

Random tracking title

Till Raab, ..., ..., Jan Benda

November 9, 2018

Abstract

work in progress

Introduction

Across animal species we can observe varriouse social structures and species dependent behaviours.

8 **Materials and Methods**

9 **Field site**

10 The study site was a small stream called Rio Rubiano in the Colombian part of tropical grassland plain “Los Llanos”
 11 near San Martin, Province Meta. The recording site was an easy to access part of the Rio Rubiano near the Finca
 12 Altamira (3°76′52.70″N, 73°67′53.41″W) which also served as accommodation. The river bed consists of rocks
 13 with a diameter ranging from a few to up to 50 cm and the riverbank consists mainly of soil, rocks and the roots of
 14 the surrounding vegetation. The very location the recording equipment was installed was a part of the river where
 15 the river width is approximately 9 m and water depth is around 20 cm (distance between water surface and stone
 16 layer on the riverbed). The temperature of the clear water of Rio Rubiano fluctuated between 23 and 27 °C on a
 17 daily basis and showed a conductivity ranging from 2 μ S/cm to 7 μ S/cm. Data acquisition started in April 2016, i.e.
 18 during the start of the rainy season.

19 **Field monitoring system**

20 The recording system used to obtain our data is similar to the one used by (Henninger et al., 2018) in the Republic
 21 of Panamá. It consists of a custom-build 64-channel electrode and amplifier system (npi electronics GmbH, Tamm,
 22 Germany) powered by 12 V car batteries. Signals detected by the electrodes (low-noise headstages embeded in
 23 epoxy resin (1 \times gain, 10 \times 5 \times 5 mm)) were amplified by the main amplifier (100 \times gain, 1st order high-pass
 24 filter 100 Hz, low-pass 10 kHz) before being digitalized with 20 kHz per channel with 16-bit amplitude resolution
 25 using a custom build computer with two digital-analog converter cards (PCI-6259, National Instruments, Austin,
 26 Texas, USA). Data acquisition and storage for offline analysis were managed by a custom software written in C++
 27 (<https://github.com/bendalab/fishgrid>). The maximum of 64 electrodes mounted on 8 PVC tubes were arranged in
 28 an 8 by 8 electrode grid (50 cm spacing) covering an area of 350 \times 350 cm. All 64 electrodes were used throughout
 29 the whole recording period. Each PVC tube, equipped with 8 electrodes got tied to a rope crossing the river,
 30 forming a structure allowing small shifts in electrode distance but being resilient to destruction by rapidly changing
 31 environmental factors, i.e. rising water levels after heavy rainfall.

32 **Extraction of EOD frequencies**

33 Different species of wavetype weakly electric fish and even individuals within the same species show individual
 34 EOD frequencies. We use fast fourier transformations of 95% overlapping data snippets of 2 seconds to evaluated
 35 the frequency composition of each recording channel, i.e. recording electrode, consistent throughout our recordings.
 36 Since EODs of wavetype weakly electric fish are composed of a fundamental discharge frequency and its harmonics,
 37 we detect peaks in the sum of the powerspectra using the Todd algorithm (Todd and Andrews, 1999) and assign
 38 them in groups of fundamental and harmonic frequencies. Fundamental EOD frequencies as well as their powers
 39 in the powerspectra of the different electrodes and the respective time of detection is stored for further analysis.

40 **EOD frequency tracking**

41 To obtain individual EOD signal traces we developed a tracking algorithm based on the difference of individual
 42 EOD signal parameters. We converted actual EOD frequency difference and EOD field differences of potential
 43 EOD signal pairs into normed error values and calculated total signal errors using a cost-function (eq. (1)) repre-
 44 senting the likelihood of potential EOD signal pairs to originate from the same individual, whereby EOD signal
 45 pairs showing smaller errors represent signal pairs more likely originating from the same individual.

$$sig.error = \frac{2}{3} * \Delta field + \frac{1}{3} * \Delta EODf \quad (1)$$

Results

We successfully recorded the electric signals of a population of weakly electric fish for 14 consecutive days. In the attempt to track the electric and spatio-temporal behaviour of single individuals of the observed population we developed an algorithm capable of tracking individual behaviour without the necessity of tagging, even when single parameters of the EOD signals, e.g. EOD frequencies, are temporarily very similar between individuals.

Tracking of individual EODs

The developed and applied tracking algorithms relies on two EOD signal parameters, the EOD frequency difference and the overall EOD field structure similarity of potential EOD signal pairs. Potential EOD signal pairs are constrained by a maximum EOD frequency difference and a maximum time difference between the respective EOD signals. For each potential EOD signal pair, EOD frequency difference and EOD field structure difference is determined and evaluated in a cost-function (see Methods: eq. (1)) resulting in total signal errors. EOD signal pairs are assigned to individual fish based on their total signal error in the two-staged sorting algorithm.

