Evolang X

Workshop

on

Signals, Speech and Signs

Vienna, April 14, 2014

Proceedings

edited by

Bart de Boer
&
Tessa Verhoef

Table of Contents

Bart de Boer & Tessa Verhoef
Introduction to the Workshop on Signals, Speech and Signs ..... p. 1

Nathaniel Clark & Marcus Perlman
Breath, Vocal and Supralaryngeal Flexibility in a Human-Reared Gorilla ..... p. 5

Mark Dingemanse, Tessa Verhoef & Seán Roberts
The Role of Iconicity in the Cultural Evolution of Communicative Signals ..... p. 11

Hannah Little & Kerem Eryılmaz
The Effect of Physical Articulation Constraints on the Emergence of Combinatorial Structure ..... p. 17

Seán Roberts & Connie de Vos
Gene-Culture Coevolution of a Linguistic System in Two Modalities ..... p. 23

Joana Rosselló
On the Separate Origin of Vowels and Consonants ..... p. 29

Marieke Schouwstra, Katja Abramova, Yasamin Mota Medi, Kenny Smith & Simon Kirby
From Silent Gesture to Artificial Sign Languages ..... p. 35

Sławomir Wacewicz, Przemysław Żywiczyński & Sylwester Orzeschowski
Emergence of Low-Level Converstational Cooperation: The Case of Nonmatching Mirroring of Adaptors ..... p. 41

Andrew Wedel & Benjamin Martin
A Laboratory Model of Sublexical Categroy Evolution ..... p. 47

Bodo Winter
Neutral Spaces and the Evolvability of Spoken Language ..... p. 53

INTRODUCTION TO THE WORKSHOP ON SIGNALS, SPEECH AND SIGNS

BART DE BOER

AI-lab, Vrije Universiteit Brussel, Pleinlaan 2
Brussels, 1050, Belgium

TESSA VERHOEF

Center for Research in Language, University of California, San Diego
9500 Gilman dr., La Jolla, CA 92093, USA

1. Aims of the workshop

This workshop aims to bring together researchers interested in the physical signals that are used to convey language and the potential precursors of these signals. The intention of the workshop is not so much to present entirely new results – the main conference would be excellent for that – but to find out which open questions remain, what new approaches would be possible and where (interdisciplinary) cooperations could be useful. Although the content of the workshop is exploratory and perhaps speculative in this respect, the science on which new ideas have to be based will play a central role. One of the workshop's main themes will be to look for new empirical ways to test ideas that have so far received no attention or have only been speculated about.

The focus of the workshop is on physical signals for several reasons. First of all, physical signals are the most directly observable aspect of language. This makes it relatively easy to compare such signals between languages, between modalities and between species. However, another exciting property of the physical signals is that they are pre-symbolic and continuous. Of course, they may be used to express symbolic and categorical information, but this may not be necessarily inherent to the signals. Rather, symbolic or categorical structure needs to be imposed on the signal by the cognitive systems processing them. This transformation from continuous, sub-symbolic signals into categorical, symbolic information is something that humans are very good at (and something which is crucial to language), while other species (even evolutionarily closely related ones) appear to be less skilled at this. Therefore, this is an aspect of

language processing that is very relevant from an evolutionary point of view. Related to this is that linguistic signals have elaborate combinatorial structure, whereas non-linguistic communicative signals tend not to, or to a much smaller extent. How we are able to deal with combinatorial structure is also an important open question in the evolution of language, and one whose answer may have repercussions outside the domain of signals, as very comparable cognitive mechanisms may be needed to process the compositional structure of syntax.

Before giving a (very brief) overview of the contributions to the workshop, we will give an equally brief overview of what we think are important open questions in the study of the evolution of speech. Our hope is that the workshop can help to elucidate these questions.

2. Evolution of signals, speech and signs

A lot of work has been done on the evolution of the vocal tract. Although the debate is far from settled, it has become more or less accepted by most parties that the more crucial evolutionary changes were probably cognitive.

An open question, however it how our abilities differ exactly from those of apes. What exactly are the vocal abilities of apes? What is their neurological basis and is this different for modern humans. Which existing ape behaviours are most closely related to modern language? Ape gestures (orofacial or manual) or vocalizations? It appears that apes have more complex gestural repertoires, but the truth is that even their vocal repertoires are poorly understood.

The role of sign language in the evolution of language is also an open question. Did language start as pure signs? But then why did it ever change into a vocal system? It appears that many researchers appear to favour a mixed system. In any case, we need to investigate what the precursors to linguistic sign could have been and how they changed into linguistic signals. Related to this question is what precisely the role of iconicity was in the evolution of language. Are iconic signals really necessary for getting a language off the ground?

In answering these questions it is important to consider the interaction between individual learning, cultural evolution and biological evolution. All these processes interact and it may be difficult to determine what role each of them plays in explaining observed (linguistic) behaviour. Fortunately, the experimental paradigm of iterated learning or experimental semiotics helps to tease the effects of cultural processes and individual cognitive biases apart. However, this paradigm has only been applied to continuous signals in very few instances. In addition, computer models have successfully been used to gain

insight into complex interactions between different processes. Both of these methods are well represented in the workshop.

Although these are a lot of questions, there are certainly more open issues, and we do not expect that they will be easy to answer. However, carefully considering what techniques we have and what evidence is available or can be gathered, should allow us to find ways to answer these questions empirically.

3. Contributions

The contributions to the workshop form an interdisciplinary mix of different research methods and address a wide range of relevant research questions.

From linguistics, there is the contribution by Rosello, which investigates possible evolutionary scenarios by which vowels and consonants could have evolved from pre-existing behaviours. From biology there is the contribution by Clark and Perlman, investigating behaviours in a Gorilla that may be related to precursors of speech. Wacewicz et al. study non-verbal behaviour in the visual/postural domain that is potentially pre-linguistic: mirroring behaviour, and propose how to study this experimentally. Schouwstra et al. and Roberts and de Vos' contributions stem from the study of sign language. Schouwstra et al.'s contribution uses an experimental paradigm to investigate the transition from a system of gestures to a conventionalized system that looks much more like a sign language. Roberts and de Vos investigate, using a computer model, the interaction between genes for deafness and the emergence of sign language in a population. Winter also uses a computational model, but investigates the emergence of robustness in systems of signals. Little and Eryılmaz combine computer models with cognitive experiments to investigate how articulatory constraints may influence emergence of structure in speech. The contributions by Dingemanse et al. and Wedel and Martin present other experimental investigations of the emergence of structure in communicative signals, but they do not focus on articulatory constraints, but rather on how signals change over time. Whereas Wedel and Martin look at what happens to real phonemes, Dingemanse et al. look at the structure of signals in artificial languages. Moreover, Wedel and Martin look at how signals change in repeated interactions between the same participants, Dingemanse et al. look at how signals change over experimental "generations".

These contributions represent a rich subset of possible approaches and address a large number of the open questions mentioned above. We hope the interaction between the contributors will result in new directions of research to investigate the evolution of humans' ability to deal with linguistic signals.

BREATH, VOCAL, AND SUPRALARYNGEAL FLEXIBILITY IN A HUMAN-REARED GORILLA

NATHANIEL CLARK
Psychology, University of California, Santa Cruz
1156 High St, Santa Cruz, CA, 95064 USA

MARCUS PERLMAN
Cognitive and Information Sciences, University of California, Merced
5200 N Lake Rd, Merced, CA, 94534 USA

"Gesture-first" theories dismiss ancestral great apes' vocalization as a substrate for language evolution based on the claim that extant apes exhibit minimal learning and volitional control of vocalization. Contrary to this claim, we present data of novel learned and voluntarily controlled vocal behaviors produced by a human-fostered gorilla (*G. gorilla gorilla*). These behaviors demonstrate varying degrees of flexibility in the vocal apparatus (including diaphragm, lungs, larynx, and supralaryngeal articulators), and are predominantly performed in coordination with manual behaviors and gestures. Instead of a gesture-first theory, we suggest that these findings support multimodal theories of language evolution in which vocal and gestural forms are coordinated and supplement one another.

1. Introduction

Theories of language evolution frequently take as a starting point the assumed fact that nonhuman primates, including the great apes, lack the ability to exercise volitional control over their vocal and breathing-related behavior or to learn new behaviors (e.g., Corballis, 2002; Tomasello, 2008). In this paper, we present video evidence documenting eight types of learned vocal and breathing behaviors produced by Koko, a human-reared gorilla (*G. gorilla gorilla*), which are predominantly performed in coordination with manual gestures and actions. Along with accumulating evidence of vocal and breathing flexibility across the great apes, the strong starting assumption of great ape vocal inflexibility is clearly untenable. We discuss the ramifications of its falsification for theories of language evolution, specifically in favor of multimodal accounts.

1.1. *Vocal and breathing behavior*

Human speech production requires fine-grained control over a complex production apparatus, from the diaphragm through the lips. Technically speaking, *vocalization* refers only to a sound produced through vibration of the larynx, excluding sounds produced through the vocal tract that employ different mechanisms (e.g., whistling), and non-audible behaviors that demonstrate control over aspects of the production apparatus (e.g., blowing out a candle). We use the broader term *vocal and breathing behavior* (VBB) to refer to behaviors that employ any part of the speech production apparatus. Particular VBBs vary in their articulatory demands, reflected, for example, in voiceless blowing (control over breath and lips) versus voiced grunts (control over breath and larynx). The broad set of behaviors described in the current study illustrate the important point that the diaphragm, lungs, larynx, and supralaryngeal articulators are not a homogenous system. Given the different demands for different behaviors, the extent of control over different effectors will vary, and will recruit different neural systems.

1.2. *Flexibility in great apes' VBB*

Among "gesture first" theories of language evolution (e.g. Arbib, Liebal, and Pika, 2008; Call & Tomasello, 2007; Corballis, 2002), there is a common assumption that the ape homologue to the human speech apparatus is a poor substrate for language evolution. These theories build on the claim that ape vocal calls are innate and stimulus-driven, and that apes lack voluntary control of the larynx (vocal chords). The preferred evolutionary scenario is one in which speech supplants gestures at a later stage in language evolution, rather than vocal and manual modalities being interconnected throughout their evolutionary history (cf. McNeill, 2012).

However, contrary to the assumption of inflexible breathing and vocalizations, a large body of evidence shows that great apes are capable both of learning new VBBs and exerting voluntary control over them. Notable examples of captive and human-reared apes include Bonnie, a whistling orangutan (Wich et al, 2009), Kanzi, a bonobo who acquired four novel peep vocalizations (Taglialatela et al., 2003), and Viki (Hayes and Hayes, 1951), a chimpanzee who learned to produce 4 amodally voiced English words. Leavens, Russell, and Hopkins (2010) reported captive chimpanzees adjusting their communicative signals: the chimpanzees used visual signals when a human faced them, but auditory signals when the experimenter turned away. These included novel learned vocalizations like raspberries and elongated grunts, both of which have not been observed in wild chimpanzee populations.

In addition to these observations of captive and human-reared apes, observations of wild animals also contradict the claim that breathing and vocalizations are inflexible. One major research area in support of VBB flexibility is fieldwork observing dialectal variation or different vocal traditions across great ape communities. Van Schaik and colleagues (2003) reported regional variations in wild orangutans' production of raspberries. Dialectal variation has also been observed in the pant-hoot calls of several communities of chimpanzees (e.g. Crockford et al., 2004). The differential use of calls across communities, particularly as ecological and genetic factors have been ruled out, indicates that wild great apes can socially learn to modify existing VBBs, and may even socially learn new VBBs.

Another research area supporting VBB flexibility in wild great apes is fieldwork exploring the tactical suppression and production of calls. Wild chimpanzees have been particularly well studied, with evidence of tactical suppression observed during territorial patrols near other chimp communities (Goodall, 1986), and interactions between individuals of the same community (Laporte and Zuberbuhler, 2010). Further, a recent experiment on wild chimpanzees' alarm calls shows that individuals only call when other group members have neither seen the snake nor been in hearing range of previous calls (Crockford et al., 2012). Chimpanzees gave an alarm call less than half the time, indicating that voluntary production may be a more parsimonious explanation than voluntary suppression. Overall, the tactical deployment of calls suggests that wild great apes may exert volitional control over their VBBs.

