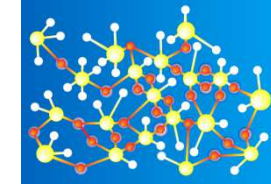




International
Neural
Network
Society

EANN2013

14th International Conference on Engineering Applications of Neural Networks



13-16 September, 2013, Halkidiki, Greece

Audio data fuzzy fusion for source localization

M. Malcangi

*Università degli Studi di Milano
Department of Computer Science
DSP&RTS Research Laboratory
Milan - Italy*

malcangi@di.unimi.it

@ a glance

- ✓ This work addresses the problem of **audio source localization** in multiple speakers indoor scenarios.
- ✓ Three **different direction of arrival** (DOA) algorithms are applied to measure the angular position of the primary audio source with respect to a reference microphone.
- ✓ A **fuzzy logic-based method** is applied to fuse the crisp measurements.

Sound Source Localization

What

- The **sound is the most important information media** in the interaction between human beings and the physical world.
- Among the several information that the sound embeds, the **spatial location of the sources** is of primary importance for executing auditory tasks such as:
 - ✓ *attention-less monitoring of the surrounding environment*
 - ✓ *the source localization*
 - ✓ *the beamforming*

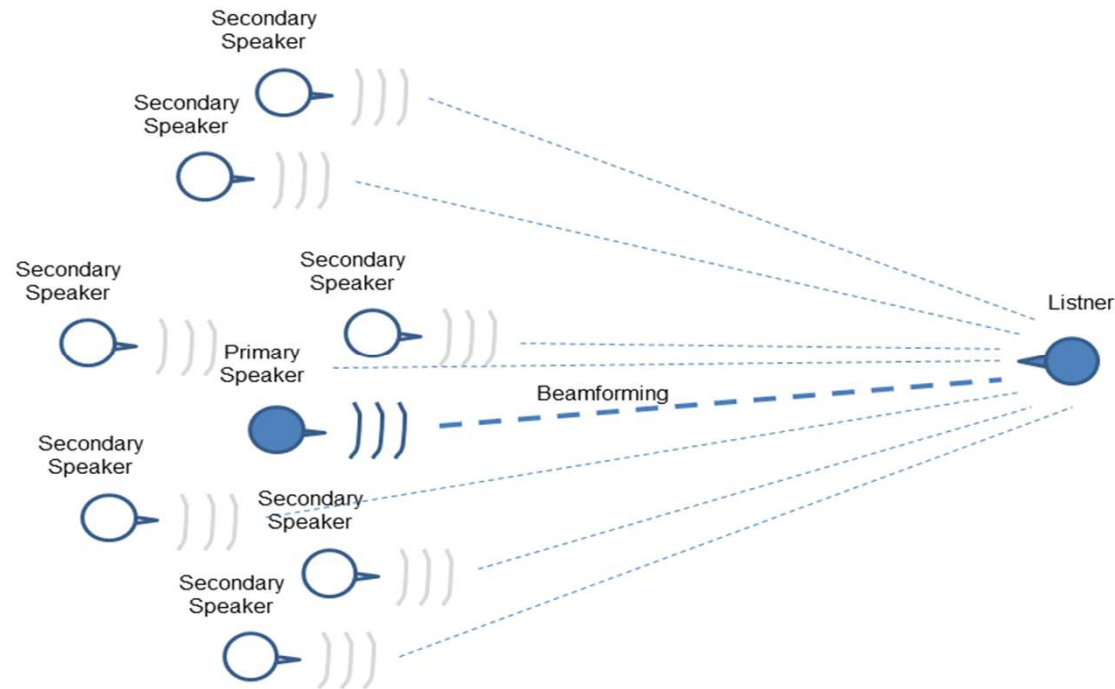
Sound Source Localization

Why

- **Audio-based tasks of human beings**, such as speech recognition, speaker tracking, and speaker identification use **sound source location and data fusion strategies** to successfully interact and communicate with other human beings.
- To enable the **interaction of a human being with the machine in a natural way**, a similar functionality needs to be developed in terms of sound source localization capabilities and data fusion strategy.
- Many audio-based applications, such as the automatic speech recognition (**ASR**) and audio/visual **ASR (AV/ASR)**, the automatic speaker identification (**ASI**), the audio noise cancellation (**ANC**), and the indoor audio navigation (**IAN**), **can be improved using SSL**.

Sound Source Localization

How it works



Human beings are able **to recognize** which sound is which, **to locate** where each sound is coming from and to recognize which sounds can be ignored, **beamforming** towards the sound source target

Sound Source Localization

How it works

- ✓ The sound source localization (SSL) is a **signal processing task** based on various signal processing methods for time delay estimation applied to signals received by an array of microphones.
- ✓ Using **proper microphone geometry**, 2D and 3D sound source location can be detected and the related measurements applied to recognize and isolate the sound source.
- ✓ **Beamforming** can be executed on the sound source, either by moving the primary microphone in the direction of the sound source, or executing application specific signal processing algorithms on the captured sound.

Sound Source Localization

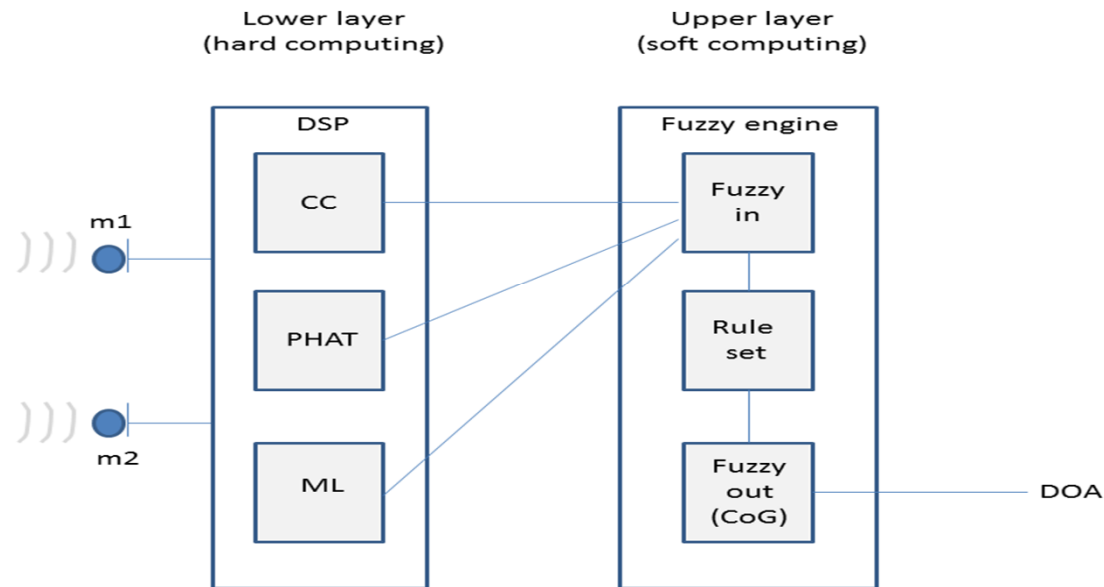
What's new

The sound source localization (SSL) proposed system **fuzzy fuses the sound time delay measurements** executed by a set of three algorithms processing at the same time the sound signal captured by a couple of microphones.

The purpose is to implement a **basic SSL capability** useful to design a **front-end between an array of microphones and the sound processing application.**

SSL System framework

The system consists of **two layers**: the lower layer implements three subsystems, each one implementing a **sound time delay measurement**. The upper layer implements a fuzzy logic-based inferential engine tuned **to fuse the decision** of the three sound delay measurements.



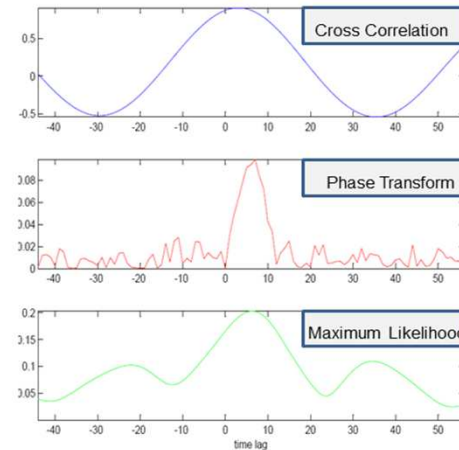
A two level audio source localization system approach is **robust and reliable** because each module operates independently from the other, and the fuzzy logic inferential engine has the capability to evaluate qualitatively the performance of each of the DOA measurement subsystems.