EOD signal tracking parameter: Δ -EODf (EOD frequency difference)

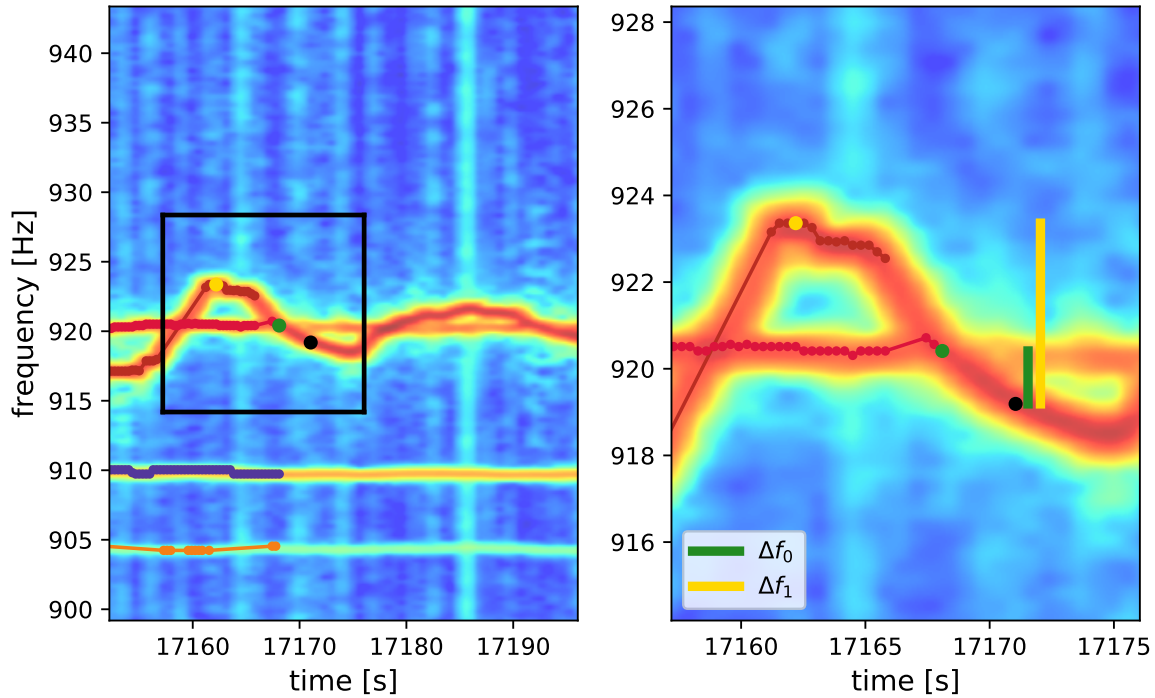


Figure 1: Spectrogram and part of EOD frequency traces from EOD signals of multiple fish including two EOD frequency traces crossing in the course of an EOD frequency rise. Single EOD signals marked in green and yellow form potential candidates to be connected to the EOD signal marked in black. The yellow and green bar represent the respective EOD frequency errors Δf_0 and Δf_1 , as part of the tracking algorithm.

Since, the EOD frequency of wave-type weakly electric fish represent one of the most stable oscillating signals known across natural systems and, thus, keeps stable for long periods of time it seems unambiguous to use this

signal parameter as main tracking criterion. However, tracking individual EOD frequencies over long periods of time gives rise to further challenges. When wave-type weakly electric fish produce communication signals, such as EOD frequency rises, EOD frequency traces of different individuals frequently cross each other, sometimes multiple times within a short time window (see fig. 4 & 6). This specially occurs in larger groups of wave-type weakly electric fish where fish density is increased and, consequently EOD frequency differences between individuals are decreased. Crossing EOD frequency trace give rise to several algorithmic problems. Detecting both peaks during the Powerspectrum analysis in the very moment of the EOD frequency traces crossing often fails resulting in missing datapoints for one EOD frequency trace. Consequently, correct tracking of crossing EOD frequency traces using only EOD frequency as tracking parameter is at chance level.

EOD signal tracking paramete: $\Delta\text{-F}$ (EOD field difference)

