Integrating Smooth Motion Assumptions with RANSAC-based Sound Source Localization

Jens Gulin†*, Kalle Åström†, Amir Aminifar†
* Sony Europe B.V., Sweden † Lund University, Sweden

LUND UNIVERSITY

SONY

Motivation & Contribution | Smooth motion leads to better source sound localization

Time-difference-of-arrival (TDOA) estimates from cross-correlation are noisy in real environments, and a robust multilateration method must handle outliers. In this work, the smooth motion assumption is explored in different stages of a RANSAC-based (Random Sample Consensus) implementation. The evaluation is done on real recordings from the public LuViRA dataset, giving the first 3D baseline result on the dataset. Each of the proposed steps is shown to reduce the localization error.

RESULTS | Indicating successful outlier removal

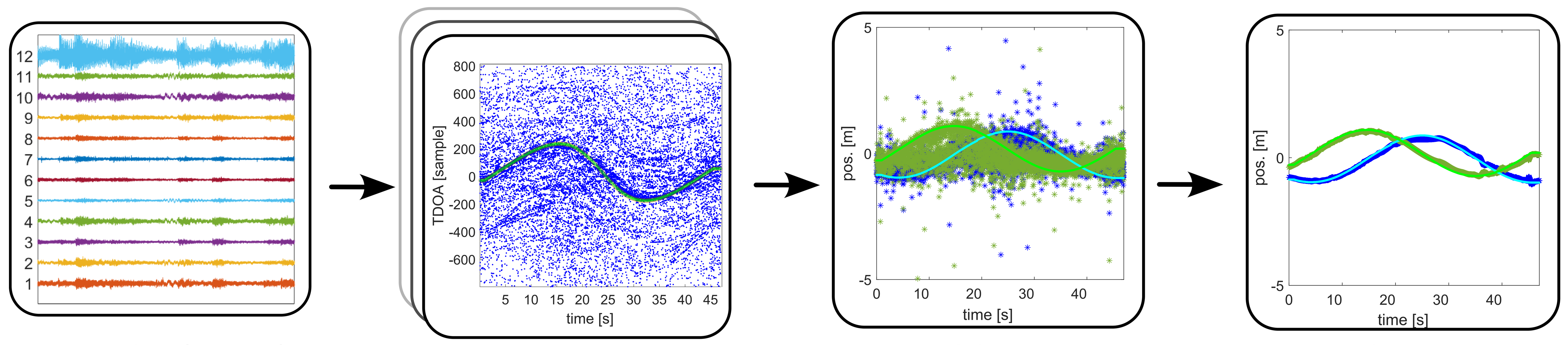
Table 1: Mean average error (MAE) and the improvement over baseline.

Method	*MAE* ↓	*gain* ↑
Baseline (A)	118.0	–
ABC	31.1	74%
ABCD	28.9	76%
ABCDO	28.8	76%
ABOCDO	**24.6**	**79%**

Using RANSAC on larger time windows and allowing TDOA to influence the smoothing step seems to be fruitful. The summary in Table 1 shows MAE improvements of 79% compared to the baseline (robust trilateration, but no smoothing).

With ABCD the median error of each trajectory is mostly well below 15 cm. Yet ABOCDO offers an additional improvement of 15% for MAE. The median smoothing (O) at the end has little impact, but the early O shows the effect of removing more outliers earlier in the chain.

Evaluated on 11 trajectories (53923 estimates, each with 11 microphones, a single sound source and additional background noise) from the LuViRA dataset [2].The TDOA estimations used are from GCC-PHAT.



Input Synchronized microphone recordings → Extract TDOA estimates for each microphone pair → Use multilateration to calculate initial locations → Use initial estimates to calculate optimized locations

Proposals | Our smooth motion assumption

Outlier avoidance (B)

- Increase the voting window, for a stronger chance of votes in the vicinity of the true location.
- RANSAC selection directly from the 3D estimates s avoids TDOA outliers and extraneous solutions from the minimal solver.
- Even a large selection window has a low runtime penalty.
- With medium size windows, favorable conditions are likely.

Smoothness optimization (C)

- Smoothness optimization proposed in [3], not RANSAC-based.
- Penalizes rapid velocity changes over s and lmeasurement errors among the TDOA inliers.

Linear motion optimization (D)

- Voting assumes a locally linear motion, not only a relaxed margin, iteratively starting from a zero-velocity assumption.
- With a large voting set, a good linear fit should be relatively independent of a single point.
- Preliminary results indicate that a zero-size selection window is sufficient.

Standard filter method (O)

- Considering the estimated locations (denoted s) without regard to the underlying TDOA measures.
- Essentially a median smoothing of each dimension separately.
- Has worst-case pitfalls, but ok without too many outliers.

Insights | RANSAC window of opportunity

RANSAC [1] can fit parameters to a known model even with outliers present. The method randomly samples the *selection set* repeatedly to find the hypothesis that fits (within a margin) the most samples, using the *voting set* when measuring consensus.

For step A the hypotheses are taken only from the current time frame, and only the current time frame gets to vote. However, in a low velocity setting, **the true source location of adjacent time frames is likely very close** to that of the current frame. Allowing a voting window of "size" 25 frames before and 25 frames after will **thus strengthen the voting set** consensus, still disregarding any votes (frames) outside the set margin. The same argument holds for higher velocity or low acceleration if the window size is small.

With a consistent inlier ratio over time, a **larger selection window does not** increase the *chance* of a successful random selection, but the higher inlier *amount* may still make it worthwhile as long as enough iterations are allowed. Assuming outliers are mostly random, the promise of RANSAC is that there is a low risk of outliers to get a majority vote.

References

[1] Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography
M. A. Fischler and R. C. Bolles. Commun. ACM 1981.

[2] LuViRA dataset validation and discussion: Comparing vision, radio, and audio sensors for indoor localization
I. Yaman et al. JISPIN 2024.

[3] Extension of time-difference-of-arrival self calibration solutions using robust multilateration
K. Åström, M. Larsson, G. Flood, and M. Oskarsson. EUSIPCO 2021.

This work was partially supported by the strategic research project ELLIIT, and by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM