LOCATA CHALLENGE - OVERVIEW OF EVALUATION MEASURES

Christine Evers[1], Heinrich W. Löllmann[2], Heinrich Mellmann[3], Alexander Schmidt[2], Hendrik Barfuss[2], Patrick A. Naylor[1], and Walter Kellermann[2]

[1]Imperial College London, Dept. Electrical & Electronic Engineering
[2]Friedrich-Alexander University Erlangen-Nürnberg
[3]Humboldt-Universität zu Berlin

Email: info@locata-challenge.org
Web: www.locata-challenge.org

ABSTRACT

This document provides an overview of the evaluation measures used for the LOCATA Challenge. The content of this document is an excerpt from a manuscript submitted to IWAENC 2018. Further information about the LOCATA Challenge and its data corpus can be found in [1].

1. EVALUATION MEASURES

This section provides an overview of performance measures used for evaluation of the LOCATA Challenge. As all estimates must be associated with an Identifier (ID), the estimates resulting from localization and tracking algorithms alike are referred to as "tracks" in the following.

1.1. Nomenclature

Consider a single recording of duration Δ_{rec}, including a maximum number of N_{max} sources. Each source $n \in \{1, \ldots, N_{\text{max}}\}$ corresponds to $A(n)$ periods of activity of duration $\Delta(a, n) = t_{\text{end}}(a, n) - t_{\text{srt}}(a, n)$ for $a \in \{1, \ldots, A(n)\}$, where $t_{\text{srt}}(a, n)$ and $t_{\text{end}}(a, n)$, respectively, mark the start and end time of the activity period. The corresponding time indices are $\{\ell_{\text{srt}}(a, n), \ell_{\text{end}}(a, n)\} \geq 0$. Each activity period corresponds to an utterance of speech, which is assumed to include voiced and unvoiced segments.

For each time step t, a localization and/or tracking algorithm estimates the source positional information in a given coordinate system, resulting in $K(t)$ track state estimates, $\mathbf{x}_i(t), i = 1, \ldots, K(t)$ for each $t \in \{1, \ldots, \Delta_{\text{rec}}\}$. Each state estimate is associated with a unique track ID. We note that the evaluation measures detailed in the following are independent of any specific coordinate system, i.e., $\mathbf{x}_i(t)$ is used in a general form and could describe, for example, the source azimuth, elevation, or 3D Cartesian position.

Gating combined with association algorithms [2] leads to an assignment of each track to the nearest ground-truth source position. In practice, tracks are subject to estimation errors. Furthermore, the assignment is subject to ambiguities [2–4] (see Fig. 1):

Valid track: Track associated to one source. $\Delta_{\text{valid}}(a, n)$ and $L_{\text{valid}}(a, n)$ respectively denote the duration and the number of time steps in which source n is associated with a valid track during activity period a.

Redundant track: Associated with a source that is already associated with one or more tracks.

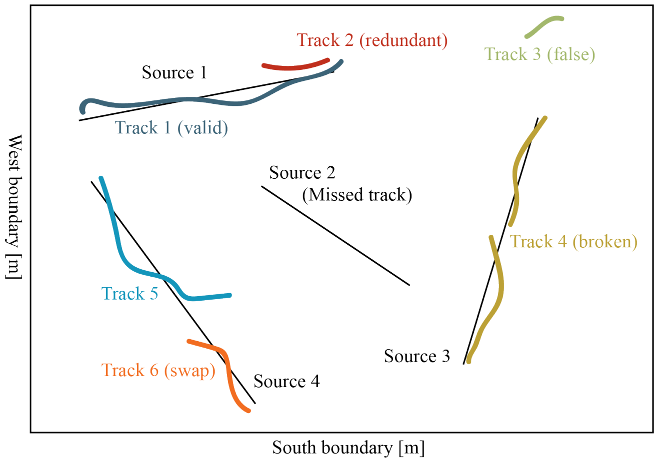

Fig. 1. Tracking ambiguities. Colors indicate unique track IDs.

False track: Not associated with any source.
Missing track: No track associated with a particular sources
Broken track: A valid track whose association is temporarily "interrupted", i.e., the track is not associated with any source for one or more time steps.
Track swap: Source n is associated with track k at $t-1$, and with track $j \neq k$ at t.

To provide a balanced evaluation, the LOCATA Challenge relies on multiple, complementary performance measures in order to reflect track quality, in addition to the accuracy, in terms of ambiguity, completeness, continuity, and timeliness [3].

1.2. Fixed Track-to-Source Association

A one-to-one assignment between sources and tracks is established using the association algorithm in [5]. Based on the track-to-source association, the following evaluation measures are used.

1.2.1. Absolute Track Accuracy

The angular errors [6] of the tracked source azimuth and elevation respectively, as well as the absolute error in the estimated range are evaluated for each recording, source, activity period, and time step. The Euclidean distance between the ground-truth and tracked Cartesian positions is also evaluated.

