Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR)

Edited by

Roger K. Moore[1], Serge Thill[2], and Ricard Marxer[3]

1   **University of Sheffield, GB, `r.k.moore@sheffield.ac.uk`**
2   **University of Skövde, SE, `serge.thill@his.se`**
3   **University of Sheffield, GB, `r.marxer@sheffield.ac.uk`**

Abstract

This seminar was held in late 2016 and brought together, for the first time, researchers studying vocal interaction in a variety of different domains covering communications between all possible combinations of humans, animals, and robots. While each of these sub-domains has extensive histories of research progress, there is much potential for cross-fertilisation that currently remains underexplored. This seminar aimed at bridging this gap. In this report, we present the nascent research field of VIHAR and the major outputs from our seminar in the form of prioritised open research questions, abstracts from stimulus talks given by prominent researchers in their respective fields, and open problem statements by all participants.

1   Executive Summary

Serge Thill
Ricard Marxer
Roger K. Moore

Almost all animals exploit vocal signals for a range of ecologically-motivated purposes. For example, predators may use vocal cues to detect their prey (and vice versa), and a variety of animals (such as birds, frogs, dogs, wolves, foxes, jackals, coyotes, etc.) use vocalisation to mark or defend their territory. Social animals (including human beings) also use vocalisation to express emotions, to establish social relations and to share information, and humans beings have extended this behaviour to a very high level of sophistication through the evolution of speech and language – a phenomenon that appears to be unique in the animal kingdom, but which shares many characteristics with the communication systems of other animals.

Also, recent years have seen important developments in a range of technologies relating to vocalisation. For example, systems have been created to analyse and playback animals calls, to investigate how vocal signalling might evolve in communicative agents, and to interact with users of spoken language technology (voice-based human-computer interaction using speech technologies such as automatic speech recognition and text-to-speech synthesis). Indeed, the

structures. This stands to reason then for vocal species, that there is some connection to how vocalizations are made as well as, why and what meaning the vocalizations have. Part of my research agenda is to better understand these connections.

- **Developmental differences within clades:** Within monophyletic groups, there is variation between different species vocal development patterns. While vocal production and comprehension is innate in some species, other species require a sensorimotor learning style and others require something in between. I am developing methods to make inferences of vocal complexity based upon a pre-existing understanding of how various species learn to meaningfully vocalize. My goal in this pursuit is to determine what environmental and genetic factors are important for developing a more complex way of communicating vocally.

- **Potential for linguistic structure:** Cognitive studies of animals have provided some insight into what certain species are capable of. My work strives to further this understanding by viewing this problem from a cross-disciplinary approach. Instead of focusing on making a direct connection to human language, I first examine what meaningful connections individuals within single species are making with conspecifics. Once these connections are explored, I then expand my view of the interactions to include individuals from other species that may come into regular contact with my focal species. My goal in this approach is to first understand what meaningful communication may be going on within a community with which the species has coevolved in before making a larger leap towards how that may relate to us.

5.8 Statement by Robert Eklund

Robert Eklund (Linköping University, SE)

Personal statement

Given a background in Speech Technology (I worked on the first concatenative speech synthesizer for Swedish, the first commercial ASR system for Swedish (now Nuance) and the first open prompt human–computer support system in Scandinavia (Telia 90 200) it has, for a long time been "natural" for me to think in terms of interaction, and concepts like agents, avatars, Theory of Mind and interface design (auditory and visual) have all been part of parcel of my work activities during the period 1994 to (roughly) 2012.

For completely unrelated reasons I started to expand my research interests into animal vocalizations in the year 2009 when I made a recording of a cheetah purring [1] and these activities did then snowball into a five-year-long project where me and colleagues will study human–cat interaction with focus on prosody/melodic aspects [2].

My Stimulus Talk during the Dagstuhl conference did not focus on or describe my previous research on the topic (cheetah, lion and domestic cat vocalizations) but instead raised some "larger issues" concerning "cross-species" (with a wide definition of 'species', including robots) communication.