SSL System

How it works

Among several time delay measurement algorithms, three of them have been selected as independent estimator methods:

- **cross-correlation (CC)**
- **phase transform (PHAT)**
- **maximum likelihood (ML)**

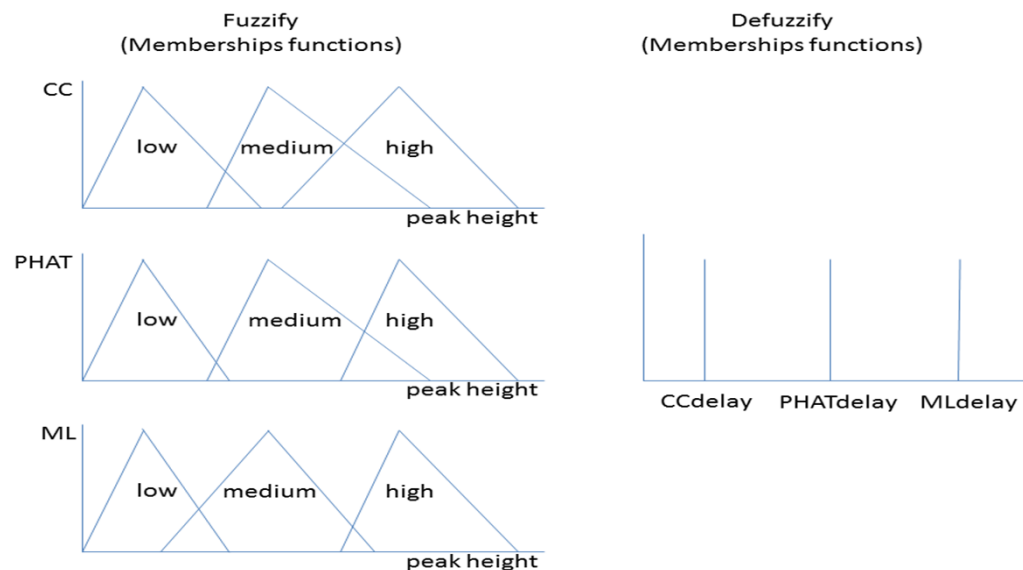


Cross correlation, phase transform, and maximum likelihood functions combine the signal captured from a pair of microphones

Fuzzy logic

How it fuses

The three methods perform well in **low noise and low reverberant rooms**, highlighting a unambiguous position of the maximum, but if noise increases and reflections occur, then each method decrease its performance. To improve the performance of the delay estimation, **the three algorithms run independently each other**, measuring the time delay of a sound frame while the upper layer executes the **fusion of the time delay measured** for the previous sound frame.



Membership functions to fuzzify the measurements and defuzzify the decision

Fuzzy logic

How it fuses

A set of rules has been compiled to fuse the information at decision level.

The decision rules look like this:

IF

CCpeak1 IS High AND
PHATpeak1 IS High AND
MLpeak1 IS High AND
CCpeak12 IS Medium AND
PATHpeak12 IS High AND
MLpeak12 IS Medium

THEN

Delay is PATHdelay

Only two parameters have been used in the rule set, peak1 and peak12. Peak1 is the primary prominent peak amplitude. Peak12 is the secondary prominent peak amplitude, closer to the primary. Basic ruleset consists of 9 rules, but **more rules are required when more secondary sound sources are close to the primary sound source.**

In this first release of the system, the rules are **hand tuned** at compile time, but an **adaptive tuning process** will be implemented to take in count the evolving nature of sound scenarios.

