

# Speaker Position Estimation in Vehicles by Means of Acoustic Analysis

Alexej Swerdlow, Timo Machmer, Benjamin Kühn, Kristian Kroschel

*Institut für Nachrichtentechnik (INT), Universität Karlsruhe (TH)*

*Kaiserstr. 12, 76131 Karlsruhe, Germany, Email: swerdlow@int.uni-karlsruhe.de*

## Introduction

The interaction between man and machine via acoustic analysis systems gains more and more in importance for next-generation cars. Thereby, the seat positions of car occupants are of peculiar interest. Generation of some position specific properties for seat and air conditioning settings, but also for manipulations of entertainment systems, is the typical situation where this information can be useful.

An in-car speaker localization system presented on the DAGA 2007 [1] enabled the detection of passengers by means of a time delay based method and a microphone array in two steps. First, the time delay of arrival (TDOA) of sound signals in a pair of spatially separated microphones had to be estimated. Then the estimated TDOA in combination with the known microphone array geometry could be used for the localization of the sound source in the environment.

The current paper evaluates an improved one-step approach for sound source localization using SRP-PHAT for the position estimation. Furthermore, different microphone array configurations are investigated to find optimal array geometry and consider requirements of the automotive industry concerning microphone positioning possibilities inside modern car models.

## Source Position Estimation

Today two approaches for acoustic localization are mainly used. The first one is based on the estimation of the time difference of arrival (TDOA) of sound signals in a pair of spatially separated microphones. The most common technique for the determination of TDOAs is the generalized cross correlation (GCC). The GCC function  $R_{ij}^{(g)}(\tau)$  is defined as

$$R_{ij}^{(g)}(\tau) = \int_{-\infty}^{+\infty} \psi_{ij}^{PHAT}(\omega) X_i(\omega) X_j(\omega)^* e^{j\omega\tau} d\omega, \quad (1)$$

where  $X_i(\omega)$  and  $X_j(\omega)$  are the Fourier-Transforms of given microphone signals.  $\psi_{ij}$  is a weighting function which intends to decrease the noise and reverberation influences and tries to emphasize the GCC peak at the true TDOA. For real environments, the Phase Transform (PHAT) technique has shown the best performance [2]. The PHAT weighting function is defined as

$$\psi_{ij}^{PHAT}(\omega) = \frac{1}{|X_i(\omega) X_j(\omega)^*|} \quad (2)$$

and can be regarded as a whitening filter.

The other well known technique for the acoustic localization is the so called Power Field (PF), also known as Steered Response Power (SRP) [3]. In this approach, beamforming is used to focus a microphone array to a specific spatial area. SRP scans the environment and searches for the spatial position with the highest acoustic power in order to find the exact position of a sound source.

Combination of both techniques mentioned before leads to a method called SRP-PHAT, which fuses the stability of the SRP against reverberations and the efficiency of the GCC method. Additionally, it enables us to build a real-time capable system. At a given time  $t$ , SRP-PHAT is computed as

$$P(t, \mathbf{s}) = \frac{1}{|M_p|} \sum_{(i,j) \in M_p} R_{ij}^{(g)}(t, \tau_{ij}(\mathbf{s})), \quad (3)$$

where  $\tau_{ij}(\mathbf{s})$  denotes the theoretical delay between the microphones in pair  $(i, j)$  for the assumed spatial source position  $\mathbf{s} = (s_x, s_y, s_z)$ .  $M_p$  represents a given set of microphone pairs. To estimate the source position  $\hat{\mathbf{s}}(t)$  at time  $t$ , the position of the maximal value in  $P(t, \mathbf{s})$  has to be found in a given search space  $\mathbf{S}$ :

$$\hat{\mathbf{s}}(t) = \arg \max_{\mathbf{s} \in \mathbf{S}} P(t, \mathbf{s}). \quad (4)$$

## Seated Position Determination

The seated position inside the vehicle  $\hat{\sigma}(t)$  at time  $t$  is determined in the following way:

$$\hat{\sigma}(t) = \arg \min_{\sigma} \|\mathbf{s}_{\sigma} - \hat{\mathbf{s}}(t)\|_2, \quad (5)$$

where  $\mathbf{s}_{\sigma}$  is the spatial head position of the person sitting on seat  $\sigma$ .

An estimation of the speaker position is deemed correct if the calculated position is located within the proper seat of the speaker.

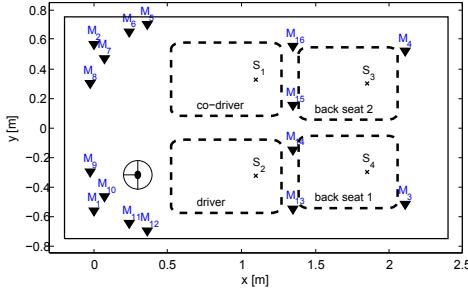
## Experimental Setup

At the first step, we used recordings and the array configuration, which we presented on DAGA 2007 [1]. This strategy allowed us to compare the performance of two different localization methods (GCC-PHAT and SRP-PHAT).

