Unravelling active sensing in groups with the *Ushichka* dataset: Supplementary Information

Thejasvi Beleyur, Holger R Goerlitz

Document updated: 2021-01-02

## Audio-video synchronisation and recording

The audio-video recording was run by a callback function that repeatedly collected input audio and output signals to and from the soundcards in the form of data buffers (1024-2048 samples). The input buffer consisted of array, while the output buffer consisted of a array. When a recording was triggered, all input buffer data were concatenated and saved into an audio file. When not triggered, the input buffer data was only used to measure audio level. The output buffer consisted of three channels. The first ouptut channel was a constantly output 25Hz square wave sent to the thermal camera control box. This 25Hz square wave was the common signal which synchronised the frame capture of all three thermal cameras. The signals in the second and third channels were dependent on whether a recording had been triggered or not. When a recording was not triggered, only zeros were sent to the second and third channel’s data buffer. When a recording was triggered, the second channel’s data consisted of a 20 kHz sine wave. The high-frequency sine wave triggered the simultaneous video recording in all three thermal cameras, and thus ensured frame-level synchronisation in the video files. On triggering, the third channel’s data buffer was an attenuated version of the square wave signal being sent in channel 1. This attenuated square wave served as an inter-soundcard synchronisation signal. Despite the use of the inbuilt Word clock signal from the master sound card, the audio data received across devices was observed to be jittered by a varying number of samples in each recording. The audio data from within a single sound card was however verified to be sampled synchronously. This thus required the use of a inter-sound card synchronisation channel for each sound card added to the system . frame capture.

The automated audio-video recording triggering was controlled by the fieldrecorder\_trigger module.

## Microphone calibration

### Array position self-calibration

To automatically estimate the positions of the microphones placed all around the cave speaker playbacks were performed after each recording session as described in (Batstone et al. 2019). Briefly, a speaker (Peerless XT25SC90-04) with a reusable heating pad attached to it was moved through the recording volume. The playback consisted of frequency modulated sweeps between 16-96kHz of various types (linear, hyperbolic, V-shaped). The sweeps were played back at a constant interval of one second (Figure 1). The signal was generated from the primary sound card in use (Fireface UC/802), and amplified through a car audio amplifier (Basetech AP-2100). Playbacks were only done past 6:00 am to prevent disturbing the resident bat population. Moreover, at this time of the day, there are almost no bats flying in the recording volume.

The speaker-playback method has yielded successful automatic microphone position estimations as documented in more detail in (Batstone et al. 2019). (Batstone et al. 2019) shows a successfully estimated example for the microphone positions one one night. The microphone positions for the remaining nights still remain to be estimated. The advantage of having a constant time interval between consecutive playbacks is that the number of parameters estimated reduces, and thus eases the self-calibration estimation.

Additional audible speaker playbacks were also performed from a small portable speaker. However, this data does not have a constant offset, and may thus be handled by the broader set of parameter estimations described in (Zhayida et al. 2016).

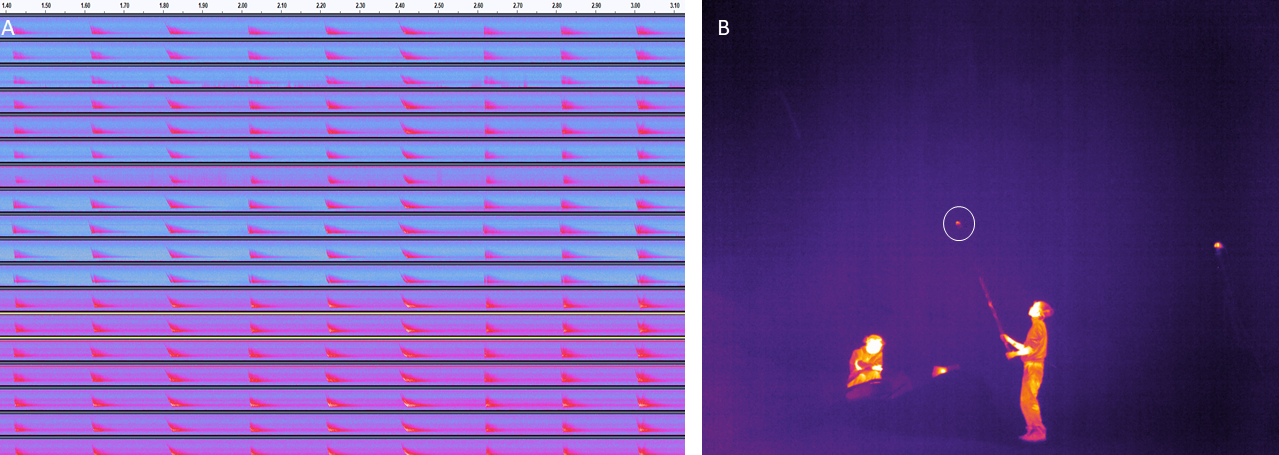


Figure 1: Data and methods behind the microphone self-calibration method used in Batstone et al. 2019. A)Spectrogram of a few of the constant interval sweeps played back from a speaker moved around the recording volume. B) An image taken from the synchronised video recording captured during speaker playback. The white circle highlights the position of the speaker.The speaker was attached to a rod was moved in such a manner that playbacks were recorded from all over the recording volume. The speaker’s position was made visible on the thermal video by a warming pad attached behind it, and thus also allows 3D video tracking of the speaker to be done if necessary.