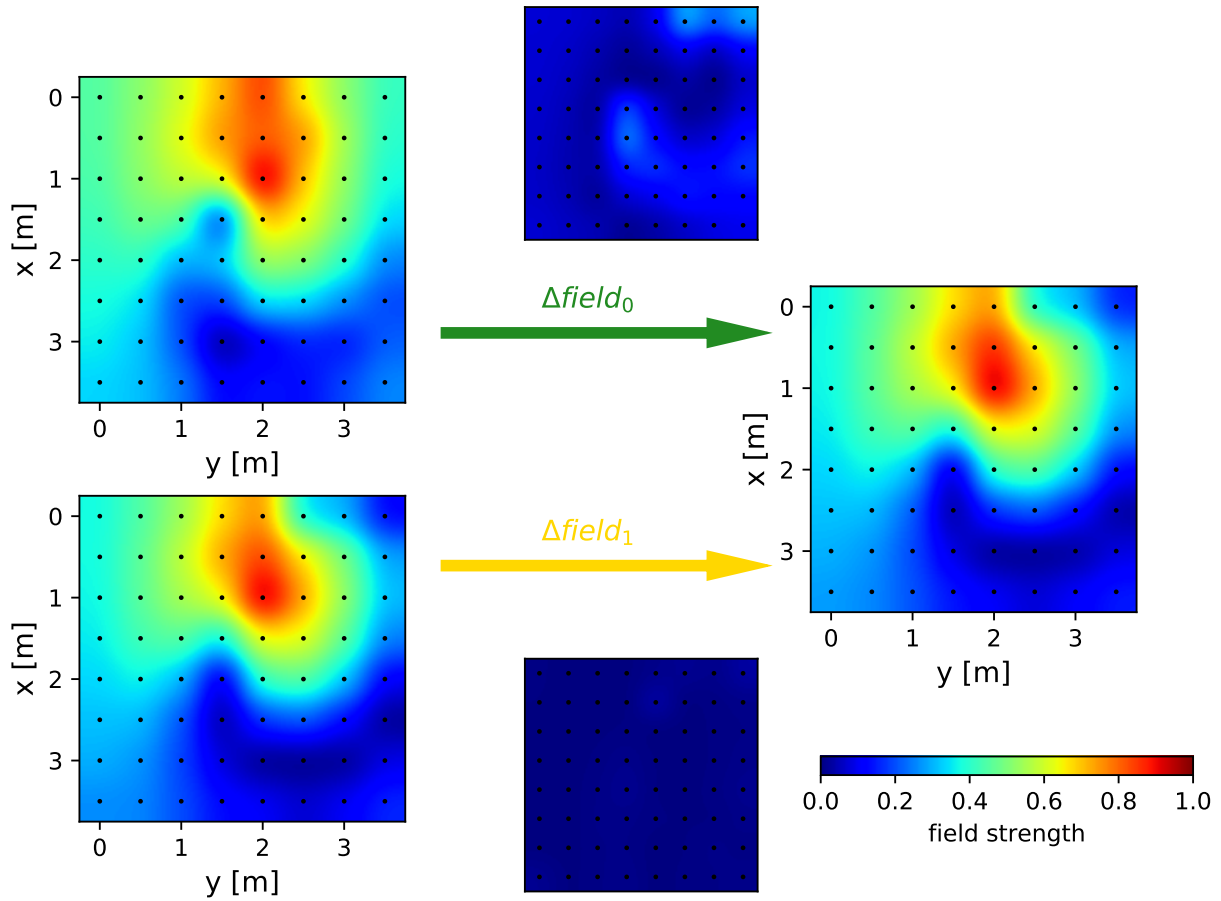


Figure 2: Electric field properties deduced from EOD frequency power on different electrodes of potential partners in the course of tracking, respective to fig.1. Field properties on the left (green and yellow signal in fig.1) are compared to the field properties on the right (black signal in fig.1). The centered top field difference results from the subtraction of the field properties of the green and black signal in fig.1, the centered bottom respectively the difference between the yellow and black signal. Δfield_0 and Δfield_1 represent the root-mean-square-error of the respective diffe differences.

The quasi-sinusoidal EODs of weakly electric fish creates individual dipole field surrounding each fish. Due to our multi-electrode recoding setup we are able to reconsturct the individual electric fields by extracting the strength of individual EOD frequencies in the powerspectra of the data recorded by the different electrodes, i.e.

at multiple locations within our electrode-grid. These two-dimensional representations of the electric fields vary between individuals depending on their very location within the electrode-grid. We use the root-mean-square-error of two EOD fields, i.e. the EOD field difference, of two EOD fields to compare the EOD field extents and, thus, obtain an additional tracking parameter independent from EOD frequency enabling us to even track crossing EOD frequency traces, e.g. during the event of EOD frequency rises.

Error values composed from Δ -EODf and Δ -Field

We convert the absolute values of EOD frequency difference and EOD field difference into relative error values, both ranging from 0 to 1, and calculate the total signal error using a cost-function (see Methods eq. (1)) representing the likelihood of two signals to originate from the same individual.