Afterwards, we evaluated two further microphone array configurations using SRP-PHAT only. For this purpose, real experiments were carried out in a exemplary up-to-date car that was parking in a typical public road.

All recordings were taken with the same male speaker, pronouncing 30 short car control commands (e.g. *Radio on!*, *Travel direction!*, etc.) from four different positions ( $S_1, \dots, S_4$ ) according to four seat possibilities within the car. Per seat position, five head rotation angles ( $-90^\circ, -45^\circ, 0^\circ, +45^\circ, +90^\circ$ ) were examined.

The sampling frequency was 48 kHz; the recorded speech signal was analyzed in frames of 8192 samples. The source position was estimated by means of a 2D grid search with grids of 2 cm and a total grid dimension of 1.50 m x 2.40 m (according to the car dimension).



**Figure 1:** Schematic illustration of the experimental setup with microphones M1-M16 and speaker positions S1-S4.

Figure 1 shows the experimental setup with positions of all microphones placed inside the car, and four possible speaker positions as well. Three different array configurations are given in Table 1.

**Table 1:** Used array configurations and corresponding microphones.

Array configuration	Used microphones
AC1	M1-M4
AC2	M5-M12
AC3	M6, M8, M9, M11, M13-M16

## Results and Discussion

In the following, experimental results are presented. Table 2 compares localization accuracy of two localization methods mentioned before, for the case of using the same audio data from [1] and array configuration AC1. As can be seen, SRP-PHAT shows a significant rise of the localization accuracy compared with GCC-PHAT, especially for both front-seat positions. Localization results for two

**Table 2:** Percentage of correct speaker localization for all seat positions and array configuration AC1. Comparison of localization accuracy between GCC-PHAT and SRP-PHAT, based on 1500 measurements per speaker position.

Speaker position	GCC-PHAT	SRP-PHAT
driver	76,71	96,21
co-driver	73,09	97,24
back seat 1	93,84	97,31
back seat 2	97,12	97,46

further array configurations (AC2 and AC3) are summarized in Table 3 and Table 4. In case of AC2, where microphones on the dashboard only are used, the localization accuracy is lower for some head rotation angles of back-seat passengers. In AC3, four microphones on the dashboard and further four microphones in the headrest of both front seats enable reliable localization for all seat positions and all head rotation angles.

**Table 3:** Percentage of correct speaker localization for AC2, based on 1000 measurements per seat and head rotation angle.

Position	$-90^\circ$	$-45^\circ$	$0^\circ$	$+45^\circ$	$+90^\circ$
driver	98,87	99,79	99,42	99,89	97,75
co-driver	100,0	100,0	99,17	99,31	91,80
back seat 1	88,61	90,26	98,90	43,15	41,05
back seat 2	56,67	26,02	95,78	90,90	90,81

**Table 4:** Percentage of correct speaker localization for AC3, based on 500 measurements per seat and head rotation angle.

Position	$-90^\circ$	$-45^\circ$	$0^\circ$	$+45^\circ$	$+90^\circ$
driver	86,12	98,47	92,63	83,33	70,20
co-driver	85,99	91,48	94,48	94,62	87,54
back seat 1	92,31	95,28	96,75	75,36	86,47
back seat 2	85,96	97,58	90,11	93,37	91,24

## Conclusion

Speaker position estimation in vehicles by means of the acoustic analysis is a promising task. In this paper, we presented an in-car localization system, which we call CarOPE (**C**ar **O**ccupants **P**osition **E**stimation). It is able to determine the seated positions of car occupants. Due to its real-time capability, it is achievable to integrate the system in next-generation car models.

In future research, we will investigate the influence of the environmental noise which inevitably occurs while driving. In order to identify persons inside the vehicle, further work will also combine the localization of car occupants with an acoustic classification method.

## Acknowledgment

This work has been supported by the German Science Foundation DFG within the Sonderforschungsbereich 588 *Humanoid Robots*.

## References

- [1] Swerdlow A., Kroschel K., Machmer T.: Speaker localization in vehicles via acoustic analysis. DAGA 2007, Stuttgart, Germany, March 2007
- [2] Knapp, C. H., Carter, G. C.: The generalized correlation method for estimation of time delay. IEEE Trans. on Acoustics, Speech and Signal Processing, 24(4): 320-327, August 1976.
- [3] DiBiase, J. H., Silverman, H. F., Brandstein, M. S.: Microphone Arrays, chapter Robust Localization in Reverberant Rooms. Springer, 2001.