2. The current study

Previous research makes a strong case for learning and volitional control of VBBs in the genera *Pan* and *Pongo*. We extend this case to the genus *Gorilla*, spanning another branch in the hominid family. We examined a video corpus spanning 3 years of interaction between a human-reared gorilla and its human caregivers, and found more than 400 tokens of novel VBBs distributed over 125 sessions. These comprise 8 categories of VBB, which exhibit several dimensions of contrast used in human phonology, including voicing (voiced and voiceless), place (labial, linguolabial, glottal), manner (stop, fricative), lip roundedness (rounded, unrounded) and nasality (present or absent).

2.1. *Koko's VBB and implications for language evolution*

Table 1 presents a description and the frequency of Koko's VBBs, which demonstrate an impressive range of flexibility across the various effectors of the speech apparatus. She performed these behaviors in a variety of contexts, and

they appear to be under her volitional control, with the majority of instances produced spontaneously without elicitation and some (e.g., playing wind instruments) often without any apparent social attention or expectation of reward. Although these behaviors have sometimes been subject to training and reinforcement over the years, they are not the result of rigorous operant conditioning, and some appear to contain elements of imitation (e.g. talking on the phone, huffing on eyeglasses). Given the contested status of laryngeal control, it is worth noting that approximately 25% of VBBs involved voicing, and approximately half involved glottal frication. While Koko's unique ontogeny cannot be overlooked, it is clear that a substantial degree of laryngeal control is learnable by non-human great apes.

Table 1. Frequency and Description of VBB Categories

Category	*# of sessions*	*Description*	*Active articulators*
Blow/huff (transitive)	15	Sometimes voiced glottal fricative w/ object-directed gesture, optional lip rounding	Glottis, (lips)
Blow/huff (intranstive)	27	Same as above but voiceless and rounded & w/ object-less manual gesture	Glottis, lips
Raspberry	17	Voiceless linguolabial fricative produced with tongue folded through lips	Lips, tongue
Cough	14	Glottal plosive, with gesture towards mouth	Glottis
Blow nose	5	Nasal frication achieved through manual pressure on nasal passage	Velum
Phone	11	Voiced glottal fricative while cradling phone-like object against ear/cheek	Glottis
Clean glasses	12	Voiceless glottal fricative w/ unrounded lips, directed at glasses, then rubbing them	Glottis
Play instrument	24	Blowing into a recorder, harmonica, or other instrument	Lips

More than 95% of Koko's VBBs were accompanied by manual gestures or routines involving the manual manipulation of objects. As McNeill (2012) notes, the close coordination of vocal and manual modalities is a hallmark of human communication, and theories of language evolution must explain this fact. The evidence provided here shows that non-human great apes share our ability to intertwine these modalities, underscoring the suitability of the vocal modality as a substrate of language evolution. But despite Koko's impressive coordination of vocal and manual modalities, it's clear that her flexibility in these behaviors is less than that of humans. Linguolabial fricatives, the most complex supralaryngeal articulation she performs, were never accompanied by manual behaviors, perhaps because of the difficulty in coordinating the hands while also coordinating breath, tongue, and lips.

While this data stem from a single individual with a highly unusual life history, when combined with data from other hominids, it is clear that the strong assumption of vocal inflexibility is definitively false: great apes both learn new VBBs and exert volitional control over them. Moving forward, we emphasize two main points. First, researchers must treat the evolution of vocal control with more anatomical nuance, considering separately the control of breathing, the larynx and various supralyaryngeal articulators. Second, researchers of language evolution must consider the vocal and manual modalities together as the substrate of language evolution. Speech did not supplant gesture; rather, they have always been supplementary.

References

Arbib, M., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the origin of language. *Current Anthropology, 49*(6): 1053-1063.

Call, J. & Tomasello, M. (2007). *The Gestural Communication of Apes and Monkeys*. Mahwah, NJ: Lawrence Erlbaum Associates.

Corballis, M. (2002). *From Hand to Mouth*. Princeton: PUP.

Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group-specific calls: A case for vocal learning? *Ethology, 110*, 221-243.

Crockford, C., Wittig, R., Mundry, R., & Zuberbuhler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology, 22*(2): 142-146.

Goodall, J. (1986). *The chimpanzees of Gombe.* Cambridge: HUP.

Hayes, K. & Hayes,C. (1951). The intellectual development of a home-raised chimpanzee. *Proceedings of the American Philosophical Society, 95*(2): 105-109.

Laporte, M. & Zuberbuhler, K. (2010). Vocal greeting behavior in wild chimpanzee females. *Animal Behavior, 80*(3): 467-473.

Levens, D., Russell, J., & Hopkins, W. (2010). Multimodal communication by captive chimpanzees (*Pan troglodytes*). *Animal Cognition, 13*(1): 33-40.

McNeill, D. (2012). *How Language Began.* Cambridge: CUP.

Taglialatela, J., Savage-Rumbaugh, S., & Baker, L. (2003). Vocal production by a language-competent *Pan paniscus*. *International Journal of Primatology, 24*(1): 1-17.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge: MIT.

van Schaik C.P, et al. (2003) Orangutan cultures and the evolution of material culture. *Science. 299*:102–105.

Wich, S., et al. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates, 50*: 56-64.

THE ROLE OF ICONICITY IN THE CULTURAL EVOLUTION OF COMMUNICATIVE SIGNALS

MARK DINGEMANSE
Language and Cognition department, Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands (mark.dingemanse@mpi.nl)

TESSA VERHOEF
Center for Research in Language, University of California, San Diego La Jolla, CA 92093-0526 USA (tverhoef@ucsd.edu)

SEAN ROBERTS
Language and Cognition department, Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands (sean.roberts@mpi.nl)

1. Introduction

The languages of the world vary in the extent to which they utilise iconic signals, in which there is a perceived resemblance between form and meaning. Sign languages make common use of iconicity, for instance by mapping motion in the world to motion in the signing space (Taub 2001). Spoken languages may also make extensive use of iconicity, for instance by depicting intensity or aspectual meanings in ideophones or sound-symbolic words, as in Japanese, Siwu, or Quechua (Dingemanse 2012). However, how iconicity emerges in a language, how it relates to the affordances of the medium of communication, or how it may bootstrap communication systems is unclear. One obvious suggestion is that the ease of mapping a semantic domain onto the signalling medium is a factor that affects the emergence of iconic signals. For example, mapping spatial relations in the world onto spatial relations in the sign space is easy to produce and to comprehend, whereas mapping spatial relations in speech is not so easy.

Here we explore this suggestion using an artificial communication game. Pairs of participants were asked to communicate about a set of meanings using whistled signals. We designed the meaning space so that some meanings would be easy to map onto the medium of communication and some would be difficult to map. The communication game was iterated, so that a pair was trained on the signals used by the previous pair. In this way we could observe how the communication system evolved over time.

We predicted that iconic signals would be more likely to emerge for the

easily mappable meanings, and that easily mappable meanings would be communicated with greater accuracy. In contrast, conventionalised and possibly compositional signals would be more likely to emerge for non-mappable meanings. What is less clear is how the two types of signal would interact. Iconic signs might form part of the building blocks for conventionalised signs, or perhaps a compositional system would eventually replace the iconic one. There may be founder effects that determine the amount of iconicity in a system, which might be analogous to the variation we see in spoken languages. It is also not clear how iconic signals would change over time. On the one hand, they should be easy to learn and easy to extrapolate, but there is also evidence that signals that combine iconic mappings with arbitrary features are *less* easy to learn than non-iconic signals (Ortega & Morgan, 2010). Iconic signals may not be subject to the same kind of drift as arbitrary signals because their transparent form-meaning mapping allows learners to regenerate them from scratch. This experiment explores some of these possibilities.

2. Methods

We use an iterated learning experiment with communication (e.g. Tamariz et al., 2012) to explore how iconicity affects the evolution of signals in a whistled language (e.g. Verhoef et al. 2012).

Materials

Participants communicated about artificial meanings. Each meaning was a picture of a well known animal facing either left or right (see figure 1). There were two 'mappable' animals and two 'non-mappable' animals. The mappable animals had shapes that were assumed to be easily mappable to the medium of communication (the slide whistle). The non-mappable animals had shapes that were assumed to be more difficult to map onto the medium of communication.

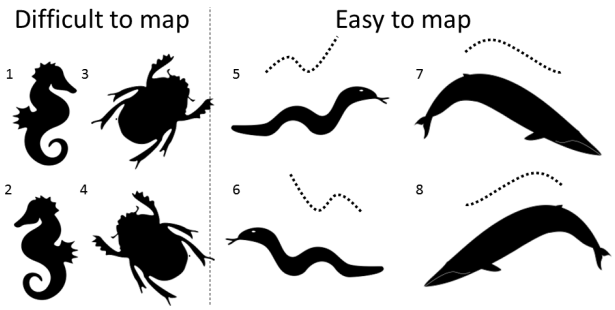


Figure 1. The meanings in the artificial language, consisting of 4 animals in two orientations. Meanings 1 to 4 are difficult to map onto the slide whistle space. Meanings 5 to 8 are easy to map onto the slide whistle space. The suggested mapping from meaning to tone contour is given above meanings 5 to 8. Note that animal *and* orientation are conveyable in iconic ways.

Procedure

Pairs of participants played a communication game via a touch-sensitive pad. In each round, one participant was chosen as the 'speaker' and the other as the 'listener'. The speaker was presented with a target meaning to communicate to the listener. The pad allowed the participants to communicate using a digital slide whistle. Moving a finger across the pad from left to right made a signal going from a low tone to a high tone.

The listener listened to the speaker's signal and was presented with a randomly ordered array containing the target meaning and 5 distractor meanings. The listener then guessed the target meaning. The pair were told whether they were correct and shown the target and the guessed meaning. After each round the speaker and listener roles were switched. Participants completed 16 rounds (each meaning twice) in a random order.

Pairs in later generations underwent a training phase before the guessing game where they saw meanings and heard the last signal used for that meaning by the previous pair in the previous generation. Participants only saw a random half of the previous meanings. This procedure differs from many iterated learning experiments because the initial input set of signals was not created by the experimenters but emerged in the interaction of the first pair.

3. Preliminary results

We ran a pilot experiment of 4 chains of between 8 and 10 generations. Participants were recruited at a museum in Utrecht and included children and adults. Easily mappable meanings were guessed correctly in 33% of trials, while non-easily mappable meanings were guessed in 22% of trials ($t = 2.9$, $p = 0.003$). We used a mixed effects logit model to predict communicative success based on the mappability of the target, the orientation, the generation, the age of the participant and the interaction between mappability and generation. The animal depicted in the meaning and the chain number were entered as random effects.

We found no main effects, but there was a significant interaction between mappability and generation ($z=2.4$, $p=0.02$). This suggests that while bootstrapping a linguistic system may not be easier with easily mappable meanings, signals for easily mappable meanings evolve to fit the communicative needs faster than signals for meanings that are not easy to map (see figure 2).

4. Discussion and future work

We used an iterated learning paradigm to explore how iconic mappings between meanings and signals can be used during the initial stages of language emergence. The results suggested that how easy a meaning can be mapped to an articulation space can affect the cultural evolution of a language.

Although in the beginning of a chain, there seems to be no difference in the proportion of correct responses for the two types of meanings, after some generations of transmission and use a clear effect appears. This is interesting,

since the possibility of using iconic signals was present from the beginning. In a further analysis of the data we want to explore possible reasons for the later emergence of success in communicating easily mappable meanings. It may take time for participants to coordinate on their strategy, leading to clashes in the earliest trials that are avoided only when participants converge on the same strategy. Participants in later chains have the advantages of a learning phase which serves to create the common ground required for quick strategic convergence. A possible iconic strategy may therefore need to be used more systematically and occur in a pattern before it actually makes learning and recall easier. Such systematic patterns in the use of strategies are expected to emerge through cultural evolution and social coordination. We are currently in the process of analysing the signals used in the experiment to assess to what extent iconic mappings were utilised. We will also analyse whether signals for easily mappable meanings are more similar across chains than signals for meanings that are difficult to map. A future version of this experiment will be conducted in a more controlled laboratory environment and will involve longer training and interaction sessions with a larger set of meanings and signals.

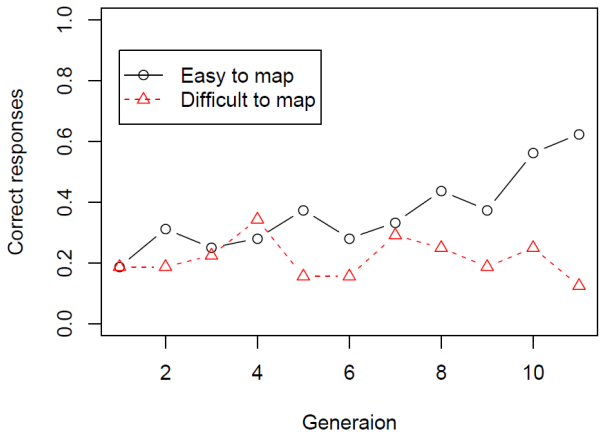


Figure 2. Proportion of correct guesses for different types of meaning over generations.

Acknowledgements

Many thanks to Shawn Bird and Marisa Casillas for help with the experiment design and implementation.

References

Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6 (10): 654–672.

Ortega, G., & Morgan, G. (2010). Comparing child and adult development of a visual phonological system. *Language interaction and acquisition*, 1(1),

p.67-81.

Taub, Sarah F. 2001. *Language from the body: iconicity and metaphor in American Sign Language*. Cambridge/NY: Cambridge University Press.

Tamariz, M., Cornish, H., Roberts, S. & Kirby, S. (2012) The effect of generation turnover and interlocutor negotiation on linguistic structure. In T. C. Scott-Phillips, M. Tamariz, E.A. Cartmill & J.R. Hurford, *The Evolution of Language: Proceedings of the 9th International Conference (EVOLANG9)*. World Scientific. p. 555

Verhoef, T., de Boer, B.G. & Kirby, S. (2012). Holistic or synthetic protolanguage: Evidence from iterated learning of whistled signals. In T.C. Scott-Phillips, M. Tamariz, E.A. Cartmill & J.R. Hurford (Eds.), *The evolution of language: Proceedings of the 9th international conference (EVOLANG9)* World Scientific. p. 368-375.

THE EFFECT OF PHYSICAL ARTICULATION CONSTRAINTS ON THE EMERGENCE OF COMBINATORIAL STRUCTURE

HANNAH LITTLE & KEREM ERYILMAZ
Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Pleinlaan 2
1050 Brussels, Belgium
hannah@ai.vub.ac.be, kerem@ai.vub.ac.be

1. Introduction

Language has "duality of patterning", which is structure on both a compositional and a combinatorial level. Compositional structure is the combination of meaningful elements into bigger meaningful structures. Combinatorial structure is the phonological combination of small meaningless units into a potentially infinite number of meaningful units.

Despite "duality of patterning" being named by Hockett (1960) as one of the basic design features of human language, empirical work exploring the emergence of combinatorial structure is still very much in its infancy. Techniques to test existing hypotheses regarding the emergence of phonological structure have only recently been developed, and the strengths and weaknesses within this ongoing work are generating new hypotheses which also need to be tested. The current contribution will outline the existing hypotheses on how combinatorial structure first emerged in language before focusing on hypotheses pertaining to the modality, size and shape of the articulation space. We will then outline existing experimental and computational work which tests the effects of physical articulation constraints on the emergence of combinatorial structure, along with our own ongoing work, and the scope for future work in this area.

2. Existing Hypotheses

Hockett (1960) hypothesised that the emergence of structure on a phonological level is the result of pressures for expressivity and discriminability imposed when the number of meanings increases, as language needs a more efficient way to create new word forms. More recently, Verhoef (2012) has shown experimentally that combinatorial structure can emerge as the result of cognitive learning constraints and biases. However, recent evidence from Al-Sayyid Bedouin Sign Language, which is a newly emerging language, suggests that languages can have thousands of words without a level of phonological patterning (Sandler, Aronoff, Meir, &

Padden, 2011). In a recent paper, Del Giudice (2012) considers that the lack of phonological patterning in emerging sign languages could be because the articulation space in sign languages is much larger than that used in spoken languages, and this allows for a greater number of distinct signals without the need for combinatoriality. This hypothesis is dismissed by Del Giudice (2012) as established sign languages have been shown to have a similarly sized phoneme inventory to those found in spoken languages (Rozelle, 2003). However, this is not evidence to suggest the size of articulation space, as well as other physiological factors, are not important factors in the *emergence* of combinatorial structure in language. Hypotheses regarding the effects of the modality, shape and size of an articulation space have yet to be empirically tested which is what we aim to rectify with this contribution.

3. Experimental Work

Artificial language learning experiments are often used in evolutionary linguistics to show how structure emerges on a compositional level. Work is now appearing on emerging combinatorial structure, started by Verhoef (2012) who used signals created by slide whistles in an iterated learning paradigm. Whistled signals are ideal for the purposes of investigating the emergence of speech as they use a continuous articulatory space, but limit interference from participants' existing linguistic knowledge. In Verhoef's (2012) experiment, participants learned whistled signals and their resulting reproductions became the input for the next participant. Del Giudice (2012) has since carried out a similar iterated experiment where participants created graphical symbols using a moving stylus which limited the use of iconic representation, and found that participants did not use the entirety of the signal space as one would expect if Hockett's (1960) hypothesis were true.

To test the effects of the size of articulation space on the emergence of combinatorial structure, we extended Verhoef's (2012) experiment by running a new condition where the slide whistle was restricted with a stopper, as well as an unrestricted condition. The shape of the whistle's articulation space was kept the same, only differing in size on one dimension. Comparison of combinatoriality between conditions eliminated the problem of an articulation space having some trajectories which are more likely to be produced, which is a problem for analysis when only one condition is being tested. We show that the size of articulation space does indeed have an effect on the emergence of combinatorial structure.

There is a large scope for future experimental work on the effects of physical articulation constraints. A whole host of electronic musical instruments and digitally generated signals are enabling more easily manipulated signal spaces and easily analysable signals. Our next steps are to experimentally test the effects that modality and the dimensionality of a signal space have.

4. Computational Work

The computational work deals with four main issues: the representation of signals, the selection process through which some signals persist while others fall into disuse, the distance and similarity measures between signals, and measures of structure.

4.1. *Signal Space and Signals*

Earlier models of the evolution of combinatorial structure abstract away from the internal structure of signals, representing them as unique symbols (Nowak, Plotkin, & Krakauer, 1999). In such models, the variation in signals necessary for evolution arises from errors in probabilistic learning, and not from comparison of the signals involved. To deal with structure, many later models use signals represented as points or trajectories in an N-dimensional feature space, which may be abstract and not correspond to any actual features of an acoustic signal (de Boer & Zuidema, 2010). The current work deals exclusively with the interplay between the shape of an artificial feature space and the combinatorial structure of signals in that space, abstracting away from the acoustic nature of the features. Each signal consists of a fixed number of ordered points in the feature space, forming a trajectory.

4.2. *Signal Selection*

The signals evolve within a multiagent imitation game. Agents start with a fixed number of randomised signals, and utter them with small, random, shape-preserving mutations as described by de Boer and Zuidema (2010). All signals are further subject to environmental noise but preserve their shape. As in de Boer and Zuidema (2010), each round, a chosen performer agent utters their repertoire $\mathbb{L}$, then the imitating agents utter the closest signal they know to the performer's signal. If the imitation is closer to the original signal than any other in the performer's repertoire, the round is successful. If more imitators are successful using the performer's mutated signal than using the original signal, the performer replaces the original with the modified signal.

4.3. *Signal Distance and Confusion*

For signals represented as trajectories, the easiest distance metric is point-to-point Euclidean distance. However, this may result in overestimation of the distance between similar signals with different timings. We estimate the distance between signals using Dynamic Time Warping (Sakoe & Chiba, 1978), also used in the analysis of some experimental studies. When a signal, X, is emitted, the probability of that signal being identified correctly varies with its distance d to the original position of the signal. This probability is chosen from a Gaussian distri-

bution around X, with the spread δ (i.e. noise level), as in de Boer and Zuidema (2010).

$$f(d) = \int_{x=\frac{1}{2}d}^{\infty} \frac{1}{\sqrt{2\pi\delta}} e^{\frac{-x^2}{2\delta^2}} dx$$

The probability of perceiving the uttered signal X as $Y \in \mathbb{L}$ becomes:

$$P(Y_{perceived}|X_{uttered}) = \frac{f(d(X,Y))}{\sum_{Z\in\mathbb{L}} f(d(X,Z))}$$

4.4. *Measures of Structure*

We propose investigating the amount of structure in the agents' repertoires based on measures motivated by information theory. Specifically, we claim that for signals that can be well-represented by a few data points per signal, such as those in this study, entropy rate of an agent's repertoire is a feasible measure of combinatorial structure.

Choosing a measure of combinatorial structure is far from trivial. It is possible to assume that combinatorial building blocks have greater power to predict what comes next than non-building blocks. However, combinations of these building blocks can also have considerable predictive power. Conversely, trends that appear on very small time scales as opposed to communicatively relevant time scales (combinatorial building blocks) can be artefacts of the articulatory apparatus (or a mathematical or computational proxy). To create a balance between problems at these two extremes, we propose focusing on quantifying the predictability of the signal-generating process per unit time, instead of the predictability of individual signal occurrences. More formally, we propose using a weighted mixture of variable-depth context trees to estimate the entropy rates, given different maximum context depths (Kennel, Shlens, Abarbanel, & Chichilnisky, 2005). By looking at the changes in the estimated entropy rate under different context depths, it is possible to estimate the maximal length of the building blocks. Any part of a signal longer than the longest building block will contain at least two (possibly partial) building blocks. Building blocks have less internal variation than combinations of building blocks, since the blocks themselves do not contain combinatorial parts. Thus, a notable decrease in the estimated entropy rate at a certain depth increment, which is not followed by a comparable decrease at the next depth increment, can be used to estimate the maximum length of a building block.

Theoretically, it is also possible to have an unbounded tree that uses complete trajectories instead of bounded contexts etracted from parts of signals. However, for inventory sizes greater than three or four, such trees becomes impractical both in memory and time complexity, as the context tree can consist of A^{D^D} nodes for an alphabet of size A and a maximum depth of D, depending on the contexts observed.

5. Conclusion

We have argued that physiological constraints are important factors affecting the emergence of combinatoriality within different modalities. We have also outlined problems in existing work which use proxies for articulatory spaces to investigate the emergence of combinatorial structure, and shown how recent experimental and computational techniques can be implemented to test hypotheses pertaining to how physiological constraints can affect the emergence of combinatorial structure. The evolution of speech, as a field, is currently divided between work dealing with the emergence of phonological structure and the cognitive capacity for speech, and work dealing with human phonetic capabilities and the physiological capacity for speech. Fitch (2002) states that some researchers do not even regard phonological evolution as part of speech evolution at all. However, we show that it is important to consider phonetic capabilities when considering the emergence of combinatorial structure.

References

de Boer, B., & Zuidema, W. (2010). Multi-agent simulations of the evolution of combinatorial phonology. *Adaptive Behavior*, *18*(2), 141-154.

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.

Fitch, W. T. (2002). Comparative vocal production and the evolution of speech: Reinterpreting the descent of the larynx. In A. Wray (Ed.), *The transition to language* (p. 21-45). Oxford University Press.

Hockett, C. D. (1960). The origin of speech. *Scientific American*.

Kennel, M. B., Shlens, J., Abarbanel, H. D., & Chichilnisky, E. J. (2005). Estimating entropy rates with Bayesian confidence intervals. *Neural Computation*, *17*(7), 1531–1576.

Nowak, M. A., Plotkin, J. B., & Krakauer, D. C. (1999). The evolutionary language game. *Journal of Theoretical Biology*, *200*(2), 147–162.

Rozelle, L. G. (2003). *The structure of sign language lexicons: Inventory and distribution of handshape and location.* Unpublished doctoral dissertation, University of Washington.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *26*(1), 43–49.

Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural language & linguistic theory*, *29*(2), 503-543.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and cognition*, *4*(4), 357-380.

GENE-CULTURE COEVOLUTION OF A LINGUISTIC SYSTEM IN TWO MODALITIES

SEÁN ROBERTS AND CONNIE DE VOS

Language and Cognition Department, Max Planck Institute for Psycholinguistics
Nijmegen 6525 XD, The Netherlands
sean.roberts@mpi.nl, connie.devos@mpi.nl

Complex communication can take place in a range of modalities such as auditory, visual, and tactile modalities. In a very general way, the modality that individuals use is constrained by their biological biases (humans cannot use magnetic fields directly to communicate to each other). The majority of natural languages have a large audible component. However, since humans can learn sign languages just as easily, it's not clear to what extent the prevalence of spoken languages is due to biological biases, the social environment or cultural inheritance. This paper suggests that we can explore the relative contribution of these factors by modelling the spontaneous emergence of sign languages that are shared by the deaf and hearing members of relatively isolated communities. Such shared signing communities have arisen in enclaves around the world and may provide useful insights by demonstrating how languages evolve as the deaf proportion of its members has strong biases towards the visual language modality. In this paper we describe a model of cultural evolution in two modalities, combining aspects that are thought to impact the emergence of sign languages in a more general evolutionary framework. The model can be used to explore hypotheses about how sign languages emerge.

One of the great linguistic discoveries of the 20th century has been that our linguistic abilities are, to an extent, independent of the natural language mode through which it is expressed and understood. That is to say, sign languages parallel spoken languages in terms of the areas of the brain that are involved in production and processing, in the patterns of language acquisition, as well as the degree of grammatical diversity among them (Meier, Cormier, & Quinto-Pozos, 2002). Sign languages may emerge spontaneously in at least two types of settings. Urban sign languages often emergence in response to the congregation of deaf individuals at government institutions for the deaf, as for instance in the well-documented case of Nicaraguan Sign Language (Senghas & Coppola, 2001). Alternatively, sign languages may arise in communities with an exceptionally high incidence of (often hereditary) deafness (Zeshan & de Vos, 2012). In the latter type of setting the sign language is used by both deaf and hearing community members, engendering a high degree of social integration for deaf individuals. Such so-called shared signing communities may therefore provide unique insights into the rel-

ative contribution of biological, cultural, and social biases in the emergence of signed languages.

However, the cases of signing communities documented so far show a striking diversity in their social attitudes to deafness, demography, history, ecology and the proportion of hearing L2 speakers (Zeshan & de Vos, 2012). There are also structural differences between the languages, such as differences in phonology or spatial grammar, possibly due to different amounts of cross-modal contact. The diversity makes it difficult to make generalisations about how these factors affect the emergence of a signing community. For example, the critical mass of deaf people that is needed for a shared signing community to emerge is not known. Models can help researches think about these questions.

1. Model

We use a model adapted from Burkett and Griffiths (2010) and Smith and Thompson (2012) which simulates gene-culture co-evolution in an iterated learning framework (for a full description, see Roberts, Thompson, & Smith, 2013). Individuals are modelled as Bayesian agents who must decide what proportion of each modality to use in communication, given their prior bias and their observations of the behaviour of other agents. Since hearing communities tend to have an audible linguistic system as an important part of their communication, hearing agents have a bias favouring the auditory modality. It is obviously a weak bias, because both hearing and deaf learners can learn non-audible (signed) languages. It is also well-documented that speakers generally distribute the message over both auditive and visual forms (Enfield, 2009; Kendon, 2004). At any rate, deaf learners can be characterised as having a very strong bias towards the visual modality (learning an audible language is hard).

The agents reproduce biologically, according to a fitness function that gives a higher probability of reproduction to individuals who can socialise successfully through language. The prior bias is inherited biologically (with some chance of mutation). This means that offspring of deaf individuals will inherit the bias against audible languages (deafness is hereditary).

We can use this model to explore the emergence of deaf communities within hearing communities, or to model the competition between auditory and visual modalities. In a community of deaf individuals, we would expect a mainly non-audible language to emerge. However, what happens in a community with mixed biases where modalities might be in competition?

Since the dynamics of this kind of model are not well understood analytically, we obtain results by numerical simulation. We run the model with hearing individuals until it converges (around 200 generations). At this point, deaf individuals are introduced into the simulation who have a strong bias against learning an audible language. We can then observe how the community changes, both in terms of the number of deaf individuals, and the use of each modality. Since deaf individuals

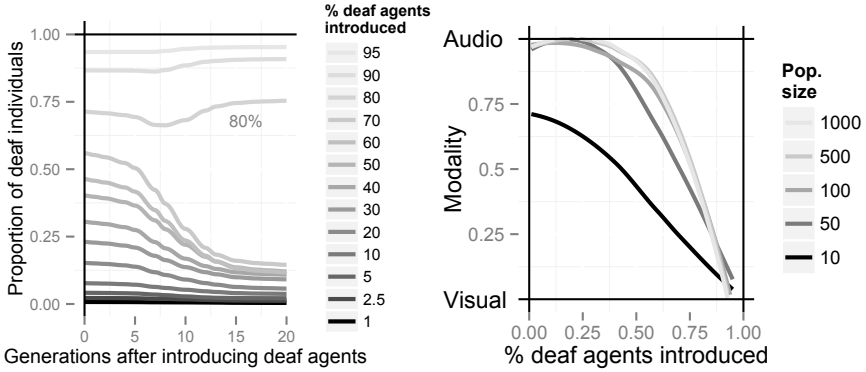


Figure 1. **Left:** Deaf individuals are introduced into a hearing population 200 generations after initialisation. The graph shows how the proportion of deaf individuals changes over generations depending on the initial number of deaf individuals introduced (lines are LOESS fits of 10 independent runs). Between 70% and 80% of the population needs to be deaf for deaf individuals to remain stable or increase. **Right:** The average modality used in a population for different population sizes, under the standard fitness function. Means are taken from 8 generations after introducing deaf individuals. Larger populations require a greater proportion of deaf individuals to affect the overall modality.

essentially cannot learn an audible language, the two aspects will be correlated. However, we also show that this is not always the case.

1.1. *Results*

The results demonstrate that in a wide range of scenarios, communities of hearing individuals using primarily audible communication are resistant to deaf individuals (see figure 1a). Shared-sign languages are unlikely to survive except when the initial proportion of deaf individuals introduced into the community is very high. The weak bias for audible languages is amplified over generations of cultural transmission so that the majority of the communication system is audible. The average modality of communication used by the population reflects the number of deaf individuals, with a large number of deaf individuals required to change the modality of the population (see figure 1b). However, in very small populations, a smaller proportion of deaf individuals may influence the modality of the language in the short-term (up to 10 generations).

These results suggest that a monolingual signing community is unlikely to emerge. However, there are conditions under which a bimodal-bilingual shared-signing community can emerge and where deaf individuals can thrive. If the ability to communicate in both modalities is prestigious within a society, then a communication system that uses both visual and auditory modalities will emerge. This is independent of the community having deaf individuals (although the presence

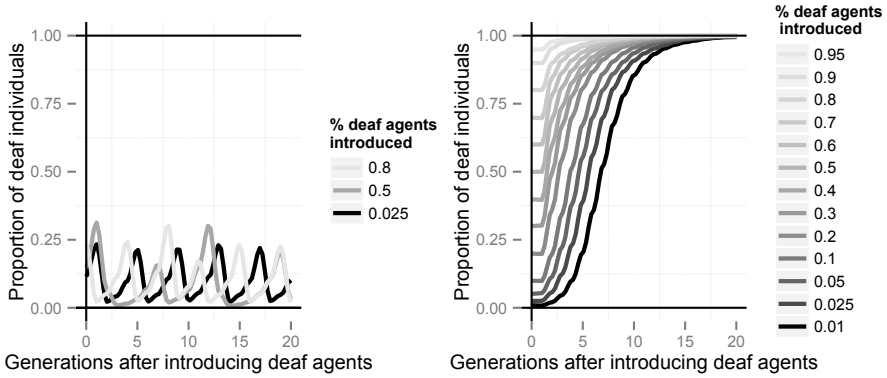


Figure 2. **Left:** Results from the model where there is a social prescription against marriage between deaf individuals. The population size matches that of the Kata Kolok community. **Right:** Results from the model using the 'parity' fitness function and a structured population of the same size as the Kata Kolok community (2189). Even very small numbers of deaf individuals introduced into the model will increase within a few generations.

of deaf individuals is an obvious motivation for the prestige of a multi-modal ability).

The social structure of the community also makes a difference. In stratified communities where agents' fitness is only derived from the communicative success between a few nearest neighbours, the community maintains a non-audible component in the language for longer. This happens because small 'enclaves' of deaf individuals can be maintained, where using a non-audible language leads to good communicative success and high probability of reproduction.

The dynamics of social interaction make a difference, too. Communities with deaf individuals are sustainable when linguistic differences lead to higher fitness (figure 2a). This can happen, for instance, if linguistic differences are perceived as resources rather than limitations (as is the case in some sign language communities). In this case, the linguistic system of the community as a whole utilises both modalities equally. The number of deaf individuals oscillates with a phase determined by the initial number of deaf individuals introduced.

Finally, if the fitness function is neutral with regards to the modality of communication (the 'parity' function, where reproduction is linked to the ability to communicate effectively, regardless of modality), the proportion of deaf individuals and non-audible language increases in small, structured societies. In fact, in this social set-up, the modality of communication is predominantly visual and the community is resistant to hearing individuals (see figure 2b). This happens because deaf select the same proportion of each modality (all visual), and so maximise their communicative fitness with other deaf individuals. Hearing individuals

are more likely to select a range of proportions of each modality, meaning that they have weaker fitness.

2. Conclusion

The extent to which modalities are exploited in communication systems depends on genetic constraints, cultural transmission and social factors. We demonstrated that the links between learning biases, modality, communicative success and the social perception of language can be complex. We hope this model can help frame the exploration of demographic differences between different types of sign languages. Future improvements could include more realistic genetic inheritance and social structures. We also hope that this paper demonstrates the relevance of shared sign languages for language evolution: given their relatively limited time depths and relative isolation, the diffusion of structural features within these communities could be charted to track their historical development.

References

Burkett, D., & Griffiths, T. (2010). Iterated learning of multiple languages from multiple teachers. In A. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The evolution of language: Proceedings of EvoLang 2010* (p. 58-65). World Scientific.

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances.* Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge University Press.

Meier, R. P., Cormier, K., & Quinto-Pozos, D. (2002). *Modality and structure in signed and spoken languages.* Cambridge University Press.

Roberts, S., Thompson, B., & Smith, K. (2013). Social interaction influences the evolution of cognitive biases for language. In E. Cartmill, H. Lyn, H. Cornish, & S. Roberts (Eds.), *The Evolution of Language: Proceedings of the 10th International Conference (EVOLANG10).* World Scientific.

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, *12*(4), 323–328.

Smith, K., & Thompson, B. (2012). Iterated learning in populations: Learning and evolving expectations about linguistic homogeneity. In T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, & J. R. Hurford (Eds.), *The Evolution of Language: Proceedings of the 9th International Conference (EVOLANG9)* (p. 227-233). World Scientific.

Zeshan, U., & de Vos, C. (Eds.). (2012). *Sign languages in village communities: anthropological and linguistic insights.* Berlin: Mouton de Gruyter.

ON THE SEPARATE ORIGIN OF VOWELS AND CONSONANTS

JOANA ROSSELLÓ

Department of General Linguistics, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, Barcelona, 08007, Spain

A non controversial claim on oral language phonology is that there are consonants (C) and vowels (V) which organize themselves into syllables. This notwithstanding, how the difference between vowels and consonants came about in evolutionary terms is unknown. Departing minimally from the frame-content theory of speech (Macneilage 1998, 2008), this work puts forward the conjecture that vowels and consonants have a different origin, neither of which traces back to primate calls: vowels —in an instance of convergent evolution—, come from vocal learners' primate song units which are analog to those of birdsong of vocal-learning birds; consonants, instead, evolved by common descent from some visual communicative displays (lip smacking, teeth-chattering, etc.). This proposal fares better than those relying on primate calls because, while avoiding their pitfalls, it automatically derives other necessary properties of speech, namely discreteness, seriality, direct cortico-laryngeal connections and repetitive babbling. Additionally, it (i) paves the way to a musical (syllabic) protolanguage, and (ii) can be a good clue on the categorical neuropsychological divide between vowels and consonants.

1.Introduction

The fact that vocal signals are made up of vowels and consonants constitutes a phylogenetic novelty that although of paramount importance not only for speech (externalization) but possibly for language (as a cognitive system) has been neglected. Of note, in this connection, is that our big public lexicons are not even imaginable without the join concurrence of both, vowels and consonants. It seems, indeed, that in linguistics, phonologists take the distinction for granted and that in the field of language evolution, the description of birdsongs in terms of syllables ($\text{syllables}_{\text{birdsong}}$) has obscured that speech syllables ($\text{syllables}_{\text{speech}}$), unlike $\text{syllables}_{\text{bird}}$, are typically made up of consonants (C) and vowels (V). Still in the evolutionary field, the often tacit commitment to the continuity hypothesis has contributed to the current situation. Fitch (2013: 434) summarizes it: "The origins of the periodic oscillations that produce the alternation of consonants and vowels that make up syllables a central feature of all spoken languages have remained mysterious, because most primate calls are produced with just a single opening of the mouth." To complete the picture, it

comes out that (neuro)psychologists seem to be the most concerned with the distinction between vowels and consonants (Caramazza et al. 2000). They go as far to claim that V and C are categorically distinct and functionally specialized.

2.The received view

The contentions that (i) syllables are present in birdsong and that (ii) speech has some kind of primate call as a precursor are both commonly accepted. However,

1.1.Syllables$_{birdsong}$ ≠ Syllables$_{speech}$

Birdsong, as speech, presents a serial organization which can be seen as possessing a syllabic frame/content mode of organization (MacNeilage 2008: 303) where the frame is the result of a beak open-close cycle. Syllables$_{birdsong}$, unlike syllables$_{speech}$, however, are usually defined acoustically rather than articulatorily —the opposite of what is found for syllables$_{speech}$. This means that units of sound are separated by silent rests. A syllable$_{birdsong}$ can contain more than one note. Crucially, the notes (the content) are the result of variations on the source (syrinx). In other words, birdsongs' content is exclusively vowel-like.

1.2.Primate calls do not lead to speech

That speech derives from non speech is indisputable but this does not mean that holistic signals are at its origin (but see Zuidema & de Boer 2009). Yet, deriving it from primate calls is virtually impossible. The pitfalls seem insurmountable. Call, in contrast to songs, are inarticulate, innate, under subcortical control and, although repressible, non structurally modifiable. By adding to certain laryngeal calls, as Fitch (2013) suggests, a co-opted visual display such as lip-smacking, which will provide the consonant (and the syllabic frame), we do not get rid of the just mentioned difficulties. Furthermore, this combination would still be in need of "a second evolutionary step" consisting of "our unique cortical-brainstem connections" (Fitch 2013: 435).

3.Primate songs + lip-smacking as the foundation of V/C distinction

Although syllables$_{birdsong}$, because of lacking consonants, do not amount to syllables$_{speech}$, songs are a much better basis for speech than calls. Primate songs are not as common as birdsongs but they do not limit to gibbons' duets either. Singing is present in 26 monogamous species of primates and has evolved four times within the taxon (Ghazanfar & Santos 2003: 7). Many properties of songs (and vocal learning animals) fit in with what we know on speech (and Sapiens). Structurally, in either song or speech, discreteness, seriality and repetitiveness in

the babbling stage are obtained. Ontogenetically, a babbling stage is innate to both non human vocal learners and humans. Neurally, all vocal learners —even mice with innate songs (see Arriaga et al. 2012)—, seem to share a neural circuitry with forebrain/cortico-bulbar-laryngeal connections to motor nuclei responsible of motor learning and fine control of vocalization. Functionally, songs (and duets in particular) reinforce pair bonding. All in all, all these commonalities suggest that a homoplasy, i.e. an instance of convergent evolution, is in place. As said in 2.1, however, songs only give us vowels.

Where do then the consonants come from? In line with recent findings (Ghazanfar et al. 2012, Fitch 2013), consonants would be originated in lip-smacking, a visual communicative display very common among primates. The main rationale for this common descent view of consonants is that $\text{syllables}_{\text{speech}}$ and lip-smacking seem to be perfectly tuned (6-hertz rhythm). Ingestive cyclicities, instead, are slower. This, by itself, makes them unnecessary as a basis for the frame in the frame-content theory. Nicely enough, having songs in the scenario would lead to the same conclusion as, in birdsongs in particular, no ingestive cyclicity (chewing, sucking, etc.) is involved, as MacNeilage (2008: 306) observes.

It is also worth to emphasize that Sapiens are vocal learners and vocal learners produce songs, not calls. Singing, in turn, automatically guarantees the existence of cortico-bulbar-laryngeal connections. By contrast, in a scenario in which calls are complemented with lip-smacking (Fitch 2013), this neural equipment calls for an extra evolutionary event. In this connection, the fact that a dorsal-laryngeal cortical connection seems exclusively human among primates (Bouchard et al 2013) needs to be qualified. As far as it is known, cortices of singing non human primates have not been examined in this regard. The prediction entailed by the present proposal is that cortico-bulbar-laryngeal connections have to be present in these species. Although the importance of these neural connections has come into question (Lieberman 2013), neglecting them does not seem justified (Brown et al. 2009).

Finally, apart from getting rid of the shortcomings listed in 2.2, resorting to songs has a further advantage, namely to provide a basis for phonology (via perhaps a musical protolanguage) completely devoid of any referential meaning. If primate calls, instead, which are stimulus-driven and perception-related, had been the point of departure to speech, a complete turnaround as far as linguistic meaning is concerned would have had to take place, which seems as much costly as implausible.

4.Further expectancies

This proposal opens some interesting avenues which will be touched on in the talk.

The first one deals with the foundational divide between vowels and consonants for which psychologists have found strong evidence. According to them (Bonatti et al. 2005), vowels are universally —not only in Semitic languages— tied to grammar, in part through prosody. Consonants, instead, are bound to lexicon. In particular, the individuation of words in continuous speech relies on them. It has been shown that in order to segment the continuous stream of (artificial) speech into words, subjects use transitional probabilities between consonants, but not between vowels. The claim goes further: the V/C divide is categorical since it has been shown that in selective impairments of either vowels or consonants, the causal factor does not depend of either the sonority value or the feature properties (Knobel & Caramazza 2007). An investigation which suggests itself from the present proposal would rely on their different neural correlates which would trace back to their different origin. Interestingly, there is recent evidence in favor of this claim. Bouchard et al. (2013: 331) not only state that "vowels and consonants occupy different regions of the cortical state-space" but also that all their findings are in accordance with gestural theories of speech production.

The second is related to the holism *vs.* discreteness issue. The contention is that it is an advantage that song provides us with a discrete origin. Speech started discrete as it was to go on. Is sign (gestural-visual modality) in contradiction with this claim? Seemingly, ABSL (Al-Sayyid Bedouin Sign Language) as presented by Sandler et al. (2011) started being holistic. Contrary to this claim, I will present some evidence that a video-recorded Deaf woman belonging to the second generation was combining discrete elements.

Finally, the plausibility of a syllabic musical protolanguage in line with Darwin (1871) who considered an analogue of birdsong as a plausible step in the way to a full-fledged language, will be examined.

References

Arriaga, G.; Zhou, E. P. & Jarvis, E. D. (2012). Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. *Plos One 7, 10,* e46610.

Bonatti, L.; Peña. M.; Nespor, M. & Mehler, J. (2005). Linguistic constraints on statistical computations. The role of consonants and vowels in continuous speech processing. *Psychological Science 18, 10,* 924-925.

Bouchard, K. E.; Mesgarani, N.; Johnson, K. & Chang, E.F. (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature 495*, 327-332.

Brown, S.; Laird, A. R; Pfordresher, P. Q.; Thelen, S. M.; Turkeltaub, P. & Liotti, M. (2009). The somatotopy of speech: phonation and articulation in the human motor cortex. *Brain Cogn. 70, 1*, 31-41.

Caramazza, A.; Chialant, D.; Capasso, R. & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature 403*, 428-430.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

Fitch, T. W. (2012). Segmental structure in banded mongoose calls. *BMC Biology* 10: 98. http://www.biomedcentral.com/1741-7007/10/98.

Fitch, T. W. (2013). Tuned to the rhythm. *Nature 494*, 434-435.

Ghazanfar, A. A.; Santos, L.R. (2003) Primate as auditory specialists. In A. A. Ghazanfar (Ed.), *Primate audition: Ethology and neurobiology* (pp. 1-11). Florida: CRC Press

Ghazanfar, A. A.; Takahashi, D. Y.: Mathur, N. & Fitch, T. W. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current Biology 22*, 1176-1182.

Knobel, M. & Karamazza, A. (2007). Evaluating computational models in cognitive neuropsychology: The case from the consonant/vowel distinction. *Brain and Language 100*, 95-100.

Lieberman, P. (2013) *The Unpredictable Species. What Makes Humans Unique*. Princeton and Oxford: Princeton University Press.

Macneilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences 21*, 499-546.

Macneilage, P. F. & Davis, B. (2000). Deriving speech from nonspeech: a view from ontogeny. *Phonetica 57*, 284-296.

Macneilage, P. F. (2008) *The origin of speech*. Oxford: Oxford University Press.

Mehler, J.; Peña, M.; Nespor, M. & L. Bonatti (2006). The "soul" of language does not use statistics: reflections on vowels and consonants. *Cortex 42*, 846-854.

Sandler, W.; Aronoff, M.; Meir, I. & Padden, C. (2011) The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory*, *29*, 503–54.

Zuidema, W. & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics, 37,* 125-144.

FROM SILENT GESTURE TO ARTIFICIAL SIGN LANGUAGES

MARIEKE SCHOUWSTRA[1], KATJA ABRAMOVA[2], YASAMIN MOTAMEDI[1], KENNY SMITH[1], SIMON KIRBY[1]

[1] *Language Evolution and Computation Research Unit, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, EH8 9AD, UK*
marieke.schouwstra@ed.ac.uk, [kenny, simon]@ling.ed.ac.uk, s0813837@sms.ed.ac.uk
[2] *Faculty of Philosophy, Radboud University Nijmegen, 6500 HD, Nijmegen, Netherlands*
e.abramova@ftr.ru.nl

1. Introduction

Language evolution can be described as the transition from something that *isn't* language to something that *is* language. This definition allows us to remain agnostic about the mechanisms (biological or cultural) involved in the emergence of language. Moreover, the definition marks the boundary between language evolution and language change: the latter is a process that takes place when there is already a language (see the description in Scott-Phillips & Kirby, 2010). Finally, language evolution is not something that only happened in pre-history: the emergence of new languages can be observed in the present day, with newly-emerging sign languages providing the best example of such a process.

In this paper we will sketch a methodology to study the transition from no-language to language. More specifically, we will show how combining different laboratory methods will allow us to observe the transition from 'silent gesture' (the behaviour observed in naive hearing participants who are asked to convey meanings while using only gesture) to artificial sign language. By allowing silent gesturing participants to interact and learn from one another via iterated learning, artificial sign languages emerge which, we will claim, share crucial properties with existing languages. Thus, the emergence of artificial sign language in the lab can help us to understand some of the mechanisms involved in the emergence of language in the human species.

2. Silent gesture: improvised communication in the lab

Silent gesture is the behaviour observed in naive participants who are asked to convey meanings (by describing simple events) while using only gesture and no speech. Constituent order in silent gesture is independent of the native language of the gesturer: Goldin-Meadow, So, Özyürek, and Mylander (2008) found that

'motion events' (such as 'captain swings pail' or 'boy tilts glass to mouth') are consistently ordered in SOV word order. Moreover, silent gesture shows structural variability based on the semantic properties of the message to be conveyed, a kind of variability that is not observed in full language: Schouwstra (2012) found that whereas motion events lead to SOV ordered strings, more abstract intensional events (such as 'man searches for guitar' or 'woman thinks of apple') are gestured in SVO order.

Silent gesture experiments can tell us something about the way in which people represent information in strings (linearly ordered messages) in the absence of language conventions. The fact that gesture sequencing is relatively consistent across participants, and independent of the dominant word order of their native language, suggests that silent gesture experiments can tell us something about cognitive biases that play a role in communication in the absence of conventional systems for constituent ordering.

3. From gesture to sign language in the lab

The communicative behaviour of silent gesturers is unidirectional: they only produce gesture sequences, but do not interpret them.[a] We will describe how the silent gesture method can be combined with the methodologies from the Iterated Learning paradigm, in order to study the evolution of silent gesture systems.

