A proposal to use distributional models to analyse dolphin vocalisation

Mats Amundin[2], Robert Eklund[3], Henrik Hållsten, Jussi Karlgren[1,4], Lars Molinder[5]

[1]Gavagai, Sweden;
[2]Kolmården Wildlife Park, Sweden;
[3]Linköping University, Sweden;
[4]KTH, Sweden;
[5]Carnegie Investment Bank, Sweden

Abstract

This paper gives a brief introduction to the starting points of an experimental project to study dolphin communicative behaviour using distributional semantics, with methods implemented for the large scale study of human language.

1. Dolphin communication

Dolphins vocalise and communicate using complex signals of at least two kinds, whistles and clicks, produced in separate systems. Most dolphin species produce whistles. Whistles can last between tens of a millisecond to several seconds and consist of continuous, narrow-band, frequency-modulated signals. Most whistles can be found in the range of 2 to 20 kHz [1, 2]. Notable among whistles are the "signature whistles" which appear to be individually specific for each dolphin [3]. All dolphins also produce pulsed signals, or "clicks". These sounds are presumably used both for communication and sensing the environment. Typically the clicks come in trains with inter-click intervals ranging from few ms to several hundred ms and have most of its sound energy above the human hearing range [4, 2, 5].

2. More data and better tools

Both clicks and whistles have been studied in detail with respect to their acoustics, their relation to dolphin behaviour, and their occurrence patterns. Recent analyses have been able to describe dolphin whistle patterns using formalisms similar to those used to describe the morphological patterns of human language in terms of regularities in the way constituent elements form patterns [6, 7]. How the constituent elements of those patterns relate to each other, has not been formally described. Doing this will require much larger data sets than before: for example the most recent pattern mining experiments are performed on no more than 25 audio files.

Recent advances in computational hardware make possible the capture, storage, and analysis of analogue signals on a scale which was unthinkable even only a few years ago. Simultaneous advances in the in-memory analysis of streaming data make new processing models technically attainable. The wide availability of human linguistic data in speech and text form has made use of the technical possibilities to build unsupervised learning and dynamic on-line analysis models for inferring emerging semantic patterns in streaming data.

3. An opportunity for distributional models

Distributional analysis was first formulated by Zellig Harris [8] and such methods have gained tremendous interest due to the proliferation of large text streams and new data-oriented learning computational paradigms. Distributional semantic models collect observations of items from linguistic data and infer semantic similarity between linguistic items based on them. If linguistic items – e.g. the words $grid$ and $distributed$ – tend to cooccur – say, in the vicinity of the word $computation$ – then we can assume that their meanings are related. The primary relations of interest are *replaceability* and *combinability* of items [9]. Distributional analysis allows us to infer similarities between fundamental units, based on their observed occurrences in various patterns through the computation of second order cooccurrence relations: not only that a precedes x with some regularity, but that a and b both frequently occur with x, even if they never occur together.

4. Aims: a thesaurus of dolphin signals

While we in the current prestudy use methodology originally developed for the analysis of human language, we refrain from claiming that dolphins communicate in ways which are human-like. The task of our project is to find a representation of the signals vocalised by dolphins which allows us to infer usage similarities between identified recurring communicative tokens in dolphin communication. This aim involves a cascade of interconnected challenges.

The general task of making sense of continuous signals, assuming that they are of a sequential nature, involves three tasks: segmenting the signal into chunks of suitable level of abstraction; identifying similarities between such chunks across situations to recognise *fundamental units* of interest, corresponding to words or morphemes in human language; and then to identify patterns of occurrence among those fundamental units, corresponding to phrases or utterances in human language to be able to establish similarity of usage of such items. The result of such a procedure is a library of patterns and a thesaurus of items.

5. Challenges: the hermeneutic circle

individuation Dolphins vocalise without visible articulation [10, 11, 12]. Separating signals from a a number of vocalising individuals at the same time without knowing where the speech from one dolphin ends and another starts will be necessary, but is a known challenge in the field: *"...Identifying the vocalizers still remains one of the greatest challenges to the study of dolphin communication signals today"* [1, 13].

feature palette Humanly obvious acoustic features such as frequency and amplitude spectrograms become more complex as the interplay between the two communicative mechanisms of whistles and clicks are taken into account. Prosodic features such as pitch, quantity, stress or overlay between whistles and click bursts can be expected to communicatively relevant as well. The features of interest to identify segments from a continuous signal are manifold and involve temporal analysis of pauses and bursts, observable changes in dynamics or amplitude of frequency and harmonics, or observation of other contiguous action on the part of the vocaliser and potentially of its peers. Previous studies, have e.g. used a categorisation of context into play, foraging, aggression, and mother–calf interaction.

segmentation and phonetic similarity Most discovery algorithms in previous work on analysis of dolphin vocalisation have used distance-based approaches to segment signals into communicative tokens by firstly manual inspection of a transposed acoustic signal or a graphical rendition of its contours and later by computationally more convenient elastic matching of the same explicit surface signal.

directionality Directionality of sounds, especially the click sounds, is used by dolphins when they address social signals to specific conspecifics. [4] Directionality is difficult to establish, and cannot be captured at all using fixed hydrophones: it will require acoustic recordings devices that can be attached to the animals; this is not to be included in this study.

distributional similarity Once a signal has been segmented into communicative tokens and a cross-situational and cross-individual similarity measure has been defined, a distributional analysis will allow for models of similarity between tokens: "token A is used much like token B. This is the key to creating a thesaurus of communicative tokens, and the main challenge of our project.

situational factors Distributional semantic models are normally constrained to the analysis of occurrences and cooccurrences of linguistic items, but there is no conceptual need to limit the analysis to words or constructions: other contextual factors are quite reasonable candidates for inclusion in the computation. In this proposed project, factors such as the presence of stimuli of interest (e.g. food, play, humans, peers, threats) might well be used as distributional features. Enriching the model to handle context is a theoretical challenge for any distributional model.

signal and grounding Our basic assumptions are that dolphins emit and perceive sequences of fundamental items in their communicative patterns, that some of the vocalisation is intended for communication between individuals, and that dolphins are able to individuate the sounds they make to each other. Our assumption is that the communicative signal is largely sequential. This may be a risky assumption in view of the two communicative mechanisms and their interaction. Our somewhat daring assumption is also that there are segmentable communicative tokens in the signal and that those tokens are composed by combinations of separable features, much as phonemes are combined into syllables and words.

meaning Going to the heart of the entire effort, the question is what dolphins communicate about. While it is likely that some referential expressions can have shareable semantics across species, it is possible or even likely that much of dolphin–dolphin communication concerns states and aspects of dolphin life which are difficult to observe and may be near impossible for humans to conceptualise. Variation in the communicative signal may encode such content, similarly to how prosodic features are used in human–human communication. Our model will start from concrete events, observable by dolphins and humans alike, there is a risk of missing such salient variation from the signal that might refer to abstractions only accessible to dolphins. Studying the communicative behaviour of another species ranges between two theoretical extremes: On the one hand we can have a overly broad notion of what constitutes a language *everything is language*. We will then interpret every observed behavioural pattern of the studied species as a negotiation or dialog between the individual and its surroundings, including other individuals. On the other hand, if we hold to the narrowest notion of language *Only human-like communication behaviour is language* then we run the risk of finding nothing or only finding crude versions of human language. As an example, should the cheetah agonistic sound sequence moaning-growling-hissing-spitting, with "paw-hit" [14] be interpreted as four distinct signals, signalling four distinct and identifiable mental states, or simply as four different "modes" of one and the same escalating mental state?

Addressing these challenges must be iterated over in turn, since the results from one will inform the processing models in both preceding and subsequent ones. After signal segmentation, we will study both similarities between those tokens as well as differences between specific individuals' uses of those tokens. The results of these studies may well force us to revisit the way we segmented the signal. It is therefore important that we capture the signals in their entire frequency spectrum with a minimum of pre-study notions as to what the relevant range of frequencies are: if the dolphins can hear it, we intend to capture it.

6. Current state of the prestudy

We are currently recording dolphins at Kolmården with a fixed hydrophone set-up, and expect to start processing the data during this year. Results will be released both as data sets and as methods and algorithms for further application in other projects. Several of the results we expect are potentially extensible to other species as well; some of the results are contributions not only to our understanding of dolphins but to our general understanding of the capacity and limits of distributional modelling.

7. References

[1] D. L. Herzing, "Making sense of it all: Multimodal dolphin communication," *Dolphin Communication and Cognition: Past, Present, and Future*, 2015.

[2] M. O. Lammers and J. N. Oswald, "Analyzing the acoustic communication of dolphins," *Dolphin Communication and Cognition: Past, Present, and Future*, vol. 107, 2015.

[3] M. C. Caldwell, D. K. Caldwell, and P. L. Tyack, "Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin," *The bottlenose dolphin*, pp. 199–234, 1990.

[4] C. Blomqvist and M. Amundin, "High-frequency burst-pulse sounds in agonistic/aggressive interactions in bottlenose dolphins, Tursiops Truncatus," in *Echolocation in bats and dolphins*, Thomas, Moss, and Vater, Eds. University of Chicago, 2004.

[5] W. Au, *The sonar of dolphins*. Springer, New York, 1993.

[6] D. Kohlsdorf, C. Mason, D. Herzing, and T. Starner, "Probabilistic extraction and discovery of fundamental units in dolphin whistles," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 8242–8246.

[7] D. Kohlsdorf, D. Herzing, and T. Starner, "Feature learning and automatic segmentation for dolphin communication analysis," *Interspeech 2016*, pp. 2621–2625, 2016.

[8] Z. Harris, *Mathematical structures of language*. Interscience Publishers, 1968.

[9] M. Sahlgren, "The distributional hypothesis," *Italian Journal of Linguistics*, vol. 20, pp. 33–54, 2008.

[10] M. Amundin and S. Andersen, "Bony nares air pressure and nasal plug muscle activity during click production in the harbour porpoise, phocoena phocoena, and the bottlenosed dolphin, tursiops truncatus," *Journal of Experimental Biology*, vol. 105, no. 1, pp. 275–282, 1983.

[11] S. Ridgway, D. Carder, R. Green, A. Gaunt, S. Gaunt, and W. Evans, "Electromyographic and pressure events in the nasolaryngeal system of dolphins during sound production," in *Animal sonar systems*. Springer, 1980, pp. 239–249.

[12] T. W. Cranford, M. Amundin, and K. S. Norris, "Functional morphology and homology in the odontocete nasal complex: implications for sound generation," *Journal of Morphology*, vol. 228, no. 3, pp. 223–285, 1996.

[13] M. Hoffmann-Kuhnt, D. Herzing, A. Ho, and M. Chitre, "Whose line sound is it anyway? identifying the vocalizer on underwater video by localizing with a hydrophone array," *Animal Behavior and Cognition*, vol. 3, no. 4, pp. 288–298, 2016.

[14] R. Eklund, G. Peters, F. Weise, and F. Munro, "An acoustic analysis of agonistic sounds in wild cheetahs," in *FONETIK 2012. Gothenburg, Sweden, May 30–June 1, 2012*. University of Gothenburg, 2012, pp. 37–40.