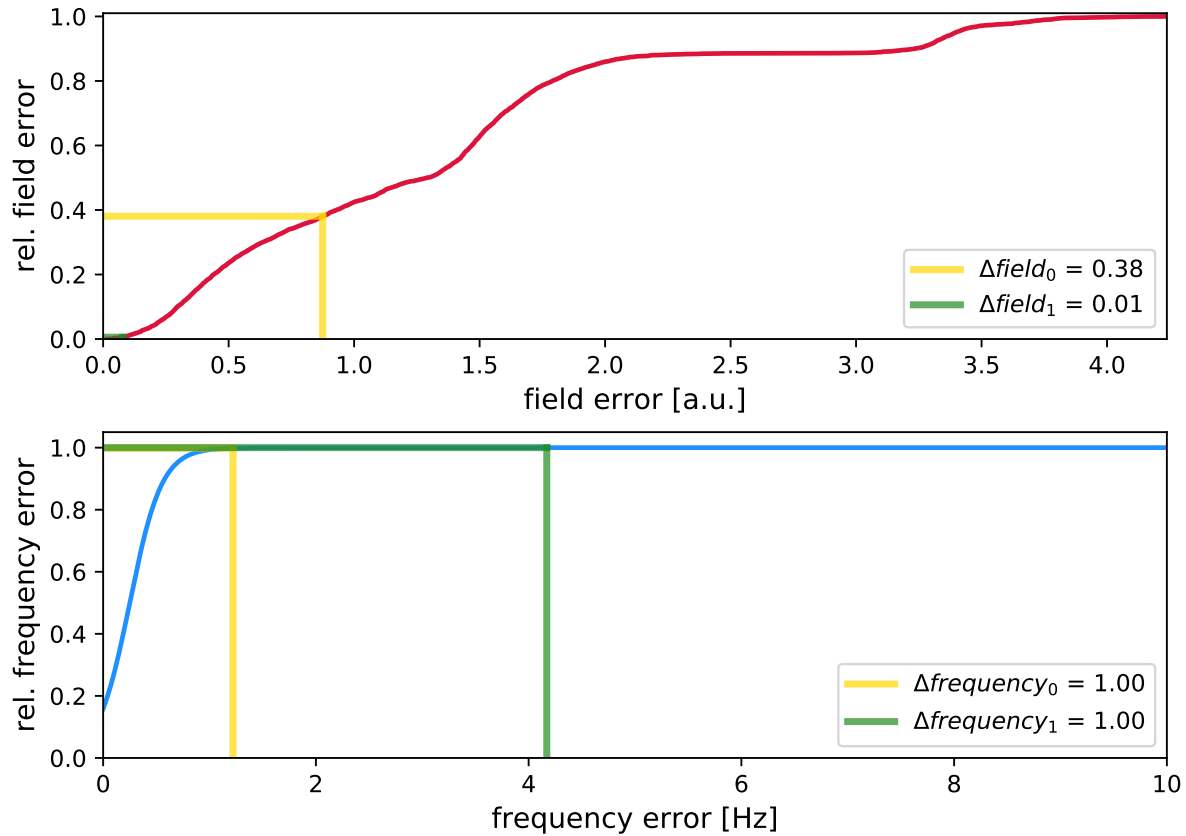


Figure 3: Determination of relative field and EOD frequency errors. Up: All possible field errors of signals within a time window of 30 s ($3 \times$ compare range) showing a maximum EOD frequency difference of 10 Hz were collected as possible field error distribution. Its cumulative sum histogram is displayed in red. Relative field errors for both, $\Delta field_0$ and $\Delta field_1$, are defined as the proportion of possible field errors smaller than the respective field errors. Smaller field errors result in smaller relative field errors and, thus, increase the likelihood of two signals belonging to one identity. Down: Relative frequency errors are calculated based on a boltzmann function. Relative frequency errors above 1 Hz already represent a special event within a EOD frequency trace and, thus, result in the maximum relative frequency error.

83 ***Frequency error determination***

84 With respect to EOD frequency differences we assume EOD frequency differences of above 1 Hz to be equally
 85 unlikely to originate from the same individual and, thus, result in the maximum relative EOD frequency error. EOD
 86 frequencies below 1 Hz result in smaller relative EOD frequency errors and are calculated from a Boltzmann-
 87 function resulting in smaller relative EOD frequency difference the lower the real EOD frequency difference is.

88 ***Field error determination***

89 Considering the difference in field structure the absolute field structure errors are highly dependent on the amount
 90 of electrodes used in the recording setup. Therefore, to estimate the relative field structure error we first estimate
 91 the distribution of possible field structure errors in a 30 seconds window around the currently datapoints of interest.
 92 These possible field structure errors are define as those field structures with a smaller EOD frequency difference
 93 than 10 Hz. Deducted from the distribution of possible field structure differences the relative field structure dif-
 94 ference of two field structures is the proportion of smaller field structure differences in the distribution of possible
 95 field structure differences.

96 ***Total error definition***

97 The absolute error between two signals is calculated using a cost-function evaluating both, relative EOD frequency
 98 error and relative field structure error. Since frequency changes of several Hz within EOD frequency traces are pos-
 99 sible due to the uttering of communication signals like EOD frequency rises and rapid spatial changes comparably
 100 uncommon (see Results) we use the cost-function displayed in equation 1 to estimate the total error value between
 101 different detected EOD signals.

102 ***Assign temporal EOD frequency traces***

103 To enable the analysis of recordings with limitless duration, the actual tracking algorithm is two-staged. First, we
 104 assign so called temporal identities for EOD signals detected in a 30 seconds window. Therefore, we calculated
 105 the total error for every possible connection within this 30 seconds window. Besides the limitation of a maximum
 106 EOD frequency difference of 10 Hz the possible EOD signal pairs are limited by a maximum compare range of 10
 107 seconds, i.e. two signals that shall be connected show a maximum time difference of 10 seconds. According to
 108 the obtained error values temporal identities are assembled starting from the smallest total error representing the
 109 best connection to the largest total error representing the least good connection. Connections that would interfere
 110 with already existing temporal identities are not made since already made connections are based on smaller total
 111 errors and therefore are more likely to be correct. The resulting temporal identities are based on the best possible
 112 connections, but only the centered 10 seconds of the 30 seconds window represent valid connections since the
 113 connections within the head and tail 10 seconds did not take into account all possible connections within \pm compare
 114 range.

115 ***Running connection***

116 The assembly of the center 10 seconds of the temporal identities, containing valid connections, and already tracked
 117 real identities, again is based on total errors between respective EOD signals. Total error values between signals
 118 of already assigned read identities and signals within the centered 10 seconds of temporal identities are identified
 119 and, again, assembled based on the total error values preferring lower total errors before larger ones. Temporal
 120 identities, which could not be assigned to a already existing real identity form a new real identity. The window
 121 to identify temporal identities is shifted by the compare range, i.e. 10 seconds, and the identification of temporal
 122 identities and their assignment to already tracked real identities continuous until the end of the recording is reached.