Iterated learning is the process by which an individual acquires a behaviour by observing a similar behaviour in another individual who acquired it in the same way (Kirby, Cornish, & Smith, 2008). This definition captures two prominent types of cultural transmission, vertical and horizontal. Vertical transmission happens when new learners come into an existing linguistic community and acquire the linguistic system of that population. Horizontal transmission occurs within generations, through interaction between peers. Both processes have been studied in laboratory experiments. Vertical transmission has been shown to result in languages which become more learnable, more compressible, and thus more systematic (Kirby et al., 2008). Horizontal transmission, when studied in a graphical communication task, leads to the emergence of communicatively functional, efficient graphical conventions (Garrod, Fay, Lee, Oberlander, & MacLeod, 2007). A combination of vertical and horizontal turnover shows that linguistic structure, the presence of regularities in the way in which complex signals are constructed to convey complex meanings, arises when both horizontal and vertical transmission are at work (Smith, Tamariz, & Kirby, 2013; Kirby, Tamariz, Cornish & Smith, submitted). These findings demonstrate that we need to develop flexible experimental methodologies that allow us to investigate the relative contributions of horizontal and vertical transmission.

[a] Although interpretation experiments have been reported (Langus & Nespor 2010, Schouwstra, 2012), in these publications production and interpretation were observed separately.

Experiments in the mixed paradigm proposed in this talk (silent gesture plus iterated learning) have a very natural starting point, beginning with the communicative gestures used when a single participant communicates solely according to his own cognitive biases. These individual-based gestures subsequently come under pressures for learnability and expressivity when participants interact with, and transmit their gestural repertoire to, other participants in dyadic, closed group and replacement designs.

Combining silent gesture and iterated learning methods yields a suite of experimental methods that we can use to study how the products of the cognitive biases of individuals, through social transmission, develop into conventionalised language systems. In other words, it offers ways to create artificial sign languages in the lab. An additional advantage of studying emerging languages in the manual modality is that it gives us the possibility to compare it directly to natural data.

4. From gesture to sign language: natural data

Recently emerged sign languages, such as Nicaraguan Sign Language (NSL, Senghas & Coppola, 2001) are a valuable source of information about language evolution in the real world, and potentially reveal mechanisms by which a fully conventionalized language emerges from earlier improvised forms of communication.

NSL is an example of a community sign language: a sign language that emerged over the past 30 years from the homesigns of deaf individuals that were put together in a group. Homesigns are spontaneous, improvised sign systems developed by deaf children who grew up in hearing families, and had no access to an existing conventional sign language. Although homesign is generally highly iconic and improvisation based, different homesign systems show some similarity in utterance structure. Like in silent gesture, semantic and pragmatic principles play a role in the organisation of utterances (Benazzo, 2009).

NSL is structurally independent of the spoken languages that surround it, and has become more richly structured and increasingly systematic over the generations. Because much is known about the social dynamics under which it emerged, it is a valuable source of information about how different kinds of social transmission shape language. Laboratory studies in which silent gesture and iterated learning are combined offer a controlled environment in which phenomena observed in natural data can be studied in further detail.

5. Back to the lab: case studies in emergent structure

We will demonstrate the validity of our experimental methodology by showing that linguistic phenomena that have been observed emerging in this natural data also arise in the laboratory context. For example, Senghas, Kita, and Özyürek (2004) have noted that later signers of Nicaraguan Sign Language develop a way of signaling complex motion events by separating manner and path. For example, a ball rolling down a hill would be expressed using a *roll* gesture followed by

a *down* gesture. Importantly, the same meaning early in the development of the language would have been expressed 'holistically' with manner and path signed simultaneously. We will show, using our iterated methodology, the same transition from holistic to compositional expression of manner and path arising in the lab. Intriguingly, we find this result does not arise universally—it is a solution to expressing events that is 'lineage specific', occurring in some runs of the experiment and not others. This is interesting because such a compositional strategy is also not universal across sign languages.

In addition to these specific syntactic properties of the emerging artificial sign systems, we will also look at the phonetics of the languages that evolve. We will give quantitative evidence (extracted directly from video) that the form of the signaling in our experiments is changing to become less pantomimic and more sign-like as the systems our participants use become conventionalized and energetically efficient. In order to quantify the efficiency of gestures, we calculate the amount of movement in each gesture video, based on pixel-by-pixel comparisons of adjacent video frames: gestures at later generations feature less movement. We can use similar techniques to quantify the extent to which a *set* of gestures exhibits systematic structure: we define the similarity between two gestures videos as the extent to which they involve similar movements (again, identified based on frame-by-frame comparison within each video), and then feed these similarity measures into standard techniques for quantifying systematic structure which we have developed for studying written miniature languages (specifically, the structure measure presented in Kirby et al., 2008).

By comparing the effects of horizontal interaction with vertical transmission, we will discuss the ways in which pressures from communication and from learning impact on the process that takes us from no language to language.

References

Benazzo, S. (2009). The emergence of temporality. In R. Botha & H. de Swart (Eds.), *Language evolution: the view from restricted linguistic systems.* LOT.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive Science*, *31*, 961–987.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *PNAS*, *105*(27), 9163–9168.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory. *PNAS*, *105*(31).

Schouwstra, M. (2012). *Semantic structures, communicative principles and the emergence of language.* LOT dissertation series. Utrecht University.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory.

Trends in Cognitive Sciences, *14*(9), 411–417.

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, *12*(4), 323–328.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in nicaragua. *Science*, *305*(5691), 1779–1782.

Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society.*

EMERGENCE OF LOW-LEVEL CONVERSATIONAL COOPERATION: THE CASE OF NONMATCHING MIRRORING OF ADAPTORS

SŁAWOMIR WACEWICZ & PRZEMYSŁAW ŻYWICZYŃSKI

Center for Language Evolution Studies (CLES),
Department of English, Nicolaus Copernicus University, Bojarskiego 1
Toruń, 87-100, Poland

SYLWESTER ORZECHOWSKI

Institute of Psychology,
Maria Sklodowska-Curie University, Pl. Litewski 5, Lublin 20-080, Poland

Like all signalling, language involves several classes of constraints, such as the physical constraints of signal production, reception and noise; and the cognitive constraints related to the content of the message or inferences in the hearer's mind. However, a third, and more fundamental, type of constraints refers to honesty and stability of signalling. In what follows, we describe a research programme, currently underway, that will address the origins of stable cooperative signalling in conversation. We aim at shedding light on the mechanisms that enable and govern cooperation at the basic, low-level, layer of the communicative interaction, and their implications for the successive layers of communicative cooperation. Secondly, in line with recent trends in the area of language evolution, we put our research on an empirical footing. We target one specific type of nonverbal behaviour for experimental investigation, i.e. we purport to test empirically the influence of nonmatching mirroring of adaptors on the flow of conversation and the formation of the disposition to cooperate. The proposed research has a novel character, since non-matching mirroring of adaptors is a hitherto unexplored phenomenon.

1. Introduction

Cooperation is a foundational feature of human linguistic communication, and one whose evolutionary bases are still an unresolved question. In conversation it is most clearly visible on the 'Gricean' level, i.e. the level of content, which is described by the Cooperative Principle and itemised by the four Gricean maxims. However, the general cooperative character of conversation extends well beyond the transmission of meaning. The underlying layer of mechanics and structuring of interaction – including phenomena such as synchronisation, turn-

taking, backchannelling or various kinds of mirroring, which are not directly related to the content of messages or inferences – shows patterns of organisation that can be described as cooperative.

We suspect the abovementioned relation to be hierarchical, with the level of mechanics/structuring being primary and forming a basis for the higher-level, Gricean cooperation (and beyond, i.e. the actual cooperation over achieving common goals in extralinguistic reality). We hypothesise that the stability of human verbal cooperative signalling depends on the low-level coordinaton mechanisms; these include *adaptor mirroring* and specifically *mirroring of non-matching adaptor behaviours*, such as e.g. head movement performed in response to hand movement. We further suspect that the level of mechanics/structuring may be primary in an evolutionary sense, i.e. may have been an evolutionary precursor for the progressively more advanced forms of cooperation.

2. Low-level coordination

What we mean by "low-level coordination" is a broad and heterogeneous class of phenomena that are not directly involved in the transmission of propositional content but facilitate focused interaction (*sensu* Goffman, 1963). We deliberately start from a possibly encompassing approach. A systematic comprehensive treatment is somewhat difficult because of the vastness of the area and multitude of traditions, and the resulting "scattered terminology" (Paxton & Dale, 2013), with partly overlapping notions such as accommodation, alignment, emulation, mimicry, synergy, etc. (see e.g. Paxton & Dale, 2013; Lakin *et al*. 2003). A more developed and principled typology is in order, but we provisionally distinguish three categories of phenomena of interest:

(i) *Alignment*, related to spatial-orientational behaviours which serve to maintain sustained interaction (such as interactants arranging themselves into an L dyadic formation or a *vis-vis* dyadic formation, cf. Kendon 2009: 5ff);

(ii) *Interactional coordination*, which refers to "the degree to which the behaviors in an interaction are nonrandom, patterned, or synchronized in both timing and form" (Bernieri & Rosenthal, 1991: 403). It can be divided into *synchrony* and *matching* (see below), and probably extended by *affect coordination* (Goffman 1967);

(iii) *Conversation-specific norms* for upholding focused interaction, which primarily concern how talk is organised into turns and how turn

transitions are effected – e.g. *local management* system, *turn-taking* rules, meeting *projectability* requirements (Sacks *et al*. 1974).

The coordinative mechanisms in question are not unique to humans or to the context of conversation, and some forms can be observed in other primates or very early in human ontogeny. For example, Meltzoff & Moore (1977) found mimicry (facial imitation) in prelinguistic infants under 1 month of age. Takahashi *et al*. (2013) report coordination in vocal exchanges in common marmosets that they compare to turn-taking and explicitly label as cooperative. But, as noted above, a more careful typology is required to assess the significance of such findings.

Importantly, low-level coordination – such as the synchronisation of adaptors – entails little cost, is easily repeatable, and can be used by the conversants to diagnose their mutual commitment to engage in future cooperation involving higher cost (e.g. sharing important information). As such, it is an interesting candidate for bootstrapping cooperative signalling in conversation.

3. Adaptor mirroring

Two major types of interpersonal coordination are distinguished – *interactional synchrony* and *behaviour matching* (Bernieri & Rosenthal 1991). Although both of these types perform a variety of roles in regulating social activities, they express one – characteristically human – motif, that is, cooperative intent. Interactional synchronisation, defined as the degree to which interactants' behaviours are temporally coordinated, plays a vital role in the organisation of the communicative process, allowing for example the smooth exchange of conversational roles. Matching – also referred to as mimicry or emulation – consists in mirroring (*sensu* adopting) the behaviours of another interactant, which may take the form of, for example, unconscious adoption of someone else's accent, tempo of speech, facial expression, posture, or mannerisms (Lakin *et al*. 2003). The main function of behaviour matching seems to be liking, rapport, and affiliation. Both these mechanisms are focused on interactants' joint goal, which is to engage in the communicative activity and to promote mutual understanding (the rapport-making function).

Adaptors are a class of behaviours or actions that are nonintentional, often nonconscious and (primarily) non-communicative, often reflecting bodily needs or arousal (Ekman & Friesen, 1969) – e.g. scratching oneself or biting the lip. They may occur in a suppressed form, usually as only the initial stage of the target action. So far, adaptor synchrony has been studied mostly with regard to

matching behaviour (see Chartrand & Bargh 1999). But preliminary results from our pilot study strongly suggest *non-matching adaptor mirroring* also occurs naturally; for example, it has been observed that postural re-alignment of one participant can elicit face rubbing or shoulder raising in the other. Interactions of that sort require a more thorough analysis as to their sources, mechanism, structure and function, with particular emphasis placed on their role in the structure of conversation, as well as their possible effect on affiliation and cooperative intent.

4. Project outline

Research in this project will be based on methods and procedures developed within linguistics (Conversation Analysis and corpus linguistics) and psychology (experimental psychology of nonverbal behaviour). Its experimental core will consist of two experiments as well as a possible third experiment. It will be followed by a theoretical elaboration of the results and their integration with the state-of-the-art language evolution research.

Experiment 1. Hypothesis: non-matching mirroring of adaptors is a process spontaneously occurring in conversation. It builds on our pilot study; it replicates Chartrand & Bargh (1999), but with the inclusion of non-matching mirroring.