Fuzzy logic

How it fuses

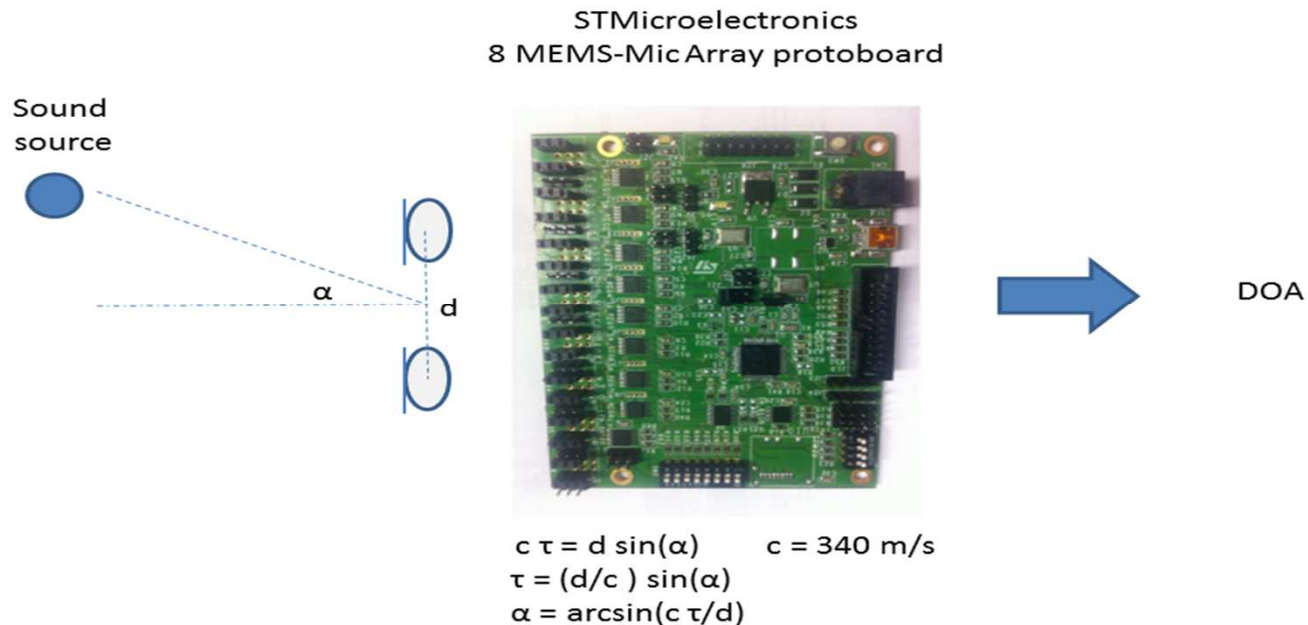
The fuzzy decision is then defuzzified by a **singleton membership function** that produces as output the crisp measurement of the estimated delay. The center of gravity (CoG) defuzzification method has been applied in its weighted average (WA) implementation for singleton membership functions:

$$\text{Crisp_out} = \frac{\sum (\text{fuzzy_out}) (\text{singleton_position})}{\sum \text{fuzzy_out}}$$

SSL System

How it performs

A set of tests has been executed in a simulated context using pure tone sound sources and short uttered frames. The simulation has been executed in Matlab environment, with an STMicroelectronics 8 MEMS microphonic array protoboard connected to Audacity acquisition and editing IDE for simultaneous 8 channel data acquisition, and a loudspeaker as sound source.



SSL System

How it performs

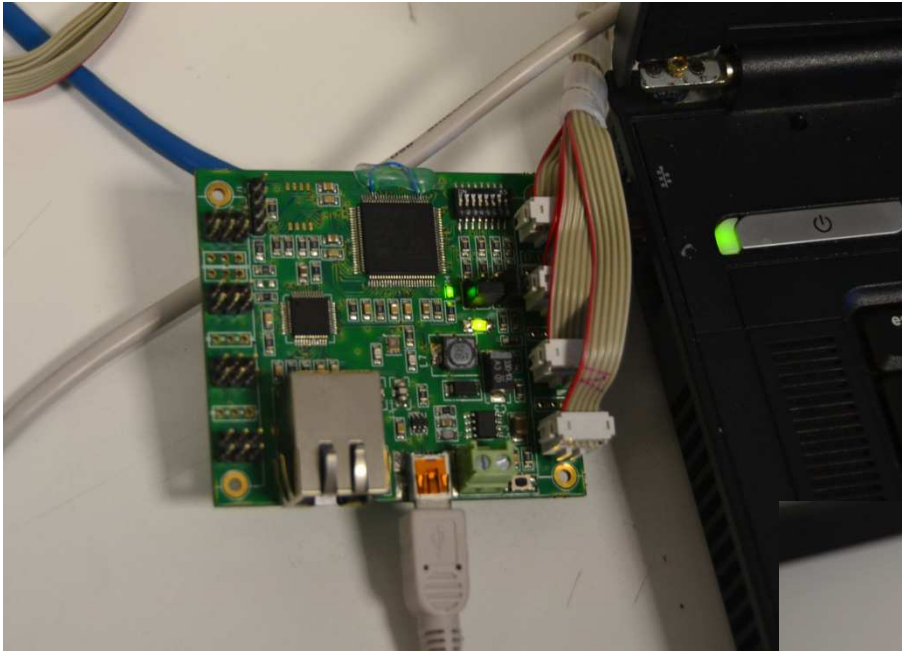
The pure tone has been played at three different frequencies (500, 1000, 2000 Hz) and positioned at three different angles (0, 15, 30 degree). The same test has been executed for a short utterance. The two tests have been executed with and without noise. The following success rate resulted:

