Unfolding how low-level elements empower evolutionary successful animal communication systems: a new and promising interdisciplinary approach

Rachael Xi Cheng[1] (presenting author), Manuel Schmitt[1], Christian Bergler[1], Elmar Nöth[1], Simone Pika[2], Volker Barth[3], Andreas Maier[1]

Killer whales rely on a variety of acoustic signals to coordinate their group activities, as underwater visibility is limited. The presence and the identification of low-level elements within their high level and evolutionary successful communication system has remained a long-standing research question. So far, a common practice in bioacoustics has been to isolate and compare a large variety of complex vocalizations manually, with little support of automated methods. This manual approach restricts the ability to cross-compare large amounts of vocalizations to identify reappearing basic units and reappearing patterns as well as the underlying mechanisms. Recent developments of deep learning methods within the human language and computer vision domain provide new automated pattern recognition methods, which within this pilot study was adapted to killer-whale communication to investigate their highly functional, yet less convolved communication system.

To investigate how a large amount of calls are organized for specific communications, first a significant number of isolated calls needs to be extracted from continuous recordings which also contain a varying amount of noises. Therefore, Convolution Neural Networks were trained on Northern Resident Killer Whale (NRKW) vocalizations in recent pilot study by Bergler and colleagues (2019). The segmentation tool allows to automatically isolate and to extract all killer whale pulsed calls within recently collected field data and within 20,000 h of archived recording with 95% accuracy. The large amount of extracted calls lays a solid foundation for higher level pattern recognition tasks.

So far, relying on hearing and visual inspection of spectrograms, researchers have classified killer whale calls manually into 44 call types. Calls that appear to the human eyes and ears almost similar but at detail show variations have been classified into the same call class (overcategorization), due to limited analysis tools. Yet the variations in the calls might play an important role within the respective communication system. For example, in a tonal language like Mandarin, the pitch/intonation of a spoken sound affects the meaning. M*ā* (with a high tone) means mother while *mǎ* (starting with a low and dropping tone and then a rising one) means horse. To detect the smallest reappearing units and how are they organized to form higher-level processes, the current study adopted semi-supervised clustering methods to systematically identify the smallest reappearing units within the killer whale calls. Specifically, Convolutional Autoencoder, a deep neural network architecture, was pretrained to learn features that best represent the discrete calls of the NRKW. Spectral clustering algorithm was subsequently applied to cluster those segmented vocalizations. Currently, further clustering on those big clusters is in progress to identify the smallest reappearing units, and on a higher level the reappearing patterns of which calls are followed by which.

Moreover, to investigate how acoustic signals are used to coordinate group activities, how the context influences their social communication, and other higher-level processes of their communication system, array recordings that enable acoustic spatial recognition of “speakers” and video documentation of behavior were collected, together with contextual information. The automated speaker identification and subsequent analyses of dialogue structures is currently conducted on 270 h of recent field recordings.

This innovative combination of methods from BioCognition, machine learning, bioacoustics, and behavior research allows to investigate and to compare fundamental elements of evolutionary successful communication systems. All methods have been developed in a generic manner to allow later adaptation to and comparison with other highly vocal and social animal species.

---

[1] Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg (FAU), Erlangen
[2] Comparative BioCognition (CBC) Group, IKW, University of Osnabrück, Osnabrück, Germany
[3] Anthro-Media, Berlin