Experiment 2. Hypothesis: the degree of mirroring is correlated with the degree of disposition to cooperate. The degree of mirroring will be calculated through segmentation and BAP. The degree of disposition to cooperate will be calculated *via* the public goods/social dilemma game paradigm.

Experiment 3. Hypothesis: the mirroring of adaptors is partly independent of the focus of visual attention. The assumed goal of this experiment is to test the assumption of the automatic character of mirroring.

The experimental procedures will consist in: collecting and analysing an audio-visual corpus; annotating the registered behaviours with BAP (The Body Action and Posture Coding System); segmentation of the stream of behaviours; microanalysis (slow-motion behavioural analysis); analysis of conversational structures focused on the use of turn-taking rules, adjacency pair formats, preference phenomena, and pre-sequences. The above steps will be followed by a statistical analysis and evolutionary interpretation.

5. Conclusion

Human language is unique in nature as a cheap but honest cooperative signalling system. Based on evolutionary logic and available evidence from the linguistic and psychological study of conversation, we suspect that this cooperative

character rests on a scaffolding of lower-level mechanisms: human verbal communication depends on various forms of coordination of mostly nonverbal signals. In our project, we will test the influence of one such mechanism, mirroring of non-matching adaptors, on the dynamics of conversational interactions. We see that as a first step in the direction of empirical study of this proposed dependence.

Acknowledgements
This research was supported by grant UMO-2012/07/E/HS2/00671 from the Polish National Science Centre.

References

Bernieri, F. J. & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In: R.S. Feldman & B. Rime (eds.). *Fundamentals of nonverbal behavior*. Cambridge: CUP, 401-432.

Chartrand, T. L. & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893–910.

Ekman, P. & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*, 1, 49-98.

Goffman, E. (1963). *Behavior in Public Places: Notes on the Social Organization of Gatherings.* New York: Free Press

Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. New York: Doubleday.

Kendon. A. (2010). Spacing and Orientation in Co-present Interaction. *Development of Multimodal Interfaces: Active Listening and Synchrony Lecture Notes in Computer Science*, 5967, 1-15.

Lakin J. L., Jefferis V. E., Cheng C. M., & Chartrand T. L. (2003). The Chameleon Effect as social glue: Evidence for the evolutionary significance of nonconscoius mimicry. *Journal of Nonverbal Behavio*r, 27 (3), 145-162.

Meltzoff, A. N., Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science*, 198 (4312), 74-78.

Paxton, A. & Dale, R. (2013). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior research methods*, 45, 329–343.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A Simplest Systematic for the Organization of Turn-Taking in Conversation. *Language*, 50(4), 696–735.

Takahashi, D. Y., Narayanan, D. Z., Ghazanfar, A. A. (2013). Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys. *Current Biology*, 23, 2162–2168.

A LABORATORY MODEL OF SUBLEXICAL SIGNAL CATEGORY EVOLUTION

ANDREW WEDEL, BENJAMIN MARTIN
Department of Linguistics, University of Arizona
Tucson, Arizona, 85721, USA

1. Introduction and Background

Human languages are characterized by multiple, nested levels of encoding, such as the division between categories that carry meaning such as words, and the smaller inventory of sublexical, largely meaningless signal categories that can be combined in multiple arrangements to form words (Ladd 2012). Given this relationship, the function of word categories in the transmission of information is dependent on a language perceiver's ability to distinguish sublexical categories within the larger linguistic percept.

A long-standing question is how the inventory of sublexical categories evolves over many cycles of language usage and acquisition. A range of theoretical work proposes that the maintenance of this inventory over generations is causally grounded in the transmission of information in usage (e.g., Trubetzkoy 1939, Martinet 1955, King 1967, Zuidema & de Boer 2009, Wedel 2012), rather than through some directly innate mechanism (e.g., Ni Chiosain & Padgett 2009). Previous modeling work has shown that the well-established perception-production feedback loop in language usage should allow any bias toward selective preservation of signal-quality to influence the evolution of the signal-category inventory over generations (Wedel 2004, Blevins & Wedel 2009, Wedel 2012; cf. work in *iterated learning* (e.g. Kirby 1999)). If signal-quality is preferentially maintained in relation to the role of that signal in communicating word-identity, we expect the evolution of signal inventories to preferentially preserve the categories that play a larger role in distinguishing word categories.

This hypothesis is supported by recent work showing that sublexical sound category loss is significantly, inversely correlated with the number of words distinguished by that category (also known as *minimal pairs*; Wedel et al. 2012).

For example, the /ɔ ~ ɑ/ vowel distinction in English distinguishes very few minimal word pairs; an example of a minimal pair like this is *caught* ~ *cot*. Correspondingly, the distinction between /ɔ/ and /ɑ/ has been lost in many North American dialects of English such that *cot* and *caught* are now homophonous in those dialects. Conversely, sound categories that distinguish many words appear to be especially protected from loss (Wedel and Jackson, in prep). Findings from recent experimental (Baese & Goldrick 2009, see also e.g., Eisner & McQueen 2005, Kraljic & Samuels 2005, Verhoef et al. 2012) and corpus studies (Wedel & Sharp in prep) are also consistent with the hypothesis that a perceptual cue to the identity of a given word is hyperarticulated if it plays a large role in distinguishing that word from a similar word, and conversely, a perceptual cue that plays a smaller role tends to be reduced.

However, the causal mechanism(s) more directly underlying selective hyperarticulation remains unknown (reviewed in Baese & Goldrick 2009, Wedel 2012). In response, we have developed a laboratory model of naturalistic speech to investigate sound change in response to communicative pressure. Here, we report an investigation suggesting that word pairs do not need to directly compete in context in order to induce hyperarticulation of perceptual cues. This question is relevant because in actual usage, minimal pairs are rarely similarly probable in the same discourse context.

2. Methods

The laboratory model is based on a map-task in which two participants take turns instructing each other to draw a path through a set of landmarks on a map. Each of the landmarks on the map is an object with a monosyllabic English name. The set of landmarks were chosen to provide examples of two kinds of easily measured phonetic contrasts: initial stop-consonant voicing (as in *peach* ~ *beach*), and vowel height (e.g, *chick* ~ *check*). Participants' speech was recorded through head-mounted microphones, and the relevant phonetic measures were subsequently made using Praat (Boersma & Weenink 2013).

A major cue to the voicing distinction in initial stops in English (i.e., p~b, t~d, k~g) is the ratio of the length of the burst to the entire stop length (Lisker & Abramson 1964). All else being equal, the longer the relative burst length the greater the percept of voicelessness, while conversely the shorter the relative burst length, the greater the percept of voicing. The burst/stop-length ratio for each stop token was normalized by z-scoring within each word, within each participant. Two phonetically-close vowel pairs were also compared, /ɪ ~ ɛ/ and /æ ~ ʌ/. Formants from the central portions of vowel tokens were measured with

Praat, and the Euclidean distance was calculated between a given vowel token and the average F1 and F2 values for the comparison vowel, for that participant. These distances were normalized as above. One set of maps consisted of landmarks with no minimal pairs in English in the relevant sounds. As an initial baseline, each pair of participants worked through ten of maps with no minimal pairs, split up evenly between two successive days. (Each different map had a different subset of landmarks, arranged differently, with different paths; five maps provided about one hour of conversation.) A prediction of the model is that the measured phonetic cues should become *less* distinctive over the two days, because these cues contribute little to distinguishing these words within the task. Each pair of participants then did a second set of 10 maps on another two subsequent days, where the second set of maps provided one of two different degrees of lexical competition. In the Direct Competition set, lexical minimal pairs (e.g., peach ~ beach, chick ~ check) were both present in the map, and members of each pair were immediately adjacent to each other in half of the individual maps, placing a premium on clear articulation of the phonetic cue. In the Indirect Competition set, the members of each minimal pair were present only in alternating maps, so that clear articulation of the relevant phonetic cues had no direct role in context, yet both minimal pairs were pronounced each day.

3. Results and Discussion

As predicted, in both pairs the phonetic cues of interest are reduced in the initial no-minimal pair condition on the second day, relative to the first. Figure 1 shows the relative shift in burst/length ratio from Day 1 to Day 2 for voiced and voiceless stops; note that the ratio grows larger (i.e., more voiceless-like) for the voiced stops, and conversely grows smaller for voiceless stops. The vowels also pairs also reduce, becoming less distinctive on the second day relative to the first. Linear mixed-effects modeling (Barr et al. 2013) indicates that this pattern is statistically significant for these participants.

For both the Direct and Indirect Competition conditions in the second set of maps for the participant pairs, the opposite occurs: on the second day, each phonetic contrast has become *greater*, and when the data is pooled across the set of participants, this is again statistically significant; Figure 2 shows the change in burst/length ratio for stops, and Figure 3 shows an interaction plot for vowel-vowel distances comparing the first set of maps without minimal pairs, to the second set of maps with minimal pairs, pooling over the Direct and Indirect Competition conditions. There is no visual or statistical evidence in this dataset that the Direct and Indirect Competition conditions produce different degrees of

phonetic cue hyperarticulation. This initial exploration suggests that multi-day trajectories of phonetic reduction and hyperarticulation in response to the existence lexical competitors can be investigated in the laboratory. Further, the finding of strong hyperarticulation in the Indirect Competition condition suggests that lexical minimal pairs do not need to directly compete within context in order to induce hyperarticulation. This is consistent with a model for sublexical contrast maintenance deriving from competition in articulatory planning, rather than through listener-orientation (reviewed in Baese & Goldrick 2009). We are currently carrying out additional studies in which lexical competitors are not present in the task at all, to ask whether the simple existence of lexical minimal pairs within the language is sufficient to prevent reduction.

Fig. 1

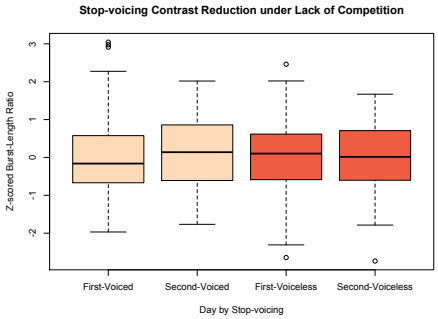


Fig. 2

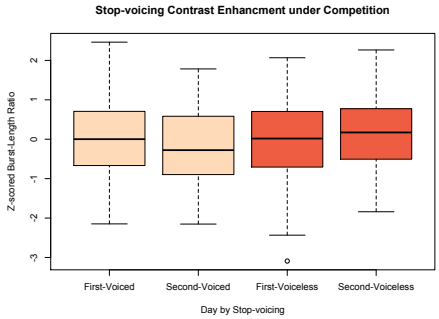
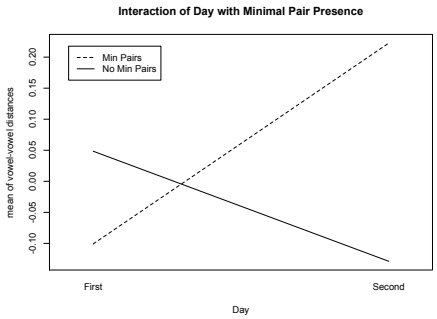


Fig. 3



References

Baese, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. Language and cognitive processes, 24 , 527-554.

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language 68, 255-278.

Blevins, J. & Wedel, A. (2009). Inhibited Sound Change: An Evolutionary Approach to Lexical Competition. Diachronica 26: 143-183.

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.56, retrieved 15 September 2013 from http://www.praat.org/

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. Perception & Psychophysics, 67 , 224-238.

King, R. (1967). Functional Load and Sound Change. Language, 43, 831- 852.

Kirby, S. (1999). Function, selection and innateness: The emergence of language universals. Oxford: Oxford University Press.

Kraljic, T. & Samuel, A. (2005). Generalization in perceptual learning for speech. Psychonomic Bulletin and Review 13, 262-268.

Ladd, D. R. (2012). What is duality of patterning, anyway? Language and Cognition 4, 261–273.

Lisker, L. and Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. Word 20, 384-422.

Martinet, A. (1952). Function, structure, and sound change. Word, 8 , 1-32.

Ni Chiosain, M., & Padgett, J. (2009). Contrast, comparison sets, and the perceptual space. In S. Parker (Ed.), Phonological argumentation: Essays on evidence and motivation (chap. 4). London: Equinox.

Son, R. J. J. H. van, & Pols, L. C. W. (2003). How efficient is speech? In E. H. Berkman (Ed.), Proceedings of the institute of phonetic sciences. Amsterdam.

