amplitude vectors for both 'X' and 'Y' axes, $Ix = real(W^* * X)$ and $Iy = real(W^* * Y)$, where W^* stands for the conjugated W.

This project has the ability to calculate multiple sound sources simultaneously (different directions) in case the events are described in different frequency bands.

The angles are calculated by adding the amplitudes in each STFT to each angle respectively. This return the probability of a certain angle representing an event in the analysed frame. As it might have been expected this latter step shows us a graph with too much noise so we pass it through a smoothing algorithm. The result is depicted in next figure.

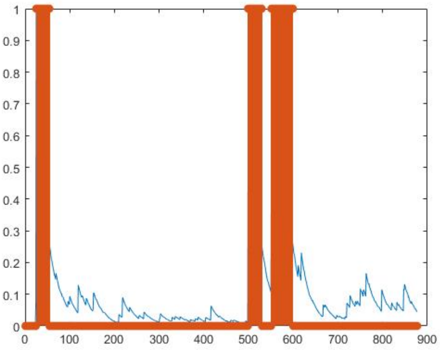

Fig. 1. Detection of sound events.

After this, we calculate the angles in each frame. As we can see the signal still contains noise so we will not take in consideration small peaks, applying a certain threshold in which the peaks are considered as real or false positives.

The amplitude of the peak is relevant, as mentioned before. That way, when we show the angle which the event points to, we show different width for each angle. The thicker the angles the more accurate estimation of direction for a sound event coming from that angle.

Fig. 1 shows in blue the graph obtained by subtraction between spectral adjacent samples of the signal and the orange the time intervals in the frame that correspond to sound events.

A histogram is determined with the sum of the amplitudes that exist for each angle. This calculation allows to obtain the probability for that time if the event is located at a given angle. However, the results show high levels of noise. Thus, it is necessary to apply a smoothing function. Thus, it is used a spline function. The result is shown in Fig.2.

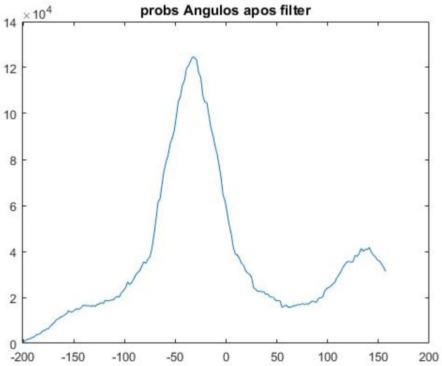

Fig. 2. Angle distribution after smoothing technique.

As can be shown it is much easier to identify peaks in this graph. After applying the smoothing technique, the angles corresponding to each frame are calculated. There are several small 'peaks' in the graph, even resulting from the large amount of noise of the signal used for testing. Thus, a threshold is applied to limit the amplitude.

After this step the diagram that identifies the angle to which the audible event points to is shown.

Being that there is some margin of error in this algorithm, this diagram can be read in two ways. It may be that the frame analysed contains two sound events (a louder sound than the other) or that this same frame has only one angle and the other can actually be a false positive, being that in the last case se the more likely it would be that the rake real were the widest (represented in blue) and not the dash to orange.

This process is repeated for the number of times that there are frames of audio to analyse in a constant cycle that if you want 'run' 24 hours per day.

References

[1] Dimoulas C., Avdelidis K., Kalliris G. andPapanikolaou G., "Sound Source Localization and B-Format Enhancement Using Sound Field Microphone Sets", in 122nd AES Convention,Preprint Number: 7091, May 2007

[2] Bugalho, M., Portelo, J., Trancoso, I., Pellegrini, T., Abad, A.: Detecting audio events for semantic video search. In: Proc. Interspeech 2009

[3] Alan Dufaux. "Detection and Recognition of Impulsive Sound Signals". Institute of Microtechnology, University of Neuchatel, Switzerland. PhD thesis 2001

[4] Li, D., Sethi, I. K., Dimitrova, N. and McGee, T. Classification of general audio data for content-based retrieval, Pattern Recognition Letters, 22, 533-544, (2001).