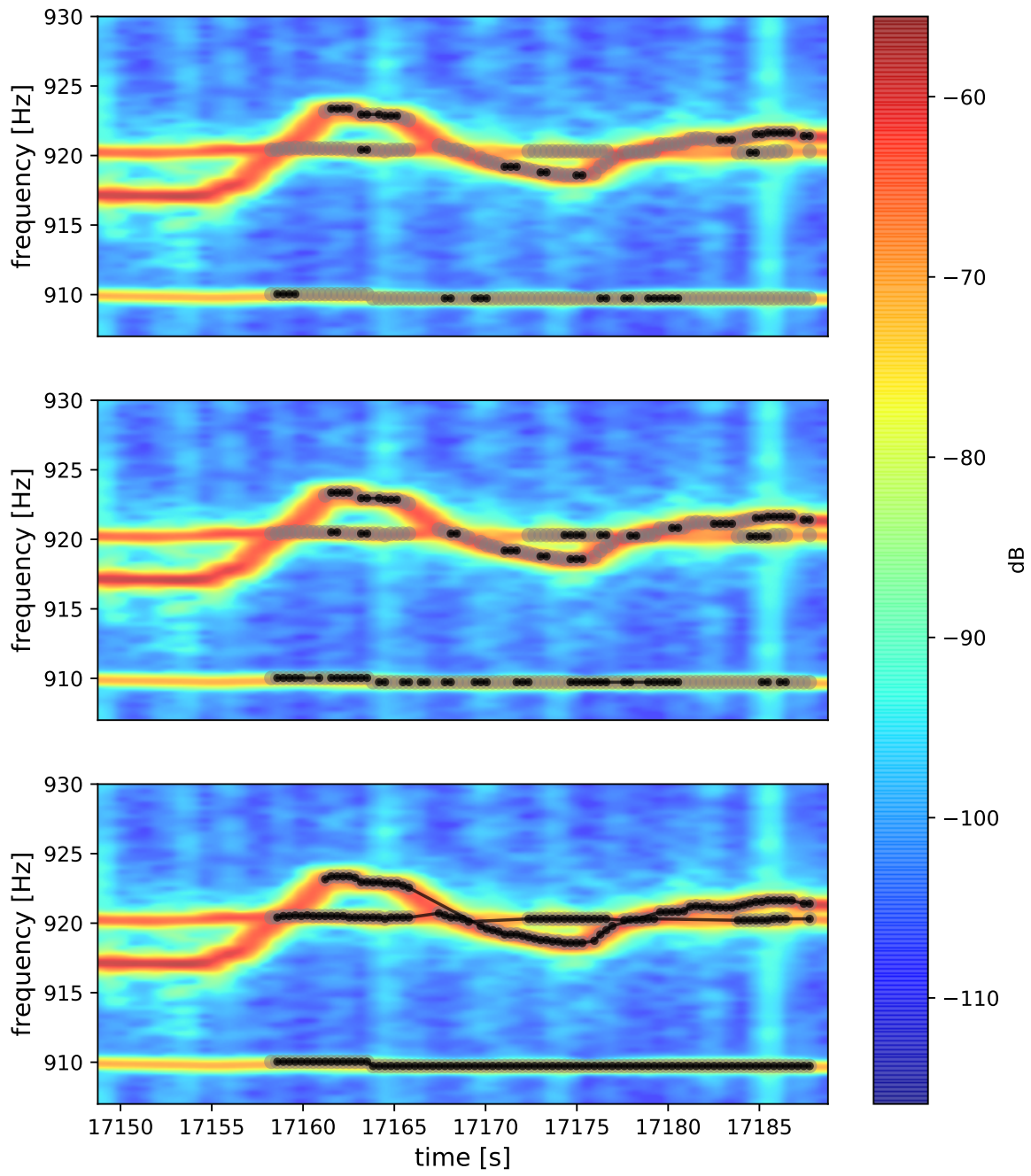


Figure 4: —

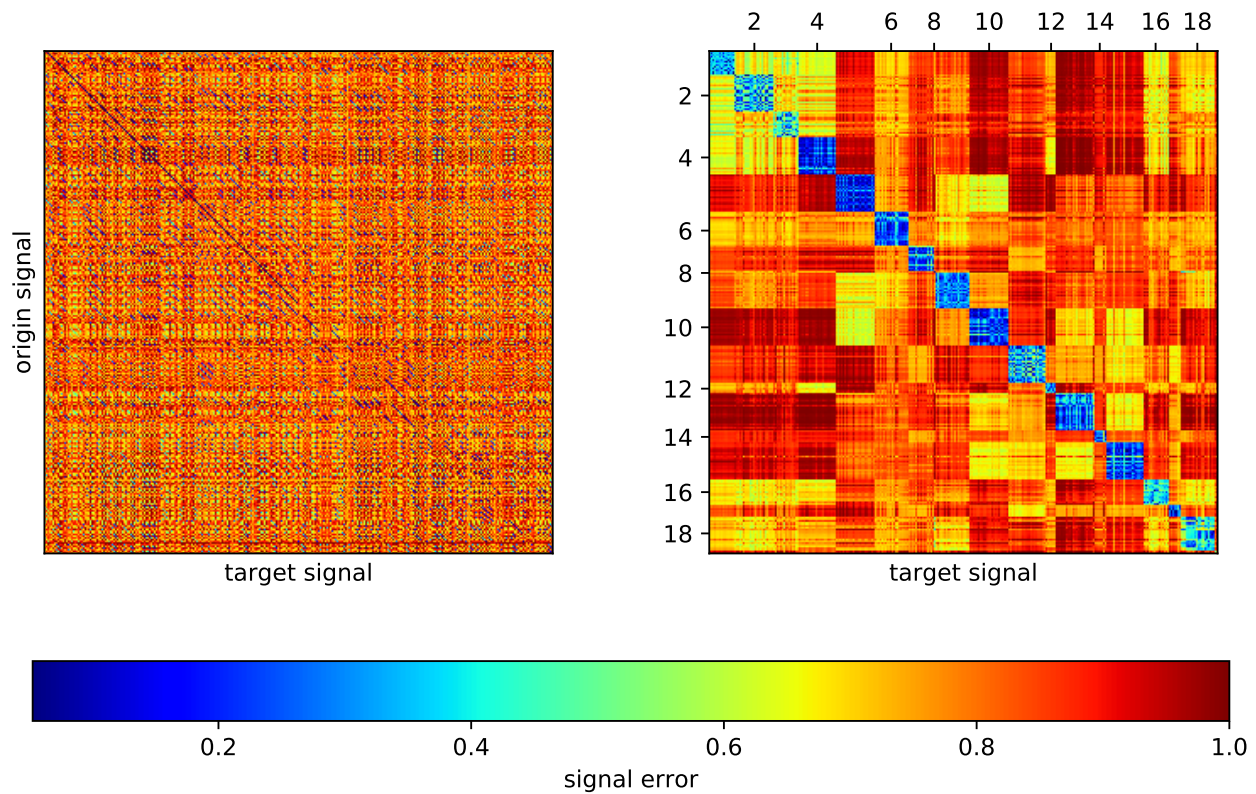


Figure 5: Pre sorting of EOD signals into temporal identities. A-C: Detected origin signals within a 30 s time window can be assigned to target signals with a maximum time lag of 10 s. Signals are assigned to each other based on their respective error values. The resulting temporal identity traces are formed based on the lowest possible error values. D: Error values of possible origin signals and target signals sorted by their temporal occurrence. E: Error values of possible origin signals and target signals sorted by their temporal identity. Only every second identity is displayed indicated for plotting reasons.

d-EODf and d-Field of same identity signals vs. non-same identity signals

To determine the utility of the used tracking parameters and, thus, our tracking algorithm we performed post-hoc analysis on the assigned EOD frequency traces after visual examination. The EOD frequency errors within subjects within the compare range of 10 s is significantly lower than EOD frequency errors between subjects (Abb.7). However, the events of EOD frequency rises cause an overlap of EOD frequency errors of within subjects signal comparison and between subject signal comparison.

Assessing field errors, within subject signals comparisons also resulted in significantly lower Field error values compared to between subject Field errors. However, the distribution of within subject errors and between subject field errors did not differ as much as the respective distributions of EOD frequency errors. Interestingly, within subject field errors stayed unaltered low in comparison to between subject field errors within the 10 s compare window.

Since EOD frequency error like Field error within subjects are mostly independent from time we could exclude the time factor from the tracking algorithm.

Roc analysis

To further validate the use of the tracking parameters we carried out ROC analysis evaluating the discriminability of within and between subject field errors, EOD frequency errors and total errors. AUC values decreased for all