Trubetzkoy, N. (1939). Grundzüge der phonologie. Prague, Czech Republic: Travaux du Cercle Linguistique de Prague.

Verhoef, T., de Boer B. & Kirby, S. (2012). Holistic or synthetic protolanguage: Evidence from iterated learning of whistled signals. In T.C. Scott-Phillips, M. Tamariz, E.A. Cartmill & J.R. Hurford (Eds.), The evolution of language: Proceedings of the 9th international conference (evolang9) (pp. 368-375). Hackensack NJ: World Scientific.

Wedel, A. (2012). Lexical contrast maintenance and the development of sublexical contrast systems. Language and Cognition, 4: 319-355.

Wedel, A., Kaplan A., and Jackson, S. (2013). Lexical contrast constrains phoneme merger: a corpus study. Cognition, 128: 179–186.
Zuidema, W. & de Boer, B. (2009). The evolution of combinatorial phonology. Journal of Phonetics, 37(2), 125-140.

NEUTRAL SPACES AND THE EVOLVABILITY OF SPOKEN LANGUAGE

BODO WINTER

Cognitive and Information Sciences,
University of California, Merced, 5200 North Lake Rd.
Merced, 95340, U.S.A.

1. Neutral spaces

Many systems have to resist changes from within and without. One way in which this is achieved is via neutrality (Wagner, 2005). For example, in biology, Kimura's neutral theory of molecular evolution (1983) states that most genetic mutations are effectively neutral with respect to evolutionary fitness; most mutants are not "seen" by natural selection. This makes biological systems robust against mutations. In general, biological systems frequently occupy ***neutral spaces*****,** which are collections of "equivalent solutions to the same biological problem" (Wagner 2005: 195).

Spoken language is another system that has to resist internal and external perturbations. For speech communication to be effective in a noisy world, it needs to be robust (Winter & Christiansen, 2012). And, just as with biological systems, one way to achieve robustness is via neutrality: If speech sounds occupy neutral spaces, underlying variation may have little or no effect on the outcome of communication. At least two phonetic phenomena create such neutral spaces:

First, ***quantality***, which refers to non-linear mappings of articulatory input to acoustic output (Stevens, 1989). Quantality says that there are regions of articulatory space where variation has no discernible acoustic effect (in Fig. 1a, regions I and III). Take, for example, /s/ as in *sell*, and / ʃ / as in *shell*. If one slowly moves one's tongue from /s/ to / ʃ /, there is a sudden transition between the two sounds, with large regions that render equally good instantiations of either /s/ or / ʃ /.

A second phenomenon is ***categorical speech perception***, which refers to non-linear mappings between acoustics and perception (for review, see Harnad, 1990). Take, for example, voicing (e.g., ***b****ear* vs. ***p****ear*), for which voice onset time (the time between the release of a stop and the beginning of the following vowel) is a crucial cue. If we manipulate voice onset time to create a continuum

between the words *bear* and *pear*, participants hear either one word, or the other, with a sudden transition at the category boundary (see Fig. 1b).

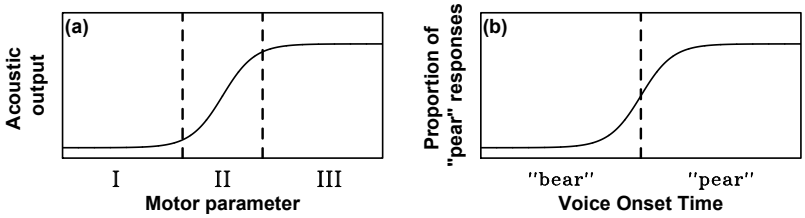


Figure 1. Schematic representations of (a) quantality and (b) categorical perception.

Neutrality unifies both quantality and categorical speech perception, because variation in an underlying parameter is neutral with respect to communicative outcomes. Neutrality assures that most perturbations result in linguistically equivalent signals.

Intuitively, one might think that robustness to noise could mean that a system cannot change easily. At first sight, the "requirements to be both robust and adaptive appear to be conflicting" (Whitacre, 2010: 1). In fact, though, robustness and evolvability are not mutually exclusive. Instead, they may even enhance one another (Wagner, 2005; Whitacre, 2010). The following simulation demonstrates this.

2. Simulation

The goal of the simulation is to show that non-linearity leads speech signals to have *less* communicatively relevant variability (i.e., more robustness), but *more* underlying, cryptic variability. As any evolutionary system needs variation for subsequent change (including sound systems, Wedel, 2006), this underlying variability can be seen as "fodder" for elvovability.

In the simulation, 100 linguistic signals are initiated. Each signal is a value drawn from a uniform distribution with the range [-10,10]. For quantality, this represents the range of possible motor inputs. For categorical perception, this represents the range of possible acoustic inputs. The input is transformed either non-linearly (see Fig. 2a) or linearly (as if no neutrality existed, see Fig. 2b).

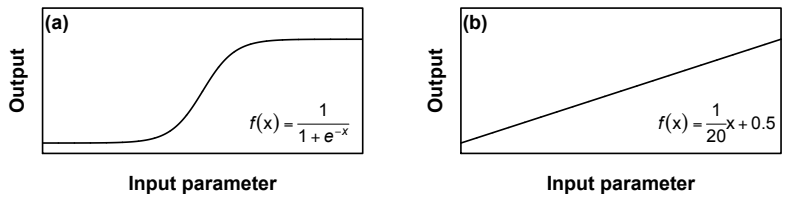


Figure 2. (a) Non-linear transform (logistic function) and (b) linear transform.

Non-linearity is implemented via the logistic function (shown in Fig. 2a). This function mirrors categorical perception curves and quantally divided acoustic spaces. The linear function (Fig. 2b) was chosen to keep inputs between -10 and 10 constrained to outputs within the range [0,1].

Change is implemented the following way: Signals are biased towards conformity, as if agents were imitating each other. One could imagine the 100 signals to be 100 slightly different phonemes (e.g., /s/) used in the same word (e.g., *sell*) by 100 different agents. The agents try to converge on the same output value for this word, that is, they try to pronounce /s/ as similarly as possible to what others say. Such an artificial conformity bias can be implemented via any clustering algorithm that finds the most frequent cluster in the output space.

In the present simulation, k-means clustering is used as one particular clustering algorithm. A two cluster solution is sought. Signals that are not classified as belonging to the more frequent cluster are adjusted upwards if they are below the centroid, and downwards if they are above the centroid. A crucial component of the model is that clustering acts on output space, but adjustments are done in input space. Figure 3 shows two representative runs.

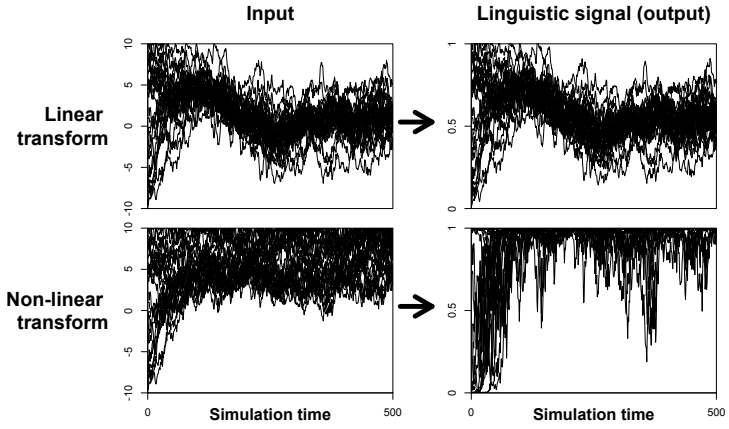


Figure 3. An example simulation of linear and non-linear simulation runs of 30 signals evolving over 500 simulation steps in underlying parameter space (left column) and output space (right column).

1,000 linear and 1,000 non-linear simulations with 500 time steps each show that non-linear transforms create more *output* ***stability*** in the linguistic signal, as well as more underlying *input* ***variability*** (see Fig. 4). At the 500th time step, non-linearly transformed signals have *higher* underlying variability (as measured by standard deviations over all signals) than the linearly

transformed ones (t(1998)=54.18, p<0.0001). For output variability, non-linear signals have *lower* values (t(1998)=45.5, p<0.0001). In these simulations, underlying parameter values are bounded to be within [-10,10]. This invites the concern that there are artificial biases due to boundary conditions (see, e.g., Bullock, 1999). However, an equivalent simulation run without restricting inputs produces qualitatively similar results.

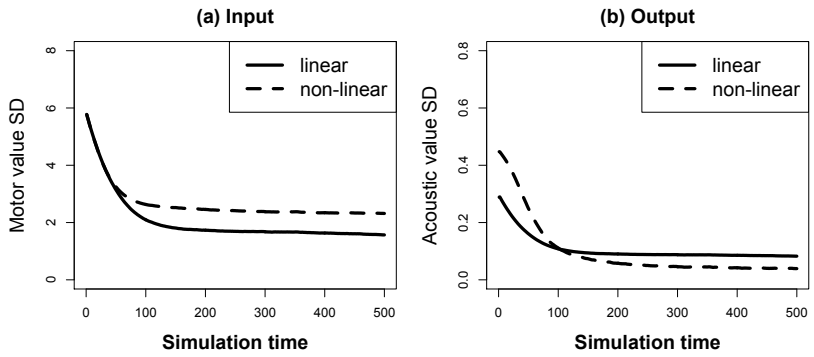


Figure 4. Standard deviations of motor input values and acoustic output values over simulation time, for simulations with linear and logistic transformation.

The simulation demonstrates that evolving signals have more cryptic underlying variability if the conformity bias acts on non-linearly transformed spaces, hence, they have more "fodder" for subsequent evolution. At the same time, signals have less communicatively relevant variability, making the underlying variation more neutral. Thus, a biological aspect of the speech apparatus (quantality) and a cognitive aspect of the language users (categorical perception) create neutral spaces that drive robustness and evolvability of spoken language.

What are the ultimate origins of these non-linearities? Categorical speech perception has been reported for many non-human animals (see reviews in Harnad, 1990), including monkeys. It thus seems safe to assume that early humans already had the capacity to divide a signal space into categories. Through historical sound change, categorical speech perception boundaries may shift, as is evidenced by the fact that different languages have strikingly different voice onset times to distinguish between voiced and voiceless stops (Lisker & Abramson, 1963). Thus, for categorical speech perception, it is realistic to assume that cryptic variation may surface when conditions change, such as when the category boundary between two sounds shifts as a result of historical change.

This is different from quantality. The quantal nature of speech is determined by vocal tract physiology and therefore, it cannot be changed throughout a speaker's lifetime. This means that the non-linearity for quantality is rigid, and underlying variation in articulation cannot surface. While cultural evolution may drive signaling systems to live within the quantal regions of motoracoustic space (because they afford a high degree of motor variability), the fact that these quantal regions exist may need to be explained via biological evolution. This would thus represent another way in which the physiology of the vocal tract is optimized for speech. However, the rigid nature of quantality means that for this phonetic phenomenon, the cryptic variability demonstrated in the above simulations does not impact evolvability—in contrast to the cryptic variability in categorical speech perception.

To conclude, this paper argues that non-linear phenomena in speech create neutrality, which is key to understanding how speech communication can be robust and at the same time evolvable. The robustness of speech is not only an *explanandum* in language evolution research—something that needs to be explained evolutionarily—but it is also a driver of language evolution.

References

Bullock, S. (1999). Are artificial mutation biases unnatural?. In D. Floreano, J.-D. Nicoud & F. Mondada (Eds.), *Advances in Artificial Life: Fifth European Conference on Artificial Life* (pp. 64-73). Berlin: Springer.

Harnad, S. R. (Ed.). (1990). *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge, UK: Cambridge University Press.

Lisker, L., & Abramson, A. S. (1963). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–46.

Wagner, A. (2005). *Robustness and evolvability in living systems*. Princeton: Princeton University Press.

Wedel, A. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, 23, 247-274.

Whitacre, J. M. (2010). Degeneracy: a link between evolvability, robustness and complexity in biological systems. *Theoretical Biology and Medical Modelling*, 7, 1-17.

Winter, B., & Christiansen, M.H. (2012). Robustness as a design feature of speech communication. *Proceedings of the 9th International Conference on the Evolution of Language* (pp. 384–391). New Jersey: World Scientific.