**1.1**

We know that correlation is a measure of how much a given signal correlates with, or resembles, another. Cross correlation is also a measure of similarity of two signals, as a function of displacement of one with respect to the other. Looking at the equation, we can see that it has a similar nature to convolution, only differing in that it does not require the other signal to be flipped, but only time-shifted.

The asymptotic running time is O(n log n).

**1.3**

GCC-PHAT stands for General Cross Correlation with Phase Transform. This is used for computing the time delay between the reference channel and another channel, for a given segment. The GCC-PHAT is estimated as the delay that causes the cross correlation between the two segments to be maximum.

One of the most significant factors to keep in mind is the analysis window size in relation to the segment size – a small analysis window degrades the cross-correlation estimation, while a large window reduces the resolution of changes in the time delay of arrival (TDOA).

The maximum value of the inverse Fourier Transform of the GCC PHAT gives the estimated delay for a particular segment, but this maximum value is avoided in three cases, when the maximum could be due to a spurious noise, when two or more speakers overlap each other, and when the processed segment comprises a lot of acoustic data that isn’t speech.

Observation, compare with 1.2, include two plots for two types of correlation

**2.1**

Given that for time delay t,

Δt = Δx. Sin (Ф)/c

Sin (Ф) = (Δt . c)/ Δx

Given values Δx = 0.25m, c = 343 m/s, fs = 45kHz

Also given Δt = m.Ts = m. (1/fs)

* Sin (Ф) = (m x 0.00002 x 343)/0.25 = m x 0.00030489

Changing Δx to 0.5 m,

* Sin (Ф) = (m x 0.00002 x 343)/0.5 = m x 0.015244

For a frequency of 900 Hz, and Δx = 0.25m,

* Sin (Ф) = (m x 0.0011 x 343)/0.25 = m x 1.52
* Ф = (m x 1.52)

Since Sin Ф can take a max value of 1,

* Ф = (1)
* The maximum angle where sound can be located is

**2.2**

Figure with speaker image

The location of the speaker for the given case is unique in 3D. A microphone array is set up in a way that allows detection over a particular range, in a given direction. The system here is a triangular microphone array with three sensors collecting speaker information, which leads to triangulation of speaker location to a unique position in the grid covered by the microphone array. If the speaker was the move, then speaker localisation methods would have needed to be implemented, which is not the case here.