1.2.2. Track Ambiguity

The track IDs provided by the participants are used to evaluate the number of valid, $K_{\text{valid}}(t)$, false, $K_{\text{false}}(t)$, missing, $K_{\text{miss}}(t)$, broken, $K_{\text{broken}}(t)$, and swapped, $K_{\text{swap}}(t)$, tracks for each time step t. To further quantify ambiguities, the following measures are incorporated in the challenge evaluation:

Probability of detection (p_d) [2]: A measure of track completeness, comparing the number of time stamps during which an active source was detected to the total number of time stamps in activity period a:

$$p_d(a, n) = \frac{L_{\text{valid}}(a, n)}{\ell_{\text{end}}(a, n) - \ell_{\text{srt}}(a, n) + 1}.$$

False Alarm Rate (FAR) [3]: A measure of track ambiguity, comparing the number of false tracks to the number of time steps per recording.

Track Latency (TL) [3]: A measure of track timeliness, indicating the delay between the onset and detection of source n in activity period a.

Track Fragmentation Rate (TFR) [7]: A measure of track continuity, indicating the number changes in the track IDs assigned to an active source.

We note that the one-to-one assignment obtained using [5] avoids redundant tracks.

1.3. Permutations of Track-to-Source Associations

Measures using fixed source-to-track association potentially penalize the tracker for ambiguities in the assignment process during evaluation between the source ground truth and the tracks. The Optimal Subpattern Assignment (OSPA) distance relaxes penalties for association ambiguities by evaluating all permutations of track-to-source associations at each time step. Moreover, the OSPA distance [8, 9] provides a combined measure of the accuracy, continuity, and FAR:

$$\text{OSPA}(\mathbf{X}(t), \mathbf{Y}(t)) \triangleq$$

$$\left[\frac{1}{K(t)} \min_{\pi \in \mathbf{\Pi}_{K(t)}} \sum_{n=1}^{N(t)} d_c(\mathbf{y}_n(t), \mathbf{x}_{\pi(n)}(t))^p + (K(t) - N(t)) c^p \right]^{\frac{1}{p}},$$

where $\mathbf{X}(t) \triangleq \{\mathbf{x}_1(t), \dots, \mathbf{x}_{K(t)}(t)\}$ denotes the set of $K(t)$ track estimates; $\mathbf{Y}(t) \triangleq \{\mathbf{y}_1(t), \dots, \mathbf{y}_{N(t)}(t)\}$ denotes the set of $N(t)$ sources active at t; $1 \leq p < \infty$ marks the order parameter; c is the cutoff parameter; $\mathbf{\Pi}_{K(t)}$ denotes the set of permutations of length $N(t)$ with elements $\{1, \dots, K(t)\}$; $d_c(\mathbf{y}_n(t), \mathbf{x}_{\pi(i)}(t)) \triangleq \min(c, d(\mathbf{y}_n(t), \mathbf{x}_{\pi(i)}(t)))$; and $d(\cdot)$ is the measure of accuracy (see Sec. 1.2.1). The impact of the choice of p and c is discussed in [10].

2. REFERENCES

[1] H. W. Löllmann, C. Evers, A. Schmidt, H. Mellmann, H. Barfuss, P. A. Naylor, and W. Kellermann, *IEEE-AASP Challenge on Source Localization and Tracking: Documentation for Participants*, Apr. 2018, www.locata-challenge.org.

[2] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood (Massachusetts), USA: Artech House, 1999.

[3] O. E. D. Ronald L. Rothrock, "Performance Metrics for Multiple-Sensor Multiple-Target Tracking," *Proc. SPIE*, vol. 4048, Jul. 2000.

[4] X. R. Li and Z. Zhao, "Evaluation of Estimation Algorithms Part I: Incomprehensive Measures of Performance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 4, pp. 1340–1358, Oct. 2006.

[5] H. W. Kuhn, "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, Mar. 1955.

[6] C. Evers and P. A. Naylor, "Acoustic SLAM," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 26, no. 9, pp. 1484–1498, Sep. 2018.

[7] A. A. Gorji, R. Tharmarasa, and T. Kirubarajan, "Performance Measures for Multiple Target Tracking Problems," in *Proc. of Intl. Conf. on Information Fusion*, Chicago (Illinois), USA, Jul. 2011.

[8] B. Ristic, *Particle Filters for Random Set Models*, 1st ed. New York: Springer, 2013.

[9] B. Ristic, B.-N. Vo, and B.-T. Vo, "A Metric for Performance Evaluation of Multi-Target Tracking Algorithms," *IEEE Trans. on Signal Processing*, vol. 59, no. 7, pp. 3452–3457, Jul. 2011.

[10] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A Consistent Metric for Performance Evaluation of Multi-Object Filters," *IEEE Trans. on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.