	CC	CCwithPHAT&ML	FuzzyFusion
Tone test noiseless	100%	100%	100%
Speech test noiseless	95%	97%	97%
Tone test noisy	85%	88%	93%
Speech test noisy	75%	77%	91%

The tests have confirmed that **CC supported by PATH and ML works well than CC alone** and that **fuzzy fusion improves the performance** of DOA measurements, mainly in noisy contexts.

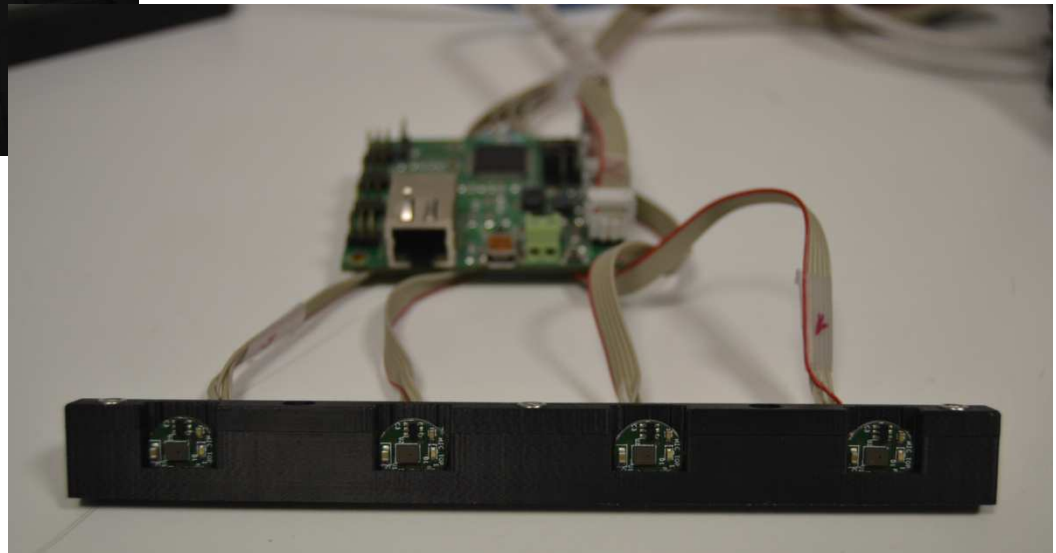
SSL System

How it is made



DAQ board

MEMS Microphone array



SSL System

How it applies



MainWindow

Connect Port Number 4 Ping

Close

RECOGNITION

Stat Recognition

Stop Recognition

Word Recognized:
NOT REC

TRAINING

Avanti

Listen to word

LOCALIZATION

-50

angle getting
angle getting
angle getting
angle getting
angle getting
angle getting
angle getting
angle getting
angle getting
angle getting

Word List:
Update List Format Flash

Start Streaming 4 mics 16000 Stop Streaming

Conclusions

- ✓ The **sound source localization is a very complex task** which, at its core, involves the interaction among multiple microphones.
- ✓ **Smart data fusion based on a fuzzy logic solution is an effective approach** to manage and fuse the gathered data, mainly because a multiple level data fusion can be implemented.

Conclusions (cont.)

- ✓ To improve the reliability of the system, **future developments will concern the reengineering of the fuzzy logic-based data fusion**, splitting the fuzzy data fusion engine in a multi-level hierarchy, the lower level for feature fusion and the upper level for decision fusion.
- ✓ **Each pair of microphones will work like a sub-module to compose a specific geometry to match a specific application.** To this purpose, a third fuzzy fusion level layer will be developed to fuse the DOA decisions of the microphones pairs.

Thank you for your attention
(any question?)

Mario Malcangi

*Università degli Studi di Milano
Department of Computer Science
Via Comelico 39 – 20135 Milano - Italy
DSP&RTS Research Laboratory
(Digital Signal Processing & Real-Time Systems)*

Please, address any further question to:

malcangi@di.unimi.it