A NEW ARTIFICIAL SIGNAL-SPACE PROXY FOR INVESTIGATING THE EMERGENCE OF STRUCTURE AND CATEGORIES IN SPEECH

Hannah Little, Kerem Eryılmaz & Bart de Boer

Artificial Intelligence Laboratory, Vrije Universiteit Brussel
hannah@ai.vub.ac.be

ABSTRACT

In order to experimentally investigate emergence of structure in speech, one needs a non-discretised signal space that is removed from participants' existing linguistic experience to prevent interference. We present a novel approach that makes use of an infrared sensor device. Participants generate signals using their hands in relation to the device, which generates audio feedback. The signalling space can be manipulated to have different sizes and to make use of different dimensions (e.g. distance to the device, or position with respect to it), and the nature of the audio feedback can also be manipulated. This paper will give a brief review of previous methods used as proxies for signalling spaces and outline why further innovation is required. We will also describe the new approach, drawing on examples, and outline future applications and possibilities within the field of phonetics and phonology, specifically in relation to the emergence of structure.

Keywords: Signal Space Proxies, Artificial Languages, Structure Emergence, Language Evolution

1. INTRODUCTION

In recent years, in the field of language evolution, there has been a trend to use human participants in experiments in the lab involving artificial language learning, cultural transmission and/or communication. These experiments first focussed on exploring the emergence of structure on a morphosyntactic level using signals composed from already discretised signals (e.g. [6]), however, there is a growing body of work which is interested in the emergence of combinatorial structure, or phonological structure, and phonemic categories. Some work has used experimental tasks using existing spoken language to explore mechanisms behind language change (e.g. [12]). However, to focus specifically on emergence, this work necessarily has to use artificial proxies for signalling or articulation spaces, in order to prevent interference from people's pre-existing linguistic knowledge and linguistic modality. The space must also be continuous, in order to observe the emergence of discrete building blocks from the continuous signal. Several proxies have already been used in order to investigate the emergence of phonological patterning in the laboratory, including slide whistles and computational auditory and/or graphical interfaces. Here we introduce a new approach that, as we will argue below, is more flexible than previous approaches. So far in our research, we have used our approach to investigate how the topology (i.e. size and dimensionality) of a signalling space, and the way signals map to a meaning space, can affect the emergence of structure, and the nature of the emerged structure. How the topology and dynamics of a signal space affect the emergence of structure can inform theories of how linguistic modality plays a role in phonetic and phonological evolution. It is important to understand the effects that linguistic modalities, and the proxies used for linguistic modalities, have on the emergence of phonetic and phonological structure, before attributing emerged patterns to cognitive effects. After summarising our current work, we will outline the approach's future potential and present ideas for further study.

2. EXISTING SIGNAL-SPACE PROXIES

The ideal signal-space proxy to investigate the emergence of combinatorial categories and structure is one which prevents interference from pre-existing linguistic knowledge, whilst having a continuous space from which discrete elements can emerge. It is also desirable for proxies to be manipulatable in their size and dimensionality, in order to identify artifacts in any emerging stucture from the physical aspects of a particular proxy being used. Also, the ease with which structure can be measured within the signals produced is highly desirable.

2.1. Graphical Interfaces

Experiments which use graphical interfaces to investigate the emergence of linguistic structure become difficult to design, as participants are very familiar with presenting content graphically, both using written language and creating iconic representations via drawing. This familiarity prevents good experimental conditions to demonstrate how structure might emerge in the absence of influence from pre-existing knowledge, whether this is linguistic or otherwise. Further to this, graphical interfaces make iconic representations very easy to achieve, which will undoubtedly affect the emergence of any structure. In order to combat some of these issues, Galantucci [4] developed an approach which used a graphical interface, but had constraints on what participants could do using the apparatus. The interface was a stylus which has a constant downwards drift so it was impossible for participants to create images in a way that is intuitive. Using this approach, Galantucci has conducted social coordination experiments to look at how communication systems emerge [4], and more specific experiments have been done which have looked at how duality of patterning might have emerged [8], how iconicity affects the emergence of structure in signals [9], and there have also been experiments which have used the apparatus in an iterated learning paradigm, where participants' outputs were fed to a new participant in transmission chains [2]. All of these experiments have made a contribution to our understanding of the emergence of structure under different pressures. However, it is very difficult to quantify structure within this approach, and results are often based on measures for conventionalisation rather than structure. It is also difficult to manipulate the topology or dynamics of the signal space, causing researchers to manipulate the meaning space, rather than the signal space, to affect mappability [9]. Though one study did manipulate the rapidity of fading on signals using this proxy [5].