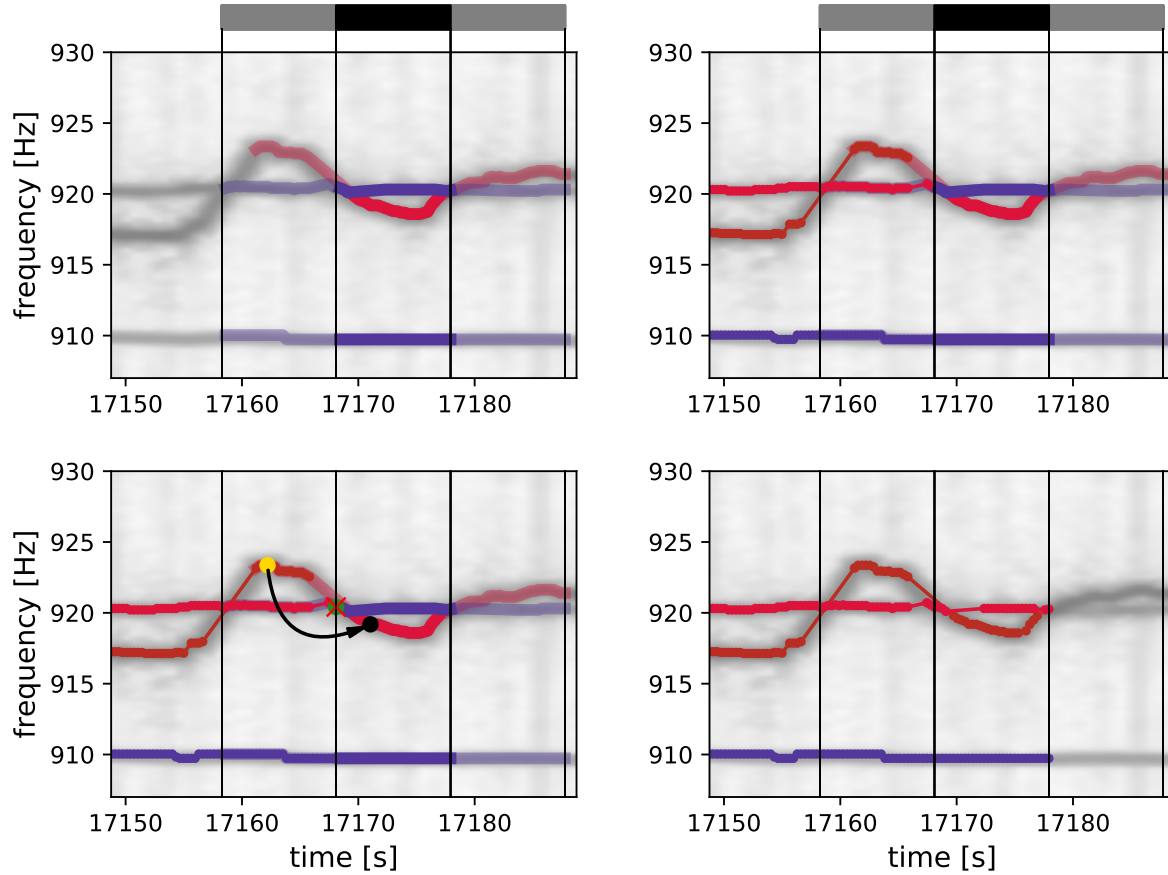


Figure 6: Running connection. A: Temporal identities assigned via lowest error value connections. The black and grey bars together indicate the whole data snippet temporal identities have been assigned before. Connections within the grey areas are assumed to be unsteady since signals within compare range but outside the gray are could point to those signals as well but are neglected in this particular temporal identity assignment. Connections within the black bar area indicate valid connections since every signal possible connected to those signals lie within the temporal assignment range. B: Already assigned identities, i.e. potential connection partners for the temporal identities to assigne, are shown. C: The yellow and green signals (already discussed in fig. 1 and fig. 2) represent the origin signals of the respective identities to best fit to a signal of one temporal identity, indicated in black. The yellow-to-black error is lower than the green-to-black error. D: Signals of temporal identities got assigned to signals of already assigned identities based on their lowest error value to each other.

error with increasing time lag. However, AUC values for all errors stayed on a very high level. Interestingly, while AUC values concerning field error and EOD frequency error dropped after a certain time lag to a base level, the AUC value for the combined total error stayed comparatively high and enables us to reject the usage of time lag as additional parameter for the calculation of the total error.