### Manual inter-mic measurements

On all nights, manual measurements of a subset of all possible inter-mic distances were also performed with a laser range finder (1mm accuracy, Bosch GLM 50C). The issue with manual measurements is that they can be time-consuming, especially when the number of microphones is large, and moreover they can be error-prone. Especially when distances between microphones were more than around 4-5 metres, shaking hands caused inaccuracies in pointing. Another issue with manual measurements is that it requires a direct line-of-sight between a pair of microphones. Automatic position self-calibration is likely to be less affected by the issues that plague manual measurements, but requires more technical equipment and know-how. The algorithms in (Batstone et al. 2019) use randomly chosen estimates to seed the position estimation process. The manual inter-mic measurements can be fed into the estimation algorithms to yield higher accuracy position estimates.

### Microphone frequency response and directionality calibration

### Calibration runs and details

Each microphone was calibrated by attaching it to a cardboard sheet to mimic the cave wall. This allowed quantification of the frequency +and directionality response as the microphones would actually be recording in the cave.

The playback signal was a combination of pure tones (10-95 kHz, in 1 kHz steps) and a 10 ms long Tukey-windowed linear sweeps from 15-95 kHz. The linear sweep was repeated five times. . The same playback was recorded at a variety of azimuth and elevation angles. Only one side (left/right or up/down) was recorded as symmetry was assumed across the azimuth and elevation. The calibration exercise remains to be completed for all microphones in the recording system (as of 2020-10-27).

#### Experiment dates:

1. 2018-05-16 : Most probably done in Seewiesen. All SANKEN mics calibration run with on-axis recordings. GRAS microphone recording also made for reference.
2. 2018-06-04 and 2018-06-05: Calibration experiment attempt (incomplete). Calibration experiments aborted because of variation in the microphone stand and inconsistency in rotation. If TB remembers correctly, the self-made L-stand could not be fixed properly and did not return to the same position after a series of rotations. This was the reason only a few mics were recorded, and then aborted because the flight room had to be cleared up for either Toni or Laura the next day!
3. 2018-07-16: Complete run of all microphones present in microphone array. Microphone-to-speaker distance was 1.12m. Experiment was done in the larger flight room if TB remembers correctly. No speaker compensation was applied. Microphones calibrated were SANKEN9, SANKEN10, SANKEN11, SANKEN12, SMP1, SMP2, SMP3, SMP3C, SMP4, SMP5, SMP6, SMP7C, SMP8.

## Camera calibration

The three thermal cameras were calibrated using the ‘wand’ procedure described in . Two halogen lamps (G4, 20W, 12V) were attached to a wooden stick and connected to a 12V DC car battery. This formed the ‘thermal wand’. The small size of the halogen lamps meant that the lamps appeared as points on the video and were easy to localise in the thermal camera data. The wand was moved systematically within the recording volume to ensure all parts where bats flew were covered.



Figure 2: Image showing the thermal wand used as the common calibration object. The two incandescent lights forming either end of the wand are encircled in red.

Preliminary Wand scores from calibrations done using the easyWand package gave a ‘wand score’ of , indicating a relatively good calibration. In addition, gravity calibration was also performed by throwing a reusable click-activated heating pad. These reusable heating pads are also ideal for use in long-distance calibrations where the wand needs to have a slightly larger size so it is visible in the video.

### Additional video recordings

1. For every recording night, the positions of microphones were ‘pointed’ at with the thermal wand and these recordings can be used as a reference to align the acoustic and video tracking data. The microphone positions can also be used as background points for the Wand calibrations.
2. The speaker playbacks used in [Array position self-calibration](#array-position-self-calibration) were recorded simultaneously with audio and video (See Figure 1). The speaker positions can also be used as background points for Wand calibrations.

Each camera (K1,K2,K3) was placed in a relatively constant configuration, with K1 typically being placed close to the entrance, K2 in the middle, and K3 towards the side tunnel leading into the Rue de Paris.

### Digitisation and 3D tracking

The DLTdv5 package was used to digitise wand points and generate the 3D tracks.

### Camera intrinsic and extrinsic parameter estimation

The easyWand package provided the intrinsic and extrinsic parameters required to track bats in 3D.

## Acknowledgements

TB would like to thank Pranav Khandelwal for extensive help setting up the thermal camera calibration workflow and Hedrick Tyson for providing helpful feedback and intrinsic parameter data for the cameras.

## References

Batstone, K., G. Flood, T. Beleyur, V. Larsson, H. R. Goerlitz, M. Oskarsson, and K. Åström. 2019. “Robust Self-Calibration of Constant Offset Time-Difference-of-Arrival.” In *ICASSP 2019 - 2019 Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp)*, 4410–4.

Zhayida, Simayijiang, Simon Segerblom Rex, Yubin Kuang, Fredrik Andersson, and Kalle Åström. 2016. “An Automatic System for Acoustic Microphone Geometry Calibration Based on Minimal Solvers.” *CoRR* abs/1610.02392. <http://arxiv.org/abs/1610.02392>.