2.2. Whistles

Tessa Verhoef has done several experiments which use slide whistles as a proxy for an articulation space (e.g. [11] and [10]). Her experiments involve participants learning pre-recorded whistles and trying to reproduce them from memory. This has been implemented in an iterated learning experiment to see if learning biases within transmission chains could influence the emergence of structure within an inventory of whistles. Verhoef has implemented this in conditions with meanings [10] and without meanings [11]. This approach is a fun and ingenious way to get participants to learn and reproduce signals, and the results have clearly shown that within transmission chains, structure emerges and signals more learnable. However, the output from these experiments is acoustic and the relationship between stopper-movement and signal-pitch is not linear, which means that analysing signals is not a straightforward process. Also, the signal space is difficult to manipulate in its shape. An attempt has been made to affect the size of the signal space using slide whistles [7], thought the results from these attempts are difficult to measure and interpret.

2.3. Digital Signals

Since Verhoef's work with slide whistles, a computational alternative has been developed which can work with a mouse on a screen, or via touch pads on tablets. One experiment used tablets to conduct social coordination tasks within larger iterated learning experiments where participants created signals by placing their finger on a digital slide whistle on a tablet [3]. This experiment explored whether some meanings being more easily mappable than others would facilitate communicative success. This has undoubtedly solved the problem of having data which can be analysed straight away (indeed, the researchers analysed some of the data as it was being produced in real time), however, this system is still only operating with one signal dimension (pitch), and so analysis of signal-spaces with more complex topologies is still an area which needs to be explored.

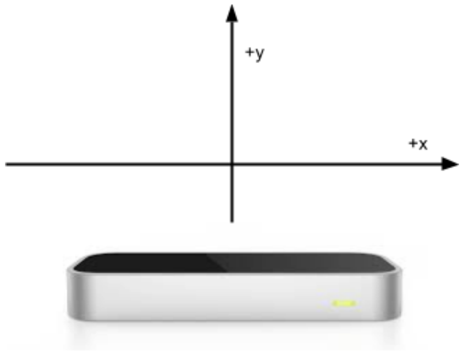
3. OUR PROXY

Our approach uses a "Leap Motion" device (Fig. 1), which is an infrared sensor that detects hand placement in the space above it. Signals can be produced within either a 1, 2 or 3 dimensional physical space, which manipulates auditory feedback, either in pitch, volume, or in any other way the experimenter can program. Participants interact with a user interface, which gives them full instructions on how to use the device throughout, and time to get familiar with the device, and how to create a mapping between hand position and the auditory feedback. Participants are to use one hand at a time, and they must sit back from the device, so that their upper body does not interfere with the signal.

3.1. Our experiments

In our signal-creation experiments so far, signals were generated which either differed in pitch, volume, or both. When a signal could be altered in both its pitch and volume, participants achieved this by moving one hand within an associated two dimensional signal space, e.g. a participant moving their hand up and down might alter the volume, while moving their hand left and right might manipulate

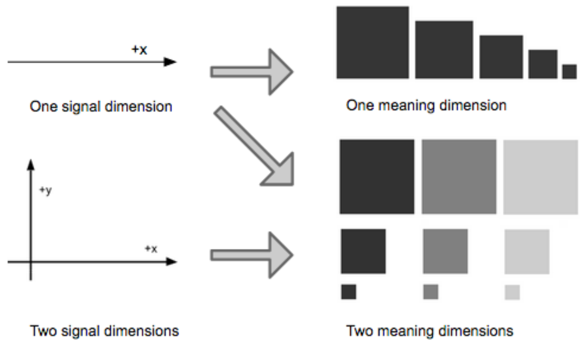
Figure 1: The leap motion, illustrating the 2 dimensional signalling space we used in our experiments.



the pitch. Participants could hear auditory feedback through headphones but had no visual feedback, except for being able to see their hand position in front of them. Within our experiment, we manipulated the dimensionality of this signal space, so participants could either manipulate signals by moving their hand within a horizontal dimension (x), vertical dimension (y) or both (Fig. 1).