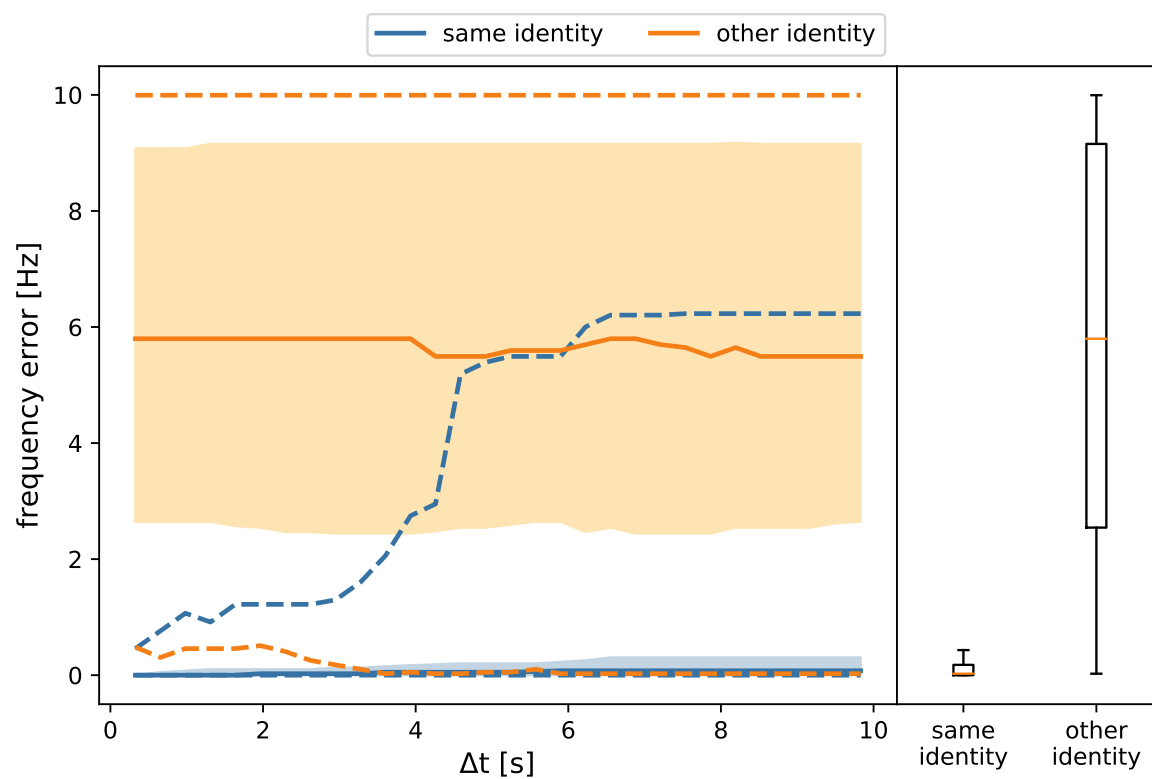


Figure 7: Random caption

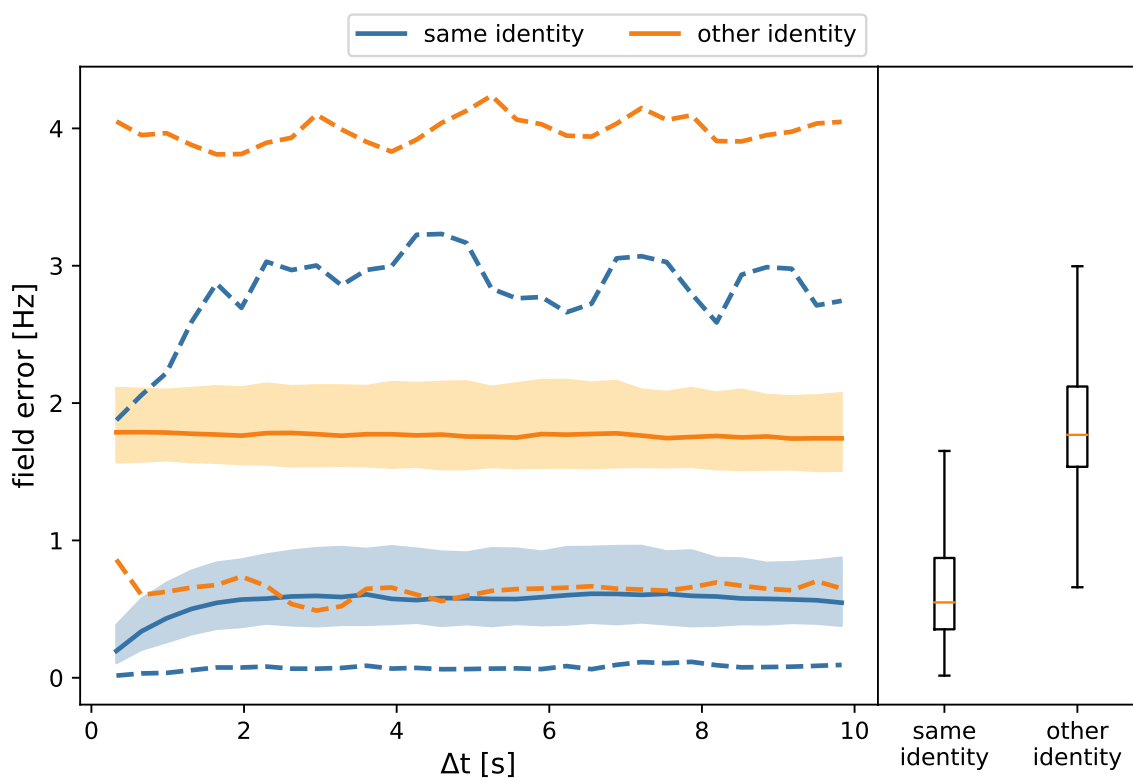


Figure 8: Random caption

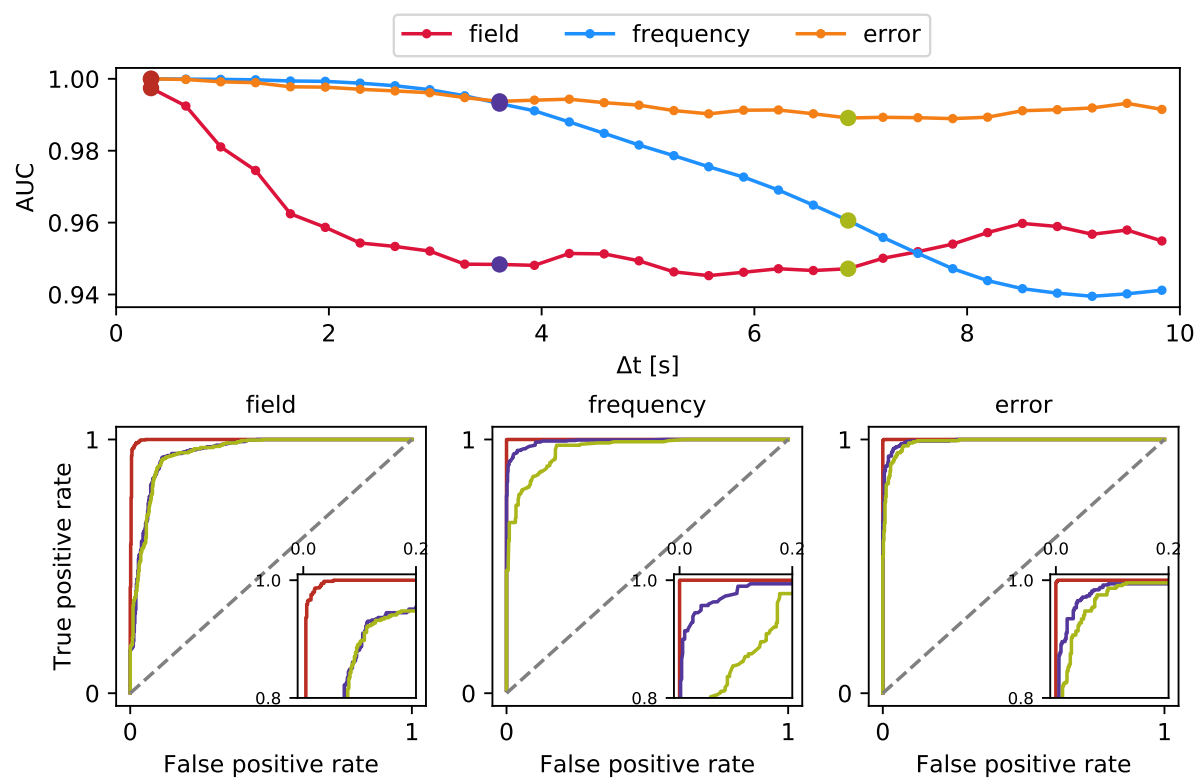


Figure 9: Random caption

References

- Henninger, J., Krahe, R., Kirschbaum, F., Grewe, J., and Benda, J. (2018). Statistics of natural communication signals observed in the wild identify important yet neglected stimulus regimes in weakly electric fish. *Journal of Neuroscience*.
- Todd, B. S. and Andrews, D. C. (1999). The identification of peaks in physiological signals. *Computers and Biomedical Research*, 32(4):322 – 335.