These will shortly be described below:

---

[1] http://www.youtube.com/watch?v=ZFvULxbN3NM
[2] http://meowsic.info

Personality issues

The literature is replete with studies of personality (and was crucial in e.g. how to put together submarine crews during WWII). However, such studies are not constrained to human but several studies of personality in different species of felids are also to be found (see Bibliography), partly for husbandry reasons. My issue-to-raise here is to what extent individual personalities play a role when humans interact with other species.

New form of "uncanny valley"?

In 1970 Masahiro Mori published a paper title "The Uncanny Valley" (in Japanese translation) [1] where he described a dip in the easiness with which we approach and regard humanoids. If these are completely not like us (like 1930s teddy bears or cartoon characters) we have no problem, which is also the case if there are very similar to us. However, if something is "eerily" similar to us – not completely not like us, but not completely not like us, either – we get a spooky feeling around them. My question here is whether this can occur in the auditory domain, too. If computers sound very much like machines, or whether animals respond to or signal to us, in ways that are definitely not human-like, we (obviously) have no problem. But what happens when either robots or animals start communicate with us in very human-like manners – both voice-quality and content-wise: will this created another/a new form of more abstract uncanny valley?

Was Wittgenstein right?

Wittgenstein famously stated that "if a lion could speak we would not understand him". This obviously played on the idea that the lion world is so basically different from the human world that there is no way that we could understand the lion's worldview. (Note that this argument has also been forwarded within anthropology when studying other – most often non-Western cultures.) But is this necessarily true? Although undeniably true that a lot has happened since humans lived "on the savannah", we still most likely share the same basic emotions, and are governed by them. This should, in my view, provide some solid common ground for mutual understanding.

Health effects?

To spend time with a pet, or even robots, is beneficial from a health perspective. Will this effect be enhanced by improved communication with animals or robots? Or will a potential new uncanny valley effect reverse this?

Symbol mapping?

The cheetah is particularly famous for its agonistic moan–growl–hiss–spit+paw hit sequence[3] (most felids exhibit this, minus the paw hit). How to interpret this sequence? As one agonistic sequence that qualitatively changes character as it escalates, or as four different "symbols", all with their own intrinsic meaning? The basic question is: to what extent can we use the standard linguistic toolbox when we describe animal vocalizations?

---

[3] http://www.youtube.com/watch?v=bBIf5g2Fp1U&feature=youtu.be

Language learning?

That several species of animals are capable of language learning – and consequently also dialectal variation – has been known since Aristotle [3]. What can we learn about our own acquisition of language, phylogenetically, from the study of language learning in animals?

Role of hearing?

Animals vary a lot when it comes to hearing abilities, both frequency-wise and from a source location point of view (see Bibliography below). To what extent do we need to take other species' hearing abilities into account when trying to communicate across species? Case in point: the Beluga whale described in [2] who deliberately made an effort to vocalize outside its comfort zone when addressing humans.

Motherese?

It is well-known that humans – at least in the western world – make use of what is sometimes called "motherese" when they address infants (or small children). This speech style is characterized by en exaggerated prosody and simplified phone and word repertories. It is also known that humans use the same "trick" when addressing their pet animals. Does this have any benefits on the animal side of things, or is it simply something that we do semi-automatically for our own benefit?

Summing it all up

There are, obviously, loads of things to consider when expanding our knowledge on how animals communicate, and on how we as humans can improve our communication with those animals. Although not exactly the same, there is considerable overlap in our communication with robots (and animated agents and/or avatars) and there is no doubt in my mind that there will be vast cross-fertilization between all those fields in the future. And I hope to be part of this!

Web resources

- http://roberteklund.info
- http://ingressivespeech.info
- http://purring.info
- http://meowsic.info

References

1   Mori, Masahiro, *The Uncanny Valley.* Energy vol 7, no 4, 33–35, 1970.
2   Ridgway, Sam, Donald Carders, Michelle Jeffries & Mark Todds, *Spontaneous human speech mimicry by a cetacean.* Current Biology, vol 22, no 20, R860–R861, 2012.
3   Zirin, Ronald A., *Aristotle's Biology of Language.* Transactions of the American Philological Association, vol 110, 325–347, 1980.