Speaker Localization in Vehicles via Acoustic Analysis

Alexej Swerdlow, Kristian Kroschel, Timo Machmer
Institut für Nachrichtentechnik (INT), Universität Karlsruhe (TH), Kaiserstr. 12, 76128 Karlsruhe, Germany
Email: swerdlow@int.uni-karlsruhe.de

Introduction

The interaction between man and machine, among others via acoustic analysis systems, gains more and more in importance for vehicles that will come on the market in the next years. Thereby, the position of the user within the vehicle is of peculiar interest. If it were feasible to appoint a specific seat from where the car is being controlled, it would be possible to parameterize the selected control instructions with some position specific properties. Demonstrative examples therefore are the seat and air conditioning settings within the vehicle, but also manipulations of entertainment systems.

The position of the speaker is also of importance in order to regulate the release of the restraint system and airbags. In contrary to the optic system, which only works with lighting or rather daylight, the acoustic system can also be used at dark. This is a basic advantage, since the sitting position of the passenger typically only changes when he speaks, for instance, with other occupants of the car.

A sound localization system, which was developed in our research group[1], enables the detection of audio signal sources, among others also speakers, by using 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 is to be estimated. Then the estimated TDOA in combination with the known microphone array geometry is used for the localization of the sound source in the environment.

Up to now, the method was deployed for localization in a kitchen environment of the collaborative research center *Humanoid Robots*. The efficiency of the localization method in a car is being investigated in this paper. Thereby, we confine us to estimate the azimuth angle. Two microphone pairs P_{12} and P_{34} were placed within the vehicle: one in the front section, on the dashboard (P_{12}) and the other one, in the back of the car, on the rear shelf (P_{34}).

Time Delay Estimation

For a given pair of spatially separated microphones M_i and M_j, the microphone signals $x_i(t)$ and $x_j(t)$ for a signal $s(t)$, coming from a remote sound source in a vehicle, can be modelled mathematically as

$$x_i(t) = h_i(t) * s(t) + n_i(t) \quad (1)$$
$$x_j(t) = h_j(t) * s(t - \tau_{ij}) + n_j(t), \quad (2)$$

where τ_{ij} represents the relative signal delay of interest, $*$ signifies the convolution operator, $h_i(t)$ is the acoustic impulse response between the sound source and the i^{th} microphone and the additive term $n_i(t)$ summarizes the channel noise in the microphone system as well as environmental noise for the i^{th} sensor. This noise $n_i(t)$ is assumed to be uncorrelated with $s(t)$ and $n_j(t)$.

The most popular approach for determining the TDOAs is the Generalized Cross Correlation (GCC) method[2]. The relative time delay τ_{ij} is estimated as the time lag with the global maximum peak in the GCC function $R_{ij}^{(g)}(\tau)$:

$$\hat{\tau_{ij}} = \arg\max_{\tau} R_{ij}^{(g)}(\tau). \quad (3)$$

This GCC function $R_{ij}^{(g)}(\tau)$ is defined as

$$R_{ij}^{(g)}(\tau) = \int_{-\infty}^{+\infty} \psi_{ij}(\omega) X_i(\omega) X_j(\omega)^* e^{j\omega\tau}\, d\omega \quad (4)$$

with $X_i(\omega)$ the Fourier transform of $x_i(t)$.

The weighting function ψ_{ij} intends to decrease noise and reverberation influence and tries to emphasize the GCC peak at the true TDOA τ_{ij}. For real environments, the *Phase Transform (PHAT)* technique has shown the best performance [3]. The PHAT weighting function is defined as

$$\psi_{ij}^{PHAT}(\omega) = \frac{1}{|X_i(\omega) X_j(\omega)^*|}. \quad (5)$$

As shown in [4] the absolute value of the first maximum peak in the GCC function can be used very efficiently to evaluate the reliability of the actual TDOA estimate. This criterion allows a reliability scoring of individual estimates and can be used to reject erroneous measurements. The higher the value of the first peak in the GCC function is, the higher is the probability that the TDOA was estimated correctly.