Participants were given a meaning space, and were asked to generate signals for each meaning. The meaning spaces also had different numbers of dimensions. So, for example, a set of squares might only differ in size, or in shades of grey, or both (Fig. 2). In our experiments, participants are subjected to different signal creation phases where the signal dimensionality matched with the meaning dimensionality, or mismatched, i.e. they had to describe a two dimensional meaning space with only a one dimensional signal space (Fig. 2). This experimental design aims to create a proxy for the situation we see in linguistic modalities (gesture or speech) in the real world. Linguistic modalities (or signal spaces) differ drastically in topology, and some signal space topologies may be more similar to the structure of objects and events in the world. This similarity will facilitate straightforward signal-meaning mappings and iconicity in linguistic systems where it is possible, and perhaps facilitate discretisation and structure where it is not possible. Experiments such as ours aim to reduce the complexity of the structure of a signal space, and the complexity of the structure of the world, in order to identify the effects of the physical aspects of modalities and to understand how we can extrapolate the results from our experiments to systems emerging in different modalities. This is something which has not been explored with other experimental signal-space proxies, but which is paramount to the discussion related to experimental studies looking at the emergence of phonological patterning in linguistic systems. We must understand the effects which the topology and dynamics of a signal space has, before we can start attributing observed effects to more cognitive mechanisms.

Figure 2: Dimensionality of the signal space and meaning space either matches or mismatches.



4. PRODUCTION AND PERCEPTION

When humans acquire language, they must solve the problem of how the shapes they make with their tongue and vocal tract map on to the phonemes of the language they are acquiring. Participants in artificial language learning experiments must overcome the same mapping problem, especially when an unfamiliar signal space is used. Participants must map between the articulation space (what they are physically manipulating), the signalling space (the audio signals being produced) and the perceptual space (what they are perceiving in their brain). It is important for instance, in Verhoef's experiments ([10], [11]), using slide whistles, to keep a distinction between the absolute position of where the plunger is on the slide whistle, the non-linear pitch of the whistle made as a result of this, and participants' perception of that whistle. Using our approach, we could have any mapping we wanted between the signalling space and the nature of the feedback given. Participant feedback in pilots indicated that people could much more intuitively manipulate non-linear scales. However, despite this, the output data from our approach was the absolute hand position of participants (represented as coordinates), rather than the non-linear pitch or volume values. We have the values for hand position in our log files, so it is straightforward to attain values pertaining to pitch and volume using the same algorithms we used to generate the signals the participants heard.

5. MEASURING STRUCTURE

With continuous non-discretised data, measuring structure is very difficult. In previous studies, measures of structure have ranged from entropy measures, to manual discretisation and bigram counts. The output data from our experimental setup is a long list of coordinates from each signal created. We can use this data to generate standard descriptive statistics, such as the standard deviation of coordinates, which might give us some basic intuition about the amount of movement within a signal, and also things like the mean coordinates for each signal trajectory and the duration of the signal, which were used to see how one participant discretised the space in figure 3. We can also create probability values of individual signals, for within subject analysis, using values from a participant's entire signal repertoire. However, whether signals being more predictable makes them more "structured" is an open question in the literature.

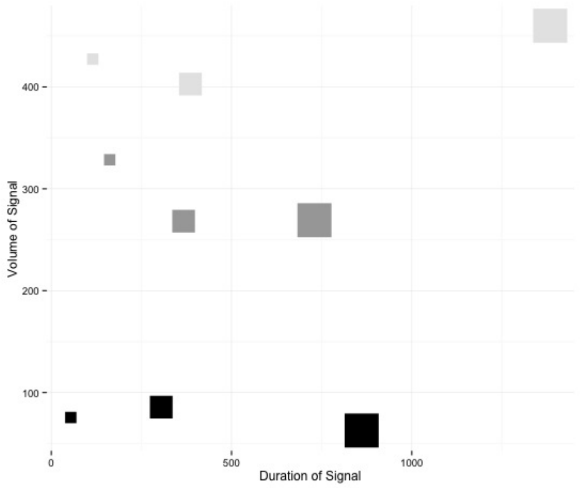


Figure 3: Mean coordinate on the volume axis for each signal plotted against the number of data frames for each signal. Size and shade of the square represented by each signal are represented by the size and shade of the points on the graph.

