Supplementary Information: The *Ushichka* dataset: a multi-channel, multi-sensor dataset to unravel active sensing in groups

Thejasvi Beleyur, Holger R. Goerlitz

Document updated: 2021-08-11

# 1 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 an 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[[1]](#footnote-20). The audio data from within a single sound card was however verified to be sampled synchronously.

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

# 2 Automatic microphone position estimation

The Structure-From-Sound approach (Zhayida et al. 2016) to automatically estimate microphone positions relies on a common sound being recorded simultaneously on multiple channels. The time-differences-of-arrival between channel pairs are used to estimate the positions of the microphones. Zhayida et al. (2016) use a continuous sound source such as a song or speech signal, while we chose to use discrete signal playbacks.

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 an attached reusable heating pad was moved through the recording volume (Figure 2.1). The playback signals changed over the course of the field season. Playbacks from 19th June to 14th July 2018 used an upward swept hyperbolic sweep between 16-96 kHz, 8ms in duration played with a constant interval of one second. Playbacks from 21st June 2018 onwards used ‘multi-chirp’ signals. Multi-chirp playbacks consisted of a packet of 9 sweeps made by a combination of three time-frequency structures (linear, hyperbolic, V-shaped) and three durations (6, 12, 24 ms). Each sweep occured at the end of a 200ms segment, and thus one packet of 9 sweeps was 1.8s long. Multiple packets were played with a 1.6s inter-packet silence between them. Playback signals were generated from the sound card in use (Fireface UC or Fireface 802), and amplified through a car audio amplifier (Basetech AP-2100). Playbacks were only conducted past 6:00 am to prevent disturbing the resident bat population (at this time of day, there are almost no bats flying in the recording volume). See 6 for further details.

Batstone et al. (2019) showed succesful position estimation using discrete playbacks with ground-truthed microphone positions from one night in the Ushichka dataset. The microphone position estimations reported by Batstone et al. (2019) are within the range of 4 cm of ground-truthed positions. This error-range may perhaps be sufficient for acoustic tracking given that the bat-microphone distances are much larger (in the order of 0.5-5m).

Additional audible speaker playbacks were also performed from a small portable speaker. However, this data is yet to be analysed.



Figure 2.1: An image taken from the synchronised video recording captured during playbacks conducted for automatic mic position estimation. The red circle highlights the position of the speaker.The speaker was attached to the end of a rod. The rod was moved in such a manner that playbacks were recorded from all over the recording volume. The speaker was made visible on the thermal video by a warming pad attached behind it, and thus also allows 3D video tracking of its position.

# 3 Microphone frequency response and directionality calibration

Inital mic frequency response and directionality measurements were done on 11th January 2019 for a subset of 15 microphones (4 SANKEN CO-100 and 11 SMP Knowles 0410’s). The microphone was placed at a distance of 1.11 m from the speaker. Both microphone and speaker were aligned and placed on tripod stands away from the ground to avoid ground-reflections. The playback signal was a combination of pure tones (10-95 kHz, in 1 kHz steps, Tukey-windowed) and linear sweeps (10 ms, 15-95 kHz, Tukey-windowed). 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.

Each microphone was attached to a L-shaped holder that was centred around a rotation point. The tip of the SANKEN CO-100’s were set to be over the rotation point. The Knowles SPU 0410 microphones were attached to a cardboard sheet to mimic the effect of the cave wall. This allowed replication of the frequency and directionality response the microphones would have had in the cave. The cardboard sheet was aligned to coincide with the rotation point.

For reference, a calibration microphone (GRAS 40BF, GRAS Sound and Vibration, Denmark) attached to a pre-amplifier (GRAS Type 26AC) and amplifier (GRAS Type 12HF) was placed in the same location as the microphones under calibration and the calibration playbacks were recorded.

# 4 Camera calibration

The three thermal cameras were calibrated for their extrinsic (position and orientation) and intrinsic (focal length and distortion) parameters using the ‘wand’ procedure and ‘easyWand’ package described in Theriault et al. (2014). Two halogen lamps (G4, 20W, 12V) were attached to the ends of a wooden stick (lamp-to-lamp distance of 59cm) and connected in parallel to a 12V DC car battery, forming the ‘thermal wand.’ The small size of the hot halogen lamps meant that the lamps appeared as points on the video and were easy to localise in the video data. The wand was moved systematically within the recording volume to cover all parts where bats flew.