Determination of Sound Source Position

Using the estimated time delay, the angle α_{ij} to the sound source in interval $[-\frac{\pi}{2}; \frac{\pi}{2}]$ can be determined, whereas the peak of the angle is the center between the two microphones M_i and M_j of the microphone pair P_{ij}. The estimation of the angle takes place independently for each microphone pair. In order to determine the position of the sound source within the vehicle, the intersection point of both rays of the estimated angles α_{12} (front microphones) and α_{34} (rear microphones) must be calculated. The peak of angle α_{12} is defined as point of origin of the coordinate system. Since the mobility of the vehicle's occupants is limited, the ascertained position estimation can

be allocated with one of the four classes (driver's seat, front passenger seat, seat behind the driver and seat behind the front passenger). Additionally, there is going to be a reject-class, which will contain all estimations that could not definitely be appointed to any other seat within the vehicle. This assumption of a maximum of four occupants is done without loss of generality. An expansion for five or even more occupants is conceivable.

Experimental Setup

As already mentioned, two microphone pairs of omnidirectional electret condenser microphones were used for data recording. Real experiments were carried out in an exemplary up to date car that was parking in a typical public road. The distance of the microphone pair P_{12} amounts 1.14 m, that of P_{34} 1.04m. Different utterances of German sentences (altogether 5104 words) from two speakers were played back by a loudspeaker. The loudspeaker was placed in four different positions $(S_1, \ldots, S_4)$ according to four seat possibilities in the car. The sampling frequency was 16 kHz. The recorded speech signals were analyzed in frames of 32 ms to assure quasi-stationarity. For this data segmentation a Hamming window with a 50% overlap was applied. According to [4] we decided to use $m = 0.35$ as the threshhold for the value of the first peak in the GCC function (4). An estimation of a speaker position is deemed correct if the calculated position is located within the proper seat of the speaker. Figure 1 shows the experimental setup using the example of the speaker on the back seat behind the co-driver. It contains the estimated speaker position as the intersection point of two rays.

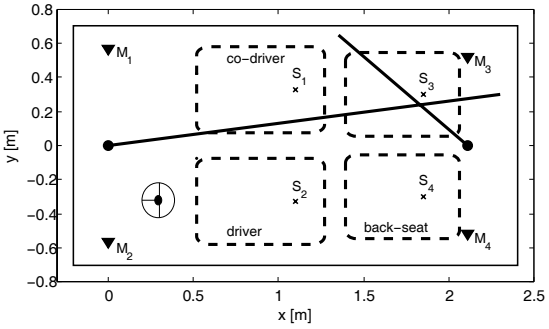

Figure 1: Schematical illustration of the experimental setup

Results and Discussion

In the following, the analysis of the localization accuracy is presented. The results are summarized in Table 1 and Table 2 separately for each of the four possible seating positions and averaged over the two speakers.

It is noticeable that the quality of the localization is remarkably influenced by the seating position. The recognition on the front seats is, by tendency, worse than on the back seats. This results from the relatively imprecise localization of the front speakers by the microphone pair P_{34}. That is conditioned by the specific sound wave propagation within the vehicle. The microphone pair P_{12} on the dashboard receives mainly direct waves, whereas the microphones M_3 and M_4 receive primarily the already reflected signal, because they are situated behind the speakers on the rear shelf. The estimations of the angle α_{12}, delivered by the microphone pair P_{12}, is considerably precise for each of the four seating positions. Microphone pair P_{34} on the other hand, gives accurate estimations for the occupants on the back seats only.

speaker position	**front-seat**	
	driver side	**co-driver side**
correct	76,71	73,09
rejected	17,71	22,07
incorrect	5,58	4,84

Table 1: Percentage of estimation results for front-seats

speaker position	**back-seat**	
	driver side	**co-driver side**
correct	93,84	97,12
rejected	5,32	1,72
incorrect	0,84	1,16

Table 2: Percentage of estimation results for back-seats

Conclusion

Acoustic speaker localization in a real car environment is a promising task. The simple localization system with four microphones presented in our paper is the first step in this direction. Further investigations will consider other microphone positioning to improve the reliability of the TDOAs and. A post processing unit promises better and more accurate determination of the speaker position. In future research, measurements should be done while driving to investigate the influence of the environmental noise. Furthermore a real time speaker tracking system is conceivable. Further work will also combine localization estimations with a classification method which enables to identify the localized person[5].

References

[1] Bechler, D.: Acoustic Speaker Localization Using a Microphone Array. Degree thesis, Karlsruhe, 2006

[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.

[4] Bechler, D., Kroschel, K.: Confidence scoring of time difference of arrival estimation for speaker localization with microphone arrays, 13. Konferenz Elektronische Sprachsignalverarbeitung ESSV, Dresden, 2002.

[5] Kroschel, K., Bechler, D.: Demonstrator for Automatic Text-independent Speaker Identification. DAGA '06, Braunschweig, 2006, p. 102