Further to the above, we have also explored more sophisticated methods for measuring and modelling structure, and have developed a method using Hidden Markov Models (HMM) to model structure by how many "states" are required in order to predict signal recognition success. We focus on modelling the individual signal generating apparatus in the form of a finite-state, multivariate HMM with continuous (Gaussian) emissions. HMM states and their emission distributions correspond to areas in the signal space that are regularly visited in a predictable context of preceding and following states. This is a simple analogue for the underlying phonological structure of a surface form. By training HMMs on signal repertoires, we can estimate the most likely vocabulary of states, or the most likely "phonological" alphabet.

Once such a model is trained, it can be used, in combination with the signal repertoire and the meaning space, to index iconicity in the repertoire. Iconicity suggests that there is a topological mapping between the signal and meaning spaces [1], i.e. two meanings that are similar to one another will also have similar signals denoting them. To measure iconicity within our results, we can use Viterbi paths produced by the HMM to reduce the signals to a discrete sequence of integers representing the emitting states. Then, a pair-wise distance matrix can be generated for the signal repertoire, using Levenshtein distances as the distance metric. Then, the meanings the repertoire maps to can also put into a pair-wise distance matrix among themselves. Finally, Mantel's test of matrix correlation can be run between the two distance matrices to obtain an index of how iconic the repertoire is, given the corresponding meaning space, in the form of a correlation coefficient.

6. CONCLUSION AND FUTURE WORK

We have presented a new experimental approach to generating signals for investigating the emergence of structure from a continuous signalling space. This approach can be used in experiments designed to investigate the effects of physical signalling constraints, signal-meaning mappings and cognitive mechanisms, as well as communication and cultural transmission.

With this new approach, a body of work is in development which explores issues, such as manipulating the dimensionality of the signal space or the nature of the auditory feedback, which couldnât be investigated using other approaches developed so far. Further to this, new methods for measuring structure within continuous signal streams are being developed, the usage for which may stretch far beyond analysing the signals from these experiments.

We plan to further utilise our approach in future signal creation experiments, social coordination tasks, and iterated learning experiments. These experiments will contribute to understanding and separating the physical and cognitive mechanisms involved in how structure emerges from a continuous space, both in speech, and in other modalities.

7. REFERENCES

[1] De Boer, B., Verhoef, T. 2012. Language dynamics in structured form and meaning spaces. *Advances in Complex Systems* 15(3), 1150021–1â1150021–20.

[2] Del Giudice, A. 2012. The emergence of duality of patterning through iterated learning: Precursors to phonology in a visual lexicon. *Language and cognition* 4(4), 381–418.

[3] Dingemanse, M., Verhoef, T., Roberts, S. 2014. The role of iconicity in the cultural evolution of communicative signals. de Boer, B., Verhoef, T., (eds), *EvoLangX: Workshop on Signals, Speech and Signs* 11–17.

[4] Galantucci, B. 2005. An experimental study of the emergence of human communication systems. *Cognitive science* 29(5), 737–767.

[5] Galantucci, B., Kroos, C., Rhodes, T. 2010. The effects of rapidity of fading on communication systems. *Interaction Studies* 11(1), 100–111.

[6] Kirby, S., Cornish, H., Smith, K. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31), 10681–10686.

[7] Little, H., De Boer, B. 2014. The effect of size of articulation space on the emergence of combinatorial structure. Cartmill, A., Erica, Roberts, S., Lyn, H., Cornish, H., (eds), *The Evolution of Language: Proceedings of the 10th international conference (EvoLangX)* volume 10. World Scientific 479–481.

[8] Roberts, G., Galantucci, B. 2012. The emergence of duality of patterning: Insights from the laboratory. *Language and cognition* 4(4), 297–318.

[9] Roberts, G., Galantucci, B. 2014. The effect of iconicity on the emergence of combinatorial structure: an experimental study. Cartmill, A., Erica, Roberts, S., Lyn, H., Cornish, H., (eds), *The Evolution of Language: Proceedings of the 10th international conference (EvoLangX)* volume 10. World Scientific 503–505.

[10] Verhoef, T., Kirby, S., de Boer, B. 2013. Combinatorial structure and iconicity in artificial whistled languages. *Cooperative Minds: Social Interaction and Group Dynamics Proceedings of the 35th Annual Meeting of the Cognitive Science Society* volume 35 3669–3674.

[11] Verhoef, T., Kirby, S., De Boer, B. 2014. Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics* 43, 57–68.

[12] Wedel, A., Martin, B. 2014. A laboratory model of sublexical signal category evolution. de Boer, B., Verhoef, T., (eds), *EvoLangX: Workshop on Signals, Speech and Signs* 47–53.