Figure 4.1: 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.

Wand points were digitised using the DLTdv package (Hedrick 2008) and preliminary Wand scores from calibrations done using the easyWand package gave a ‘wand score’ of , indicating a relatively good calibration. The Wand score indicates the variation in estimated distance between the two wand points across the calibrated volume, where a low score indicates a good calibration. In addition, gravity calibration was also performed by throwing a reusable click-activated heating pad.

# 5 Additional video recordings of relevance in *Ushichka*

Aside from bat flight and wand video recordings, a series of other video recordings were also performed. These video recordings serve to provide at least three functionalities. The first function is to align audio and video tracking data into a common co-ordinate system. The second function is that the video recordings provide ‘background points’ (sensu Theriault et al. (2014)) to assist camera parameter estimations. The third functionality is that the mic position recordings can be used to align the LiDAR and video coordinate systems. Since most of the microphones were placed on the walls of the cave, these positions represent points on the wall surface that can be matched to the LiDAR scan.

1. Mic positions : For every recording night, the positions of microphones were ‘pointed’ at with the thermal wand. These recordings are particularly useful to align the acoustic and video tracking data into a common coordinate system.
2. Speaker playback recordings: The speaker playbacks for mic position estimation described in 2 were recorded simultaneously with audio and video (See Figure 2.1). While these recordings could be used to align audio and video, they will be less accurate as the heating pad is placed at a slight distance from the speaker. The speaker playback recordings are thus better used to add ‘background points’ to improve camera parameter estimations.

# 6 Recording nights in the recording volume

Table 6.1: Table showing the date of recording session and type of playback. The multi-chirp sweeps consisted of linear, hyperbolic and V-shaped sweeps each with short,medium and long durations of 6,12,24ms length respectively. Sessions with manually triggered recordings were triggered by an experimenter listening for bat calls on a bat detector while watching the thermal camera video feed. Automatically triggered recordings were set to start when the RMS of an incoming audio buffer crossed a threshold value.

|  |  |  |  |
| --- | --- | --- | --- |
| Session Date | Playback signal type | Duration (ms) | Recordings triggered |
| 2018-06-19 | Hyperbolic | 8 | Manually |
| 2018-06-21 | Hyperbolic | 8 | Manually |
| 2018-06-22 | Hyperbolic | 8 | Automatically |
| 2018-07-14 | Hyperbolic | 8 | Automatically |
| 2018-07-21 | Multi-chirp | 6,12,24 | Automatically |
| 2018-07-25 | Multi-chirp | 6,12,24 | Automatically |
| 2018-07-28 | Multi-chirp | 6,12,24 | Automatically |
| 2018-08-14 | Multi-chirp | 6,12,24 | Automatically |
| 2018-07-28 | Multi-chirp | 6,12,24 | Automatically |
| 2018-08-02 | Multi-chirp | 6,12,24 | Automatically |
| 2018-08-14 | Multi-chirp | 6,12,24 | Automatically |
| 2018-08-17 | Multi-chirp | 6,12,24 | Automatically |
| 2018-08-18 | Multi-chirp | 6,12,24 | Automatically |
| 2018-08-19 | Multi-chirp | 6,12,24 | Automatically |

# 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)*, 4410–14.

Hedrick, Tyson L. 2008. “Software Techniques for Two-and Three-Dimensional Kinematic Measurements of Biological and Biomimetic Systems.” *Bioinspiration & Biomimetics* 3 (3): 034001.

Theriault, Diane H., Nathan W. Fuller, Brandon E. Jackson, Evan Bluhm, Dennis Evangelista, Zheng Wu, Margrit Betke, and Tyson L. Hedrick. 2014. “A protocol and calibration method for accurate multi-camera field videography.” *Journal of Experimental Biology* 217 (11): 1843–48. <https://doi.org/10.1242/jeb.100529>.

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>.

1. The jitter may have been introduced by impedance mismatch and internal reflections caused by the use of 50 Ohm BNC cables and T-splitters. Recent experiments with the same Fireface devices using 75 Ohm cables and an impedance matched signal splitter lead to constant synchronisation between multiple Fireface soundcards [↑](#footnote-ref-20)